

## **2002 NYTS Sample Design**

The 2002 NYTS was designed to support estimation of tobacco use, attitudes towards smoking, and exposure to the tobacco use of others among the national population of public and private school students in grades 6-12. The 2002 NYTS over-sampled African Americans, Hispanics, and Asians to allow separate analyses of these groups. Over-sampling of African American, Hispanic, and Asian students was facilitated by increasing the probabilities of selection for PSUs with high minority populations, and for schools with high minority enrollment. To adjust for varying probabilities of selection, including those resulting from nonresponse and over-sampling, a weighting factor was applied to each student record. Weights were post-stratified to national estimates of enrollment within the cross-classification of race/ethnicity and grade.

A partial panel design was used to select the sample of schools. The partial panel design consists of two independent samples of schools. The first sample of schools (drawn sample) was selected using a multi-stage sample design similar to that used in the 2000 NYTS and intended to be representative of all schools in the U.S. The second sample was selected from the list of schools that participated in the 2000 NYTS.

The sampling frame for the drawn sample consisted of all public and private schools in the U.S. All schools that participated in the 2000 NYTS formed the sampling frame for the 2002 panel school sample. Both sampling frames were stratified by Census region and level of exposure to the “truth” tobacco counter-marketing campaign. The first-stage sampling units, or Primary Sampling Units (PSUs) consist of large counties or groups of contiguous counties for the drawn sample. Within the school panel sample, PSUs consist of schools.

The sample was post-stratified by whether each school was selected from the drawn or panel samples. Post-stratification was used to establish further strata within the original strata in order to account for the partial panel design and improve estimates of variance. Initial post-stratification resulted in 6 strata that contained one PSU. Some statistical packages (e.g., STATA) require that each stratum contain at least two PSUs in order to estimate variance. Analysts that encounter difficulty estimating variances with STATA are encouraged to use an alternative set of stratum and PSU variables (PSEUDO\_STRAT and PSEUDO\_PSU) that have been recoded in order to allow variance estimation.

The pseudo survey design variables were created as follows: All but one of the 6 strata with one PSU occurred within the drawn sample since counties, a larger unit than schools, are the PSUs in the drawn sample. Schools were therefore designated as the PSUs in these cases, which is allowable under conventional statistical properties. One stratum (WEST1) from the panel sample contained a single PSU. Since the PSU is already designated at the lowest possible unit (schools) in the panel sample, this stratum was collapsed with another stratum within the same truth level and within the panel sample. To retain the most geographic integrity, this stratum was combined with MWEST1.

The accompanying Excel tables illustrate the original and pseudo strata, their values, and the type of PSU within each stratum. For example, stratum “Mwest1P” identifies observations drawn from the panel sample, in the Midwest, and in the lowest level of truth exposure (level 5 is the highest).