University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2004 *Paper* 35

Does Weighting for Nonresponse Increase the Variance of Survey Means?

Rod Little[*] Sonya L. Vartivarian[†]

[*]University of Michigan, rlittle@umich.edu

[†]University of Michigan, slvartiv@umich.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/umichbiostat/paper35

Copyright ©2004 by the authors.

Does Weighting for Nonresponse Increase the Variance of Survey Means?

Rod Little and Sonya L. Vartivarian

Abstract

Nonresponse weighting is a common method for handling unit nonresponse in surveys. A widespread view is that the weighting method is aimed at reducing nonresponse bias, at the expense of an increase in variance. Hence, the efficacy of weighting adjustments becomes a bias-variance trade-off. This note suggests that 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.

2nd Draft April 30, 2004

Does Weighting for Nonresponse Increase the Variance of Survey Means?

Roderick J. Little and Sonya Vartivarian
University of Michigan

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

ABSTRACT

Nonresponse weighting is a common method for handling unit nonresponse in surveys. A widespread view is that the weighting method is aimed at reducing nonresponse bias, at the expense of an increase in variance. Hence, the efficacy of weighting adjustments becomes a bias-variance trade-off. This note suggests that 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.

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

COBRA
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/umichbiostat/paper35

available from external sources, such as a Census. The weight is then computed as the inverse of the ratio of the number of respondents in a cell to the population count in that cell.

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. (More generally, with variances that vary across sampling strata, deviations from Neyman allocation lead to a loss of precision from weighting). 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. This idea is well established in the survey sampling literature. For example, Kalton and Kasprzyk (1986) write:

> "As with population weighting adjustments, the aim of sample weighting adjustments is to reduce the bias that nonresponse may cause in survey estimates. An effect of sample weighting adjustments is, however, to increase the variance of the survey estimates. There is therefore a trade-off to be made between bias reductions and variance increase."

In another widely cited paper, Kish (1992) writes:

> "*Increased variances* can result from weighting for random, or haphazard, or irregular differences in selection probabilities, when these are not "optimal". For example, the inequalities due to frame problems or nonresponses are generally of this kind" {italics due to author}.

Kish also presents a simple formula for the proportional increase in variance, say L, under the assumption that the variance of the observations is approximately constant:

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive
Hosted by The Berkeley Electronic Press

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

where cv is the coefficient of variation of the respondent weights. This formula is a reasonable approximation for outcomes that are weakly associated with the adjustment cell variable, and one of us has used this formula in attempting to model variance increase from weighting adjustments that include a substantial nonresponse component (Little et al. 1997). However, even when (1) is a good approximation for the increase in variance, it does not reflect the potential reduction in mean squared error from bias reduction, and hence does not allow the bias/variance tradeoff to be explicitly assessed. We propose a refinement of Eq. (1) that captures both bias and variance components, and reflects the variance reduction that occurs when the adjustment cell variable is strongly associated with the outcome.

If nonresponse weights are formed as the inverse of the response rates within adjustment cells, then increasing the number of these cells by including more covariates typically results in more variable weights, and may result in an increase in the variance of the weighted estimates. For example, Kalton and Kasprzyk (1986) write:

> "A large variance in the weights can arise from segmenting the sample into many weighting classes with only a few sampled elements in each. When the weighting classes are small, their response rates are unstable, and this gives rise to large variation in the weights. To avoid this effect, it is common practice to limit the extent to which the sample is segmented… These procedures avoid the increase in variance associated with the use of extreme weights, but they may lead to increased bias"

We agree that excessive over-segmentation of the sample is not a good idea and does tend to increase the variance of survey estimates, but 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 will be shown below.

More generally, the point of this article is to show that nonresponse weighting does *not* necessarily result in increased variance; indeed the situations where nonresponse weighting is most effective in reducing bias are precisely the situations where the weighting tends to reduce, not increase, variance. 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

COBRA
Collection of Biostatistics
Research Archive
http://biostats.bepress.com/umichbiostat/paper35

constants; see, for example, Robins, Rotnitsky and Zhao (1994). That the argument that weighting increases variance is an oversimplification can be seen in the case of post-stratification, a weighting method that is often used for nonresponse adjustment, and results in estimators with reduced variance if the post-stratifiers are associated with the outcome (Holt and Smith 1979). Post-stratification is included in our analysis as a special case.

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 $M = m, X = c, m = 0,1;\ c = 1,..., C$, $n_{+c} = n_{0c} + n_{1c}$ denote the number of sampled individuals in cell c, $n_0 = \sum_{c=1}^{C} n_{oc}$ 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 m of Y, the unweighted mean

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

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

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

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

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

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

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

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

$$\Pr(X = c) = f p_{0c} + (1 - f) p_{1c} = p_c\,,$$

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

$$m_0 = \sum_{c=1}^{C} p_{0c} m_{0c}\,,\ m_1 = \sum_{c=1}^{C} p_{1c} m_{1c}\,,$$

respectively, and the overall mean of *Y* is $m = f m_0 + (1 - f) m_1$.

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

$$b(\bar{y}_w) = \sum_{c=1}^{C} p_c (m_{0c} - m_c)\,,$$

where p_c and m_c are the population proportion and mean of *Y* in cell *c*. This can be written as

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

where $\tilde{m}_0 = \sum_{c=1}^{C} p_c m_{0c}$ is the respondent mean "adjusted" for the covariates, and $m = \sum_{c=1}^{C} p_c m_c$ 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

$$V(\bar{y}_w) = (1 + l) s^2 / n_0 + \sum_{c=1}^{C} p_c (m_{0c} - \tilde{m}_0)^2 / n\,, \tag{5}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/umichbiostat/paper35

where $l = \sum_{c=1}^{C} p_{0c} \left((p_c / p_{0c} - 1)^2 \right)$ is the population analog of the variance of the nonresponse weights $\{w_c\}$, which is the same as L in Eq. (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 variation 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

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

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

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

where:

$$b(\bar{y}_0) = b(\bar{y}_w) + m_0 - \tilde{m}_0 , \tag{8}$$

is the bias and

$$V(\bar{y}_0) = s^2 / n_0 + \sum_{c=1}^{C} p_{0c} (m_{0c} - m_0)^2 / n_0 , \tag{9}$$

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

$$\begin{aligned} &\Delta = mse(\bar{y}_0) - mse(\bar{y}_w) = B + V_1 + V_2, \text{ where} \\ &B = (m_0 - \tilde{m}_0)^2 + 2(m_0 - \tilde{m}_0)(\tilde{m}_0 - m), \\ &V_1 = \sum_{c=1}^{C} p_{0c} (m_{0c} - m_0)^2 / n_0 - \sum_{c=1}^{C} p_c (m_{0c} - \tilde{m}_0)^2 / n, \\ &V_2 = -l s^2 / n_0 \end{aligned} \tag{10}$$

Eq. (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 $m \le \tilde{m}_0 < m_0$ or $m_0 < \tilde{m}_0 \le m$, 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

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

sampled cases in each cell c, then $\tilde{m}_0 = m$ and weighting eliminates the bias of the unweighted mean. The bias adjustment is

$$m_0 - \tilde{m}_0 \simeq \sum_{c=1}^{C} p_{0c}(w_c - 1)(m_{0c} - m_0),$$

ignoring differences between the weights and their expectations. This is zero to $O(1)$ if either nonresponse 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 $m_{0c} \approx m_0$ 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, $m_{0c} = m_0$ for all c, $V_1 = 0$, and weighting increases the variance, since V_2 is negative. Eq. (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 nonresponse but is a good predictor of an outcome, then V_2 tends to be small, since the residual variance s^2 of Y is small relative to the variation of the means across the adjustment cells. The term V_1 is then positive, since $\sum_{c=1}^{C} p_{0c}(m_{0c} - m_0)^2 \simeq \sum_{c=1}^{C} p_c(m_{0c} - \tilde{m}_0)^2$, and the divisor $1/n$ in the (negative) second term in V_1 is smaller than the divisor $1/n_0$ in the first (positive) term. Thus weighting in this case tends to have no impact on the bias, but reduces variance. This contradicts the notion that weighting increases variance. The above-mentioned "super-efficiency" that results from estimating nonresponse weights from the sample is seen by the fact that if the data are missing completely

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive
http://biostats.bepress.com/umichbiostat/paper35

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 s^2, and V_1 is generally positive by a similar argument to (d). The term $\sum_{c=1}^{C} p_{0c}(m_{0c} - m_0)^2$ may deviate more from $\sum_{c=1}^{C} p_c(m_{0c} - \tilde{m}_0)^2$ 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) Eq. (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).

Table 1 about here

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 variables is related to nonresponse.

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, $s^2 = \sum_{c=1}^{C}(n_{0c} - 1)s_{0c}^2/(n_0 - C)$ the pooled within-cell variance, and

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$s_0^2 = \sum_{i=1}^{n_0} (y_i - \overline{y}_0)^2 / (n_0 - 1)$ the total sample variance of the respondent values. We use the following approximately unbiased expressions, under the assumption that the data are MAR:

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

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

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

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

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

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

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

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 Eqs. (6) and (7) and their sample-based estimates (11) and (13) with the empirical MSE over repeated samples.

Tables 2-4 about here

3.1 Superpopulation Parameters

The simulation set-up is described in Tables 2 and 3. The sample is approximately uniformly distributed across the adjustment cell variable *X*, which has $C = 10$ cells. The distribution of *M* given *X* was chosen to model high and low associations, as shown in Table 4.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/umichbiostat/paper35

In row 1 the association is high and $L = 1.064$; in row 2 the association is low and $l = 0.001$. The conditional distribution of the outcome *Y* given $M = m, X = c$ is simulated as

$$[Y \mid M = m, X = c] \sim N(b_0 + b_1 X, s^2),$$

for three sets of values of (b_1, s^2) chosen to represent outcomes *Y* with high, medium and low associations with X (Table 3). The intercept b_0 is chosen so that the overall mean of *Y* is $m = 26.3625$ for each scenario. The overall response rate is approximately 52%. We choose 1000 replicate samples for each combination of scenarios in Tables 2 and 3. We exclude samples where $n_{0c} = 0$ for all *c*, resulting in 134, 120 and 131 omitted replicates corresponding respectively to High, Medium and Low association between *Y* and *X* for the smaller sample size *n* = 400.

3.2 Comparisons of Bias, Variance and Mean Squared Error, and Their Estimates

Summaries of empirical bias and root MSE's are reported in Table 4. In this table the analytical root MSE's in Eqs. (6) and (7) and the sample-based estimated root MSE's of Eq. (11) and (13), averaged over the 1000 replicate samples, can be compared with the empirical root MSE over the replicate samples. We also include the estimated root MSE of the weighted mean based on Kish's rule of thumb (1):

$$m\hat{s}e_{Kish}(\bar{y}_w) = (1+L)s_Y^2 / n_0, \text{ where } s_Y^2 = \sum_{i=1}^{n_0} (y_i - \bar{y}_0)^2 / (n_0 - 1), \quad (15)$$

Following the suggestion of Oh and Scheuren (1983), we include in Table 4 the average empirical bias and root MSE 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 root MSE of the mean before deletion of cases due to nonresponse

In the four cases corresponding to cells 2 and 4 in Table 1, with moderate or high correlation between *Y* and *X*, $\bar{y}_w$ has lower root MSE than $\bar{y}_0$; the root MSE for $\bar{y}_w$ is bolded

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

for these cases in Table 4. When the association between *M* and *X* is high, the improvement in MSE from weighting is attributable to removal of the bias of $\bar{y}_0$. When the association between *M* and *X* is low, $\bar{y}_0$ is no longer biased, but $\bar{y}_w$ has improved precision. These situations illustrate cases where the variance is reduced, not increased, by weighting. In these cases, the analytic estimates of MSE and sample-based estimates are close to the empirical estimates, while Kish's formula overestimates the MSE, as predicted by the theory in Section 2. The overestimation is greater for the larger than for the smaller sample size.

There are two cases where the correlation between *Y* and *X* is low and the association between *M* and *X* is high, corresponding to cell 3 of Table 1. In these cases $\bar{y}_w$ has higher MSE than $\bar{y}_0$, and the MSE for $\bar{y}_0$ is bolded for these cases. These cases illustrate situations where the weighting increases variance, with no compensating reduction in bias. Finally, there are two cases corresponding to cell 1 of Table 1 where the correlation between *Y* and *X* is low and the association between *M* and *X* is low. In these cases the root MSE for $\bar{y}_w$ and $\bar{y}_0$ are similar. Note that for cases in cell 1 and 3, root MSE's from Kish's formula are similar to those from the analytical formula in Section 2 and empirical estimates based on this formulae, and all these formulae are close to the empirical root MSE.

The composite method that chooses $\bar{y}_w$ or $\bar{y}_0$ based on the estimated MSE always chooses the superior estimator $\bar{y}_w$ for the simulations in cells 2 and 4, where the bias of $\bar{y}_0$ inflates its MSE. For simulations in cell 1 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 3 that are not favorable to weighting, the composite estimator has lower root MSE 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.

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

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/umichbiostat/paper35

same weights are usually are 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 et al. 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

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

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 an earlier draft.

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 1992,* 339-344.

Horvitz, D. *G*., and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite population, *J. Am. Statist. Assoc.* **47**, 663-685.

Holt, D. and Smith, T. M. F. (1979). Post stratification, *J. Roy. Statist. Soc. 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.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/umichbiostat/paper35

Little, R.J.A. (1993). Post-stratification: a modeler's perspective, *J. Am. Statist. Assoc.* **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. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data,* 2nd edition. Wiley, New York.

Oh, H.L. & Scheuren, F.S. (1983). Weighting adjustments for unit nonresponse, in *Incomplete Data in Sample Surveys*, Vol. 2, Theory and Bibliographies, W.G. Madow, I. Olkin & D.B. Rubin, eds., Academic Press, New York. pp. 143-184.

Robins, J.M., Rotnitsky, 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. & 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*, Wiley, New York.

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 2002.*

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

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

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

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Table 2. Association Between Adjustment Cell *X* and Missingness *M*

Association Between M and X		X	1	2	3	4	5	6	7	8	9	10
1.	High	$pr(M=0 \mid X=c)$	0.06	0.1	0.4	0.45	0.5	0.55	0.6	0.65	0.9	0.98
		$p_{0c} = pr(X=c \mid M=0)$	0.0106	0.0191	0.0765	0.0863	0.0961	0.1059	0.1157	0.1255	0.1743	0.1900
2.	Low	$pr(M=0 \mid X=c)$	0.5	0.515	0.52	0.525	0.53	0.535	0.54	0.545	0.55	0.555
		$p_{0c} = pr(X=c \mid M=0)$	0.0869	0.0968	0.0980	0.0992	0.1004	0.1015	0.1026	0.1037	0.1050	0.1060

Table 3. Parameters for $[Y \mid M=m, X=c] \sim N(b_0 + b_1 X, s^2)$

Association Between Y and X		b_1	s^2	r^2
1.	High	4.75	46	≈ 0.8
2.	Medium	3.7	122	≈ 0.48
3.	Low	0	234	0

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Table 4. Summaries of Estimators based on 1000 Replicate Samples for *C* = 10 Adjustment Cells, restricted to Sample Replicates with $n_{cR} > 0$ for all *c*. Values are multiplied by 1000.

Association With Adjustment Cells Based on X				***Respondent Mean***				***Weighted Mean***					***Before Deletion Mean***		***Composite Mean***	
Cell	(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	6955	7024	7055	6974	0	**1057**	1410	956	988	-38	795	0	1057
			2000	7008	7020	7006	7015	-2	**424**	608	427	434	12	342	-2	424
	High	Medium	400	5376	5471	5536	5404	-33	**1264**	1510	1216	1297	-21	776	-33	1264
			2000	5424	5441	5466	5466	-41	**561**	650	545	559	-30	338	-41	561
3	High	Low (0)	400	56	**1070**	1056	1275	96	1658	1613	1518	1631	28	793	83	1410
			2000	-11	**464**	473	567	-26	698	698	679	699	-19	337	-25	620
2	Low	High	400	476	1148	1113	1178	40	**823**	1050	823	828	30	764	40	823
			2000	376	587	614	595	-11	**361**	465	368	368	-3	333	-11	361
	Low	Medium	400	350	1106	1095	1134	13	**927**	1063	925	939	-16	762	13	927
			2000	287	565	563	559	-20	**429**	470	413	414	-22	353	-20	429
1	Low	Low (0)	400	-30	1038	1050	1055	-30	1053	1064	1050	1076	-30	752	-30	1040
			2000	-2	472	469	469	-1	474	470	469	471	-8	343	-1	472

---

[1] Computed using Eq. (7)

[2] Computed using Eq. (11)

[3] Computed using Eq. (15)

[4] Computed using Eq. (6)

[5] Computed using Eq. (13)

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive