

A Simulation Study of Alternative Weighting Class Adjustments for Nonresponse when Estimating a Population Mean from Complex Sample Survey Data

Brady T. West¹

¹Michigan Program in Survey Methodology, Institute for Social Research,
Ann Arbor, MI

Abstract

Results are presented from a simulation study of alternative weighting class adjustments for nonresponse when estimating a population mean from complex sample survey data, in an effort to extend the previous work of Little and Vartivarian (2003, 2005) to a complex sample survey setting involving stratified cluster sampling from a finite target population. A total of 30 simulations were performed, varying based on five different parameters: 1) the relationship of an auxiliary variable X available for respondents and nonrespondents with the survey variable of interest Y in the population of interest; 2) the relationship of the auxiliary variable X with the probability of unit nonresponse, P ; 3) the use of base sampling weights according to a complex sample design when estimating response rates within weighting classes defined by X and/or the sampling strata (vs. no nonresponse adjustment at all); 4) the expected response rate for each sample across repeated sampling (75% or 25%); and 5) the relationship of the design strata with the probability of unit nonresponse, P . Each simulation examines the empirical bias and the empirical root mean squared error (RMSE) of a particular weighted estimator of the population mean for Y , with nonresponse adjustments to the base sampling weights computed in weighting classes defined by the auxiliary variable X and the sampling strata in selected simulations. Results from the simulations suggest that the use of weighted response rates within weighting classes defined by an auxiliary variable X and the sampling strata can be beneficial when working with survey data collected from stratified cluster samples, particularly when response rates are low and the auxiliary and stratum variables are correlated with both the survey variable of interest Y and response propensity.

Key Words: Complex Sample Survey Data, Unit Nonresponse, Nonresponse Weighting Adjustments, Missing Data

1. Introduction

Unit nonresponse, or the failure to obtain any survey measures on a sampled unit in a survey research project, is a critical breakdown in the survey measurement process and a ubiquitous problem facing survey research organizations and independent survey researchers worldwide. Unit nonresponse can arise from an inability to contact the sampled unit (non-contact) or a refusal on the part of the sampled unit to participate in the survey (Groves and Couper, 1998a), and recent work has shown that the rates of non-contacts and refusals to official government surveys are steadily increasing over time in a uniform manner worldwide (de Leeuw and de Heer, 2002). Although additional recent

empirical work has shown that an increase in the nonresponse rate of a survey does not necessarily imply that the nonresponse bias of estimates based on that survey will increase (e.g., Groves, 2006; Groves and Peytcheva, 2008), unit nonresponse engenders the potential for estimates based on the respondents to a survey to be biased, especially if non-respondents have distinctive values on the survey variables of interest. Estimates based on survey respondents will only be unbiased if unit nonresponse is arising completely at random, making the respondents a random sample of all sample subjects.

Although survey research organizations presumably take all possible steps to prevent unit nonresponse in the first place (Groves and Couper, 1998b), nonresponse error due to unit nonresponse is inevitable, and the disturbing worldwide trend of increasing survey nonresponse rates underscores the need for survey researchers to use statistically sound methods of post-survey adjustment for repairing nonresponse errors in surveys. A method commonly used in practice to adjust for nonresponse is the weighting class adjustment method. This method relies on a somewhat weaker assumption regarding the missing data mechanism relative to the missing completely at random (MCAR) assumption, or that unit non-respondents are missing at random (MAR). In this case, missing data arise as a function of other observed variables but not the survey variable being measured (e.g., a respondent is unable to recall an event due to older age, but not because of the event itself), and other observed variables are used to define weighting classes that are distinctive in terms of response propensity. The missing data within the classes are “ignorable” (Rubin, 1976) in that the respondents represent a random sample of all sample elements within the weighting class (i.e., the non-respondents within the class are only randomly different from the respondents). The weighting classes are generally formed based on auxiliary variables (possibly measured on the sampling frame) available for both respondents and non-respondents. Respondents within each weighting class have weights applied to their survey measures that are equal to their base sampling weight (according to a complex sample design) multiplied by the inverse of either the *unweighted* or the *weighted* response rate (if applicable, given unequal probability of selection arising from a complex sample design) within the weighting class, such that the respondents represent all of the non-respondents within each weighting class. The choice of using unweighted or weighted response rates when applying nonresponse adjustments within the weighting classes in the complex sample design setting is a central research question motivating the work in this paper.

Recent work has provided survey researchers with sound guidance regarding more appropriate applications of this relatively simple weighting class method. Little and Vartivarian (2003) empirically demonstrate that the use of weighted response rates within weighting classes is either incorrect (resulting in biased estimates if sample design variables are related to survey nonresponse) or unnecessary (if design variables are not related to survey nonresponse). They argue that adjustment classes should be formed based on variables related to both nonresponse and probability of selection (design variables, or variables related to differential probability of selection), and that *unweighted* nonresponse adjustments based on this weighting class method will reduce the bias and maintain the efficiency of weighted estimators computed using the survey data. However, the authors only consider the case of stratified sampling, where probabilities of selection differ based on two sampling strata. Further, Little and Vartivarian (2005) emphasize the need to choose auxiliary variables to form adjustment classes that are predictive of both response propensity and the outcome variable of interest for nonresponse bias reduction (and potentially reduction in the variance of estimates as well). In short, the authors demonstrate that choosing auxiliary variables

only predictive of response propensity and not predictive of the survey measures of interest will actually increase the variance of survey estimates without reducing the nonresponse bias, which is an unfavorable situation. However, their empirical results are based on simple random sampling, and only unweighted response rates are considered for developing the weighting class adjustments (p. 162).

As suggested by Little and Vartivarian (2005, p. 167), "...It would be of interest to see to what extent the results can be generalized to complex sample designs involving clustering and stratification." The present study aims to extend these results by using simulations to repeatedly select complex stratified cluster samples from simulated populations where survey variables of interest have known mean and covariance parameters (using stratified probability proportionate to estimated size, or PPeS, selection of clusters, to introduce unequal probability of selection). The simulations artificially introduce unit nonresponse in each sample according to a MAR mechanism, and then evaluate the alternative weighting class adjustments tested in Little and Vartivarian by applying the nonresponse adjustments to the base sampling weights reflecting unequal probability of selection according to the complex sample design. The potential results are extended by considering both unweighted and weighted response rates based on the stratified cluster sampling when performing the adjustments, and evaluating the utility of using weighted response rates when the weighting classes are formed based on auxiliary variables *and* design variables that are related to response propensity. Empirical properties of the various weighted estimators for a population mean (bias and RMSE) are presented to evaluate the alternative approaches (in addition to an approach ignoring adjustments for nonresponse) and provide additional guidance on appropriate weighting class adjustments for unit nonresponse in the complex sample design setting.

2. Methods

2.1 Overview

A total of 30 simulations were performed in this study, each involving 1,000 repeated random selections of 1,000 sample elements from a hypothetical population of elements with known mean and covariance parameters. In line with this study's goal of extending previous theoretical results in this area, a complex multistage sampling scheme was used to select the repeated samples of size 1,000 for each simulation. The first stage of sampling involved the PPeS selection of two primary sampling units (PSUs) from each of 50 strata defining the hypothetical population. The second stage of sampling involved the simple random sampling (without replacement) of 10 elements within each of the sampled PSUs. Twenty-four (24) of the simulations varied according to four parameters:

- The relationship of an auxiliary variable X , measured for both respondents and non-respondents, with the survey variable of interest Y in the hypothetical population of elements (high or low);
- The relationship of the auxiliary variable X with the probability P that the survey variable Y is not observed (assumed to be unit nonresponse) on a given sample element (high or low);
- The use of base sampling weights adjusting for unequal probability of selection into the sample according to the PPeS design when computing estimated response rates in weighting class adjustment cells defined by the auxiliary variable X (yes, no, or no adjustment for nonresponse);

- The expected overall response rate for the samples (75% or 25% in expectation across repeated samples).

An additional six (6) simulations were performed in the setting where the auxiliary variable X had a strong relationship with both the survey variable Y and the probability of unit nonresponse P , given that this setting has the maximum potential for reduction of both bias and variance in the estimator of a mean through the nonresponse adjustments (Little and Vartivarian, 2005). In these simulations, response propensity varied as a function of both the auxiliary variable X and the strata used to select the complex sample, and the weighting classes were formed based on both X and groups defined by the sampling strata. These simulations were performed in an attempt to extend the work of Little and Vartivarian (2003) to a setting involving cluster sampling within strata.

All simulations were programmed and performed using the SAS software (Version 9.1.3), and specifically the SURVEYSELECT and SURVEYMEANS procedures. SAS code used to implement the simulations is available from the author upon request.

2.2 Definitions of the Hypothetical Populations

Two hypothetical populations of elements were artificially constructed for the simulations. Both populations were defined by the exact same number of strata (50), and the exact same number of PSUs within each stratum. Each stratum had a randomly determined number of PSUs defined, with the number of PSUs in each stratum ranging from 2 to 51 (each population had the same number of PSUs within a given stratum; for example, in Stratum 1, there were 11 PSUs in both populations). Each population was defined to have exactly 76,918 elements. The number of population elements within each PSU was also randomly determined, such that there were at least 10 elements within each PSU. The number of population elements within the PSUs ranged from 10 to 109.

Values for the population elements on the survey variable of interest Y were simulated from a superpopulation model according to the following two-step process. In the first step, values on the auxiliary variable X for the population elements were randomly selected from a standard normal distribution with mean 0 and variance 1. In the second step, values on the survey variable of interest Y were computed based on one of two superpopulation models. In the first population, the superpopulation model defined a relationship between the auxiliary variable X and the survey variable Y :

$$Y_{ijk} = \beta_0 + 3X_{ijk} + S_k + u_{jk} + \varepsilon_{ijk} \quad (1)$$

In this notation, the index i refers to an element, the index j refers to a PSU, and the index k refers to a sampling stratum. The S_k term is a fixed effect associated with a given stratum, equal to the integer code for the stratum (1, 2, ..., 50) divided by 25. These fixed stratum effects were introduced to simulate the variance between strata that might be observed in a real-world setting, where different strata (by design) tend to be homogeneous within and heterogeneous between in terms of the survey variable of interest (Y). The u_{jk} term is a random variable representing the random effect of PSU j nested within stratum k , and the ε_{ijk} term is a random variable representing random error associated with the observation of Y for the i -th element in the j -th PSU within the k -th stratum. The values of the random PSU effects within a stratum and the random errors were randomly selected from the following distributions:

$$\begin{aligned} u_{jk} &\sim N(0, \sigma_{PSU}^2) \\ \varepsilon_{ijk} &\sim N(0, \sigma^2) \end{aligned} \quad (2)$$

In the superpopulation model used to define values for the first population, the variance of the random PSU effects within a stratum (σ_{PSU}^2) was defined to be 0.1, and the variance of the random measurement errors (σ^2) was defined to be 1 (corresponding to a marginal intra-PSU correlation of approximately 0.09 for the Y values). Values of Y were computed for each element in the population after a value of X had first been randomly selected and values of u_{jk} and ε_{ijk} had been randomly (and independently) selected. Fixing the intercept parameter (β_0) to be 10, the resulting population of values on the Y variable had a mean equal to 11.1286795, and a standard deviation equal to 3.2368377.

In the second population, values of Y were computed using the same two-step process, only using a superpopulation model where X did not have a relationship with Y :

$$Y_{ijk} = \beta_0 + S_k + u_{jk} + \varepsilon_{ijk} \quad (3)$$

The variance of the random PSU effects was once again defined to be 0.1 in this model. To ensure that the marginal univariate distribution of Y was identical to that of the first population (mean equal to 11.1286795, standard deviation equal to 3.2368377), the variance of the random measurement errors (σ^2) was defined to be 10.014115. This increase was necessary to introduce the variance in Y that was explained by the predictor variable X in the first population (approximately 89%). Further, the intercept parameter was fixed to be 9.9930072 so that the mean of the Y values would be equal to 11.1286795 for this population as well. This approach is similar to that used by Little and Vartivarian (2005) in their simulation study, only extended to include complex design features that might be encountered in real-world survey research.

2.3 Complex Sample Design

Each simulation in this study involved 1,000 repeated selections of probability samples of size $n = 1,000$ from one of the two hypothetical populations described above. The probability samples were each selected according to a complex multistage sample design. In the first stage of selection, two PSUs were randomly selected from each of the 50 design strata with probability proportionate to estimated size (PPeS). Because the population sizes of each PSU were known in the two hypothetical populations, a simpler probability proportionate to size (PPS) design (in combination with proportionate allocation of the sample of size $n = 1,000$ across the strata) could have been used to achieve an equal probability of selection mechanism (*epsem*) design. In an effort to simulate more realistic sample design settings, where PSU sizes are often estimated using more readily available measures of size (MOS) for the PSUs (e.g., MOS for U.S. counties based on Census 2000 data), a small amount of random noise was added to each true population size for the PSUs. Specifically, a random value was sampled from the $N(0,1)$ distribution, multiplied by 5, rounded, and then added to the true population size of each PSU to compute an estimated MOS. Two PSUs were then selected from each stratum with probability proportionate to the estimated size represented by this MOS.

In the second stage of selection, it was assumed that once a PSU was randomly sampled within a stratum at the first stage, the true size of the PSU would be available. Exactly 10 population elements were randomly selected from each PSU sampled at the first stage,

using simple random sampling without replacement (an equal allocation design, with 20 sample elements allocated to each of the 50 strata). The second stage probabilities of selection for the population elements within each PSU were computed based on the *true* population sizes of each PSU assumed to be available at the second stage. For example, if there were actually 50 population elements in a given PSU, the probability of selection for elements within that PSU was $10/50 = 0.2$. This design was used to introduce unequal probabilities of selection for sample elements across the different PSUs. Based on this two-stage sample design, the probability of selection for elements from randomly sampled PSU j within stratum k can be defined as follows:

$$f_{jk} = \frac{2MOS_j}{\sum_{j \in k} MOS_j} \times \frac{10}{B_j} \quad (4)$$

In this notation, B_j refers to the true population size of PSU j . The base sampling weight w_{ijk} for each of the 1,000 individual sample elements was then computed as the inverse of this probability of selection:

$$w_{ijk} = \frac{1}{f_{jk}} \quad (5)$$

2.4 Introduction of Missing Data

In the first 24 simulations, the sample elements were classified into deciles based on their values on the variable X after the selection of each sample. Values on the variable X were assumed in this study to be available for all 1,000 sampled elements regardless of response to the survey, meaning that X can be thought of as an auxiliary variable available on the sampling frame that is measured for the entire population (e.g., Bethlehem, 2002). Then, depending on the missing data parameters defining the simulation being studied (expected sample response rate: 75% or 25%; association of X with the probability of missing data on Y : high or low), missing data were artificially introduced on the survey variable of interest Y within each of the 10 deciles according to a missing at random (MAR) mechanism.

In each simulated sample, each of the 1,000 sample elements first had a value U randomly sampled from the UNIFORM(0,1) distribution. The 100 sample elements in each of the 10 decile groups for X would therefore have a random sample of values for U , sampled from this uniform distribution. Then, depending on the expected sample response rate for the simulation (75% or 25%) and the association of X with the probability of having missing data on Y for the simulation (High or Low), missing values were assigned on the variable Y in each decile of X according to the rules in Table 1.

Table 1: Rules defining observation of Y for sample elements in the simulations (simulations defined by expected overall response rates of 75% or 25%, and associations of X with the probability of unit nonresponse P , denoted by “High” or “Low”). U refers to a random variable sampled from the UNIFORM(0,1) distribution.

Decile of X	75% Response Rate		25% Response Rate	
	High $X \rightarrow P^*$	Low $X \rightarrow P$	High $X \rightarrow P$	Low $X \rightarrow P$
1 (Low)	$U > 0.475$	$U > 0.250$	$U > 0.975$	$U > 0.750$

2	$U > 0.425$	$U > 0.250$	$U > 0.925$	$U > 0.750$
3	$U > 0.375$	$U > 0.250$	$U > 0.875$	$U > 0.750$
4	$U > 0.325$	$U > 0.250$	$U > 0.825$	$U > 0.750$
5	$U > 0.275$	$U > 0.250$	$U > 0.775$	$U > 0.750$
6	$U > 0.225$	$U > 0.250$	$U > 0.725$	$U > 0.750$
7	$U > 0.175$	$U > 0.250$	$U > 0.675$	$U > 0.750$
8	$U > 0.125$	$U > 0.250$	$U > 0.625$	$U > 0.750$
9	$U > 0.075$	$U > 0.250$	$U > 0.575$	$U > 0.750$
10 (High)	$U > 0.025$	$U > 0.250$	$U > 0.525$	$U > 0.750$
Number of Respondents				
Mean**	750.17	749.94	249.55	250.20
SD**	13.06	13.91	13.02	13.59

* P = Probability of unit nonresponse.

** Computed based on 4,000 samples selected under the specified simulation conditions (four simulated sets of 1,000 samples of size $n = 1,000$ under the conditions specified in Table 1, varying based on the relationship of X with Y and whether or not estimated response rates within the deciles were weighted).

An indicator variable R_{ih} was also defined according to this process, equal to 1 if a sample element i within weighting class h responded on Y and 0 otherwise; this variable was used for the various weighting class adjustments (see *Weighting Class Adjustments*). Note that this process of artificially introducing missing data by first randomly generating a sample of 1,000 values from a UNIFORM(0,1) distribution introduces random variance in the response rates around the overall expectations for the sample. For example, in the “Low” settings for the 25% Response Rate condition above, response rates in the 10 deciles will randomly vary around 25%, with the overall expectation across repeated sampling being 25%. This introduces some slight variance between the deciles in terms of the response rates in the “Low” condition, and also introduces slight variance in the response rates for each sample around the overall expectation, reflecting expected real-world variance in response rates across hypothetical repeated sampling.

In the remaining six simulations, missing data were introduced as a function of the sample design strata as well as the auxiliary variable X in a similar manner. These simulations only considered the setting where the auxiliary variable X had a strong relationship with both the survey variable of interest Y and the probability of unit nonresponse P , given that this setting has the strongest potential for reduction of both bias and variance in the estimator of the mean. These additional simulations allowed for an examination of the behavior of the alternative estimators when the sampling strata were associated with the probability of selection (by definition), the survey variable of interest Y (via the superpopulation model), and the probability of unit nonresponse P . To maintain the same number of weighting classes (10) given the samples of size 1,000, the weighting classes were defined by combinations of two groups based on the median of X and five groups of design strata (strata 1 through 10, strata 11 through 20, etc.). The 10 resulting groups were therefore distinct in terms of values on Y and response propensity, and this was verified by fitting logistic regression models to the resulting response indicator with the group based on X and the stratum group as two independent categorical predictors.

2.5 Weighting Class Adjustments

In 20 of the 30 simulations, the base sampling weights w_{ijk} for cases responding on the survey variable of interest Y were adjusted to account for the nonresponse of other

sampled cases in their respective weighting classes defined by the auxiliary variable X and the sampling strata. These adjustments to the base sampling weights in a given weighting class were defined in one of two ways, depending on the simulation being considered. Weighted estimates of the response rates within a given weighting class h were computed as follows, using the response indicator variable R_{ih} defined above:

$$r_{2h,weighted} = \frac{\sum_{i \in h} w_{1ijk} R_{ih}}{\sum_{i \in h} w_{1ijk}} \quad (6)$$

Unweighted estimates of the response rates within a given weighting class h were computed as follows:

$$r_{2h,unweighted} = \frac{\sum_{i \in h} R_{ih}}{n_h} \quad (7)$$

Then, given the 10 estimated response rates for each weighting class based on (6) or (7), depending on the simulation, the base sampling weights for the sampled cases in a given weighting class h were adjusted as follows, to compute a final analysis weight w_{ijk} for each respondent:

$$w_{ijk} = w_{1ijk} \times \frac{1}{r_{2h,weighted}} \quad (8)$$

or

$$w_{ijk} = w_{1ijk} \times \frac{1}{r_{2h,unweighted}}$$

These final sampling weights were then used to compute weighted estimates of the population mean based on the data collected for the respondents:

$$\bar{y}_w = \frac{\sum_k \sum_{j \in k} \sum_{i \in j} w_{ijk} y_{ijk}}{\sum_k \sum_{j \in k} \sum_{i \in j} w_{ijk}} \quad (9)$$

In 10 of the simulations, no adjustments for nonresponse were applied to the base sampling weights, and weighted estimates of the population mean were computed using the base sampling weights for the respondents only:

$$\bar{y}_0 = \frac{\sum_k \sum_{j \in k} \sum_{i \in j} w_{1ijk} y_{ijk}}{\sum_k \sum_{j \in k} \sum_{i \in j} w_{1ijk}} \quad (10)$$

Given prior knowledge of the true population means, the empirical properties of these alternative weighted estimators were evaluated based on the simulations by computing the empirical bias, variance, and root mean squared error for each of the estimators as described in the following section.

2.6 Computation of Empirical Root Mean Squared Error (RMSE)

Each simulation involved the selection of 1,000 complex samples of size $n = 1,000$ from one of the hypothetical populations described earlier. The empirical root mean squared error (RMSE) for a weighted estimator of the known population mean defined by one of the 30 simulations (denoted here using the general notation of \bar{y}) was computed as follows, based on the 1,000 samples (where t represents a sample index):

$$BIAS = E(\bar{y}) - \bar{Y} = \frac{\sum_t \bar{y}_t}{1000} - \bar{Y}$$

$$VARIANCE = E[\bar{y} - E(\bar{y})]^2 = \frac{\sum_t [\bar{y}_t - E(\bar{y})]^2}{1000} \quad (11)$$

$$MSE = E[\bar{y} - \bar{Y}]^2 = BIAS^2 + VARIANCE$$

$$RMSE = \sqrt{MSE}$$

Results presenting empirical estimates of the bias and the RMSE of each estimator based on the parameters defining each of the 30 simulations follow.

3. Results

Table 2 presents results from the 12 simulations where the expected response rate for the samples was 75% and the probability of unit nonresponse P was independent of the sample design strata.

Table 2: Simulation results, with expected sample response rate = 75%. Results in this table are based on 12 simulations, each defined by a particular combination of the three parameters defining the table. In each simulation, 1,000 complex samples of size $n = 1,000$ were selected from a population with properties defined by the far left column of the table.

Association of Y with X	Weighting Class Adjustments →	Base Weights Only (None)		Unweighted Response Rates		Weighted Response Rates	
	Association of P with X	Emp. Bias	Emp. RMSE	Emp. Bias	Emp. RMSE	Emp. Bias	Emp. RMSE
High	High	0.5570	0.5747	-0.0060	0.1319	0.0024	0.1279
High	Low	0.0012	0.1469	0.0040	0.1256	0.0024	0.1283
Low	High	0.0036	0.1471	-0.0046	0.1390	0.0021	0.1455
Low	Low	0.0093	0.1448	-0.0073	0.1452	-0.0003	0.1456

The results in the first row of Table 2 clearly indicate the substantial reductions in bias and variance for the estimator of a population mean that are possible when the auxiliary variable used to define weighting classes for nonresponse adjustments is strongly related

to both the survey variable of interest and response propensity, supporting results reported by Little and Vartivarian (2005). These results also provide support for the “common cause model” described by Groves (2006), where a common correlate of both the survey variable of interest and response propensity can introduce a substantial covariance between the survey variable and response propensity, leading to increased nonresponse bias. A failure on the part of a survey organization or an analyst to consider a weighting class adjustment that corrects for this type of nonresponse error will result in an estimator with substantially increased bias and variance. This is clearly illustrated by the simulation results in the first row for the weighted estimator without any nonresponse adjustments applied to the base sampling weights.

Further, the results in Table 2 suggest that the use of weighted response rates for computing the weighting class adjustments for nonresponse can reduce the bias and variance of the estimator even further, in the setting where the weighting classes are correlated with both the survey variable of interest and response propensity. The additional reductions in bias and variance are not substantial in the first row of Table 2, but these results are replicated below for the simulations with lower expected response rates. In addition, as reported by Little and Vartivarian (2005), the use of weighting classes related to the survey variable of interest but not response propensity (the results in the second row of Table 2) does not tend to reduce bias, but does reduce the variance of the estimator. The results in the final two rows suggest that the use of weighting classes unrelated to the survey variable of interest is not entirely beneficial in terms of reductions in bias or variance in this high response rate setting.

Table 3 presents results from the 12 simulations where the expected response rate for the samples was 25% and the probability of unit nonresponse was independent of the sample design strata.

Table 3: Simulation results, with expected sample response rate = 25%. Results in this table are based on 12 simulations, each defined by a particular combination of the three parameters defining the table. In each simulation, 1,000 complex samples of size $n = 1,000$ were selected from a population with properties defined by the far left column of the table.

Association of Y with X	Weighting Class Adjustments → Association of P with X	Base Weights Only (None)		Unweighted Response Rates		Weighted Response Rates	
		Emp. Bias	Emp. RMSE	Emp. Bias	Emp. RMSE	Emp. Bias	Emp. RMSE
High	High	1.6619	1.6765	0.0535	0.3650	0.0341	0.2483
High	Low	0.0003	0.2531	0.0062	0.1873	-0.0011	0.1518
Low	High	0.0096	0.2412	-0.0030	0.3580	-0.0071	0.3654
Low	Low	-0.0026	0.2326	0.0048	0.2412	-0.0006	0.2449

The results in Table 3 deserve special consideration, given the declining trends in survey response rates worldwide and recently published evidence of consistently lower response rates in web surveys relative to other survey data collection modes (Manfreda et al., 2008). Of primary note is the tremendous reduction in both the bias and the variance of the weighted estimator once again found when the weighting classes are related to both the survey variable of interest and response propensity, and the additional reduction in both bias and variance possible when using weighted response rates within the weighting

classes for the nonresponse adjustments. The use of weighted response rates for the nonresponse adjustments appears to once again reduce the variance of the estimators in both cases where the weighting classes are related to the survey variable of interest; in this setting, with an expected 25% response rate, the estimators with base weights adjusted by weighted estimates of the response rates within the weighting classes have the lowest RMSE.

Readers should also note in this lower response rate setting that the use of weighting classes related to response propensity but not the survey variable of interest (row 3 of Table 3) can actually increase the variance of the estimators, a result that was noted by Little and Vartivarian (2005). Finally, the use of weighting classes unrelated to both the survey variable of interest and response propensity provides little benefit, but also does not significantly harm the bias or variance properties of the estimator.

Collectively, the results of the simulations in Tables 2 and 3 provide support for the emerging awareness in the survey nonresponse literature that lower response rates do not necessarily correlate with increased nonresponse bias. Weighted estimators of the population mean with no adjustments applied to the base sampling weights for nonresponse have very little bias in all settings except for the case where an auxiliary variable is correlated with the survey variable of interest and response propensity (i.e., the “common cause model” setting where response propensity and the survey variable of interest are correlated). For all other cases, these results demonstrate that the precision of the weighted estimators can be improved by choosing weighting classes associated with the survey variable of interest, and by using weighted estimates of the response rates within the weighting classes for the nonresponse adjustments.

Table 4 presents results from the six simulations where the auxiliary variable X had a strong association with both Y and P , and the probability of unit nonresponse P was defined to be a function of the sample design strata in addition to the auxiliary variable X .

Table 4: Simulation results, where X had a strong association with both Y and P , and the probability of unit nonresponse P was defined to be a function of the sample design strata in addition to the auxiliary variable X . Results in this table are based on 6 simulations, each defined by a particular combination of the two parameters defining the table (application of nonresponse adjustments and expected response rate). In each simulation, 1,000 complex samples of size $n = 1,000$ were selected from the population where X had a ‘strong’ relationship with Y .

Weighting Class Adjustments →	Base Weights Only (None)		Unweighted Response Rates		Weighted Response Rates	
	Emp. Bias	Emp. RMSE	Emp. Bias	Emp. RMSE	Emp. Bias	Emp. RMSE
Expected Response Rate						
75%	0.1828	0.2311	0.0048	0.1303	-0.0063	0.1329
25%	0.4958	0.5475	0.0150	0.2690	0.0221	0.2321

The results in Table 4 suggest that the use of weighted response rates in this setting may result in more bias than the use of unweighted response rates, confirming the findings of Little and Vartivarian (2003). However, there was evidence that the use of weighted response rates may lead to an additional reduction in the variance of the estimator, in addition to the overall RMSE. In the setting of a 25% expected response rate, the RMSE was substantially reduced when applying the weighting class adjustments for

nonresponse, and even more so when using weighted response rates within the weighting classes. The empirical RMSE values for estimators using weighted and unweighted response rates were roughly similar in the 75% expected response rate setting.

4. Discussion

This simulation study sought to extend previous work regarding optimal weighting class adjustments for repair of nonresponse errors associated with sample estimates of population means. Specifically, the study considered extensions of previous work in this area to the case of complex sample designs, involving stratified cluster sampling of elements from a finite target population of interest. The results support findings previously reported by Little and Vartivarian (2005): when defining weighting classes for post-survey nonresponse adjustment, the most important auxiliary variables measured for respondents and non-respondents are those that are related to the survey variable of interest (for which a mean is being estimated). More importantly, the results of the simulations suggest that the use of weighted response rates within the weighting classes used to form nonresponse adjustments can offer improvements in terms of the precision and bias of estimators of population means in the complex sample setting (relative to the use of unweighted response rates), provided that the weighting classes are defined by auxiliary variables related to both the survey variable of interest and response propensity.

These results are only partially consistent with those presented by Little and Vartivarian (2003), who only considered a stratified sample design with no cluster sampling. In the first set of simulations performed in this study, the probability of responding (or providing a value on the survey variable of interest Y) was by design unrelated to the design variables (the strata and PSUs), and Little and Vartivarian suggested that the use of weighted response rates for nonresponse adjustments was unnecessary in this case. The results presented in this study suggest that the use of weighted response rates within adjustment cells may be beneficial when stratified cluster samples are selected and the adjustment cells are defined based on auxiliary variables correlated with both the survey variable of interest and response propensity. The largest benefits in terms of reduced bias and increased precision were found in the case of low response rates (25%), which are unfortunately becoming more commonplace in survey research.

This study also found that when weighting classes are defined by auxiliary variables and sample design variables related to both the survey variable of interest Y and response propensity, the use of weighted response rates for computing nonresponse adjustments within the weighting classes resulted in slightly increased bias of the estimators (relative to the use of unweighted response rates). These findings are consistent with those reported by Little and Vartivarian (2003), but this study also found evidence of an additional reduction in the overall RMSE of a weighted estimator in this setting when the expected response rate is low (25%). Collectively, these results suggest that the use of weighted response rates to define nonresponse adjustments within weighting classes can be beneficial in complex sample surveys with lower response rates, provided that the auxiliary and design variables used to form the weighting classes are correlated with the survey variable of interest and response propensity.

There are certainly important limitations of this study worth mentioning. These results apply to a single survey variable of interest, and many surveys are multi-purpose in nature, with analytic interest directed toward the estimation of multiple descriptive

parameters, including means and proportions, and analytic parameters, including regression coefficients and odds ratios. As suggested by Little and Vartivarian (2005), an appropriate approach in practice would be to estimate the MSEs of alternative weighted estimators of a population mean, and use the estimator shown to have a lower MSE (a “composite” estimator). This study did not develop sample-based estimators of the MSE for a given estimator of a population mean in the complex sample setting, and those developments would be an important extension of this work. Further, the extension of these results to analytic statistics such as regression coefficients is still an open area of research. All of these extensions present the possibility that different analyses of public-use survey data sets may require different sets of survey weights, which could complicate matters for users of the data. The development of both technical documentation and computer software that simplifies this process of determining optimal nonresponse weighting adjustments for a given analytic objective should therefore be a top priority for survey organizations and developers of statistical software.

Regarding survey practice, the results of this study suggest that those responsible for performing post-survey adjustments to base sampling weights in an effort to repair nonresponse errors should form weighting classes based on auxiliary and sample design variables related to both the survey variables of interest and the propensity of responding, which is not a new finding in this literature. The results of the present study further underscore the need to collect more auxiliary variables for all elements in the target population, possibly by assembling rich frames or collecting observations of field researchers during listing operations. Important correlates of survey variables of interest reported in previous survey research on a particular topic would seem to be the most important measures to collect prior to actual sample selection. Analysts generally do not know what variables will tend to correlate with survey variables of interest and response propensity, but having more variables to choose from will be helpful in making a decision about how to determine nonresponse adjustments. The results of this study provide new evidence that the use of weighted response rates can be beneficial when performing weighting class adjustments for nonresponse in complex sample design settings involving stratified cluster sampling.

References

- Bethlehem, J.G., 2002, Weighting Nonresponse Adjustments Based on Auxiliary Information (Chapter 18 in Groves et al., *Survey Nonresponse*)
- de Leeuw, E., and de Heer, W., 2002, Trends in Household Survey Nonresponse: A Longitudinal and International Comparison (Chapter 3 in Groves et al., *Survey Nonresponse*)
- Groves, R.M., 2006, Nonresponse Rates and Nonresponse Bias in Household Surveys (Public Opinion Quarterly, 70(5), 646-675)
- Groves, R.M., and Couper, M.P., 1998, A Conceptual Framework for Survey Participation (Chapter 2 in *Nonresponse in Household Interview Surveys*, New York, Wiley, 25-46)
- Groves, R.M., and Couper, M.P., 1998, How Survey Design Features Affect Participation (Chapter 10 in *Nonresponse in Household Interview Surveys*, New York, Wiley, 269-293)
- Groves, R.M., and Peytcheva, E., 2008, The Impact of Nonresponse Rates on Nonresponse Bias: A Meta Analysis (Public Opinion Quarterly, 72(2), 167-189)

- Little, R.J., and Vartivarian, S., 2003, On Weighting the Rates in Nonresponse Weights (Statistics in Medicine, 22(9), 1589-1599)
- Little, R.J., and Vartivarian, S., 2005, Does Weighting for Nonresponse Increase the Variance of Survey Means? (Survey Methodology, 31(2), 161-168)
- Manfreda, K.L., Bosnjak, M., Berzelak, J., Hass, I., and Vehovar, V., 2008, Web Surveys Versus Other Survey Modes – A Meta-Analysis Comparing Response Rates (International Journal of Market Research, 50(1), 79-104)
- Rubin, D.B., 1976, Inference and Missing Data (Biometrika, 63, 581-592)