

Does Weighting for Nonresponse Increase the Variance of Survey Means?

Roderick J. Little and Sonya Vartivarian¹

Abstract

Nonresponse weighting is a common method for handling unit nonresponse in surveys. The method is aimed at reducing nonresponse bias, and it is often accompanied by an increase in variance. Hence, the efficacy of weighting adjustments is often seen as a bias-variance trade-off. This view is an oversimplification – nonresponse weighting can in fact lead to a reduction in variance as well as bias. A covariate for a weighting adjustment must have two characteristics to reduce nonresponse bias – it needs to be related to the probability of response, and it needs to be related to the survey outcome. If the latter is true, then weighting can reduce, not increase, sampling variance. A detailed analysis of bias and variance is provided in the setting of weighting for an estimate of a survey mean based on adjustment cells. The analysis suggests that the most important feature of variables for inclusion in weighting adjustments is that they are predictive of survey outcomes; prediction of the propensity to respond is a secondary, though useful, goal. Empirical estimates of root mean squared error for assessing when weighting is effective are proposed and evaluated in a simulation study. A simple composite estimator based on the empirical root mean squared error yields some gains over the weighted estimator in the simulations.

Key Words: Missing data; Nonresponse adjustment; Sampling weights; Survey nonresponse.

1. Introduction

In most surveys, some individuals provide no information because of noncontact or refusal to respond (*unit nonresponse*). The most common method of adjustment for unit nonresponse is weighting, where respondents and nonrespondents are classified into adjustment cells based on covariate information known for all units in the sample, and a nonresponse weight is computed for cases in a cell proportional to the inverse of the response rate in the cell. These weights often multiply the sample weight, and the overall weight is normalized to sum to the number of respondents in the sample. A good overview of nonresponse weighting is Oh and Scheuren (1983). A related approach to nonresponse weighting is post-stratification (Holt and Smith 1979), which applies when the distribution of the population over adjustment cells is available from external sources, such as a Census. The weight is then proportional to the ratio of the population count in a cell to the number of respondents in that cell.

Nonresponse weighting is primarily viewed as a device for reducing bias from unit nonresponse. This role of weighting is analogous to the role of sampling weights, and is related to the design unbiasedness property of the Horvitz-Thompson estimator of the total (Horvitz and Thompson 1952), which weights units by the inverse of their selection probabilities. Nonresponse weighting can be viewed as a natural extension of this idea, where included units are weighted by the inverse of their inclusion

probabilities, estimated as the product of the probability of selection and the probability of response given selection; the inverse of the latter probability is the nonresponse weight. Modelers have argued that weighting for bias adjustment is not necessary for models where the weights are not associated with the survey outcomes, but in practice few are willing to make such a strong assumption.

Sampling weights reduce bias at the expense of increased variance, if the outcome has a constant variance. Given the analogy of nonresponse weights with sampling weights, it seems plausible that nonresponse weighting also reduces bias at the expense of an increase in the variance of survey estimates. The idea of a bias-variance trade-off arises in discussions of nonresponse weighting adjustments (Kalton and Kasprzyk 1986, Kish 1992, Little, Lewitzky, Heeringa, Lepkowski and Kessler 1997). Kish (1992) presents a simple formula for the proportional increase in variance from weighting, say L , under the assumption that the variance of the observations is approximately constant:

$$L = cv^2, \quad (1)$$

where cv is the coefficient of variation of the respondent weights.

Equation (1) is a good approximation when the adjustment cell variable is weakly associated with the survey outcome. However, since it approximates variance rather than mean squared error, it does not measure the potential nonresponse bias reduction that is the main objective of weighting, and it does not apply to outcomes

1. Roderick J. Little, University of Michigan, U.S.A. E-mail: rlittle@umich.edu; Sonya Vartivarian, Mathematica Policy Research, Inc. 600 Maryland Ave SW, Suite 550, Washington, D.C. 20024-2512. E-mail: SVartivarian@Mathematica-MPR.com.

that are associated with the adjustment cell variable, where nonresponse weighting can in fact reduce the variance. The fact that nonresponse weighting can reduce variance is implicit in the formulae in Oh and Scheuren (1983), and is noted in Little (1986) when adjustment cells are created using predictive mean stratification. It is also seen in the related method of post-stratification for nonresponse adjustment (Holt and Smith 1979).

Variability of the weights per se does not necessarily translate into estimates with high variance: an estimate with a high value of L can have a smaller variance than an estimate with a small value of L , as is shown in the simulations in section 3. Also, the situations where nonresponse weighting is most effective in reducing bias are precisely the situations where the weighting tends to reduce, not increase, variance, and Equation (1) does not apply. This differs from the case of sampling weights, and is related to “super-efficiency” that can result when weights are estimated from the sample rather than fixed constants; see, for example, Robins, Rotnitzky and Zhao (1994).

We propose a simple refinement of Equation (1), namely Equation (14) below, that captures both bias and variance components whether or not the adjustment cell variable is associated with the outcome, and hence is a more accurate gauge of the value of weighting the estimates, and of alternative adjustment cell variables. In multipurpose surveys with many outcomes, the standard approach is to apply the same nonresponse weighting adjustment to all the variables, with the implicit assumption that the value of nonresponse bias reduction for some variables outweighs the potential variance increase for others. Our empirical estimate of mean squared error allows a simple refinement of this strategy, namely to restrict nonresponse weighting to the subset of variables for which nonresponse weighting reduces the estimated mean squared error. This composite strategy is assessed in the simulation study in section 3, and shows some gains over weighting all the outcomes. As noted in section 4, there are alternative approaches that have even better statistical properties, but these lead to different weights for each variable and hence are more cumbersome to implement and explain to survey users.

2. Nonresponse Weighting Adjustments for a Mean

Suppose a sample of n units is selected. We consider inference for the population mean of a survey variable Y subject to nonresponse. To keep things simple and focused on the nonresponse adjustment question, we assume that units are selected by simple random sampling. The points made here about nonresponse adjustments also apply in

general to complex designs, although the technical details become more complicated.

We assume that respondents and nonrespondents can be classified into C adjustment cells based on a covariate X . Let M be a missing-data indicator taking the value 0 for respondents and 1 for nonrespondents. Let n_{mc} be the number of sampled individuals with $M = m, X = c, m = 0, 1; c = 1, \dots, C, n_{+c} = n_{0c} + n_{1c}$ denote the number of sampled individuals in cell $c, n_0 = \sum_{c=1}^C n_{0c}$ and $n_1 = \sum_{c=1}^C n_{1c}$ the total number of respondents and nonrespondents, and $p_c = n_{+c} / n, p_{0c} = n_{0c} / n_0$ the proportions of sampled and responding cases in cell c . We compare two estimates of the population mean μ of Y , the unweighted mean

$$\bar{y}_0 = \sum_{c=1}^C p_{0c} \bar{y}_{0c}, \tag{2}$$

where \bar{y}_{0c} is the respondent mean in cell c , and the weighted mean

$$\bar{y}_w = \sum_{c=1}^C p_c \bar{y}_{0c} = \sum_{c=1}^C w_c p_{0c} \bar{y}_{0c}, \tag{3}$$

which weights respondents in cell c by the inverse of the response rate $w_c = p_c / p_{0c}$. The estimator (3) can be viewed as a special case of a regression estimator, where missing values are imputed by the regression of Y on indicators for the adjustment cells. We compare the bias and mean squared error of (2) and (3) under the following model, which captures the important features of the problem. We suppose that conditional on the sample size n , the sampled cases have a multinomial distribution over the $(C \times 2)$ contingency table based on the classification of M and X , with cell probabilities

$$\Pr(M = 0, X = c) = \phi \pi_{0c}; \Pr(M = 1, X = c) = (1 - \phi) \pi_{1c},$$

where $\phi = \Pr(M = 0)$ is the marginal probability of response. The conditional distribution of X given $M = 0$ and n_0 is multinomial with cell probabilities $\Pr(X = c | M = 0) = \pi_{0c}$, and the marginal distribution of X given n is multinomial with index n and cell probabilities

$$\Pr(X = c) = \phi \pi_{0c} + (1 - \phi) \pi_{1c} = \pi_c,$$

say. We assume that the conditional distribution of Y given $M = m, X = c$ has mean μ_{mc} and constant variance σ^2 . The mean of Y for respondents and nonrespondents are

$$\mu_0 = \sum_{c=1}^C \pi_{0c} \mu_{0c}, \mu_1 = \sum_{c=1}^C \pi_{1c} \mu_{1c},$$

respectively, and the overall mean of Y is $\mu = \phi \mu_0 + (1 - \phi) \mu_1$.

Under this model, the conditional mean and variance of \bar{y}_w given $\{p_c\}$ are respectively $\sum_{c=1}^C p_c \mu_{0c}$ and $\sigma^2 \sum_{c=1}^C p_c^2 / n_{0c}$. Hence the bias of \bar{y}_w is

$$b(\bar{y}_w) = \sum_{c=1}^C \pi_c (\mu_{0c} - \mu_c),$$

where π_c and μ_c are the population proportion and mean of Y in cell c . This can be written as

$$b(\bar{y}_w) = \tilde{\mu}_0 - \mu, \tag{4}$$

where $\tilde{\mu}_0 = \sum_{c=1}^C \pi_c \mu_{0c}$ is the respondent mean “adjusted” for the covariates, and $\mu = \sum_{c=1}^C \pi_c \mu_c$ is the true population mean of Y . The variance of \bar{y}_w is the sum of the expected value of the conditional variance and the variance of its conditional expectation, and is approximately

$$V(\bar{y}_w) = (1 + \lambda) \sigma^2 / n_0 + \sum_{c=1}^C \pi_c (\mu_{0c} - \tilde{\mu}_0)^2 / n, \tag{5}$$

where $\lambda = \sum_{c=1}^C \pi_{0c} ((\pi_c / \pi_{0c} - 1)^2)$ is the population analog of the variance of the nonresponse weights $\{w_c\}$, which is the same as L in Equation (1) since the weights are scaled to average to one. The formula for the variance of the weighted mean in Oh and Scheuren (1983), derived under the quasi-randomization perspective, reduces to (5) when the within-cell variance is assumed constant, and finite population corrections and terms of order $1/n^2$ are ignored. The mean squared error of \bar{y}_w is thus

$$\text{mse}(\bar{y}_w) = b^2(\bar{y}_w) + V(\bar{y}_w). \tag{6}$$

The mean squared error of the unweighted mean (2) is

$$\text{mse}(\bar{y}_0) = b^2(\bar{y}_0) + V(\bar{y}_0), \tag{7}$$

where:

$$b(\bar{y}_0) = b(\bar{y}_w) + \mu_0 - \tilde{\mu}_0, \tag{8}$$

is the bias and

$$V(\bar{y}_0) = \sigma^2 / n_0 + \sum_{c=1}^C \pi_{0c} (\mu_{0c} - \mu_0)^2 / n_0, \tag{9}$$

is the variance. Hence the difference (say Δ) in mean squared errors is

$$\Delta = \text{mse}(\bar{y}_0) - \text{mse}(\bar{y}_w) = B + V_1 - V_2, \text{ where}$$

$$B = (\mu_0 - \tilde{\mu}_0)^2 + 2(\mu_0 - \tilde{\mu}_0)(\tilde{\mu}_0 - \mu),$$

$$V_1 = \sum_{c=1}^C \pi_{0c} (\mu_{0c} - \mu_0)^2 / n_0 - \sum_{c=1}^C \pi_c (\mu_{0c} - \tilde{\mu}_0)^2 / n,$$

$$V_2 = \lambda \sigma^2 / n_0 \tag{10}$$

Equation (10) and its detailed interpretation provide the main results of the paper; note that positive terms in (10) favor the weighted estimator \bar{y}_w .

- (a) The first term B represents the impact on MSE of bias reduction from adjustment on the covariates. It is order one and increasingly dominates the MSE as the sample size increases. If $\mu \leq \tilde{\mu}_0 < \mu_0$ or $\mu_0 < \tilde{\mu}_0 \leq \mu$, then weighting has reduced the bias of the respondent

mean, and both of the components of B are positive. In particular, if the missing data are missing at random (Rubin 1976, Little and Rubin 2002), in the sense that respondents are a random sample of the sampled cases in each cell c , then $\tilde{\mu}_0 = \mu$ and weighting eliminates the bias of the unweighted mean. The bias adjustment is

$$\mu_0 - \tilde{\mu}_0 \approx \sum_{c=1}^C \pi_{0c} (1 - w_c) (\mu_{0c} - \mu_0),$$

ignoring differences between the weights and their expectations. This is zero to $O(1)$ if either non-response is unrelated to the adjustment cells (in which case $w_c \approx 1$ for all c , or the outcome is unrelated to the adjustment cells (in which case $\mu_{0c} \approx \mu_0$ for all c). Thus a substantial bias reduction requires adjustment cell variables that are related both to nonresponse and to the outcome of interest, a fact that has been noted by several authors. It is often believed that conditioning on observed characteristics of nonrespondents will reduce bias, but note that this is not guaranteed; it is possible for the adjusted mean to be further on average from the true mean than the unadjusted mean, in which case weighting makes the bias worse.

- (b) The effect of weighting on the variance is represented by $V_1 - V_2$.
- (c) For outcomes Y that are unrelated to the adjustment cells, $\mu_{0c} = \mu_0$ for all c , $V_1 = 0$, and weighting increases the variance, since V_2 is positive. The variance part of equation (10) then reduces to the population version of Kish’s formula (1). Adjustment cell variables that are good predictors of nonresponse hurt rather than help in this situation, since they increase the variance of the weights without any reduction in bias; but there is no bias-variance trade-off for these outcomes, since there is no bias reduction.
- (d) If the adjustment cell variable X is unrelated to non-response, then λ is $O(1/n)$ and hence V_2 has a lower order of variability than V_1 . The term V_1 tends to be positive, since $\sum_{c=1}^C \pi_{0c} (\mu_{0c} - \mu_0)^2 \approx \sum_{c=1}^C \pi_{0c} (\mu_{0c} - \tilde{\mu}_0)^2$, and the divisor n in the second term is larger than the divisor n_0 in the first term. Thus weighting in this case tends to have no impact on the bias, but reduces variance to the extent that X is a good predictor of the outcome. This contradicts the notion that weighting increases variance. The above-mentioned “super-efficiency” that results from estimating non-response weights from the sample is seen by the fact that if the data are missing completely at random, then the “true” nonresponse weight is a constant for all responding units. Hence weighting by “true” weights

leads to (2), which is less efficient than weighting by the “estimated” weights, which leads to (3).

- (e) If the adjustment cell variable is a good predictor of the outcome and also predictive of nonresponse, then V_2 is again small because of the reduced residual variance σ^2 , and V_1 is generally positive by a similar argument to (d). The term $\sum_{c=1}^C \pi_{0c} (\mu_{0c} - \mu_0)^2$ may deviate more from $\sum_{c=1}^C \pi_c (\mu_{0c} - \tilde{\mu}_0)^2$ because the weights are less alike, but this difference could be positive or negative, and the different divisors seem more likely to determine the sign and size of V_1 . Thus, weighting tends to reduce both bias and variance in this case.
- (f) Equation (9) can be applied to the case of post-stratification on population counts, by letting n represent the population size rather than the sample size. Assuming a large population, the second term in V_1 essentially vanishes, increasing the potential for variance reduction when the variables forming the post-strata are predictive of the outcome. This finding replicates previous results on post-stratification (Holt and Smith 1979; Little 1993).

A simple qualitative summary of the results (a) – (f) of section 2 is shown in Table 1, which indicates the direction of bias and variance when the associations between the adjustment cells and the outcome and missing indicator are high or low. Clearly, weighting is only effective for outcomes that are associated with the adjustment cell variable, since otherwise it increases the variance with no compensating reduction in bias. For outcomes that are associated with the adjustment cell variable, weighting increases precision, and also reduces bias if the adjustment cell variable is related to nonresponse.

Table 1
Effect of Weighting Adjustments on Bias and Variance of a Mean, by Strength of Association of the Adjustment Cell Variables with Nonresponse and Outcome

Association with nonresponse	Association with outcome	
	Low	High
Low	Cell 1	Cell 3
	Bias: --- Var: ---	Bias: --- Var: ↓
High	Cell 2	Cell 4
	Bias: --- Var: ↑	Bias: ↓ Var: ↓

It is useful to have estimates of the MSE of \bar{y}_0 and \bar{y}_w that can be computed from the observed data. Let $s_{0c}^2 = \sum_{i \in c} (y_i - \bar{y}_{0c})^2 / (n_{0c} - 1)$ denote the sample variance of respondents in cell c , $s^2 = \sum_{c=1}^C (n_{0c} - 1) s_{0c}^2 / (n_{0c} - C)$ the pooled within-cell variance, and $s_0^2 = \sum_{i=1}^{n_0} (y_i - \bar{y}_0)^2 / (n_0 - 1)$, the total sample variance of the respondent

values. We use the following approximately unbiased expressions, under the assumption that the data are MAR:

$$m\hat{s}e(\bar{y}_0) = \hat{B}^2(\bar{y}_0) + \hat{V}(\bar{y}_0), \tag{11}$$

where $\hat{V}(\bar{y}_0) = s_0^2 / n_0$ and

$$\hat{B}^2(\bar{y}_0) = \max\{0, (\bar{y}_w - \bar{y}_0)^2 - V_d\}$$

$$V_d = (n_1 / n)^2 \left[\sum_{c=1}^C p_{1c} (\bar{y}_{0c} - \bar{y}_0^{(1)})^2 / n_1 + \sum_{c=1}^C p_{0c} (\bar{y}_{0c} - \bar{y}_0)^2 / n_0 + s^2 \sum_{c=1}^C (p_{1c} - p_{0c})^2 / n_{0c} \right], \tag{12}$$

where $\bar{y}_0^{(1)} = \sum_{c=1}^C p_{1c} \bar{y}_{0c}$, and V_d estimates the variance of $(\bar{y}_w - \bar{y}_0)$ and is included in (12) as a bias adjustment for $(\bar{y}_w - \bar{y}_0)^2$ as an estimate of $B^2(\bar{y}_0)$, similar to that in Little *et al.* (1997). Also

$$m\hat{s}e(\bar{y}_w) = \hat{V}(\bar{y}_w) = (1 + L) s^2 / n_0 + \sum_{c=1}^C p_c (\bar{y}_{0c} - \bar{y}_w)^2 / n. \tag{13}$$

Subtracting (11) from (13), the difference in MSE’s of \bar{y}_w and \bar{y}_0 is then estimated by

$$D = L s^2 / n_0 - (s_0^2 - s^2) / n_0 + \sum_{c=1}^C p_c (\bar{y}_{0c} - \bar{y}_w)^2 / n - \hat{B}^2(\bar{y}_0). \tag{14}$$

This is our proposed refinement of (1), which is represented by the leading term on the right side of (14).

3. Simulation Study

We include simulations to illustrate the bias and variance of the weighted and unweighted mean for sets of parameters representing each cell in Table 1. We also compare the analytic MSE approximations in Equations (6) and (7) and their sample-based estimates (11) and (13) with the empirical MSE over repeated samples.

3.1 Superpopulation Parameters

The simulation set-up for the joint distribution of X and M is described in Table 2. The sample is approximately uniformly distributed across the adjustment cell variable X , which has $C = 10$ cells. Two marginal response rates are chosen, 70%, corresponding to a typical survey value, and 52%, a more extreme value to accentuate differences in methods. Three distributions of M given X are simulated to model high, medium and low association.

The simulated distributions of the outcome Y given $M = m, X = c$ are shown in Table 3. These all have the form

$$[Y | M = m, X = c] \sim N(\beta_0 + \beta_1 X, \sigma^2).$$

Table 2
Percent of Sample Cases in Adjustment Cell *X* and Missingness Cell *M*

a. Overall Response Rate = 52%												
	Association Between <i>M</i> and <i>X</i>	<i>X</i>	1	2	3	4	5	6	7	8	9	10
1.	High	<i>M</i> = 0	0.55	1.00	4.01	4.52	5.04	5.55	6.06	6.58	9.14	9.96
		<i>M</i> = 1	8.69	9.00	6.01	5.53	5.04	4.54	4.04	3.54	1.02	0.20
2.	Medium	<i>M</i> = 0	2.77	3.50	4.01	4.52	5.04	5.55	6.06	6.58	7.11	7.62
		<i>M</i> = 1	6.47	6.50	6.01	5.53	5.04	4.54	4.04	3.54	3.05	2.54
3.	Low	<i>M</i> = 0	4.62	5.15	5.21	5.28	5.34	5.40	5.45	5.52	5.58	5.64
		<i>M</i> = 1	4.62	4.85	4.81	4.77	4.73	4.69	4.65	4.60	4.57	4.52

b. Overall Response Rate = 70%												
	Association Between <i>M</i> and <i>X</i>	<i>X</i>	1	2	3	4	5	6	7	8	9	10
1.	High	<i>M</i> = 0	0.55	3.00	6.51	7.04	7.55	8.07	8.59	9.11	9.64	9.96
		<i>M</i> = 1	8.69	7.00	3.51	3.02	2.52	2.02	1.52	1.01	0.51	0.20
2.	Medium	<i>M</i> = 0	4.44	5.30	5.81	6.33	6.85	7.37	7.88	8.40	8.93	9.45
		<i>M</i> = 1	4.80	4.70	4.21	3.72	3.22	2.72	2.22	1.72	1.22	0.71
3.	Low	<i>M</i> = 0	6.19	6.85	6.91	6.98	7.05	7.11	7.17	7.24	7.31	7.37
		<i>M</i> = 1	3.05	3.15	3.11	3.07	3.02	2.98	2.93	2.88	2.84	2.79

Table 3
Parameters for $[Y | M = m, X = c] \sim N(\beta_0 + \beta_1 c, \sigma^2)$

Association Between <i>Y</i> and <i>X</i>	β_1	σ^2	ρ^2
1. High	4.75	46	≈ 0.80
2. Medium	3.70	122	≈ 0.48
3. Low	0.00	234	0.00

Three sets of values of (β_1, σ^2) are simulated to model high, medium and low associations between *Y* and *X*. The intercept β_0 is chosen so that the overall mean of *Y* is $\mu = 26.3625$ for each scenario.

A thousand replicate samples of size $n = 400$ and $n = 2,000$ were simulated for each combination of parameters in Tables 2 and 3. Samples where $n_{0c} = 0$ for any *c* were excluded, since the weighted estimate cannot be computed; in practice some cells would probably be pooled in such cases. The numbers of excluded simulations are shown in Table 4.

Table 4
Numbers of Replicates Excluded Because of Cell with no Respondents

Association of <i>M</i> and <i>X</i>	Association of <i>Y</i> and <i>X</i>	Response Rate	
		52%	70%
High	High	134	113
	Medium	120	117
	Low	131	104
Medium	Low	1	0

3.2 Comparisons of Bias, Variance and Root Mean Squared Error, and their Estimates

Summaries of empirical bias and root MSE's (RMSE's) are reported in Table 5. The empirical RMSE's of the weighted mean can be compared with the following estimates, which are displayed in Table 5, averaged over the 1,000 replicates: The estimated RMSE based on Kish's rule of thumb Equation (1), namely:

$$m\hat{s}e_{\text{Kish}}(\bar{y}_w) = (1 + L)s_Y^2 / n_0,$$

$$\text{where } s_Y^2 = \sum_{i=1}^{n_0} (y_i - \bar{y}_0)^2 / (n_0 - 1); \quad (15)$$

The analytical RMSE from Equations (6) and (7); and the estimated RMSE from Equations (11) and (13).

Following the suggestion of Oh and Scheuren (1983), we include in the last two columns of Table 5 the average empirical bias and RMSE of a composite mean that chooses between \bar{y}_w and \bar{y}_0 , picking the estimate with a lower sample-based estimate of the MSE. The empirical bias relative to the population parameter is reported for all estimators. We also include the bias and RMSE of the mean before deletion of cases due to nonresponse.

Table 5a shows results for simulations with a response rate of 52%. Rows are labeled according to the four cells in Table 1, with medium and high associations combined. For each row, the lower of the RMSE's for the unweighted and weighted respondent means is bolded, indicating superiority for the corresponding method.

The first four rows of Table 5a correspond to cell 4 in Table 1, with medium/high association between *Y* and *X* and

medium/high association between M and X . In these cases \bar{y}_w has much lower RMSE than \bar{y}_0 , reflecting substantial bias of \bar{y}_0 that is removed by the weighting.

The next two rows of Table 5a corresponding to cell 3 of Table 1, with medium/high association between Y and X and low association between M and X . In these cases \bar{y}_0 is no longer seriously biased, but \bar{y}_w has improved precision, particularly when the association of Y and X is high. These are cases where the variance is reduced, not increased, by weighting. The analytic estimates of RMSE and sample-based estimates are close to the empirical RMSE estimates, while Kish's rule of thumb overestimates the RMSE, as predicted by the theory in section 2.

The next two rows of Table 5a correspond to cell 2 of Table 1, where the association between Y and X is low and the association between M and X is medium or high. In these cases, \bar{y}_w has higher MSE than \bar{y}_0 . These cases illustrate situations where the weighting increases variance, with no compensating reduction in bias. The last row corresponds to cell 1 of Table 1, with low associations between M and X and between Y and X . The unweighted mean has lower RMSE in these cases, but the increase in RMSE from weighting is negligible. For the last three rows of Table 5a, RMSE's from Kish's rule of thumb are similar

to those from the analytical formula in section 2 and empirical estimates based on this formulae, and all these estimates are close to the empirical RMSE.

The last two columns of Table 5a show empirical bias and RMSE of the composite method that chooses \bar{y}_w or \bar{y}_0 based on the estimated RMSE. For the simulations in the first 6 rows, the composite estimator is the same as \bar{y}_w , and hence detects and removes the bias of the unweighted mean. For simulations in cell 1 (the last row) the composite estimator performs like \bar{y}_w or \bar{y}_0 , as expected since \bar{y}_w and \bar{y}_0 perform similarly in this case. For simulations in cell 2 that are not favorable to weighting, the composite estimator has lower RMSE than \bar{y}_w , but considerably higher than that of \bar{y}_0 , suggesting that for the conditions of this simulation the empirical MSE affords limited ability to pick the better estimator in individual samples.

Nevertheless, the composite estimator is the best overall estimator of the three considered in this simulation.

Table 5b shows results for the 70% response rate. The pattern of results is very similar to that of Table 5a. As expected, differences between the methods are smaller, although they remain substantial in many rows of the table.

Table 5a
Summaries of Estimators Based on 1,000 Replicate Samples for $C = 10$ Adjustment Cells, Restricted to Sample Replicates with $n_{0c} > 0$ for all c . Response Rate of 52%. Values are Multiplied by 1,000

Cell	Association with Adjustment Cells Based on X			Unweighted Mean				Weighted Mean					Before Deletion Mean		Composite Mean	
	(M, X)	(Y, X)	n	emp. bias	emp. rmse	analytical rmse ¹	est. rmse ²	emp. bias	emp. rmse	Kish rmse ³	analytical rmse ⁴	est. rmse ⁵	emp. bias	emp. rmse	emp. bias	emp. rmse
4	High	High	400	6,955	7,024	7,055	6,974	0	1,057	1,410	956	988	-38	795	0	1,057
			2,000	7,008	7,020	7,006	7,015	-2	424	608	427	434	12	342	-2	424
4	High	Medium	400	5,376	5,471	5,536	5,404	-33	1,264	1,510	1,216	1,297	-21	776	-33	1,264
			2,000	5,424	5,441	5,466	5,466	-41	561	650	545	559	-30	338	-41	561
4	Medium	High	400	3,664	3,794	3,809	3,754	-4	816	1,071	835	842	6	741	-4	816
			2,000	3,703	3,731	3,700	3,712	7	369	473	373	374	4	337	7	369
4	Medium	Medium	400	2,838	3,006	3,042	2,991	-18	938	1,095	954	970	-9	747	-18	938
			2,000	2,864	2,900	2,898	2,893	-2	426	483	426	428	6	335	-2	426
3	Low	High	400	476	1,148	1,113	1,178	40	823	1,050	823	828	30	764	40	823
			2,000	376	587	614	595	-11	361	465	368	368	-3	333	-11	361
3	Low	Medium	400	350	1,106	1,095	1,134	13	927	1,063	925	939	-16	762	13	927
			2,000	287	565	563	559	-20	429	470	413	414	-22	353	-20	429
2	High	Low(0)	400	56	1,070	1,056	1,275	96	1,658	1,613	1,518	1,631	28	793	83	1,410
			2,000	-11	464	473	567	-26	698	698	679	699	-19	337	-25	620
2	Medium	Low(0)	400	9	1,042	1,053	1,077	-27	1,122	1,112	1,097	1,125	21	772	-12	1,074
			2,000	-4	474	471	480	-11	491	491	491	493	11	340	-9	481
1	Low	Low(0)	400	-30	1,038	1,050	1,055	-30	1,053	1,064	1,050	1,076	-30	752	-30	1,040
			2,000	-2	472	469	469	-1	474	470	469	471	-8	343	-1	472

¹ Computed using Equation (7)
² Computed using Equation (11)
³ Computed using Equation (15)
⁴ Computed using Equation (6)
⁵ Computed using Equation (13)

Table 5b
 Summaries of Estimators based on 1,000 Replicate Samples for $C = 10$ Adjustment Cells, Restricted to Sample Replicates with $n_{0c} > 0$ for all c . Response Rate of 70%. Values are Multiplied by 1,000

Association with Adjustment Cells based on X				Unweighted Mean				Weighted Mean				Before Deletion Mean		Composite Mean		
Cell	(M, X)	(Y, X)	n	emp. bias	emp. rmse	analytical rmse ⁶	est. rmse ⁷	emp. bias	emp. rmse	Kish rmse ⁸	analytical rmse ⁹	est. rmse ¹⁰	emp. bias	emp. rmse	emp. bias	emp. rmse
4	High	High	400	4,692	4,810	4,893	4,860	-133	1,129	1,192	889	894	-129	998	-133	1,129
			2,000	4,827	4,841	4,839	4,854	-20	400	529	398	405	-5	334	-20	400
4	High	Medium	400	3,581	3,716	3,855	3,733	-133	1,266	1,250	1,075	1,097	-128	917	-127	1,284
			2,000	3,763	3,784	3,778	3,777	-9	501	554	481	490	11	343	-9	501
4	Medium	High	400	2,666	2,812	2,878	2,837	-58	803	910	794	796	-49	772	-58	803
			2,000	2,732	2,760	2,767	2,761	-6	353	406	355	355	-9	333	-6	353
4	Medium	Medium	400	2,104	2,282	2,315	2,291	-28	833	924	854	861	-43	751	-28	833
			2,000	2,146	2,180	2,170	2,165	13	370	411	382	382	10	334	13	370
3	Low	High	400	217	906	954	980	-81	797	911	790	793	-77	771	-81	797
			2,000	312	513	506	502	2	365	405	353	353	4	349	2	365
3	Low	Medium	400	251	922	942	960	15	804	916	845	852	26	727	15	804
			2,000	224	454	472	471	-14	370	408	378	379	-15	327	-14	370
2	High	Low(0)	400	0	952	915	1,131	35	1,445	1,349	1,298	1,358	1	807	26	1,292
			2,000	-11	416	409	485	-41	608	598	580	599	-4	347	-31	535
2	Medium	Low(0)	400	22	911	910	920	24	942	936	930	946	2	757	21	925
			2,000	23	418	407	411	20	425	416	416	417	15	344	19	420
1	Low	Low(0)	400	1	914	914	912	2	917	916	914	926	-5	751	1	914
			2,000	4	402	408	408	4	403	409	408	410	6	331	4	402

⁶ Computed using Equation (7)
⁷ Computed using Equation (11)
⁸ Computed using Equation (15)
⁹ Computed using Equation (6)
¹⁰ Computed using Equation (13)

4. Discussion

The results in sections 2 and 3 have important implications for the use of weighting as an adjustment tool for unit nonresponse. Surveys often have many outcome variables, and the same weights are usually applied to all these outcomes. The analysis of section 2 and simulations in section 3 suggests that improved results might be obtained by estimating the MSE of the weighted and unweighted mean and confining weighting to cases where this relationship is substantial. A more sophisticated approach is to apply random-effects models to shrink the weights, with more shrinkage for outcomes that are not strongly related to the covariates (e.g., Elliott and Little 2000). A flexible alternative to this approach is imputation based on prediction models, since these models allow for interval-scaled as well as categorical predictors, and allow interactions to be dropped to incorporate more main effects. Multiple imputation (Rubin 1987) can be used to propagate uncertainty.

When there is substantial covariate information, one attractive approach to generalizing weighting class adjustments is to create a propensity score for each respondent based on a logistic regression of the nonresponse indicator on the covariates, and then create adjustment cells based on this score. Propensity score methods were originally

developed in the context of matching cases and controls in observational studies (Rosenbaum and Rubin 1983), but are now quite commonly applied in the setting of unit nonresponse (Little 1986; Czajka, Hirabayashi, Little and Rubin 1987; Ezzati and Khare 1992). The analysis here suggests that for this approach to be productive, the propensity score has to be predictive of the outcomes. Vartivarian and Little (2002) consider adjustment cells based on joint classification by the response propensity and summary predictors of the outcomes, to exploit residual associations between the covariates and the outcome after adjusting for the propensity score. The requirement that adjustment cell variables predict the outcomes lends support to this approach.

The analysis presented here might be extended in a number of ways. Second order terms in the variance are ignored here, which if included would penalize weighting adjustments based on a large number of small adjustment cells. Finite population corrections could be included, although it seems unlikely that they would affect the main conclusions. It would be of interest to see to what extent the results can be generalized to complex sample designs involving clustering and stratification. Also, careful analysis of the bias and variance implications of nonresponse weighting on statistics other than means, such as subclass means or regression coefficients, would be worthwhile. We

expect it to be important that adjustment cell variables predict the outcome in many of these analyses too, but other points of interest may emerge.

Acknowledgements

This research is supported by grant SES-0106914 from the National Science Foundation. We thank an associate editor and three referees for useful comments on earlier drafts.

References

- Czajka, J.L., Hirabayashi, S.M., Little, R.J.A. and Rubin, D.B. (1987). Evaluation of a new procedure for estimating income aggregates from advance data. In *Statistics of Income and Related Administrative Record Research: 1986-1987*, U.S. Department of the Treasury, 109-136.
- Elliott, M.R., and Little, R.J.A. (2000). Model-based alternatives to trimming survey weights. *Journal of Official Statistics*, 16, 191-209.
- Ezzati, T., and Khare, M. (1992). Nonresponse adjustments in a National Health Survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 339-344.
- Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association*, 47, 663-685.
- Holt, D., and Smith, T.M.F. (1979). Post stratification. *Journal of the Royal Statistical, A*, 142, 33-46.
- Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.
- Kish, L. (1992). Weighting for unequal P_i . *Journal of Official Statistics*, 8, 183-200.
- Little, R.J.A. (1986). Survey nonresponse adjustments. *International Statistical Review*, 54, 139-157.
- Little, R.J.A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association*, 88, 1001-1012.
- Little, R.J.A., Lewitzky, S., Heeringa, S., Lepkowski, J. and Kessler, R.C. (1997). An assessment of weighting methodology for the national comorbidity study. *American Journal of Epidemiology*, 146, 439-449.
- Little, R.J.A., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd edition. New York: John Wiley & Sons, Inc.
- Oh, H.L., and Scheuren, F.S. (1983). Weighting adjustments for unit nonresponse. In *Incomplete Data in Sample Surveys*, 2, Theory and Bibliographies, (Eds. W.G. Madow, I. Olkin and D.B. Rubin), Academic Press, New York, 143-184.
- Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846-866.
- Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.
- Vartivarian, S., and Little, R.J.A. (2002). On the formation of weighting adjustment cells for unit nonresponse. *Proceedings of the Survey Research Methods Section*, American Statistical Association.