

---

---

**2010 New York Youth Tobacco Survey  
SAMPLING AND WEIGHTING REPORT**

**Submitted to:**

**New York Department of Health**

**by:**

**ICF Macro (Macro International Inc.)  
September 30, 2010**

---

---

## CHAPTER 1. INTRODUCTION

The New York Youth Tobacco Survey (NY YTS) includes questions about tobacco use, exposure to environmental tobacco smoke, minors' ability to purchase or otherwise obtain tobacco products, knowledge and attitudes about tobacco, and familiarity with pro- and anti-tobacco media messages. This was the third successful statewide YTS conducted by the state of New York. The NY YTS was funded by the New York Department of Health in order to collect data sufficient to provide accurate estimates separately for students in the City of New York, for students in New York State attending schools outside New York City, and for the students in New York State as a whole.

The YTS, developed by the Centers for Disease Control and Prevention (CDC) in collaboration with the states and first conducted in four states in 1998, has been successfully completed at least once by 46 states. The National Youth Tobacco Survey (NYTS), also developed by the CDC and first conducted in 1999, provides national level estimates on a similar set of youth tobacco use topics and, with the state YTS surveys, forms the surveillance system for the assessment and monitoring of tobacco use and related behaviors among students in grades six through twelve.

The NY YTS sample was selected with a multistage stratified cluster design with sample sizes developed to meet the estimation objectives with required precision levels. The study is designed to produce regional estimates; and within each of the two regions, estimates by school level (middle, high), by gender, and by race/ethnicity (Whites, blacks and Hispanics).

The first-stage sampling frame for the NY YTS sample was obtained from national files constructed from a database obtained from MDR (a company formerly known as QED, Inc.). From this frame, 42 PSUs were drawn with replacement, with probability proportional to weighted enrollment. At the second sampling stage, 46 schools were drawn in New York City and 42 schools were drawn from the remaining portion of the state. The final stage of sampling consisted of the selection of approximately 125 students in five intact class sections across grades 6 to 12 in each participating school. The actual number of class sections selected depended on the average class size of a given school to maintain comparability across schools in the number of students sampled.

A weighting factor was applied to each student record from to adjust for non-response and for the varying probabilities of selection. The weights were adjusted so that the weighted number of students in each region and each school level matched state population enrollment figures. A total of 9,503 students in 83 schools completed NY YTS questionnaires.

Based on the study design, any findings derived from NY YTS are subject to at least two limitations. First, findings from the NY YTS apply only to youth attending middle or high school and are not representative of all youth in this age group. However, nationwide, relatively few persons aged 16 and under do not attend school. Second, despite the inclusion of novel tobacco products, the prevalence of "any current tobacco use" still might be underestimated in the survey because the questionnaire does not specifically address "roll-your-own" tobacco products.

Chapter 2 of this report is devoted to the sampling design developed for the 2010 NY YTS. Chapter 3 describes data collection procedures. Chapter 4 describes the weighting procedures adopted in the 2010 NY YTS, and Chapter 5 describes data processing procedures.

## CHAPTER 2. SAMPLING DESIGN

The 2010 NY YTS survey was re-designed to be a more efficient, streamlined survey with parameters tailored specifically for the state survey objectives. In essence, two regional sub-studies needed to be designed for the two different regions, New York City (NYC) and Balance of the State (BOS), which vary in the racial/ethnic distributions of the student population.

In particular, a NY-specific design would no longer oversample minority groups in New York City where they are more than sufficiently represented to generate group estimates for Blacks and Hispanics. Rather, the re-design would consider oversampling Whites in the region as previous cycles of the NY YTS yielded lower precision levels for this subgroup.

The re-design was focused on maximizing effective sample sizes, i.e., on maximizing the precision of survey estimates for a given sample size; or equivalently, minimizing the sample sizes needed to generate specified levels of precision.

Section 2.1 provides an overview of the sampling design. Section 2.2 describes the target sample sizes based on the key analytic objectives of the study. The objectives are framed as estimates for key subgroups or domains; sample sizes then flow from the precision required for these estimates. Section 2.3 describes the results of the simulation study leading to the final sampling design parameters related to the measure of size within each region. Section 2.4 describes selection methods used for sampling schools and classrooms.

### 2.1 Overview of the Sampling Design

As summarized in Exhibit 1, the NY YTS state sample consisted of 84 large school selections and 4 small school selections, sample sizes that account for non-response at both the school and student levels. This sample was designed to yield at least 3,000 student surveys for each grade level and for each region. Our simulation studies suggested that we would exceed these targets; we anticipated achieving about 3,500 participating students in NYC and 3,300 in BOS (total=6,800 approximately).

**Exhibit 1: Summary Sample Size Parameters - Number of Schools in each Stratum**

<b>Stratum</b>	<b>Large Schools</b>	<b>Small Schools</b>	<b>Total (Schools)</b>	<b>PSUs</b>
BOS – Urban (BOSU)	20	1	21	10
BOS – Non-Urban (BOSN)	20	1	21	10
NYC	44	2	46	22
<b>Total</b>	<b>84</b>	<b>4</b>	<b>88</b>	<b>42</b>

To achieve the target numbers of sample schools in each region with 2 large schools per PSU, we selected 22 PSUs in the NYC region and 20 PSUs in the BOS region. Note that the BOS region is divided into two regional sub-strata, an urban portion (BOS-Urban, or BOSU) and a non-urban portion (BOS-Non-urban, or BOSN), for a total of 3 regional strata.

Sampling methods with probabilities proportional to size (PPS) were used at the first two stages of selection for PSUs and schools. The PPS sampling methods for both regions used a weighted measure of size (MOS) that permits the use of varying sampling rates for schools with different racial compositions.

First-stage sampling units, or primary sampling units (PSUs), were areas comprised of single counties or groups of smaller counties. At the second stage, sample schools were selected within each sample PSU. The measure of size (MOS) used in PPS selection was assigned to each school in the frame. An aggregate MOS was assigned to each primary sampling unit (PSU) by adding the MOS values of all schools in the PSU.

The small schools were sampled from a separate stratum comprised of small schools, a stratum formed within each primary stratum.

## **2.2 Sample Sizes and Expected Precision**

The NY YTS is designed to provide accurate estimates for these tobacco-related measures separately for students in the City of New York, for students in New York State attending schools outside New York City, and for the students in New York State as a whole, for the New York State student population of 6th through 12th graders. The study is also designed to produce estimates by school level (middle, high), by gender, and by race/ethnicity (African American and Hispanic).

Using data from previous cycles, we estimated that a sample of approximately 6,500 participating students would be needed from about 80 participating schools. We then selected 88 sample schools—4 of which are small schools—with the allocation described in Exhibit 1.

The sample was designed to support estimates with the required levels of precision for two regions rather than for the state as a whole. The two regions are defined as New York City (NYC), and the balance of the State (BOS). New York City consists of the five Boroughs (counties): Borough of the Bronx (Bronx county FIPS 36005); Borough of Brooklyn (Kings county FIPS 36047); Borough of Manhattan (New York county FIPS 36061); Borough of Queens (Queens county FIPS 36081) and the Borough of Staten Island (Richmond county FIPS 36085).

The tightest estimation precision requirements are for target racial and ethnic groups within each region. The sample design parameters—sample students and schools per region and per grade level—were derived from the following targets:

- a) 250 Whites in New York City for each grade level (HS, MS), and 500 total for the region;
- b) 250 Blacks and Hispanics in the BOS region for each grade level (HS, MS), and 500 total for the region.

The simulation studies described below suggested that these target sample sizes could be approximately achieved with the planned PPS sampling design.

### 2.3 Measure of Size for PPS Selection

To over-sample African American and Hispanics, a weighted measure of size (MOS) is used for PPS sampling. This MOS takes the form  $r_aA + r_bB + r_hH + r_oO$  where the r's are the weighting factors for the Whites, African American, Hispanic and Other populations ( $A$ ,  $B$ ,  $H$ , and  $O$ , respectively). This function can increase the chances of schools with relatively large minority enrollments entering the sample. For the 2010 NY YTS, we simplify the notation for the coefficients needed for three racial/ethnic groups of analytic interest as follows:

B= African Americans (or Blacks)

H= Hispanics

W= Whites

The effectiveness of a weighted measure of size in achieving over-sampling is dependent upon the distributions of African Americans and Hispanics in schools which varies across the two regions. For the NYC region, no over-sampling of Blacks and Hispanics is necessary; rather, it seems necessary to over-sample whites to achieve the required subgroup sample sizes within this region. For the BOS region, however, the over-sampling design takes on the more traditional form used in previous cycles. In both regions, the specific coefficients were derived to optimize the design and achieve the target sample sizes per subgroup specified in each region.

We developed a simulation study to generate optimal coefficients for the separate measures of size (MOS) associated with the PPS selection of schools in NYC, and of schools and PSUs in the BOS region. In consultation with the NY DOH staff, we developed coefficients that in both regions, balance the goals of achieving the minority group target yields while maintaining Design Effects (DEFFs) at reasonably low levels. The MOS coefficients for African Americans, Hispanics and Whites were chosen in the two regions to maximize the effective sample sizes expected for the racial/ethnic groups of interest in the two regions.

The measure of size was computed as follows. In New York City, the MOS was

$$\text{Size} = \sum_{i \in \text{grades}} ER_i (2.0 H_i + 3.0 B_i + 4.0 W_i)$$

In the Balance of State, the MOS was

$$\text{Size} = \sum_{i \in \text{grades}} ER_i (9.0 H_i + 6.0 B_i + 1.0 W_i)$$

Here,  $i$  denotes the grade,  $ER$  denotes the enrollment ratio, and  $W_i$ ,  $B_i$ ,  $H_i$  denote the  $i^{\text{th}}$  grade enrollment for the three racial/ethnic groups described above (Whites, Blacks, Hispanics).

These coefficients mean that in NYC, Blacks, Hispanics and Whites were selected with weighting factors of 3, 2 and 4, respectively. In BOS, they were selected with weighting factors 6, 9, and 1.

## 2.4 Selection Methods

*School sample selection:* In the NYC stratum, two large schools were allocated to each selected PSU each time the PSU was drawn into the sample. In other words, the school sample size for the PSU was two times the number of times the PSU was selected. In the remaining strata (BOSN and BOSU), two large schools were allocated to each of the sample PSUs.

Schools were selected from each PSU with probability proportional to the weighted measure of size of the school, using a PPS systematic sampling procedure with a random start. Prior to sampling, schools were sorted by the highest eligible grade, then lowest eligible grade, then geographic location, and finally by eligible enrollment. This sort implicitly stratified the sample by high school/middle school grades as well as by location and size.

For small schools, ICF Macro calculated the percentage of students that should come from small schools considering both data collection costs and statistical efficiency. Given that unit costs were higher among small schools, this meant that the small school sample had fewer students in the sample than their proportion in the population. Four sample PSUs were randomly sub-sampled for the sampling of the specified number of small schools ( $n=4$ ). The sub-sample PSUs were allocated two each to the two regions, i.e., two PSUs were sub-sampled in the NYC region and two PSUs were sub-sampled in the BOS region (one in each of the BOSN and BOSU sampling strata). In each sub-sampled PSU, one small school was selected with probability proportional to size, using the weighted measure of size.

Refusing schools were not replaced. To account for school non-response, the number of school selections was inflated by a factor (estimated from the expected rates of non-participation) to generate the required number of participating schools, a method that is statistically preferable to replacing schools.

*Student sample selection:* The method of selecting students varied from school to school, depending upon the organization of that school. To ensure that all students had one and only one chance of being selected to participate, classes were chosen from required core courses such as English, social studies, math, or science. Among middle school students (and in a few states, also among high school students), physical education and/or health also were considered core courses.

In a few schools, it was difficult to develop an appropriate frame. In those schools, classes were selected by using a time of day when all students had class, and randomly selecting from all classes held at this time (usually second period). In other schools, homerooms were used as the frame for class selection. Homerooms can be good choices; however, relatively few schools operate with homerooms meeting long enough for a survey to be conducted.

The target number of students selected within a given school remained constant. A target of 125 students was relaxed in those schools with fewer than 125 students at the grades of interest. The number of classes to be selected varied widely depending on the average class size obtained by dividing the total enrollment at the grades of interest by the number of class sections available in the

chosen core subject area or at the selected time period. Sample class sections were then selected with a systematic random sampling procedure. This process consistently targeted the number of classes needed to meet or exceed the 125 student requirement. Because average class sizes were employed, the number of students actually selected in each school varied from this target number (125 students).

All students in a selected classroom were given the opportunity to participate in the survey. Refusing classes and students were not replaced. As mentioned earlier, the sample sizes accounted for non-response at the student level using historical average yields per class.



## CHAPTER 3. DATA COLLECTION

The recruitment of school districts and schools required a well-coordinated system of pre-contact calls to districts largely by State Education Agency or Health Department personnel, and occasionally by local health department personnel; recruitment calls to districts and schools by highly experienced recruiters who emphasized the need for accountability in fielding the statewide tobacco education efforts; pre-contact of schools by sympathetic district personnel; use of tested "conversion" techniques; and close monitoring of recruitment efforts at all levels (state, district, and school).

### 3.1 State and District Approval Process

For the 2010 NY YTS, Richard F. Daines, M.D., Commissioner of the New York Department of Health wrote personalized support letters to all school districts represented in the New York State sample. These letters were followed by phone calls to the districts and schools to obtain their agreement to participate.

### 3.2 School and Student Participation

Of the 88 eligible sample schools selected for the New York YTS, 83 schools, or 94.3%, participated in the NY YTS.

Of the 10,811 students selected for the New York YTS, 9,503 students, or 87.9 %, participated. The combined participation rate (school x student) was 74.3%, exceeding the minimum requirement for a combined participation rate of 82.9 %.

### 3.3 Incentives

As a gesture of appreciation for contributing their time and support and for being a significant partner in the NY YTS effort, participants were offered a fairly substantial incentive, which amounted to \$1,000 per school. It was suggested to schools that the funds be used for prevention curricula and educational materials. However, no restrictions were placed on how the schools were to use these funds. The schools greatly appreciated the incentive funds. Several administrators and instructors told us the funds were used for tobacco prevention curriculum, staff training, and development of cessation programs.

The financial incentives were not viewed as "bribery," but certainly "sweetened the pot" in a number of instances, eliciting participation from schools that may have otherwise declined. NY YTS participants viewed the receipt of incentives as appropriate, and also saw the incentive as a way of building programs directly related to the content of the survey.

### **3.4 Field Data Collection Methods**

Five data collectors were recruited for the NYTS 2010, from a variety of sources. Recruitment sources included returning NY YTS data collectors and their networks, prior recruitment sources including Craigs List and other listservs, and the school health network tapped through the American School Health Association.

Data collector training was modeled after the highly successful national NYTS training. The training is intended to develop technical skills while engendering a strong commitment to the concept of the project and the project team. Over the course of training, the field staff first observed everything they would have to say or do being performed by the training team; then they acquired these skills through practice, demonstrated them to each other, and finally refined each other's performance through constructive feedback. By the end of training, the field staff acquired the requisite skills; were capable of professionally representing the New York Department of Health, the Centers for Disease Control and Prevention, and ICF Macro; were bonded to the project, the training team, and each other; and were better equipped to "stay sane on the road." All of the data collectors stayed with the project until its successful conclusion and exhibited a high level of performance.

## CHAPTER 4. WEIGHTING PROCEDURES

This chapter describes the weighting procedures implemented for the 2010 NY YTS. Section 4.1 describes the school sampling weights that reflect the probabilities of selection. Section 4.2 describes non-response adjusted weights for schools. Section 4.3 introduces student weights, and Section 4.4 describes post-stratification adjustments. Section 4.5 describes weight trimming procedures. Finally, Section 4.6 presents some weighted estimates that were computed as part of the QA process for the weighting.

### 4.1 School Sampling Weights

As described in Chapter 2, Primary Sampling Units (PSUs) are counties or groups of counties. There are 5 PSUs in the NYC region (the 5 boroughs), each selected with certainty with a multiple number of PSU selections (or “hits”). Sampling weights for PSUs were computed as the reciprocal of the probabilities of selection for the PSU. In NYC, all certainty PSUs were assigned a sampling weight of one.

Sampling weights at the school level were computed as a product of two sampling weights:

- 1) PSU-level weights,  $W_1$ , the reciprocal of the probability of selection for each PSU, and
- 2) School-level conditional sampling weights,  $W_2$ , the reciprocal of the probability of selection for each school within the given sample PSU.

In other words, overall sampling weights for schools are:

$$W_{12} = W_1 * W_2$$

Specific expressions for the sampling weights,  $W_1$  and  $W_2$ , are provided next.

The derivation of the first-stage sampling weights, denoted by  $W_1$  for simplicity, corresponds to the selection of  $n(i)$  PSUs in stratum- $i$  with probabilities proportional to  $S(i)$ . These weights, the reciprocal of the PSU probabilities of selection, are then computed as follows.

$$W_{ij}^P = \frac{S_i}{n_i S_{ij}}$$

As shown in Exhibit 1,  $n(i)=22$  in the NYC stratum and  $n(i)=10$  in each of the two BOS strata.

The second-stage sampling weights, W2, correspond to the selection of 2 schools in PSU(j) in stratum-i with probabilities proportional to S(i,j). These weights, the reciprocal of the PPS probabilities of selection, were computed as follows for school-k in PSU-j and stratum-i.

$$W_{ijk} = \frac{S_{ij}}{2S_{ijk}}$$

The sampling weight for small schools includes an additional factor associated with the sub-sampling of PSUs for the small school sample (n=4).

## 4.2 School Weight Adjustments

Adjustment classes are defined as the three regional strata defined in Section 2.1, NYC and the urban and non-urban portions of BOS. Weight adjustment factors are then separately computed in these weighting classes. The adjustment factor, F2, was computed separately within each weighting class, or regional stratum, as the ratio of the following two sums:

A= sum of the weighted measure of size (MOS), i.e. sum of W12\* Size, over all sample schools in the stratum, and

B= sum of the weighted measure of size (MOS), i.e. sum of W12\* Size, over all participating schools in the stratum.

In defining the numerator (A) and denominator (B) of this adjustment factor, recall that W12 are overall sampling weights for schools and “Size” denotes the MOS.

That is,

$$F2 = A/B.$$

Adjusted school weights were then computed as

$$WT\_SCHOOL = F2 * W12$$

For large schools, the adjustment factor is based on the sums of weighted size measure (MOS) over all schools (k) in PSU-j in stratum-i

In the numerator, the sum is over all selections; in the denominator, the sum is over participating schools only.

$$F_{i,large} = \left( \frac{\sum_k W_{ijk}^S W_{ij}^P S_{ijk}}{\sum_{k \in \text{Stratum } i \text{ respondents}} W_{ijk}^S W_{ij}^P S_{ijk}} \right)$$

A separate weight adjustment class was created for small schools. In this case, the adjustment factor is the sum over participating small schools divided by the sum over all small schools in the sample.

$$F_{small} = \left( \frac{\sum_{k \in frame} S_k}{\sum_{k \in stratum} W_{ij}^P W_{ijk}^S S_k} \right)$$

### 4.3 Adjusted Student Weights

A combined student sampling and non-response adjustment factor, F3, was computed within each cell defined by gender within each participating school as follows.

F3= number of selections divided by number of respondents

Initial student weights, prior to post-stratification, were then computed as:

$$WT3 = F3 * WT\_SCHOOL$$

We may express the weight adjustment factors (F3) as a function of the selection and response factors for r(i,j,k) respondents out of a school enrollment E(i,j,k), using the same notation for school-k within PSU-i in stratum-j.

$$W_{ijkl}^R = \frac{E_{ijk}}{r_{ijk}}$$

### 4.4 Post-Stratification Adjustments

We computed separate adjustments for private and public schools in each region. For Public School students, we adjusted by gender, race/ethnicity and grade level within each region. For Private School students, we adjusted by race/ethnicity and grade level within each region. Post-stratification adjustment factors are presented in Appendix A. As described next, Exhibits 2 and 3 present population control totals used for post-stratification for private and public schools.

Exhibit 2 provides population control totals used for post-stratification for private schools, and Exhibit 3 provides similar controls for public schools. For public schools, post-stratification dimensions included not only region and school level but also race/ethnicity (4 categories) and gender. For private school, post-strata were based on region, school level and race/ethnicity (2 categories). Control totals were obtained from the QED data augmented by data from the National Center for Education Statistics (NCES)—Common Core of Data (CCD) data for public schools and Private School Survey data for private schools.

**Exhibit 2 Control Totals for Private Schools: Population totals for each cell defined by region, school level, and race**

<i>Region</i>	<i>School Level</i>	<i>Race</i>	<i>Population Total (Enrollment)</i>
Balance of State	Middle School	Non-Whites	6479.42
Balance of State	Middle School	Whites	29057.34
Balance of State	High School	Non-Whites	9220.67
Balance of State	High School	Whites	40159.41
New York City	Middle School	Non-Whites	20043.16
New York City	Middle School	Whites	24571.86
New York City	High School	Non-Whites	21205.83
New York City	High School	Whites	31933.12

**Exhibit 3 Control Totals for Private Schools: Population totals for each cell defined by region, school level, race and gender**

<i>Region</i>	<i>Gender</i>	<i>School level</i>	<i>Race</i>	<i>Population total</i>
Balance of State	Female	Middle School	Black	23548.00
Balance of State	Female	Middle School	Hispanic	19462.62
Balance of State	Female	Middle School	Other	8296.68
Balance of State	Female	Middle School	White	146252.12
Balance of State	Female	High School	Black	31747.97
Balance of State	Female	High School	Hispanic	24672.38
Balance of State	Female	High School	Other	10668.25
Balance of State	Female	High School	White	207160.19
Balance of State	Male	Middle School	Black	24619.65
Balance of State	Male	Middle School	Hispanic	20457.74
Balance of State	Male	Middle School	Other	8721.00
Balance of State	Male	Middle School	White	152598.19
Balance of State	Male	High School	Black	32787.82
Balance of State	Male	High School	Hispanic	25905.40
Balance of State	Male	High School	Other	11143.85
Balance of State	Male	High School	White	214949.13
New York City	Female	Middle School	Black	31628.16
New York City	Female	Middle School	Hispanic	38522.85
New York City	Female	Middle School	Other	13968.08
New York City	Female	Middle School	White	13653.47
New York City	Female	High School	Black	47770.13
New York City	Female	High School	Hispanic	53779.80
New York City	Female	High School	Other	20570.37
New York City	Female	High School	White	18328.11
New York City	Male	Middle School	Black	32176.58

New York City	Male	Middle School	Hispanic	40050.36
New York City	Male	Middle School	Other	14780.68
New York City	Male	Middle School	White	13997.82
New York City	Male	High School	Black	47751.18
New York City	Male	High School	Hispanic	55181.07
New York City	Male	High School	Other	21657.94
New York City	Male	High School	White	18026.40

The demographic variables needed for post-stratification adjustments—grade level, gender and race/ethnicity—were imputed with the procedures described in Appendix B.

#### 4.5 Trimming

We used post-strata as trimming classes: region, school type, gender, grade and race.<sup>1</sup> This approach preserves the weight sum within post-strata that are desired to control and keep constant. The trimming approach limits the design effect due to unequal weighting by truncating out the largest weights (in this case, the top 5% of the weights), and re-distributing the weight within a cell so that the weight sum remains constant. Details of the trimming approach are provided in Appendix A.

---

<sup>1</sup> In the Balance of State (BOS) region, trimming cells were defined within a combined group of public and non-public schools. This approach, described in Appendix A, was used to attenuate weights in one BOS cell.

## **CHAPTER 5. DATA PROCESSING PROCEDURES**

The survey response data were cleaned and processed to produce final weighted edited data files with documentation. The editing procedures are described in Appendix C. Appendix D provides a summary description of created variables also included in the analytic data file.



# **Appendix A**

## **Post-stratification Adjustment Factors**

## **Appendix A**

### **Post-stratification Adjustment Factors**

Because post-stratification adjustment cells vary between public schools and private schools, the tables in this appendix are organized by school type. Tables A1 and A2 provide the post-stratification adjustment factors for public schools separately for the two regions (NYC and BOS). Tables A3 and A4 present the post-stratification adjustment factors for private schools separately for the two regions (NYC and BOS).

Because adjustments were too large for one cell—Private schools in the Balance of State (BOS) region—we implemented one trimming step. Then, the Balance of State region was post-stratified again to ensure that weights sum to control totals.

The trimming capped all weights at the 95<sup>th</sup> percentile; i.e., trimmed the most extreme 5% of the weight distribution overall. The capped weights were then (re)distributed across trimming cells so as to preserve the weight sums. Trimming cells mirrored the post-stratum cells with one exception: the BOS cells were collapsed by combining public and non-public schools. In other words, the BOS cells were defined by school level, gender and race (4 categories) just like in the NYC public school post-stratum.

Note that the BOS weights were post-stratified again following trimming to ensure that the weight sums equal the population enrollment totals in the public and non-public schools. The two sets of adjustment factors are shown in Tables A2 and A4.

**Table A1. Post-stratification Adjustment Factors for Public Schools: New York City**

	ADJUSTMENT FACTOR#1
<i>Middle School</i>	
Females	
Blacks	1.1969
Hispanics	0.7666
Others	0.5712
Whites	1.2120
Males	
Blacks	1.5296
Hispanics	1.0625
Others	0.7662
Whites	0.6857
<i>High School</i>	
Females	
Blacks	1.0852
Hispanics	0.9006
Others	0.5254
Whites	1.3243
Males	
Blacks	1.5167
Hispanics	1.4104
Others	0.9173
Whites	0.8945

**Table A2. Post-stratification Adjustment Factors for Public Schools: Balance of State**

	ADJUSTMENT FACTOR#1	ADJUSTMENT FACTOR#2
<i>Middle School</i>		
Females		
Blacks	1.2645	1.009
Hispanics	0.6451	1.001
Others	0.2857	1.001
Whites	0.7064	0.946
Males		
Blacks	1.2133	1.000
Hispanics	0.6334	1.000
Others	0.5318	1.000
Whites	1.0539	0.998
<i>High School</i>		
Females		
Blacks	1.1855	1.004
Hispanics	0.7047	1.001
Others	0.3593	1.002
Whites	0.7352	0.920
Males		
Blacks	1.0074	1.001
Hispanics	0.5771	1.000
Others	0.7178	1.002
Whites	1.1276	0.999

**Table A3. Post-stratification Adjustment Factors for Private Schools: New York City**

	ADJUSTMENT FACTOR
<i>Middle School</i>	
Non-Whites	0.7574
Whites	0.7554
<i>High School</i>	
Non-Whites	0.4032
Whites	1.5543

**Table A4. Post-stratification Adjustment Factors for Private Schools:  
New York City Balance of State**

	ADJUSTMENT FACTOR#1	ADJUSTMENT FACTOR#2
<i>Middle School</i>		
Non-Whites	0.2699	0.9441
Whites	11.0193	3.9081
<i>High School</i>		
Non-Whites	0.6220	0.9933
Whites	1.9248	1.0155

# **Appendix B**

## **Imputation of Weighting Variables**

## **Appendix B**

### **Imputation of Weighting Variables**

Several variables needed in the weighting process had some missing data, and needed to be imputed. Generally, imputation was performed with sequential hot-deck methods that use sorting variables and imputation of data from a neighboring record (donor). In all imputations—for gender, grade and race—the first sorting variable was School ID. This approach capitalizes on the similar characteristics of students close to each other in these variables (e.g., from a same school). In other words, such students are more similar on the average than the sample students as a whole.

For imputation of gender, a variable that was missing for only 3 records, the student records were sorted by grade as well as by School ID. For imputing grade (19 records missing), sorting by age within each school ensured that age data for neighboring records were used. For race/ethnicity, both grade and gender were used for sorting within each school.

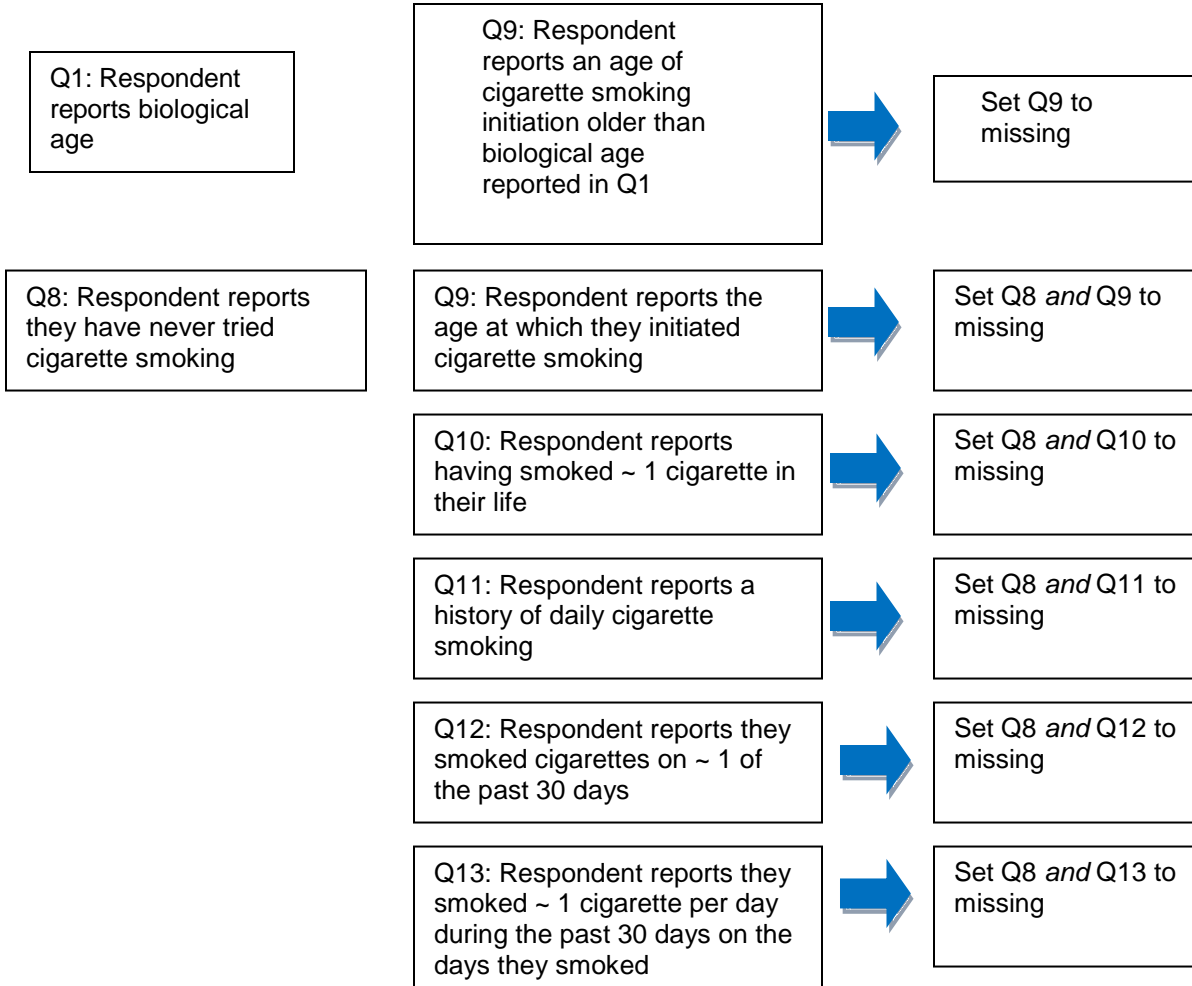
The imputation of race and ethnicity was based initially on the ethnicity variable (Q4) and on the single-race variable (Q6). When these were missing, and the multiple race variable (Q5) had to be used, we adopted the same hierarchical approach used in previous cycles. With this approach, a non-Hispanic respondent reporting multiple races is assigned to a single race in the following order: black, Asian, White, and Other.

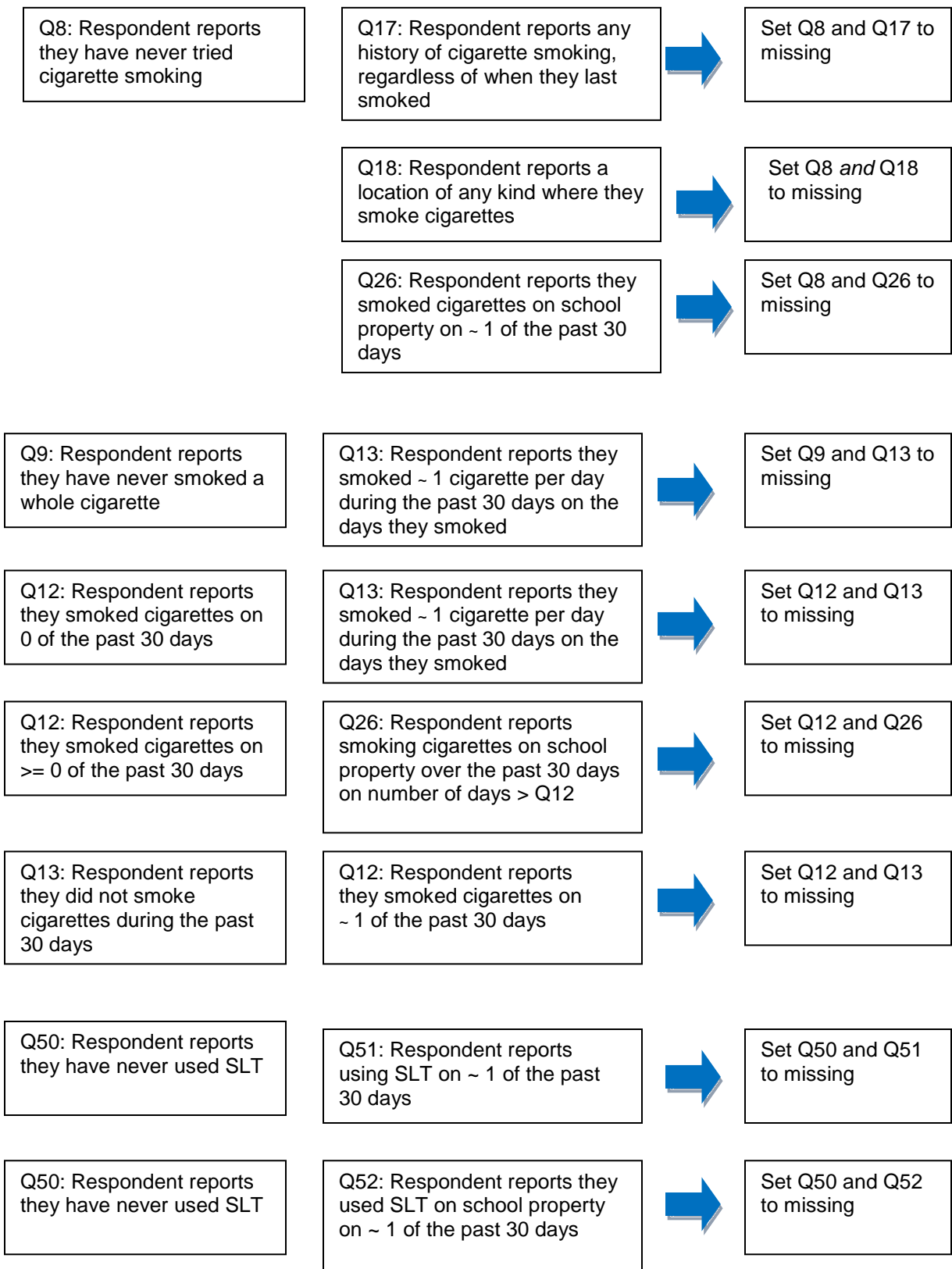
# **Appendix C**

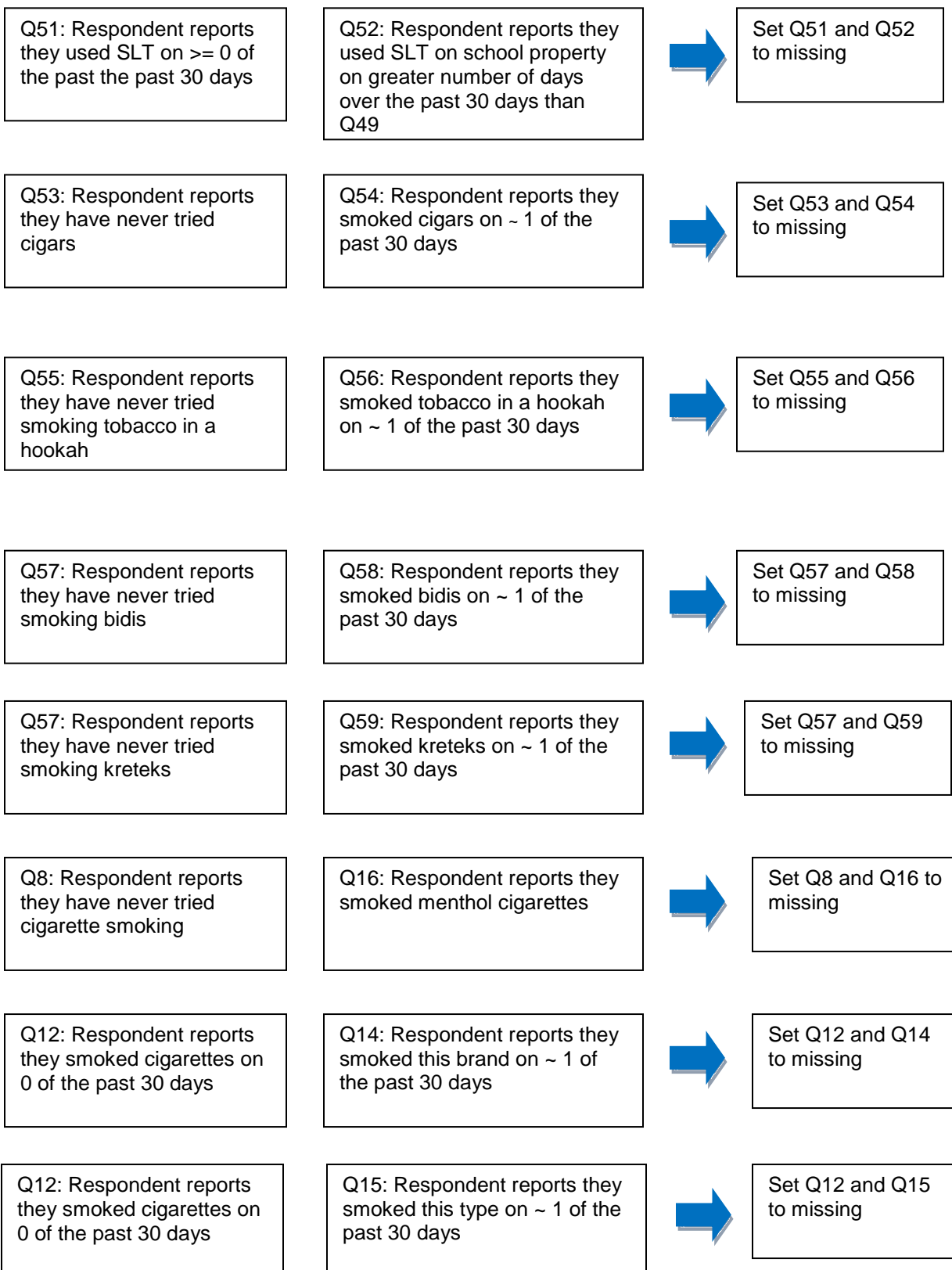
## **Edits Adopted in the 2010 NY YTS**



**Appendix C**  
**Edits Adopted in the 2010 NY YTS**







# **Appendix D**

## **Created Variables**

## Appendix D Created Variables

Variable Name: addicted

Variable Description: % addicted to cigarettes (100+ cigarettes in lifetime)

SAS Code: IF Qn10 in (8,9) THEN ADDICTED = 1; ELSE IF Qn10 NE . THEN ADDICTED = 2;

Variable Name: afcig

Variable Description: Age at which smoked first cigarette

SAS Code: IF Qn9 GE 2 THEN DO; AFCIG = Qn9; END;

Variable Name: cbidis

Variable Description: % current bidi smokers (smoked bidis 1+ days in the past 30 days)

SAS Code: IF Qn58 GE 2 THEN CBIDIS = 1; ELSE IF Qn58 NE . THEN CBIDIS = 2;

Variable Name: cfsk1

Variable Description: % never smokers who have 1+ close friends who smoke

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn65 GE 2 THEN CFSK1 = 1; ELSE IF Qn65 NE . THEN CFSK1 = 2; END;

Variable Name: cfsk2

Variable Description: % current smokers who have 1+ close friends who smoke

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn65 GE 2 THEN CFSK2 = 1; ELSE IF Qn65 NE . THEN CFSK2 = 2; END;

Variable Name: cigschp

Variable Description: % smoked cigarettes on school property 1+ days in the past 30 days

SAS Code: IF Qn26 GE 2 THEN CIGSCHP = 1; ELSE IF Qn26 NE . THEN CIGSCHP = 2;

Variable Name: krettek

Variable Description: % current kretek smokers (smoked kreteks 1+ days in the past 30 days)

SAS Code: IF Qn59 GE 2 THEN CKRETEK = 1; ELSE IF Qn59 NE . THEN CKRETEK = 2;

Variable Name: csdaily

Variable Description: % current smokers who ever smoked cigarettes daily

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn11 = 1 THEN CSDAILY = 1; ELSE IF Qn11 NE . THEN CSDAILY = 2; END;

Variable Name: cslt

Variable Description: % current smokeless tobacco users (used smokeless tobacco 1+ days in the past 30 days)

SAS Code: IF Qn51 GE 2 THEN CSLT = 1; ELSE IF Qn51 NE . THEN CSLT = 2;

Variable Name: csmoker

Variable Description: % current cigarette smokers (smoked cigarettes 1+ days in the past 30 days)

SAS Code: IF Qn12 GE 2 THEN CSMOKER = 1; ELSE IF Qn12 NE . THEN CSMOKER = 2;

Variable Name: cstore

Variable Description: % students who see ads for tobacco products when they go to a convenience store, supermarket, or gas station

SAS Code: IF Qn82 GE 2 THEN DO; IF Qn82 GE 2 AND Qn82 LE 4 THEN CSTORE = 1; ELSE IF Qn82 NE . THEN CSTORE = 2; END;

Variable Name: daily

Variable Description: % students who ever smoked cigarettes daily

SAS Code: IF Qn11 = 1 THEN DAILY = 1; ELSE IF Qn11 NE . THEN DAILY = 2;

Variable Name: defsk1

Variable Description: % never smokers who will definitely not smoke next year

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn67 = 4 THEN DEFSK1 = 1; ELSE IF Qn67 NE . THEN DEFSK1 = 2; END;

Variable Name: defsk2

Variable Description: % previous smokers who will definitely not smoke next year

SAS Code: IF SKSTATUS = 2 THEN DO; IF Qn67 = 4 THEN DEFSK2 = 1; ELSE IF Qn67 NE . THEN DEFSK2 = 2; END;

Variable Name: doch1

Variable Description: % never smokers who have had medical personnel talk to them about the danger of tobacco use

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn35 = 1 THEN DOCH1 = 1; ELSE IF Qn35 NE . THEN DOCH1 = 2; END;

Variable Name: doch2

Variable Description: % current smokers who have had medical personnel talk to them about the danger of tobacco use

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn35 = 1 THEN DOCH2 = 1; ELSE IF Qn35 NE . THEN DOCH2 = 2; END;

Variable Name: ebidis

Variable Description: % ever smoked a bidi

SAS Code: IF Qn57 = 1 or Qn57 = 3 THEN EBIDIS = 1; ELSE IF Qn57 NE . THEN EBIDIS = 2;

Variable Name: ekretek

Variable Description: % ever smoked a kretek

SAS Code: IF Qn57 = 2 or Qn57 = 3 THEN EKRETEK = 1; ELSE IF Qn57 NE . THEN EKRETEK = 2;

Variable Name: esdaily

Variable Description: % ever smokers who ever smoked cigarettes daily

SAS Code: IF ESMOKE = 1 THEN DO; IF Qn11 = 1 THEN ESDAILY = 1; ELSE IF Qn11 NE . THEN ESDAILY = 2; END;

Variable Name: eslt

Variable Description: % ever used smokeless tobacco

SAS Code: IF Qn50 = 1 THEN ESLT = 1; ELSE IF Qn50 NE . THEN ESLT = 2;

Variable Name: esmoke

Variable Description: % ever smoked cigarettes

SAS Code: IF Qn8 = 1 THEN ESMOKE = 1; ELSE IF Qn8 NE . THEN ESMOKE = 2;

Variable Name: fbidis

Variable Description: % frequent bidi smokers (smoked bidis 20+ days in the past 30 days)

SAS Code: IF Qn58 GE 6 THEN FBIDIS = 1; ELSE IF Qn58 NE . THEN FBIDIS = 2;

Variable Name: feelcig

Variable Description: Current smoker - experience craving if not smoking few hours

SAS Code: IF CSMOKER = 1 AND Qn41 GE 4 THEN DO; FEELCIG = Qn41; END;

Variable Name: fkretek

Variable Description: % frequent kretek smokers (smoked kreteks 20+ days in the past 30 days)

SAS Code: IF Qn59 GE 6 THEN FKRETEK = 1; ELSE IF Qn59 NE . THEN FKRETEK = 2;

Variable Name: fslt

Variable Description: % frequent smokeless tobacco users (used SLT 20+ days in the past 30 days)

SAS Code: IF Qn51 GE 6 THEN FSLT = 1; ELSE IF Qn51 NE . THEN FSLT = 2;

Variable Name: fsmoker

Variable Description: % frequent cigarette smokers (smoked cigarettes 20+ days in the past 30 days)

SAS Code: IF Qn12 GE 6 THEN FSMOKER = 1; ELSE IF Qn12 NE . THEN FSMOKER = 2;

Variable Name: ltsk

Variable Description: ever smokers - last time smoked

SAS Code: IF ESMOKE = 1 AND Qn17 GE 2 THEN DO; LTSK = Qn17; END;

Variable Name: ncigs

Variable Description: % current smokers who smoked 6+ cigarettes per day in the last 30 days

SAS Code: IF CSMOKER = 1 AND Qn13 GE 2 THEN DO; IF Qn13 GE 5 THEN NCIGS = 1; ELSE IF Qn13 NE . THEN NCIGS = 2; END;

Variable Name: newsmag

Variable Description: % (of those who read newspapers or magazines) who see ads for tobacco products in newspapers or magazines

SAS Code: IF Qn81 GE 2 THEN DO; IF Qn81 GE 2 AND Qn81 LE 4 THEN NEWSMAG = 1; ELSE IF Qn81 NE . THEN NEWSMAG = 2; END;

Variable Name: oskhome1

Variable Description: % never smokers who live with a smoker

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn64 = 1 THEN OSKHOME1 = 1; ELSE IF Qn64 NE . THEN OSKHOME1 = 2; END;

Variable Name: oskhome2

Variable Description: % current smokers who live with a smoker

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn64 = 1 THEN OSKHOME2 = 1; ELSE IF Qn64 NE . THEN OSKHOME2 = 2; END;

Variable Name: page

Variable Description: % current smokers who were asked to show proof of age when purchasing cigarettes

SAS Code: IF CSMOKER = 1 AND Qn23 GE 2 AND Qn1 ge 1 AND Qn1 LE 9 THEN DO; IF Qn23 = 2 THEN PAGE = 1; ELSE IF Qn23 NE . THEN PAGE = 2; END;

Variable Name: partalk1

Variable Description: % never smokers whose parents discussed dangers of tobacco

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn70 = 4 THEN PARTALK = 2; ELSE IF Qn70 NE . THEN PARTALK = 1; END;

Variable Name: pbought

Variable Description: Place bought last pack, current smoker

SAS Code: IF CSMOKER = 1 AND Qn20 GE 2 THEN DO; PBOUGHT = Qn20; END;

Variable Name: refuse

Variable Description: % current smokers who were refused a purchase of cigarettes because of their age

SAS Code: IF CSMOKER = 1 AND Qn25 GE 2 AND Qn1 GE 1 AND Qn1 LE 9 THEN DO; IF Qn25 = 2 THEN REFUSE = 1; ELSE IF Qn25 NE . THEN REFUSE = 2; END;

Variable Name: sk51

Variable Description: % never smokers who will definitely smoke 5 years from now

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn68 = 1 THEN SK51 = 1; ELSE IF Qn68 NE . THEN SK51 = 2; END;

Variable Name: sk52

Variable Description: % current smokers who will definitely smoke 5 years from now

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn68 = 1 THEN SK52 = 1; ELSE IF Qn68 NE . THEN SK52 = 2; END;

Variable Name: skbf1

Variable Description: % never smokers who would definitely not smoke if best friend offered them a cigarette

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn69 = 4 THEN SKBF1 = 1; ELSE IF Qn69 NE . THEN SKBF1 = 2; END;

Variable Name: skbf2

Variable Description: % previous smokers who would definitely not smoke if best friend offered them a cigarette

SAS Code: IF SKSTATUS = 2 THEN DO; IF Qn69 = 4 THEN SKBF2 = 1; ELSE IF Qn69 NE . THEN SKBF2 = 2; END;

Variable Name: skcar1

Variable Description: % never smokers who rode in a car with someone who was smoking cigarettes 1 or more days during the past 7 days

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn61 GE 2 THEN SKCAR1 = 1; ELSE IF Qn61 NE . THEN SKCAR1 = 2; END;

Variable Name: skcar2

Variable Description: % current smokers who rode in a car with someone who was smoking cigarettes 1 or more days during the past 7 days

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn61 GE 2 THEN SKCAR2 = 1; ELSE IF Qn61 NE . THEN SKCAR2 = 2; END;  
;

Variable Name: skoth1

Variable Description: % never smokers who think smoke from other people's cigarettes is harmful

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn63 GE 1 AND Qn63 LE 2 THEN SKOTH1 = 1; ELSE IF Qn63 NE . THEN SKOTH1 = 2; END;

Variable Name: skoth2

Variable Description: % current smokers who think smoke from other people's cigarettes is harmful

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn63 GE 1 AND Qn63 LE 2 THEN SKOTH2 = 1; ELSE IF Qn63 NE . THEN SKOTH2 = 2; END;

Variable Name: skroom1

Variable Description: % never smokers who were in the same room with someone who was smoking 1 or more days during the past 7 days

SAS Code: IF ESMOKE = 2 THEN DO; IF Qn60 GE 2 THEN SKROOM1 = 1; ELSE IF Qn60 NE . THEN SKROOM1 = 2; END;

Variable Name: skroom2

Variable Description: % current smokers who were in the same room with someone who was smoking 1 or more days during the past 7 days

SAS Code: IF CSMOKER = 1 THEN DO; IF Qn60 GE 2 THEN SKROOM2 = 1; ELSE IF Qn60 NE . THEN SKROOM2 = 2; END;

Variable Name: skstatus

Variable Description: current smoker \* previous smoker \* never smoker

SAS Code: IF ESMOKE = 1 THEN DO; IF CSMOKER = 1 THEN SKSTATUS = 1; IF CSMOKER = 2 THEN SKSTATUS = 2; end; IF ESMOKE = 2 THEN SKSTATUS = 3;

Variable Name: sltschp

Variable Description: % used smokeless tobacco on the school property 1+ days of past 30 days

SAS Code: IF Qn52 GE 2 THEN SLTSCHP = 1; ELSE IF Qn52 NE . THEN SLTSCHP = 2;

Variable Name: stayoff

Variable Description: Among those who tried to quit how long stayed off cigarettes

SAS Code: IF Qn32 GE 3 THEN DO; STAYOFF = Qn32; END;

Variable Name: cigtype

Variable Description: % current smokers who usually smoke full-regular flavor

SAS Code: IF CSMOKER = 1 AND Qn15 GE 3 THEN DO; IF Qn15 = 4 THEN TYPE = 1; ELSE IF Qn15 NE . THEN TYPE = 2; END;

Variable Name: usualcig

Variable Description: How current smoker usually get cigarettes

SAS Code: IF CSMOKER = 1 AND Qn19 GE 2 THEN DO; USUALCIG = Qn19; END;