Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Nonresponse adjustments with misspecified models in stratified designs

by Ismael Flores Cervantes and J. Michael Brick

Release date: June 22, 2016

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2016
Vol. 42, No. 1, pp. 161-177
Statistics Canada, Catalogue No. 12-001-X

Nonresponse adjustments with misspecified models in stratified designs

Ismael Flores Cervantes and J. Michael Brick[1]

Abstract

Adjusting the base weights using weighting classes is a standard approach for dealing with unit nonresponse. A common approach is to create nonresponse adjustments that are weighted by the inverse of the assumed response propensity of respondents within weighting classes under a quasi-randomization approach. Little and Vartivarian (2003) questioned the value of weighting the adjustment factor. In practice the models assumed are misspecified, so it is critical to understand the impact of weighting might have in this case. This paper describes the effects on nonresponse adjusted estimates of means and totals for population and domains computed using the weighted and unweighted inverse of the response propensities in stratified simple random sample designs. The performance of these estimators under different conditions such as different sample allocation, response mechanism, and population structure is evaluated. The findings show that for the scenarios considered the weighted adjustment has substantial advantages for estimating totals and using an unweighted adjustment may lead to serious biases except in very limited cases. Furthermore, unlike the unweighted estimates, the weighted estimates are not sensitive to how the sample is allocated.

Key Words: Nonresponse; Stratification; Sampling weights; Weighting classes reweighting.

1 Introduction

Adjusting the base weights for unit nonresponse using weighting classes is a standard approach to survey weighting, but the adjustments are not done in the same way by all researchers or survey organizations. Little and Vartivarian (2003), hereafter referred to as L&V, observed that using a nonresponse adjustment factor that is weighted by the inverse of the probability of selection appears to be the most common approach. They also pointed out that using design weights to compute a weighted nonresponse adjustment does not eliminate nonresponse bias in estimates of the mean of the population when the response mechanism is not specified correctly by the weighting adjustment model. L&V then conducted a simulation study using a simple stratified sample design to examine the effect of weighting the nonresponse adjustment factors. They concluded that weighting the nonresponse adjustment has little or no value.

Theoretical justifications for nonresponse adjustment require that either the response mechanism or the target variable must be modeled correctly to eliminate nonresponse bias; we are not aware of any theory that suggests that weighting by the inverse of the probability of selection completely eliminates bias when the model is misspecified (e.g., Kalton 1983; Little 1986; Little and Rubin 2002; Särndal and Lundström 2005). In this regard, the importance of modeling for nonresponse adjustment urged by L&V is essential for good statistical practice. However, correctly specifying a highly predictive model is an ideal that cannot be achieved in most surveys because of the complexity of the phenomenon and because powerful auxiliary variables rarely exist. The search for better auxiliary data for this modeling has fueled research into paradata, but the models using these data still have relatively poor correlations with response propensities (Kreuter,

1. Ismael Flores Cervantes and J. Michael Brick, Westat, 1600 Research Blvd, Rockville, Maryland 20850. E-mail: ismaelflorescervantes@westat.com.

Olson, Wagner, Yan, Ezzati-Rice, Casas-Cordero, Lemay, Peytchev, Groves and Raghunathan 2010). In practice, imperfect models are used and nonresponse bias is never completely eliminated.

Consequently, understanding the effects of nonresponse adjustment methods and whether there is any value to weighting the nonresponse adjustment with an incorrectly specified response model is important. Even though a message of L&V was the need to include design variables in the nonresponse modeling, some researchers appear to have concluded that weighting the adjustment has no role (e.g., Chadborn, Baster, Delpech, Sabin, Sinka, Rice and Evans 2005; Haukoos and Newgard 2007). However, L&V's conclusion that weighting the nonresponse adjustment factor is either incorrect or inefficient was based on comparisons to correctly specified models that always produce unbiased estimates. Their suggestion to condition on the design variables (in their setting the design variable was the stratum) resulted in identical weighted and unweighted estimators. Their simulations are also centered on a specific stratified sample design and they only consider estimating means. As discussed below, these are substantial limitations and the conclusions that some have drawn that weighting the adjustment is inappropriate need to be reconsidered.

Following L&V, researchers have examined the effects of weighting in other cases. Sukasih, Jang, Vartivarian, Cohen and Zhang (2009) compared nonresponse adjustments with and without weights by simulation within the context of a specific survey. West (2009) used simulation to study estimates of population means under more complex sample designs that featured clustering and differential sampling rates. Both of these studies concluded that weighting the nonresponse adjustments by the design weights was beneficial compared to using an unweighted approach, even though the differences due to weighting were not large. Kott (2012) assessed the robustness of the adjustments theoretically and described the conditions under which the various estimators for population means had greater protection against nonresponse bias; he recommended a weighted approach. Related research has been conducted on the need for weighting for estimating response propensity model coefficients (Wun, Ezzati-Rice, Diaz-Tena and Greenblatt 2007; Grau, Potter, Williams and Diaz-Tena 2006), but this line of research is sufficiently different that we do not discuss it here.

In this article, we explore the effect of weighting nonresponse adjustments when the nonresponse model is imperfect. In Section 2, we expand on the L&V results by looking at estimators for totals and domain means and totals; L&V only considered overall means. Using the same population and basic simulation setting of L&V, we also explore the effect of different sample allocation to the strata while L&V used one sample allocation. The results of the simulations presented in Section 3 show that there are important differences in the properties of the weighted and unweighted estimators and these vary by how the sample is allocated. We explain the behaviors of the estimators using simple approximations to show why they differ. Although weighting the adjustment factor does not always give estimates with lower bias and root mean square error when compared to estimates from the unweighted alternative, it has substantial benefits for estimates of totals and provides protection against large errors that may arise with an unweighted approach. As a result, we recommend a weighted approach when the true response mechanism is not fully known. Conclusions are presented in Section 4.

2 Setting

Survey weights compensate for different types of missing data – sampling or base weights adjust for those that are not sampled, noncoverage adjustment weights account for those that are not in the sampling frame, and nonresponse adjustment weights compensate for those that are sampled but do not respond. We focus on nonresponse adjustment weights and the effect of using the base weights in creating the nonresponse adjustments.

We begin with the unadjusted Horvitz-Thompson estimator of the total

$$\hat{y}_{un} = \sum_s R_i d_i y_i, \tag{2.1}$$

where d_i is the inverse of the probability of selection of unit $i, R_i = 1$ if unit i responds and $= 0$ otherwise, and the sum is over the units in sample s. The ratio mean is $\hat{\bar{y}}_{un} = \hat{y}_{un} \big/ \sum_s R_i d_i$. If all the sample data are observed and the frame is complete, then $E(\hat{y}_{un}) = Y$, and the ratio mean is consistent for $\bar{Y}$.

When there is unit nonresponse, we assume that response is a random variable and the probability or propensity of response $(\phi_i = \Pr(R_i = 1))$ is like the probability from an additional phase of sampling (Särndal, Swensson and Wretman 1992). If we assume $\phi_i > 0$ for all i, then the nonresponse bias of an estimated ratio mean under the stochastic model is

$$\text{bias}\left(\hat{\bar{y}}_{un}\right) \approx \bar{\phi}^{-1} \sigma_\phi \sigma_y \rho_{\phi,y}, \tag{2.2}$$

where $\bar{\phi}$ is the population mean of the response propensities, σ_ϕ is the standard deviation of ϕ, σ_y is the standard deviation of y, $\rho_{\phi,y}$ is the correlation between ϕ and y (Bethlehem 1988). The estimated respondent mean is unbiased if ϕ and y are uncorrelated. Brick and Jones (2008) extend these results to other types of statistics and estimators.

To reduce nonresponse bias, auxiliary variables associated with the sample can be used to support nonresponse adjustments to the base weights. The adjustments can be implemented by modeling either the distribution of ϕ or y, or both using the auxiliaries. We are specifically interested in modeling the response mechanism.

The estimated response propensities are applied as if they were the actual probabilities of responding. In other words, the nonresponse adjustment factor is the inverse of the estimated propensity of responding for sampled unit $i\left(\hat{\phi}_i\right)$. The response propensity can be estimated by a variety of methods such as logistic regression, but most surveys form mutually exclusive groups called weighting classes or response homogeneity groups which contain units with similar estimated propensities and adjust the weights in each group or class by a common factor, say $\hat{f}_c = \hat{\phi}_c^{-1}$ for all $i \in c$ (Särndal et al. 1992, and Little 1986). When this approach is used, the adjusted estimator is called a weighting class estimator and is

$$\hat{y}_{wc} = \sum_c \sum_{i \in s_c} R_{ci} d_{ci} \hat{f}_c y_{ci}, \tag{2.3}$$

where $c = 1, 2, \ldots, C$ are the nonresponse adjustment classes and $i \in s_c$ is a sampled unit in class c.

The specific issue we address is the effect of weighting the adjustment factor. The unweighted factor is

$$\hat{f}_c^{\,u} = \frac{\sum_{i \in s_c} \delta_{ci}}{\sum_{i \in s_c} R_{ci} \delta_{ci}} = \frac{n_{c+}}{r_{c+}}$$

where $\delta_{ci} = 1$ if $i \in c$ and $\delta_{ci} = 0$ if $i \notin c$, and n_{c+} and r_{c+} are the number of sampled and responding units in class c. The weighted adjustment factor is

$$\hat{f}_c^{\,w} = \frac{\sum_{i \in s_c} d_{ci}}{\sum_{i \in s_c} R_{ci} d_{ci}} = \frac{\hat{N}_c}{\hat{N}'_c},$$

where $\hat{N}_c = \sum_{i \in s_c} d_{ci}$ and $\hat{N}'_c = \sum_{i \in s_c} R_{ci} d_{ci}$. The factors correspond to the unweighted and weighted response rates, respectively. Substituting the factors into the estimator (2.3) yields two alternative estimators (2.4) and (2.5) of the total population. These are both weighting class estimators but we have changed notation to emphasize whether the weighted or unweighted response rate is used.

$$\hat{y}_{urr} = \sum_c \hat{f}_c^{\,u} \sum_{i \in r_c} d_{ci} y_{ci} = \sum_c \frac{n_{c+}}{r_{c+}} \sum_{i \in r_c} d_{ci} y_{ci}, \tag{2.4}$$

$$\hat{y}_{wrr} = \sum_c \hat{f}_c^{\,w} \sum_{i \in r_c} d_{ci} y_{ci} = \sum_c \frac{\hat{N}_c}{\hat{N}'_c} \sum_{i \in r_c} d_{ci} y_{ci}. \tag{2.5}$$

These two estimators are the building blocks for all the types of statistics that we consider in the simulation study. For example, estimators of means, domain means, and ratios are simple functions of estimators (2.4) and (2.5).

To be consistent with the structure, notation, and simulations in L&V, we restrict our study to the same population with a stratified simple random sample where two strata are defined by the binary design variable, Z, and two nonresponse adjustment classes are defined by a binary auxiliary variable, C, that cross the strata as shown in Table 2.1. We replaced the X used in L&V with C for weighting cell as introduced above to easily identify the nonresponse adjustment cell. Consistent with L&V, the population size is set at $N = 10{,}000$.

Table 2.1
Population counts by strata *Z* and nonresponse adjustment cell *C*

Sampling strata	Nonresponse adjustment cell	
	***C* = 0**	***C* = 1**
Z = 0	3,064	3,931
Z = 1	2,079	926

Source: Little and Vartivarian (2003) who used *X* instead of *C*.

The variable of interest, Y, is a binary variable with the probability that $Y = 1$ defined by a logistic model with $\operatorname{logit}(Y = 1 \mid C, Z) = 0.5 + \gamma_C (C - \bar{C}) + \gamma_Z (Z - \bar{Z}) + \gamma_{CZ} (C - \bar{C})(Z - \bar{Z})$. The response variable R is also binary with the probability of $R = 1$ generated from a logistic model with $\operatorname{logit}(R \mid C, Z) = 0.5 + \beta_C (C - \bar{C}) + \beta_Z (Z - \bar{Z}) + \beta_{CZ} (C - \bar{C})(Z - \bar{Z})$. Different populations and response propensities are generated depending on the values of $\gamma_C, \gamma_Z, \gamma_{CZ}, \beta_C, \beta_Z$ and β_{CZ} as shown in Table 2.2. We have adopted the generalized linear model notation L&V used to make comparison to their work easier. The tabled values are the same populations and response variables that L&V generated by assigning values to $(\gamma_C, \gamma_Z, \gamma_{CZ}, \beta_C, \beta_Z, \beta_{CZ})$. In the notation $[A]^B$ in Table 2.2, the population (Y) or the response propensity (R) are indicated by the superscript B while the parameters and interactions of the model for the distribution of the population or response are indicated by A inside the brackets. For example, the additive logistic model that generates the distribution of Y within the sampling stratum Z and nonresponse cell C is indicated by $[C + Z]^Y$. Similarly, models where R depends on C only, Z only or neither C nor Z are denoted by $[C]^R$, $[Z]^R$, and $[C + Z]^R$ respectively. L&V give more details on their rationale for choosing these populations and response models.

Table 2.2
Models for outcome variable, *Y*, and probability of response, *R*

Model for *Y* (Variable of interest)	Model for *R* (Response propensity)	Parameters		
		γ_C, β_C	γ_Z, β_Z	γ_{CZ}, β_{CZ}
$[CZ]^Y$	$[CZ]^R$	2	2	2
$[C + Z]^Y$	$[C + Z]^R$	2	2	0
$[C]^Y$	$[C]^R$	2	0	0
$[Z]^Y$	$[Z]^R$	0	2	0
$[\phi]^Y$	$[\phi]^R$	0	0	0

Source: Little and Vartivarian (2003).

L&V computed estimates of means that are, in our notation,

$$\hat{\bar{y}}_{urr} = \frac{\hat{y}_{urr}}{\sum_c \hat{f}_c^u \sum_{i \in s_c} R_{ci} d_{ci}} = \frac{\hat{y}_{urr}}{\sum_c \hat{f}_c^u \hat{N}_c'}, \tag{2.6}$$

and

$$\hat{\bar{y}}_{wrr} = \frac{\hat{y}_{wrr}}{\sum_c \hat{f}_c^w \sum_{i \in s_c} R_{ci} d_{ci}} = \frac{\hat{y}_{wrr}}{\sum_c \hat{N}_c}. \tag{2.7}$$

The denominators of the means are estimates of the population size N. In estimator (2.7), the denominator is a constant and equal N, but in estimator (2.6) the denominator is a random variable. In the simulation setting with the stratified simple random sample design described below, or in any design where $\sum_{i \in s} d_i = N$ for every s, the estimator (2.7) reduces to the linear estimator $\hat{\bar{y}}_{wrr} = N^{-1} \hat{y}_{wrr}$; whereas (2.6) is a ratio estimator. This is an important point we return to later.

Domain means may have properties that differ from overall means because the denominators of the weighted and unweighted domain means are both random variables. One exception is when the domains match the sampling strata and therefore both the domain sizes and stratum sizes are known. L&V did not discuss domains, so these estimates are not studied in their simulation. We create domains by randomly generating a random variable ν_i from a uniform (0, 1) distribution, and defining the membership function $\tau(a) = 1$ if $a < 0$ and $\tau(a) = 0$ if $a \geq 0$. Domain means of 50% were created by substituting $d_{ci}^* = \tau(\nu_i - 0.5)\, d_{ci}$ into expressions (2.6) and (2.7) to produce the estimators $\hat{y}_{urr,0.5}$ and $\hat{y}_{wrr,0.5}$, respectively. Weighted and unweighted estimators of domain totals $\hat{y}_{urr,0.5}$ and $\hat{y}_{wrr,0.5}$ were formed similarly. We used the same device to create 25 percent domain means and 25 percent domain totals. Since we are interested in the effect of the nonresponse adjustments in means computed as ratio estimators, other domains such as those defined close to 100 percent of the population were excluded from the analysis because the denominator of the domain means approaches the constant population total N and the mean becomes a linear estimator. Domains closer to 0 percent were excluded because of small sample sizes.

3 Findings

The simulation was done in R (R Development Core Team 2011) with 10,000 draws (L&V used 1,000 draws). We evaluated the estimators by computing the root mean squared error (rmse) and the bias of the estimates, where the bias and rmse are measured in deviations from the population quantities as done in L&V. We used the same total sample size of 312 that they used in their simulation, but with different sample allocation or relative sampling rates between strata. We replicated all 25 configurations in L&V and these results are found in Table S-1 in the supplemental materials. Table S-2 in supplemental materials also includes the 25 configurations but presents the relative bias of unweighted and weighted means and totals, as well as ratios of variances and rmse of unweighted to weighted estimates. The relative bias and ratios of variances and rmse facilitate the comparisons between the estimates. These materials include their estimated simulation errors, which are all relatively small. For those estimators and sampling rates given in L&V, our results are consistent with their published values within simulation error. We begin by examining the bias of the estimators.

3.1 Bias

There are two situations where theoretical results exist and are well-known (Little and Rubin 2002). One is when the propensity to respond is the same in all cells – missing completely at random (MCAR); MCAR corresponds to the model $[\phi]^R = (\beta_c = 0, \beta_z = 0, \beta_{cz} = 0)$ in the last row in Table 2.2. With MCAR, the unweighted and weighted adjustment factors are equal in expectation, and both produce unbiased estimates. The simulation results in Table V of L&V paper (rows 5, 10, 15, 20, and 25) confirm this observation. The second situation is when the response propensity is independent of the strata, corresponding to missing at random (MAR) with the response model $[\phi]^C = (\beta_c = 2, \beta_z = 0, \beta_{cz} = 0)$ in the third row of Table 2.2. We refer to these situations as MAR because the bias of the estimator does not depend on whether the information about Z is used in the model. Here again, the weighted and unweighted estimates are both unbiased and the adjustments are equal in expectation. The simulation results in Table V of L&V (rows 3, 8, 13, 18, and 23) confirm this observation empirically.

To focus on the situation in which the model is incorrectly specified, we do not present the simulations results for the MCAR and MAR situations in this document, but these results can be found in the supplemental materials. An important point is that even though the weighted and unweighted adjustments for the MCAR and MAR models are equal in expectation, they are not identical. Sukasih et al. (2009) simulated the two approaches under MAR models and stated a preference for the weighted approach largely due to the lower variability in the estimates of total across simulations even though both were unbiased.

As mentioned before, our simulations vary the sampling rates while keeping the overall sample size fixed at 312; L&V used a single sampling rate. When the sampling rates are the same across strata (i.e., the sample is proportionally allocated to the strata), then the sampling weights are the same for the two strata and consequently the weighted and unweighted estimators are identical. The proportional allocation sampling rate plays a visible role in our presentation because the two estimates must converge at this point.

Figure 3.1 (left panel) is a graph of the simulation results for the bias of the weighted and unweighted estimator of the total for $[CZ]^Y$ and $[C+Z]^R$. We chose this configuration (row 2 in L&V's tables) because the simulations in L&V showed the unweighted mean had lower bias and rmse than the weighted mean for this case. The horizontal axis shows the relative sampling rate computed as the ratio of the sampling rate of $Z = 0$ to $Z = 1$ or $N_0 n_0^{-1} / \left(N_1 n_1^{-1}\right)$. The relative sampling rate used by L&V was about 2.25. It is immediately apparent that the bias of the weighted estimator is essentially constant across different sampling rates while the bias of the unweighted estimator varies substantially with the relative sampling rate. The bias of the unweighted estimators of the total can be more than two times the bias of the weighted estimator for some sampling rates. Both estimators are biased for almost all relative sampling rates, and the estimator that has the lower bias depends on this rate. When the relative sampling rates are equal (proportional allocation) the unweighted and weighted estimators have the same bias, as expected. However, in practice, it is not generally possible to recognize the effect the sampling rate has on the bias and choose in advance the adjustment method to reduce the bias for a specific sample.

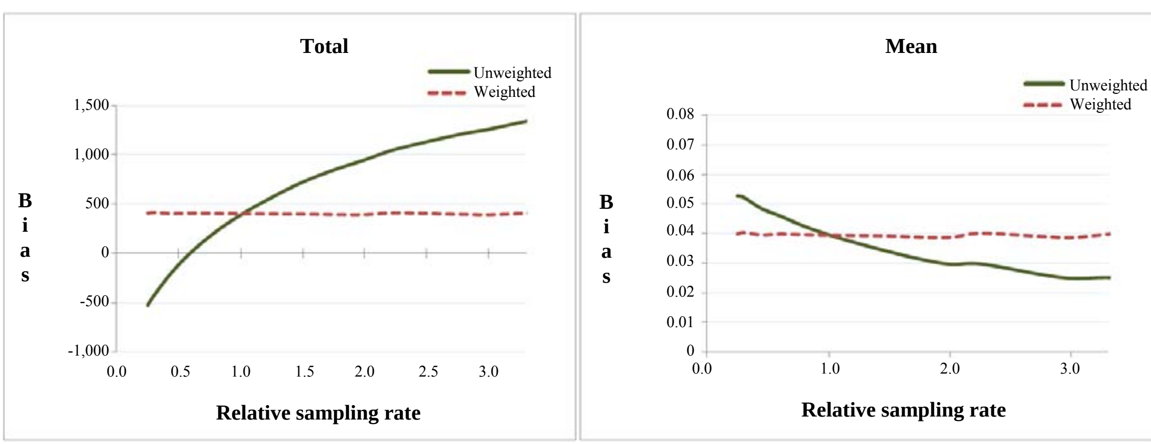

Figure 3.1 Bias for weighted and unweighted estimator for the population model $[CZ]^Y$ and response propensity model $[C+Z]^R$, where the left panel is for the total and right panel is for the mean.

To understand these findings, we applied some standard approximations that hold reasonably well in this situation (i.e., $E\left(\eta^{-1}\right) \approx E^{-1}(\eta)$). The approximate expected value for the weighted estimator is

$$E\hat{y}_{wrr} \approx \sum_z \sum_c \frac{N_c}{\left(\sum_z \phi_{cz} N_{cz}\right)} \phi_{cz} Y_{cz}, \tag{3.1}$$

where Y_{cz} is the population total in cell cz. Similarly, the approximate expected value for the unweighted estimator is

$$E\hat{y}_{urr} \approx \sum_z \sum_c \frac{\left(\sum_z N_z n_z^{-1} N_{cz}\right)}{\left(\sum_z \phi_{cz} N_z n_z^{-1} N_{cz}\right)} \phi_{cz} Y_{cz}. \tag{3.2}$$

If ϕ_{cz} is a constant (MCAR) or ϕ_{cz} is constant within weighting cells (MAR), then both estimators are unbiased to this order of approximation and consistent with known theory. When the sampling rates are the same across strata, the two estimators have the same expected value (as noted above they are identical in this case). More importantly, these approximations show the expectation of the weighted estimator is not dependent on the sampling rate, but the expectation of the unweighted estimator is. This explains the patterns shown in the Figure 3.1.

Some details of the simulation estimates for this configuration are shown in Table 3.1 for selected sampling rates. As noted above, the full simulation results for all configurations and sampling rates used to create the figures can be found in the supplemental materials. These materials include the relative biases, ratios of variances and ratios of rmse which are better indicators for assessing the impact of the adjustments on the estimates. We observed that all configurations with biased estimates of totals have biases that are lower for the weighted estimator on one side of the relative sampling rate of 1 and are higher on the other side. All configurations exhibit an approximately constant bias for the weighted estimator of the total across the relative sampling rates, but the bias of the unweighted estimator varies by relative sampling rate.

Next, we turn to estimated means – the only estimators considered by L&V. The right panel of Figure 3.1 shows the bias for the weighted estimator is again independent of the relative sampling rate while the bias of the unweighted estimator varies with the sampling rate. L&V used a sampling rate of 2.25 so this explains why they found the unweighted estimator had a lower bias for the mean in their simulation. Two points are worth noting. First, the biases for the means for both adjustment methods are all relatively small, especially when compared to the potential relative biases of the totals with the unweighted estimator in the panel on the left. Second, there is no way to identify if a particular estimate would fall on the left or right of the relative sampling rate of 1. Table 3.1 shows the estimated biases for this configuration.

The graphs also show a relationship that is somewhat surprising; the relative sampling rates where the unweighted estimator of the total has a lower bias are those where the unweighted estimator of the mean has a higher bias. In other words, the means behave differently from the totals because the unweighted mean is a ratio while the weighted mean is not. As a result, the relative bias (rb = bias/estimate) of the unweighted estimator of the mean is not equal to the relative bias of the unweighted estimator of the total (the relationship holds for the weighted estimator). We approximate the relative bias by

$$rb\left(\hat{\bar{y}}_{urr}\right) \approx \frac{1 + rb\left(\hat{y}_{urr}\right)}{1 + rb\left(\hat{N}_{urr}\right)},$$

where $\hat{N}_{urr}$ is the unweighted estimator of the total (where $y_i = 1$ for all i). This approximation holds well in this situation since $\operatorname{cov}\left(\hat{y}_{urr}, \hat{N}_{urr}\right) / E\left(\hat{N}_{urr}\right) \approx 0.$ Thus, the relative bias of the unweighted mean is reduced whenever the biases of the numerator and denominator are positively correlated.

Now, consider domain estimates – estimators not studied in L&V. The biases for the weighted and unweighted domain total estimators and the relationships with the biases of the unweighted estimators varying by the relative sampling rate are the same as observed for the overall totals (see Table 3.1). This follows because domain totals are still totals and approximations (3.1) and (3.2) still apply. The domain means are also given in the table and they too exhibit the same pattern of biases as shown in Figure 3.1 for the full sample mean. It is worth noting that the relative biases for the mean estimates (overall and for the domains) do not vary much, with most relative biases in the range of 5 to 7 percent.

Table 3.1
Bias (times 10,000), root mean square error (times 10,000) and variance of weighted and unweighted estimators of means and total of the full sample and domains, configuration $[CZ]^Y$, $[C+Z]^R$ with various sampling rates

				Relative sampling rate				
	Characteristic	**Domain**	**Adjustment**	**0.30**	**0.44**	**1.00**	**2.25**	**3.30**
Bias	Mean	Full	urr	515	491	404	301	248
			wrr	398	403	404	404	394
		50%	urr	513	501	411	307	257
			wrr	397	414	410	410	401
		25%	urr	523	498	407	298	252
			wrr	408	411	407	400	395
	Total	Full	urr	-419	-184	401	1,058	1,335
			wrr	398	403	404	404	394
		50%	urr	-214	-89	205	535	673
			wrr	194	205	206	207	200
		25%	urr	-107	-48	101	264	335
			wrr	97	98	102	101	100
Rmse	Mean	Full	urr	643	614	546	536	566
			wrr	553	547	545	587	616
		50%	urr	758	726	669	699	778
			wrr	687	671	669	728	794
		25%	urr	949	898	863	952	1,062
			wrr	895	859	863	955	1,041
	Total	Full	urr	537	376	543	1,183	1,485
			wrr	553	547	545	587	616
		50%	urr	371	311	393	714	888
			wrr	399	392	394	449	494
		25%	urr	255	233	282	451	553
			wrr	285	273	283	328	365
Variance	Mean	Full	urr	15	14	14	20	26
			wrr	15	14	14	18	22
		50%	urr	32	28	28	40	54
			wrr	32	28	28	37	47
		25%	urr	64	57	59	83	107
			wrr	64	58	59	76	93
	Total	Full	urr	11	11	14	28	43
			wrr	15	14	14	18	22
		50%	urr	9	9	11	23	34
			wrr	12	11	11	16	21
		25%	urr	5	5	7	14	20
			wrr	7	7	7	10	12

3.2 Root mean square error

Despite the small sample size used in the simulations (312 before nonresponse) and the relatively modest relative bias of the estimates for means, the bias is still a large component of the rmse. For example, the bias accounts for 56 (unweighted) to 69 (weighted) percent of the rmse for the estimate of the mean in the $[CZ]^Y$ and $[C+Z]^R$ configuration using the L&V sampling rate. With larger sample sizes that are common in large sample surveys, the bias is often the dominant component of the rmse (Brick 2013).

Figure 3.2 shows the rmse for the estimated total (left panel) and for the mean (right panel) using the same configuration used in the previous figure. The rmse for the total for the weighted estimator is approximately constant and smaller than the rmse for the unweighted estimator, except when the relative sampling rate is about 0.5 which corresponds to the region with very low bias for the unweighted estimator as shown in Figure 3.1. However, when the relative sampling rate is greater than one, the rmse of the unweighted estimator of the total is much larger than the rmse of the weighted estimator (it can be as much as twice the rmse of the weighted estimator for some sampling rates). In contrast, for the estimates of the mean shown in Figure 3.2 (right panel), the rmse of both the weighted and unweighted estimators are similar in magnitude, and the symmetry around the proportional allocation rate remains. Even though L&V point out the unweighted estimator has a lower rmse (at the relative sampling rate of 2.25), we view the rmse of both estimators to be approximately equal across the range of relative sampling rates.

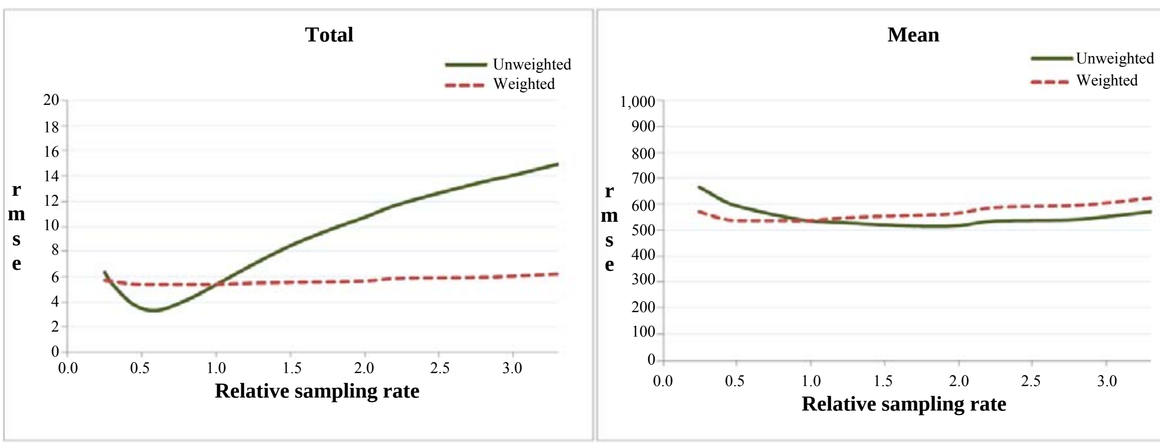

Figure 3.2 Root mean square error for weighted and unweighted estimator when $[CZ]^Y$ and $[C+Z]^R$, where the left panel is for the total (rmse is in millions) and the right panel is for the mean.

Figure 3.3 shows the rmse for the estimated 50% domain mean (left panel) and for the 25% domain mean (right panel) again using $[CZ]^Y$ and $[C+Z]^R$. Looking at the three graphs of the rmse of the means (for the overall mean, the 50% domain mean, and the 25% domain mean) reveals the effect of the ratio estimator. As the percentage in the domain decreases from 100% to 25%, the weighted estimator becomes more like an unconditional ratio estimator and the correlation between the numerator and denominator reduces the rmse of the estimate. As a result, the rmse of the weighted and unweighted estimators are very similar for the domain estimators. Even though the weighted estimator has a lower rmse at each of the relative sampling rates compared to the unweighted one for the 25% domain mean, the two estimators are essentially equivalent in terms of rmse. The slight advantage of the unweighted estimator pointed out by L&V for the full population mean for this configuration vanishes for domain means where the weighted estimator is also a ratio estimator.

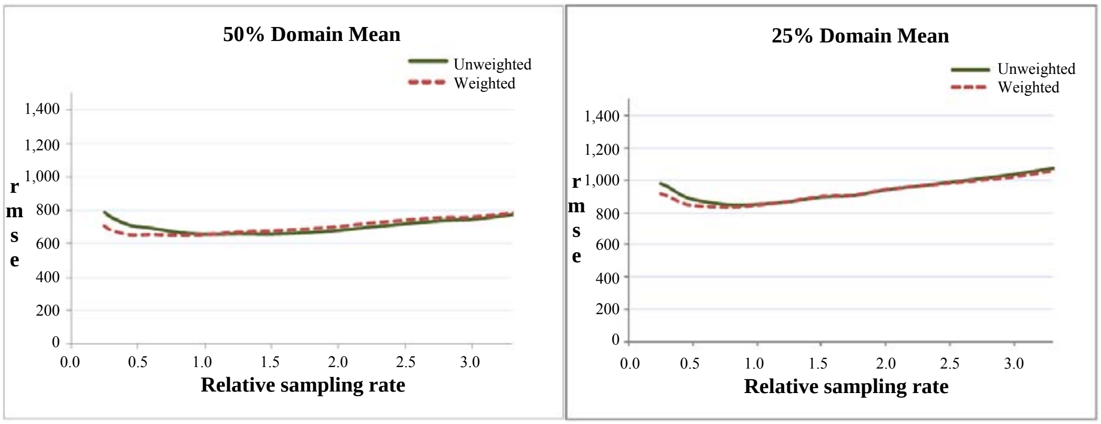

Figure 3.3 Root mean square error for weighted and unweighted estimator when $[CZ]^Y$ and $[C+Z]^R$, where the left panel is for the 50% domain mean and the right panel is for the 25% domain mean.

3.3 Variance

A general concern about nonresponse adjustment factors is that when the factors are based on a small number of respondents they may increase the variance of the estimates (Kalton 1983; Tremblay 1986). L&V suggest weighting the nonresponse adjustment factors may be responsible for greater variance inflation than using the unweighted factors. The figures above show that this did not occur in this simulation. Figure 3.4 shows the ratio of the unweighted estimator's variance to that of the weighted estimates for the full population mean and total and the 50% domain total for the $[CZ]^Y$ and $[C + Z]^R$ configuration. For the mean, the variance ratio is nearly equal to one over all the relative sampling rates showing no inflation of variance for the weighted estimator compared to the unweighted estimator. For totals, the ratio is less than unity for relative sampling rates less than 1 and greater than 1 for relative sampling rates greater than unity. The same relationship holds true for the 50% domain total. These results suggest that weighting the adjustment is not the source of large factors that can inflate the variance of the estimates. A prudent approach is to examine the size of nonresponse factors, irrespective of whether they are weighted or unweighted.

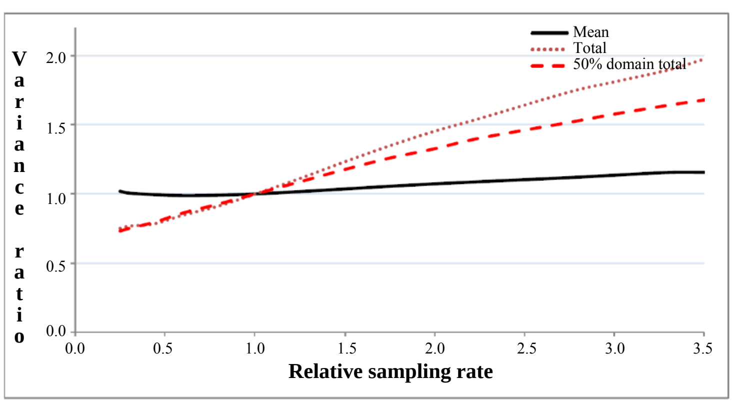

Figure 3.4 Ratio of variance of unweighted to weighted estimates of the mean, total and 50% domain total for $[CZ]^Y$ and $[C+Z]^R$.

Table 3.2 gives the simulation results for another configuration, $[CZ]^Y$ and $[CZ]^R$, that was favorable to the unweighted adjustment in L&V (the first row in their tables). In contrast, Table 3.3 gives simulation results for $[C+Z]^Y$ and $[C+Z]^R$ which is a configuration that was favorable to the weighted adjustment. The results for both of these configurations show the same general patterns as discussed above for $[CZ]^Y$ and $[C+Z]^R$.

Table 3.2
Bias (times 10,000), root mean square error (times 10,000) and variance of weighted and unweighted estimators of means and total of the full sample and domains, configuration $[CZ]^Y$, $[CZ]^R$ with various sampling rates

				Relative sampling rate				
	Characteristic	**Domain**	**Adjustment**	**0.30**	**0.44**	**1.00**	**2.25**	**3.30**
Bias	Mean	Full	urr	329	329	289	255	237
			wrr	294	299	289	298	298
		50%	urr	334	341	293	251	238
			wrr	299	311	293	294	298
		25%	urr	336	344	306	257	247
			wrr	302	314	306	299	307
	Total	Full	urr	-412	-187	287	732	901
			wrr	294	299	289	298	298
		50%	urr	-209	-91	145	367	455
			wrr	143	152	146	149	154
		25%	urr	-103	-46	72	184	230
			wrr	74	76	73	75	79
Rmse	Mean	Full	urr	530	507	476	501	533
			wrr	505	487	476	520	554
		50%	urr	684	653	616	664	732
			wrr	666	638	616	674	740
		25%	urr	911	859	832	920	1,016
			wrr	900	849	832	920	1,011
	Total	Full	urr	550	395	474	886	1,078
			wrr	505	487	476	520	554
		50%	urr	385	326	373	575	696
			wrr	394	375	373	425	475
		25%	urr	263	244	278	390	464
			wrr	285	274	278	321	361
Variance	Mean	Full	urr	17	15	14	19	23
			wrr	17	15	14	18	22
		50%	urr	36	31	30	38	48
			wrr	36	31	30	37	46
		25%	urr	73	63	61	79	98
			wrr	73	63	61	76	94
	Total	Full	urr	14	12	14	25	35
			wrr	17	15	14	18	22
		50%	urr	11	10	12	20	28
			wrr	14	12	12	16	20
		25%	urr	6	6	7	12	16
			wrr	8	7	7	10	13

Table 3.3
Bias (times 10,000), root mean square error (times 10,000) and variance of weighted and unweighted estimators of means and total of the full sample and domains, configuration $[C+Z]^Y$, $[C+Z]^R$ with various sampling rates

	Characteristic	Domain	Adjustment	Relative sampling rate				
				0.30	0.44	1.00	2.25	3.30
Bias	Mean	Full	urr	763	735	654	566	529
			wrr	665	661	654	654	652
		50%	urr	773	737	653	564	532
			wrr	677	664	653	651	656
		25%	urr	773	739	659	574	513
			wrr	679	668	659	660	636
	Total	Full	urr	-272	-8	651	1,411	1,744
			wrr	665	661	654	654	652
		50%	urr	-133	-6	326	711	875
			wrr	336	328	328	332	328
		25%	urr	-69	-2	157	359	438
			wrr	165	166	158	168	165
Rmse	Mean	Full	urr	854	818	745	699	711
			wrr	767	753	745	764	790
		50%	urr	951	901	827	816	863
			wrr	877	845	826	863	912
		25%	urr	1,101	1,046	981	1,023	1,098
			wrr	1,044	1,004	981	1,045	1,107
	Total	Full	urr	426	313	741	1,503	1,868
			wrr	767	753	745	764	790
		50%	urr	334	300	475	867	1,071
			wrr	489	470	476	529	575
		25%	urr	246	240	314	530	649
			wrr	320	316	314	372	409
Variance	Mean	Full	urr	15	13	13	17	23
			wrr	15	13	13	16	20
		50%	urr	31	27	26	35	46
			wrr	31	28	26	32	40
		25%	urr	62	56	54	73	95
			wrr	63	57	54	67	83
	Total	Full	urr	11	10	13	27	45
			wrr	15	13	13	16	20
		50%	urr	10	9	12	25	39
			wrr	13	12	12	17	22
		25%	urr	6	6	7	15	23
			wrr	8	7	8	11	14

3.4 Estimating population size

A particular type of estimate studied by Sukasih et al. (2009) is the estimate of the number of units in a population. We refer to this as an estimate of population size where the population size is just an estimate of a total where $y_i = 1$ for all i. It can be estimated for a domain by assigning all units outside the domain $y_i = 0$. In the simple stratified sample design studied here, the weighted estimator always reproduces the overall total population size $N = 10{,}000$, but the unweighted estimator does not. Since this situation clearly favors the weighted estimator, we instead examine the domain population size estimates.

Suppose we are estimating the number of units in a domain or subgroup that have a value below a percentile defined by a characteristic for the total population (e.g., national median income). This type of statistic is extremely important in surveys because estimates of the population size for domains are often key outcome statistics. For example, an estimate of this type is the total number of persons with an income below the poverty line or the low income line (Kovačević and Yung 1997).

The L&V analysis did not consider estimates for domains sizes or means, so there is not an explicit variable that can be used to define a subpopulation. To avoid complicating this analysis, we illustrate the performance of the two estimators using an artificial domain created by randomly selecting half of the population (i.e., 50% domain). Similar to the analysis in previous sections we computed weighted and unweighted totals and means for the 50% domain. Even though we know the size for this domain beforehand for this example (i.e., 50 percent of the total population), the analysis is still valid. In practice, the domain size would not be known.

When estimating a statistic such as the population size in a domain, both the weighted and unweighted estimators of domain population size are unbiased when the data are MCAR or MAR, as noted by Sukasih et al. (2009). Furthermore, the rmse errors of the weighted and unweighted estimators are approximately equal in this case as confirmed in the simulations.

When the data are not MAR, the situation may be very different. The weighted estimator of a domain population size is approximately unbiased for all relative sampling rates and all configurations, but the unweighted estimator is always biased except when it is identical to the weighted estimator (at a relative sampling rate of 1). As a consequence the rmse of the unweighted estimator for the domain size is often considerably greater than that of the weighted estimator. Figure 3.5 shows that the rmse of the unweighted estimator of the 50% domain size for $[CZ]^Y$ and $[C+Z]^R$ is substantially greater than that of the weighted estimator for most relative sampling rates (as much as twice the rmse of the weighted estimator). The only exception is when the two estimators are approximately equal (near proportional allocation).

The weighted estimator of domain sizes thus has a substantial advantage over the unweighted estimator for all of the missing data mechanisms in L&V that are not MCAR or MAR.

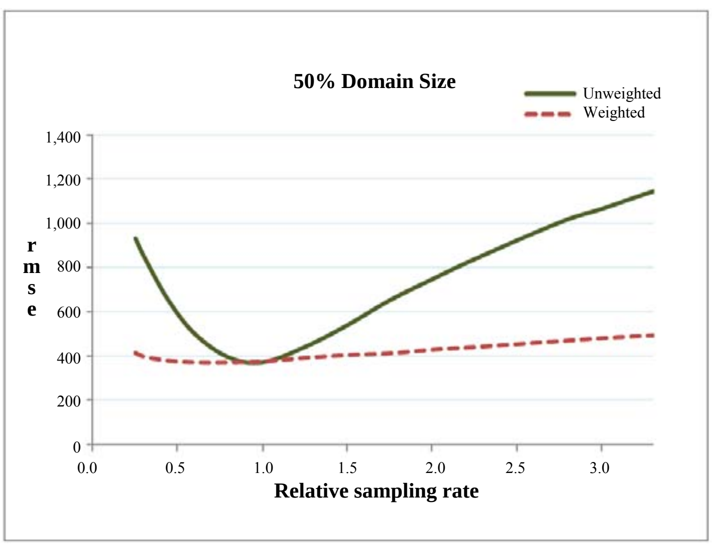

Figure 3.5 Root mean square error (rmse) for 50% domain size weighted and unweighted estimators when $[CZ]^Y$ and $[C+Z]^R$.

4 Conclusions

Nearly every survey suffers from nonresponse so the method for adjusting the base weights for unit nonresponse is an important topic. L&V appropriately noted that using design weights to compute a weighted nonresponse adjustment does not eliminate nonresponse bias when the response mechanism is not specified correctly in the weighting adjustment model. However, their simulation study suggested to at least some researchers that an unweighted adjustment might be more appropriate than a weighted adjustment more generally. The results from our evaluation, using the same setting as in L&V, contradict this perception. We explored the differences between the unweighted and weighted estimators when the adjustment model is misspecified in more detail using the L&V setting by including different sampling rates and estimates of totals and domains in addition to the means discussed in L&V.

These expanded simulations show that the unweighted and weighted adjustments do have different properties. The bias of the weighted estimator of totals means in stratified simple random sample designs is approximately constant irrespective of the sampling rate but the bias of the unweighted estimator depends on the sampling rate. In contrast, the bias of the unweighted estimator of the total is substantially larger than that of the weighted estimator for some sampling rates. For means, the bias and the rmse of the two estimators are not very different including those configurations that L&V described as favoring the unweighted estimator. The same general conclusions hold for estimates of domain means and totals as the weighted mean becomes more of a ratio estimate for domains and this influences its behavior somewhat.

We also looked at estimating domain sizes. With this type of statistic, the rmse of the weighted estimator is almost uniformly lower than the rmse of the unweighted estimator when the data are not MAR in the simulation settings. The differences are due to the bias in the unweighted estimator of the domain size, and this bias causes the unweighted estimator to have a substantially greater rmse compared to the weighted estimator for some sampling rates.

Imperfect models are used in most surveys so the nonresponse adjustment method is important. The expanded simulation findings we present show the weighted adjustment has substantial advantages for some estimates and for some sampling rates when compared to the unweighted adjustment. In particular, any survey with this design that produces estimates of totals and statistics other than just means appears to benefit by weighting the adjustment. Of course, weighting the adjustment does not remove bias; weighting does diminish the magnitude of the bias in many situations and for many of the estimators we examined. The bias of the weighted estimator also is not sensitive to the relative sampling rate, but the bias of the unweighted estimator is sensitive. The potential disadvantage of an increase in the variance of the estimate using the weighted adjustment did not arise in these simulations, and can be avoided by inspecting the adjustment factors, as should also be done with an unweighted adjustment. Finally, the results of this study highlight the potential problem of generalizing from simulations. Although simulations are valuable to demonstrate a specific point, generalizing simulation findings more broadly can be misleading especially when the findings are highly dependent on the conditions of the model being simulated.

References

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, 251-260.

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics,* 29(2), 329-353.

Brick, J., and Jones, M. (2008). Propensity to respond and nonresponse bias. *Metron-International Journal of Statistics*, LXVI, 51-73.

Chadborn, T.R., Baster, K., Delpech, V., Sabin, C.A., Sinka, K., Rice, B.D. and Evans, B. (2005). No time to wait: How many HIV-infected homosexual men are diagnosed late and consequently die? (England and Wales, 1993-2002). *Aids*, 19(5), 513-520.

Grau, E., Potter, F., Williams, S. and Diaz-Tena, N. (2006). Nonresponse adjustment using logistic regression: To weight or not to weight? *Proceedings of the Survey Research Methods Section*, American Statistical Association, 3073-3080.

Haukoos, J.S., and Newgard, C.D. (2007). Advanced statistics: Missing data in clinical research - part 1: An introduction and conceptual framework. *Academic Emergency Medicine*, 14(7), 662-668.

Kalton, G. (1983). *Introduction to Survey Sampling, SAGE University Paper 35.* Thousand Oaks, CA: SAGE Publications.

Kott, P. (2012). Why one should incorporate the design weights when adjusting for unit nonresponse using response homogeneity groups. *Survey Methodology*, 38, 1, 95-99.

Kovačević, M., and Yung, W. (1997). Variance estimation for measures of income inequality and polarization - An empirical study. *Survey Methodology*, 23, 1, 41-52.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M. and Raghunathan, T.E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, 173(2), 389-407.

Little, R.J. (1986). Survey nonresponse adjustments. *International Statistical Review*, 54, 139-157.

Little, R.J., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data (2nd Ed.).* New York: John Wiley & Sons, Inc.

Little, R., and Vartivarian, S. (2003). On weighting the rates in nonresponse weights. *Statistics in Medicine*, 22, 1589-1599.

R Development Core Team (2011). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. doi: http://www.R-project.org.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse.* Chichester, England: John Wiley & Sons, Inc.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling.* New York: Springer.

Sukasih, A., Jang, D., Vartivarian, S., Cohen, S. and Zhang, F. (2009). A simulation study to compare weighting methods for nonresponses in the National Survey of Recent College Graduates. *Proceedings of the Survey Research Methods Section*, American Statistical Association. Retrieved October 21, 2013, from www.amstat.org/sections/srms/proceedings/y2009/Files/304345.pdf.

Tremblay, V. (1986). Practical criteria for definition of weighting classes. *Survey Methodology*, 12, 1, 85-97.

West, B.T. (2009). A simulation study of alternative weighting class adjustments for nonresponse when estimating a population mean from complex sample survey data. *Proceedings of the Survey Research Methods Section*, American Statistical Association. Retrieved October 21, 2013, from www.amstat.org/sections/srms/proceedings/y2009/Files/305394.pdf.

Wun, L.-M., Ezzati-Rice, T.M., Diaz-Tena, N. and Greenblatt, J. (2007). On modelling response propensity for dwelling unit (DU) level non-response adjustment in the Medical Expenditure Panel Survey (MEPS). *Statistics in Medicine*, 26(8), 1875-1884.