NORC at the UNIVERSITY of CHICAGO

MEMORANDUM

TO: Khadar Diria

FROM: Erin Tannenbaum, Michael Latterner, Felicia LeClere

DATE: March 31, 2016

RE: Survey Design, Weight, and Standard Error Harmonization for the National Youth Tobacco Survey

---

In this memo, we review issues of importance to the harmonization of weights and design parameters for the National Youth Tobacco Survey (NYTS). NORC is working closely with our partners at the Minnesota Population Center to harmonize the NYTS tobacco items across the years for which the data are available. Inference from these harmonized variables to assess trends in tobacco use across states and time also depend on a thorough understanding of how survey design and implementation has changed through the life of the survey series. We begin this memo by discussing the role of survey weights and design on inference and follow with a detailed description of how to pool standard errors for inference across time. The final step is a discussion of how changes in the survey weights and design impact inference from harmonized data. Users of the harmonized NYTS variables should be aware of the consequences of using pooled harmonized data in the absence of harmonized weights. We outline some of the potential sources of error associated with the changes in weighting approaches and design elements.

Failure to adjust weights, when analyzing survey results across time will result in inaccurate estimates due to biases caused by (but not limited to) 1) trend breaks nationally and/or by sub-populations, 2) response rate fluctuations, 3) mode effects, 4) non-coverage biases, 5) response biases, and 6) seasonality impacts. To properly adjust weights across all NYTS years would require a comprehensive weighting methodology to take the considerable sampling, questionnaire, data collection, and weighting changes into consideration. Such an endeavor is not within the scope of this memo, however, the building blocks towards this effort are provided.

NYTS Weighting Overview

Weights are provided for use when computing health risks for all NYTS data files, including estimates of tobacco usage. We briefly describe the weighting methodology used within the NYTS below.

For each sample year of the NYTS, weights are in place to make sure the sample is representative of the population of interest and that other objectives are met. If data are not weighted, each respondent is treated as though they had an equal probability of selection. Implicit in such use are the assumptions that each record has an equal probability of being selected and that non-coverage and nonresponse are equal among all segments of the population. When deviations from these assumptions are large enough to affect the results obtained from a data set, weighting each record appropriately can help to adjust for violations of these assumptions. An additional but conceptually unrelated reason for weighting is to make the total number of cases equal to a desired population which, for NYTS data, is the number of middle school or high school students in the 50 U.S. States and the District of Columbia.

The health characteristics estimated from the NYTS pertain to students in grade 6 through 12, who attend high school or middle school in either a public, private, or Catholic school in the United States. Students enrolled in an alternative school, special education school, Department of Defense operated school, or vocational school (that only serves students who are enrolled in another school) were excluded. Additionally, students who were unable to complete the questionnaire without special assistance were excluded.

NYTS sample selection and weighting documentation are available online from 2004 to 2014 through a Methodology Report published with each survey wave. The 1999, 2000, and 2002 documentation is available upon request from the Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion. The documentation for these earlier years although not published on-line, were provided to NORC and used for this memo.

A three stage stratified sample design was employed, for which the first stage is stratified by county, urban status, and/or region. The second stage stratifies by school, and the third stage selects classes within the school. A summary of changes detailed in the overview statements are included below in "Changes by Year."

Finally, the time-span of a long-lived survey such as the NYTS, means that changes are often made to the sample design components (strata and clusters) as well as to the way survey weights are calculated (including probabilities of selection, non-response adjustments and post-stratification adjustments). These changes in sample design and survey weights can impact both point estimates in a series (e.g., such as the proportion of middle school students who smokes over time) and the variance/standard errors of those estimates. Thus, focusing only on variable harmonization is not sufficient to guarantee the integrity of statistical inference. The impact of changes in study design and sampling methods can have as profound an impact on the analysis of time trends as changes in question wording and variable code frames.

Standard Errors across Years

This memo assumes the reader is familiar with standard errors as a statistical concept as well as calculation methods to produce standard errors for the NYTS for a single year. In this section, we describe the implications of survey design changes on standard error estimation from data from multiple years of a survey.

There are several valid ways to compute estimates from data that are combined across survey years. The adjustment of the weights and standard error calculations depends on the desired final inferential population. For the sake of presentation, the following formulas assume that sampling and weighting methodologies are similar across time. In theory, one could make a weight adjustment for each estimate to be computed and for each domain, or subpopulation of interest (i.e. an individual state or area within a state).

Assume that a researcher would like to combine years of survey results to improve precision. For example, suppose $\bar{x}$ is an average derived from n consecutive years' data, i.e. $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ where the x_i are the estimates for the individual years.

If we use an appropriate formula within a particular year's NYTS to estimate the standard error, $s_{\bar{x}}$, of each year's estimate, then the standard error of $\bar{x}$, $s_{\bar{x}}$ is

$$s_{\bar{x}} = \frac{s_{\bar{x}}}{n}$$

where

$$s_{\bar{x}} = \sqrt{\sum_{i=1}^{n} s_{x_i}^2 + 2r \sum_{i=1}^{n-1} s_{x_i} s_{x_{i+1}}}$$

The correlation between consecutive years, r, is approximately 0 as the NYTS is a cross-sectional survey. That is, the probability of selecting the same person in two consecutive years is rare. As such this reduces to

$$s_{\bar{x}} = \sqrt{\sum_{i=1}^{n} s_{x_i}^2}$$

This formulation assumes, however, that the standard errors are calculated similarly across all years of the survey.

Changes by Year

Pooling standard errors across years assumes that the weight construction and calculation are comparable from year to year. We examined the methods used to construct survey weights over time in the NYTS that might introduce changes to the estimates or standard error calculations. Weights represent the inverse probability of selection of a sample element and are also generally adjusted for differential non-response and coverage error using a process of post-stratification or iterative proportionate fitting, depending on the survey year.

An overview of changes that may impact estimates across time are presented at the bottom of this memo in the form of five figures: Figure 1 outlines increases to the nonresponse rate. Figures 2 and 3 outline changes in race and ethnicity oversampling across time. Figures 4 and 5 provide a detailed look at the changes made within sampling stage 3 across time. Figure 6 outlines the number of strata selected at each sampling stage across time. Lastly, Figure 7 outlines additional changes that may impact comparisons from year to year. The changes across these tables represent a myriad of methodological changes, some of which will directly impact the estimates from the NYTS across time. The analytic impact of these changes are substantial and the breath of their impact is beyond the scope of this document. The primary goal is to alert users to the potential for substantial inferential bias.

NORC recommends creating an "adjusted" set of harmonized weights that account for known issues that could influence key estimates of interest (e.g., smoking status). There are many examples in which underlying design issues can impact inference. For example, the NYTS used various methods to oversample minority populations. Additionally in 2002, the NYTS used a panel sample design and subsequently removed the panel design in the next survey wave (2004). The inclusion of identical school respondents was touted as a way to determine changes in smoking behaviors and attitudes between 2002 and 2006. Because race and ethnicity are related to smoking behavior, the

overall percentage of students who smoke may appear to have changed as an artifact of these adjustments alone. We recommend smoothing the introduction of large changes over a longer time to avoid any incorrect conclusions.

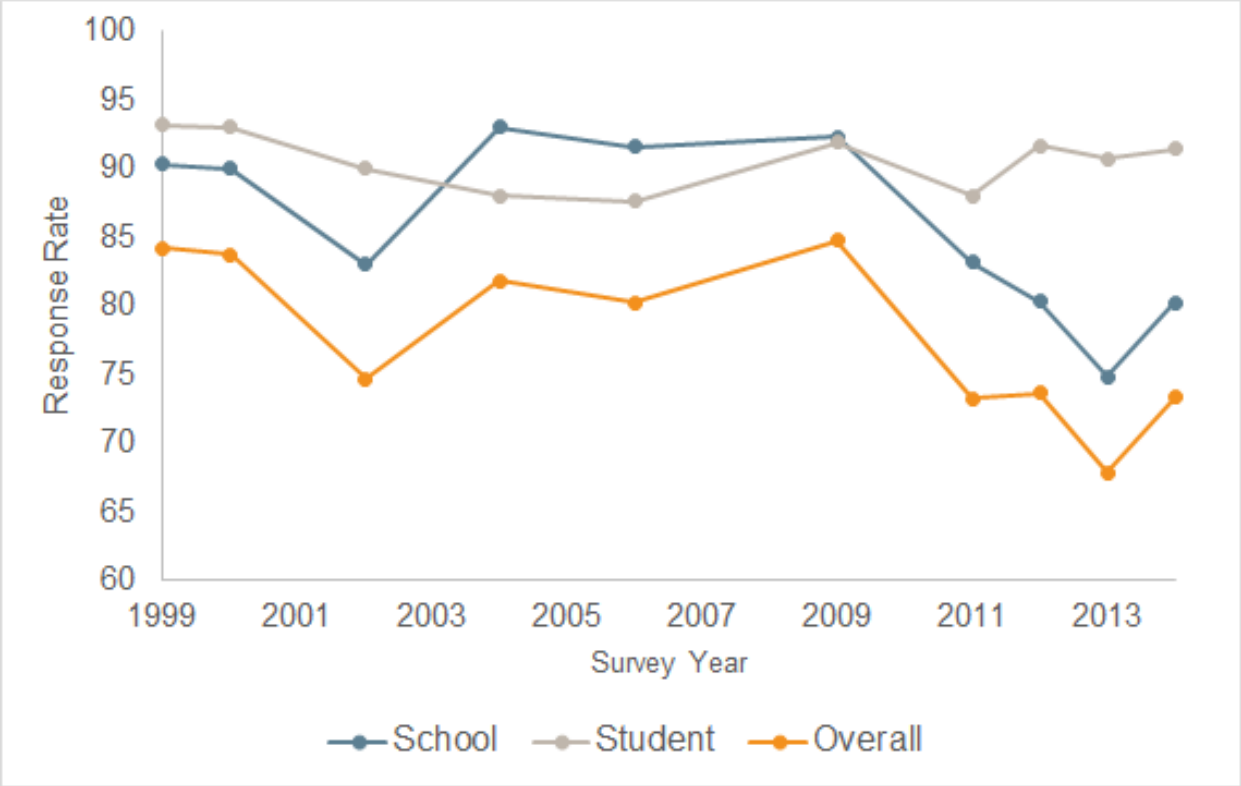
We also examined the sample methodology used over time for potential estimation issues. For example, the NYTS switched from administering surveys at the beginning of the school year (fall) to the end of the school year (spring) between 1999 and 2000. When the NYTS used a fall data collection, it was likely that fewer students had taken up smoking or tried smoking because they were as many as six months younger in age. We recommend further investigation of potential biases before the introduction of 1999 data and smoothing or adjusting the introduction of this change over a longer time.

Another major complication in harmonizing strata over time is that the Public Use Files of federal surveys with variables relevant to tobacco policy and tobacco control often have limited sampling strata identifiers. Sample design information such as PSUs and strata often provide detailed geographic identifiers that could be used with other information to re-identify respondents. Inference across Public Use Files over time will likely neglect important design changes that are not represented in the strata available to adjust standard errors for survey design effects.

Given the large number of changes across time, researchers should carefully examine the years of interest, determine all methodological changes across those years, test for substantive differences within the survey results, and adjust the estimates and weights as appropriate.

Figure 1. NYTS Response Rate by Year and Type

Note: Increases in nonresponse rates may cause nonresponse biases. These effects ideally should be measured and taken into consideration across survey waves.



Response Rate Type	Year									
	1999	2000	2002	2004	2006	2009	2011	2012	2013	2014
School	90.3	90	83	93	91.6	92.3	83.2	80.3	74.8	80.2
Student	93.2	93	90	88	87.6	91.9	88	91.7	90.7	91.4
Overall	84.2	83.7	74.7	81.8	80.2	84.8	73.2	73.6	67.8	73.3

Figure 2. NYTS Oversampling Strategy by Year

Note: The number of units per sampling stage vary by year. Thus, the probability of selection for a single school of the same size may change from sample wave to wave. Small changes in probability of selection may lead to trend breaks.

	Method Employed				
Year	**Minority Oversampled**	**Over-sampling PSUs**	**Weighted MOS**	**Unweighted MOS**	**Double Class Selection**
1999	Hispanic and Black				
2000	Hispanic, Black, and Asian/Pacific Islanders	Implicit	Yes		
2002	Hispanic, Black, and Asian/Pacific Islanders	Implicit			
2004	Hispanic, Black, and Asian	Implicit	Yes		
2006	Hispanic, Black, and Asian	Implicit			
2009	Hispanic and Black	Implicit	Yes		75% of the selected large schools, schools with the highest concentrations of Hispanics and black students for each level.
2011	Hispanic and Black	Implicit	Yes		
2012	None	Explicit		Tested	Large schools randomly selected.
2013	Black	Explicit		Yes	Any school with greater than 5 percent non-Hispanic black student enrollment.
2014	Black	Explicit		Yes	Large schools that had greater than 3 percent non-Hispanic black student enrollment.

Figure 3. NYTS Oversampling Factor by Race and Year

Below are equations provided by year within the weighted measurement of size which is a technique used to oversample minorities. Changes in sample design also impacts weights which likely impacts sub-populations and thus, breaks any potential trends within those groups.

Note: O = Other, H = Hispanic, B = Non-Hispanic Black, API = Asian, Pacific Islander, A = Asian

Year	Weighted MOS Factor				
	Full Equation	H	B	API	A
1999	1.0 O + 3.0 H + 2.0 B	3	2	[1]	NA
2000	1.0 O + 3.0 H + 2.0 B + 4.0 API	3	2	4	NA
2002	1.0 O + 10.0 H + 8.0 B + 16.0 API	10	8	16	[1]
2004	1.0 O + 9.0 H + 5.0 B + 16.0 A	9	5	NA	16
2006	1.0 O + 10.0 H + 5.0 B + 20.0 A	10	5	NA	20
2009	1.0 O + 2.0 H + 2.0 B	2	2	[1]	[1]
2011	1.0 O + 2.5 H + 3.0 B	2.5	3	[1]	[1]
2012	1.0 O + 1.0 H + 1.0 B	[1]	[1]	[1]	[1]
2013	[1.0 O + 1.0 H + 1.0 B]	[1]	[1]	[1]	[1]
2014	[1.0 O + 1.0 H + 1.0 B]	[1]	[1]	[1]	[1]

Figure 4. NYTS Stage Three Sampling Strategy Stratification Definitional Changes by Year

Note: School stratification proportions vary by year. Thus, the probability of selection for a single school of the same size may change from sample wave to wave. Small changes in probability of selection may lead to trend breaks.

Year	Percent of Schools Selected by Strata Type			
	Large	Small	Medium	Double
1999	92%	8%	NA	NA
2000	92%	8%	NA	NA
2002	93%	7%	NA	NA
2004	94%	6%	NA	NA
2006	95%	5%	NA	NA
2009	79%	11%	11%	NA
2011	85%	6%	9%	NA
2012	82%	8%	10%	NA
2013	9%	14%	NA	78%
2014	77%	14%	9%	NA

Figure 5. NYTS Stage Three Sampling Strategy Stratification Definitional Changes by Year

Note: School stratification definitions vary by year. Thus, the probability of selection for a single school of the same size may change from sample wave to wave. Small changes in probability of selection may lead to trend breaks.

Sample Design, School Designation as…	Year			
	1999 - 2006	2009 - 2012	2013	2014
Large	Able to sample 125 students	Schools with at least 50 students at each grade	Between 25 and 49 students per grade	Contained at least 56 students at each grade level per school
Medium	NA	More than 25 students but fewer than 50 students in one or more of the eligible grades for the level	NA	Contained between 28 and 55 students per grade per school
Small	Not able to sample 125 students	Schools with 25 or fewer students in one or more of the eligible grades for the level	Less than 25 students at any grade level	Contained less than 28 students per grade per school
Double	NA	NA	Contain at least 50 students at each grade level. Those selected for two classes had Non-Hispanic black enrollments over 5%	NA

Figure 6. NYTS Number of Sampling Strata by Year

Note: Number of units per sampling stage vary by year. Thus, the probability of selection for a single school of the same size may change from sample wave to wave. Small changes in probability of selection may lead to trend breaks.

Stage/ Sampling Unit	Year									
	1999	2000	2002	2004	2006	2009	2011	2012	2013	2014
1.PSU	66	165	100	91	91	80	82	100	85	93
2.School	145	360	298	289	289	200	194	244	219	220
3.Classes	5 or 6 classes	5 or 6 classes	5 or 6 classes	4 to 6 classes	4 to 6 classes	1 or 2 per grade[1]	1 or 2 per grade	1 or 2 per grade	1 or 2 per grade	1 or 2 per grade

[1] Two per grade in large high minority schools for 2009, 2011 - 2014

Figure 7. NYTS List of Substantial Changes by Year

Data Year	Description	Potential Impact
2000	Survey fielding period moved from the fall to the spring. First stage stratification change from degree of urbanization and percent of Blacks and Hispanics in the Primary Sampling Unit (PSU) to degree of urbanization and geographic location. Increased sample sizes for Texas California and New York. Race ethnicity questions changed. Weighting adjustments for other non-response adjustments changed. Weight trimming calculations altered. The threshold of weights trimmed changed from 10.0 to 3% of all weights. Weight post-stratification changes from post-stratifying to the nation to post-stratifying to Texas, California, New York, and the balance of the United States.	Rate of tobacco product usage likely increased due to changes in fielding period (students older). Changes in weights likely impact sub-populations and thus break any potential trends within those groups.
2002	Partial panel design for sample of schools. This results in substantial changes to the sampling methodology and weighting. Selection of "large" PSU's changed from selection with certainty to any PSUs larger than the sampling interval are split into a set of sub-PSUs such that the selection probability for the resulting PSU is less than 50%, with each resulting PSU having a roughly equal number of schools. Selection of PSUs changed from stratifying by degree of urbanization and geographic location to Census Region and levels of truth[sm] exposure. Selection of PSUs changed from implicit stratification by 3-digit ZIP to implicit stratification by 5-digit ZIP. Weighting changed in its entirety to include the panel survey respondents. Weights were not adjusted for national estimates before weight trimming but were instead post-stratified after weight trimming to national estimates of racial percentages by grade. Post-stratification weighting changed. Other and unknown combined with White race/ethnicity.	Changes in sampling methodology and weighting likely impact sub-populations and thus break any potential trends.

2004	Incentive changed from $1,000 per school to no mention of an incentive. Selection of PSUs changed from stratifying by Census Region and levels of truth[sm] exposure to stratifying by Census Region and MSA status. Selection of secondary sampling units (SSUs or schools) changed to include implicit stratification by first sorting the schools by highest grade, lowest grade, zip code and enrollment. Partial panel design retired. This resulted in substantial changes to the sampling methodology and weighting. NYTS oversight changed from American Legacy Foundation to CDC/OSH. Weighting changed in its entirety to exclude equations for the 2002 panel survey design. PSU Definition changed from large counties or groups of smaller adjacent counties to a county, or a group of small counties, or portions of large counties. Race/ethnicity questions changed. Post-stratification weighting changed. Other and unknown race/ethnicity first imputed using hot-deck methods. First student exclusions mentioned including those enrolled in: • regular schools who are unable to complete the questionnaire without special assistance, • alternative schools, • special education schools, • Department of Defense operated schools, • vocational schools that only serve students who are enrolled at another school.	School may choose to not participate due to the reduced incentive. A thorough bias study is needed to determine potential impacts. Changes in weights and race/ethnicity questions likely impact sub-populations and thus break any potential trends within those groups.
2006	Four Louisiana schools removed due to Hurricane Katrina emergency response. Post-stratification weighting changed. Categories changed to grade, as well as already used race/ethnicity categories. Selection of schools (SSUs) changed to include implicit stratification by first sorting the schools by grade, zip code and enrollment.	Deviations from the original sample design likely impacts trends. The Louisiana schools were not replaced and thus fewer southern schools were included in the sample for 2006. Changes in weights likely impact sub-populations and thus break any potential trends within those groups.
2009	Sampling and weighting realigned to be	Changes in sampling and

	consistent with Youth Risk Behavior Survey (YRBS). The samples were drawn to be non-overlapping samples of schools between surveys. Stratification: minority concentration and urban status. Selection of PSUs changed from stratifying by Census Region and MSA status to stratifying by non-Hispanic black or Hispanic composition and MSA status. Selection of PSUs changed from implicit stratification by 5-digit ZIP to implicit stratification by state and 5-digit ZIP. SSU definition changed to include "fragment" schools. SSU selection changed from school size to school level (middle schools, high schools), size, and proportion of minorities. Selection of students changes from clustering by class to clustering by grade and class. Weights changed to take fragment schools into consideration. Weight trimming: survey weights not adjusted to stratum population total before trimming, post stratification total used for population adjustments after trimming. Weights were not adjusted for national estimates before weight trimming but were instead post-stratified after weight trimming to national estimates of school type, sex, as well as already used grade and race/ethnicity categories.	weights likely impact sub-populations and thus break any potential trends within those groups.
2011	The Centers for Disease Control and Prevention, Office on Smoking and Health (OSH) and the Food and Drug Administration, Center for Tobacco Products (CTP) began collaborating together on the NYTS. NYTS is now fielded on an annual basis. Selection of PSUs changed from stratifying by Census Region and MSA status to stratifying by ethnic/racial composition and MSA status. School sampling weights formulas changed. Post-stratification: survey weights not adjusted to race/ethnicity whereas before it was adjusted only to race.	Changes in weights likely impact sub-populations and thus break any potential trends within those groups. Changes in questionnaire design may impact response rates as well as the quality of responses.

	The NYTS questionnaire will align to tasks assigned to OSH in odd years (e.g. 2011).	
2012	SSU selection changed from school level (middle schools, high schools), size, and proportion of minorities to school level and size only. Base weights, moved from conditional student weights to adjusted conditional student weights. Also, moved from cluster schools to linked schools. Post-stratification weights first calculated for middle school and high schools separately. The NYTS questionnaire will more closely align to tasks assigned to CTP on even years (e.g. 2012).	Changes in weights likely impact sub-populations and thus break any potential trends within those groups. Changes in questionnaire design may impact response rates as well as the quality of responses.
2013	The development and implementation was coordinated with the 2013 YRBS to avoid duplicative efforts and to minimize potential burden on school participants. Selection of PSUs changed from stratifying by ethnic/racial composition and MSA status to ethnic/racial composition, non-Hispanic black and Hispanic, and MSA status. Post-stratification race/ethnicity categories redefined.	Changes in weights likely impact sub-populations and thus break any potential trends within those groups. Changes in questionnaire design may impact response rates as well as the quality of responses.
2014	PSU selection changed by removing MOS component. SSU selection changed from probability proportional to weighted school enrollment with weighted MOS to probability proportional to eligible school enrollment. Frame source was merged with a secondary source resulting in a coverage increase among all public and non-public high schools of 23%, (15.5% increase public schools, 46% increase non-public high schools). Odd year, OSH alignment of questionnaire.	Changes in sample frame were linked to an increase in coverage which may in turn reduce coverage biases.

Additional literature sources:

Kennickell, A. B. (1999). Revisions to the SCF Weighting Methodology: Accounting for Race/Ethnicity and Homeownership. Federal Reserve Board, mimeo.