ON WEIGHTING THE RATES IN NONRESPONSE WEIGHTS.

Roderick J. Little, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor MI 48109-2029, U.S.A. E-mail: rlittle@umich.edu
Sonya Vartivarian, Department of Statistics, University of Michigan

Key Words: sampling weights, survey inference, unit nonresponse adjustment.

SUMMARY

A basic estimation strategy in sample surveys is to weight units inversely proportional to the probability of selection and response. Response weights in this method are usually estimated by the inverse of the sample-weighted response rate in an adjustment cell, that is, the sum of the sampling weights of respondents in a cell to the sum of the sampling weights for respondents and nonrespondents in that cell. We show by simulations that weighting the response rates by the sampling weights to adjust for design variables is either incorrect or unnecessary. It is incorrect, in the sense of yielding biased estimates of population quantities, if the design variables are related to survey nonresponse; it is unnecessary if the design variables are unrelated to survey nonresponse. The correct approach is to model nonresponse as a function of the adjustment cell *and* design variables, and to estimate the response weight as the inverse of the estimated response probability from this model. This approach can be implemented by creating adjustment cells that include design variables in the crossclassification, if the number of cells created in this way is not too large. Otherwise, response propensity weighting can be applied.

1. INTRODUCTION

Weighting is the standard method of nonresponse adjustment for surveys subject to unit nonresponse, where entire interviews are missing due to noncontact or refusal to answer the questionnaire. Respondents and nonrespondents are classified into adjustment cells based on covariate information recorded for both groups, and respondents in cell c are weighted by the inverse of the response rate in cell c. The sampling weight for each respondent is then multiplied by this nonresponse weight to obtain a combined weight for subsequent analysis.

Weighting for nonresponse is a natural extension of weighting for sample selection. The sampling weight (say π_i^{-1}) of a sampled unit *i* is the inverse of the probability of selection, and can be interpreted as the number of units in the population that unit *i* is "representing". In particular suppose y_i is the value of a survey variable *Y*, and *T* is the population total of *Y*. In the absence of nonresponse a natural estimator of *T* is the Horvitz-Thompson (HT) estimator[1] $\sum \pi_i^{-1} y_i$, where the sum is over sampled units. The HT estimator is an unbiased estimator of *T* with respect to the randomization distribution. Although it can have unacceptably high variance[2], it is a useful all-purpose estimator in large samples.

In the presence of nonresponse, let π_i^{-1} be the sampling weight and w_i the nonresponse weight for responding unit *i*. The product $\pi_i^{-1} w_i$ can be interpreted as the number of units in the population represented by unit *i*. An obvious extension of the HT estimator of *T* is then

$$\hat{T} = \sum \pi_i^{-1} w_i y_i \qquad (1)$$

where the sum is over units that are sampled and respond. This estimate is approximately design unbiased for *T*, provided the respondents in cell *c* are a random subsample of the sampled units in cell *c*. Since covariate information for unit nonresponse adjustments is often limited, this proviso is often a hope rather than an expectation. However, it is often plausible that unit nonresponse adjustment at least reduces the bias.

Note that w_i is a sample quantity estimated from data, unlike the sample weight π_i^{-1} which is determined by the sample design. This paper concerns the form of w_i for unequal probability samples, more precisely for samples where the sampling weights are not constant within the adjustment cells. Suppose respondent *i* falls in adjustment cell *c*. A naive choice is $w_i = 1/\hat{\phi}_c$, where $\hat{\phi}_c$ is the unweighted response rate

$$\hat{\phi}_c = r_{c+} / n_{c+}, \qquad (2)$$

and n_{c+} and r_{c+} are the number of sampled and responding individuals in cell *c*. If the sample weights are not constant within cell *c*, $\hat{\phi}_c$ is not an unbiased estimate of the population response rate in cell *c* (that is, the proportion of the population that would respond if sampled). An (at least approximately) unbiased estimate of this quantity is the weighted response rate

$$\hat{\phi}_c = \sum_{k \in R_c} \pi_k^{-1} / \sum_{k \in S_c} \pi_k^{-1}, \qquad (3)$$

where S_c and R_c denote the set of units in cell *c* that are sampled, and the set that are sampled and respond, respectively. One might compute the nonresponse weight in (1) as the inverse of the weighted rate in (3).

Should response rates be weighted, as in Eq. (3), or not weighted, as in Eq. (2)? Platek and Gray[3] discusses both methods, but draw no conclusions about which is to be preferred in practice. In a review of Census Bureau adjustment procedures, Chapman, Bailey and Kasprzyk[4] state that "nonresponse adjustment factors are usually either the inverse of the survey's unweighted nonresponse rate, or an analogous ratio based on weighted survey counts". Practice appears to favor weighted response rates. A recent enquiry to the list server for the Survey Research Methods Section of the American Statistical Association suggests that weighted response rates (3) are routinely used by the major survey research organizations. For example, the survey design for the National Health Interview Survey[5] oversamples Black and Hispanic households relative to other of races within Secondary Sampling Units (SSU's), and then computes weighted response weights (3) within adjustment cells consisting of SSU's. To judge from their description, the nonresponse weights for the National Crime Survey appear to be unweighted, but cross-sectional weighting adjustments for the Survey of Income and Program Participation are currently weighted[6].

We argue in this paper that neither of these approaches is correct. The correct approach is to use the inverse of the unweighted rate (2), for adjustment cells that condition on both covariate <u>and</u> design information. In essence, the argument is that (a) adjustment cells should be created to be homogeneous with respect to the propensity to respond; (b) if adjustment cells are created in this way, then weighting the nonresponse rates is unnecessary and inefficient, that is, it adds variance to estimates; and (c) if adjustment cells are created that are not homogeneous with respect to the propensity to respond, then weighting the response rates does <u>not</u> yield unbiased estimates of the means of population outcomes, even though the weighted response rates are unbiased estimates of the population response rates within each adjustment cell. Given (c), the right approach is not to weight the nonresponse rates, but rather to create adjustment cells based on a classification of the observed variables and the survey design variables, to control for association between survey stratifiers and nonresponse. Section 2 provides simulations in support of these statements. On the other hand joint stratification on the adjustment cell variable and *Z* may achieve reduced nonresponse bias at the expense of increased variance, if the resulting adjustment cells are too sparse; approaches to that problem are discussed in Section 3.

2. SIMULATION STUDY

A simulation study was conducted to provide more insights into the variance and bias of estimators (2), (3) and alternatives, under a variety of population structures and nonresponse mechanisms. Categorical variables were simulated to avoid the need for distributional assumptions such as normality.

<u>Description of the Population</u>: A population of size N=10,000 was generated on a binary stratifier *Z*, observed for all units of the population, a binary adjustment cell variable *X* observed for the sample, and a binary survey outcome *Y* observed only for unit respondents. Also let *S* denote the sampling indicator, observed for all units in the population, and *R* the response indicator, observed for all units in the sample. The joint distribution of these variables, say $[Z, X, Y, S, R]$, can be factorized as follows:

$$[Z, X, Y, S, R] = [Z \mid X][Y \mid Z, X][S \mid Z, X, Y][R \mid Z, X, Y, S];$$

The distributions on the right side are then defined as follows:

(i) Distribution of *Z* and *X*.

The joint distribution of [*Z*,*X*] was multinomial, with $pr(Z = X = 0) = 0.3$, $pr(Z = 0, X = 1) = 0.4$, $pr(Z = 1, X = 0) = 0.2$, and $pr(Z = X = 1) = 0.1$, yielding the population counts in Table 1.

Table 1. Population Counts of X and Z.

	Z=0	Z=1
X=0	3064	2079
X=1	3931	926

(ii) Distribution of *Y* given *X*, *Z*

Values of the survey variable *Y* were generated according to the logistic model:

$$\log itP(Y = 1 \mid X, Z) = 0.5 + \gamma_X (X - \bar{X}) + \gamma_Z (Z - \bar{Z}) + \gamma_{XZ} (X - \bar{X})(Z - \bar{Z}) \quad (4)$$

for five choices of $\gamma = (\gamma_X, \gamma_Z, \gamma_{XZ})$ chosen to reflect different relationships between *Y* and *X* and *Z*. These choices are displayed in Table 2, using conventional generalized linear model notation.

Table 2: Models for *Y* given *X*, *Z*.

Model	γ_X	γ_Z	γ_{XZ}
1. $[XZ]^Y$	2	2	2
2. $[X+Z]^Y$	2	2	0
3. $[X]^Y$	2	0	0
4. $[Z]^Y$	0	2	0
5. $[\phi]^Y$	0	0	0

Here the additive logistic model is labeled $[X+Z]^Y$, and sets the interaction γ_{XZ} to zero, whereas the model $[XZ]^Y$ sets this interaction equal to 2. Models where *Y* depends on *X* only, *Z* only or neither *X* nor *Z* are denoted by $[X]^Y$, $[Z]^Y$ and $[\phi]^Y$, respectively.

(iii) Distribution of *S* given *Z*, *X* and *Y*.

The sample cases were assumed to be selected from the population using stratified random sampling, so *S* is independent of *X* and *Y* given *Z*, that is $[S \mid Z, X, Y] = [S \mid Z]$. The probabilities of selection were π_0 = 262/6995 (about 0.04) when *Z* = 0 and π_1 = 50/3005 (about 0.02) when *Z* = 1.

(iv) Distribution of *R* given *Z*, *X*, *Y* and *S*:

Since the response mechanism is assumed ignorable and the selection is by stratified random sampling, $[R \mid Z, X, Y, S] = [R \mid Z, X]$. The latter is generated by a logistic model:

$$\log itP(R=1 \mid X, Z) = 0.5 + \beta_X (X - \bar{X}) + \beta_Z (Z - \bar{Z}) + \beta_{XZ} (X - \bar{X})(Z - \bar{Z}), \quad (5)$$

where $\beta = (\beta_X, \beta_Z, \beta_{XZ})$ takes the same values found in Table 2, with γ replaced by β. We also ran the simulation with a negative interaction term, but the results were similar. As for the distribution of *Y* given *X* and *Z*, this yields five models for the distribution of *R* given *X* and *Z*. For example, $[X+Z]^R$ refers to R being additively dependent on *X* and *Z*.

There were thus a total of 5*5 = 25 combinations of population structures and nonresponse mechanisms in the simulation study. One thousand replicate datasets were generated for each of the 25 combinations. Table 3 displays the form of nine estimators of the overall mean $\bar{Y} = N^{-1}\sum_j \sum_k N_{jk}\bar{Y}_{jk}$, which were computed for each data set. The first estimator is the weighted response rate estimator (3) based on adjustment cells *X*, labeled wrr(x), and the second estimator is the analogous unweighted response estimator (2), labeled urr(x). The next five estimators are maximum likelihood (ML) for the assumed models relating *Y* to *X* and *Z* listed in the second column of Table 3. These estimates all have the form $\hat{Y} = \sum_j \sum_k \hat{P}_{jk}\hat{Y}_{jk}$, where $\hat{P}_{jk} = (N_{j+}/N)(n_{jk}/n_{j+})$ is the ML estimate of the proportion of the population with $Z = j, X = k$, and

1. If the model for *Y* is $[XZ]^Y$, then $\hat{Y}_{jk} = \bar{y}_{jk}$;
2. If the model for *Y* is $[X+Z]^Y$, then $\hat{Y}_{jk} = \hat{\mu} + \hat{\alpha}_{1j} + \hat{\alpha}_{2k}$, predicted values from an additive logistic model fitted to the respondent data;
3. If the model for *Y* is $[X]^Y$, then $\hat{Y}_{jk} = \bar{y}_{+k}$;
4. If the model for *Y* is $[Z]^Y$, then $\hat{Y}_{jk} = \bar{y}_{j+}$;
5. If the model for *Y* is $[\phi]^Y$, then $\hat{Y}_{jk} = \bar{y}_{++}$.

It is interesting to note that neither of the weighting class estimators urr(x) and wrr(x) are ML for any of the models used to generate the data in this simulation study. On the other hand, the estimator that weights by the response rates in cells based on the classification by *Z* <u>and</u> *X* is ML for the saturated model $[XZ]^Y$; this estimator is denoted as urr(xz) in Table 3. The last two estimators in Table 3, wrr(x+z) and urr(x+z), both obtain the estimate the mean of *Y* in cell *jk* as $\hat{Y}_{jk} = \bar{y}_{jk}$. These estimators involve response rates that are predictions from an additive logistic model for R on *X* and *Z*, where for urr(x+z) the cases in the logistic regression are weighted equally, and for wrr(x+z) the cases are weighted by the inverse of the probability of selection. These methods are closely related to the response propensity stratification discussed in Section 3 below.

Table 4 shows the average root mean square error (RMSE) of the nine estimators in Table 3 over the 1000 replicate data sets, a measure that takes into account both precision and bias. Asymptotic properties of ML lead us to believe that the ML estimator for a particular assumed model will have close to the lowest RMSE when the assumed model is the same as the model used to generate the data. Table 5 displays the average bias over the 1000 replicates, defined to be the average of the difference of the estimator before deletion of cases due to nonresponse and the estimator based on respondents alone.

A crude summary of the performance of the relative performance of the methods is the RMSE averaged over all problems, shown in the last row of Table 4. Note that the best methods all stratify on both *X* and *Z*, and have similar average RMSE:

urr(xz) = 382, ml(x+z) = 380,
wrr(x+z)=383, urr(x+z) = 381.

The methods that stratify on *X* but not *Z* are much worse than these methods in overall RMSE:

urr(x) = 471, wrr(x) = 471, ml(x) = 443,

with the slightly better performance of ml(x) reflecting gains in efficiency when the model is true. The methods that stratify on *Z* but not *X* are worst of all in overall RMSE:

ml(z) = 507, ml(null)=528,

although as expected these methods show some gains of efficiency in populations where *Y* does not depend on *X*.

As expected, the ML estimate for the model used to generate the data is always best or close to best in these simulations. The estimate for the additive model $[X+Z]^Y$ is theoretically biased when the data-generating model includes the *XZ* interaction, but in these simulations the bias for the overall mean of *Y* is modest.

The unweighted response weight estimator urr(x) is biased and performs poorly when both *Y* and *R* depend on *Z*, since in these cases the stratification on

Z cannot be ignored. Note, however, that weighting the response weights, as in wrr(x), does not generally correct the bias of urr(x) in these situations: wrr(x) performs very similarly to urr(x), and in fact as we have seen its average RMSE over all problems is the same. Two interesting cases where wrr(x) *does* improve on urr(x) are where R depends on both X and Z and Y depends on X but not Z (specifically the models $[X]^{Y}$, $[XZ]^{R}$ and $[X]^{Y}$, $[X+Z]^{R}$), in rows 11 and 12 of the tables). In these cases, weighting the response rates yields unbiased response rate estimates in the cells defined by X, and the respondent mean of Y in these cells is unbiased since Y depends only on X. However, the gain in weighting the response rates in these cases is relatively minor, and (as might be predicted) ml(x) is superior to either method in these cases. Also, the practical importance of these cases is debatable: Y is likely to depend on Z as well as X, since the point of stratifying on Z is to exploit the relationship between Y and Z. The estimator urr(xz) that stratifies on both X and Z is robust under all of the models, and does much better overall than either urr(x) or wrr(x).

The estimators that base the estimated response rates on an additive logistic model, namely wrr(x+z) and urr(x+z), perform well, though neither are ML for any of the generating models. Unlike wrr(x) and urr(x), wrr(x+z) and urr(x+z) both take the design variable into account by obtaining separate estimates of the response rate for cells that stratify both on X and Z. Their performance is similar to ml(x+z) and urr(xz), with wrr(x+z) doing slightly worse overall. Weighting the logistic regressions does not appear to offer any advantage here.

3. GENERAL STRATEGIES FOR CREATING ADJUSTMENT CELLS.

For the relatively simple situations simulated in Section 2, with just two strata and two values of X, adjustment cells can be created based on the joint distribution of Z and X. In more realistic settings, the crossclassification of the survey design variables and observed survey variables yields too many adjustment cells, some of which may contain sampled cases but no respondents. For example in the Health Interview Survey[5], weighted response weights are calculated within second stage sampling unit (SSU), a variable that has many levels. Joint classification by Z and X would correspond to stratifying households within SSU according to race, which would yield many small adjustment cells, including perhaps some with no respondents. Thus a strategy is needed for reducing the number of adjustment cells. Two such strategies are discussed in this section.

Let D denote the complete set of variables recorded for both respondents and nonrespondents, including design variables and any survey variables measured for both groups. (For unit nonresponse survey variables are usually entirely absent for nonrespondents, but in panel surveys variables from earlier surveys may be available). We say that nonresponse is ignorable if the distribution of the incomplete survey variables is the same for respondents and nonrespondents with the same value of D. Formally, if R is an indicator for response or nonresponse, Y is the set of survey variables missing for nonrespondents, then nonresponse is ignorable if

$$R \perp\!\!\!\perp Y \mid D \quad (6)$$

where $\perp\!\!\!\perp$ denotes independence. Adjustments for nonignorable nonresponse are usually highly speculative, and all the methods discussed in this article effectively assume that nonresponse is ignorable. Thus we assume that (6) holds.

We have noted that adjustment cells defined by each distinct value of D may be too small and yield weights that are undefined or too unstable. Thus the problem becomes to define adjustment cells based on D that remove nonresponse bias, whilst avoiding sparse cells that lead to unstable weights, and resulting estimates with large variance. Two sensible objectives in defining adjustment cells are (a) to choose cells that are homogeneous with respect to outcome variables Y, and (b) to choose cells that are homogeneous with respect to the probability of response. Theory supporting both these choices is presented in Little[7], who considers two methods for creating adjustment cells when D is extensive: (i) *predictive mean* stratification, motivated by objective (a), groups units according to predicted means of Y given D, estimated for example by regression of Y on D based on the responding cases; (ii) *response propensity* stratification, motivated by objective (b), groups units according to their estimated probabilities of response, computed for example by logistic regression of the response indicator R on D based on sampled cases. Little[7] showed that if $\hat{Y}(D)$ denotes the predicted mean of Y given D, and $\hat{p}(D)$ denotes the predicted probability of response given D, then (with some additional conditions described in the paper), (6) implies that

$$Y \perp\!\!\!\perp R \mid \hat{Y}(D), \quad (7)$$

and

$$Y \perp\!\!\!\perp R \mid \hat{p}(D). \quad (8)$$

In particular assuming ignorable nonresponse and ignoring the effects of estimating $\hat{Y}(D)$ and $\hat{p}(D)$, weighting based on either of these methods of stratification removes nonresponse bias in estimating population means. Of these two methods, only response propensity stratification also removes bias of estimates of means for subclasses of the population[7]. This theory supports the idea of basing weights based on a model for the propensity to respond on D, where the latter includes the design variables that determine the sampling weight. This approach is closely related to the urr(x+z) method in the simulation study, which was competitive with the best methods. Weighting the logistic regression by the sampling weight, as in wrr(x+z), did not offer any advantage in our simulations, and by analogy with the simpler case of stratification on x alone, we do not expect any advantages of weighting in more complex situations.

ACKNOWLEDGEMENTS

This research was supported by CDC grant UR6/CCU517481. We appreciate helpful comments from Michael Brick, David Judkins and Trivellore Raghunathan.

REFERENCES

1. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1952; 47:663-685.
2. Basu, D. An essay on the logical foundations of survey sampling, Part 1. In *Foundations of Statistical Inference*, Holt, Rinehart and Winston: Toronto, 1971; pp 203-242.
3. Platek R., Gray GB. Imputation methodology. In *Incomplete Data in Sample Surveys*, *Vol. 2: Theory and Bibliographies*, WG. Madow, I Olkin and DB Rubin, eds. Academic Press: New York, 1983; pp 255-294.
4. Chapman DW, Bailey L, Kasprzyk D. Nonresponse adjustment procedures at the U.S. Bureau of the Census. *Survey Methodology* 1986; 12:161-180.
5. Botman SL, Moore TF, Moriarty CL, Parsons VL. Design and Estimation of the National Health Interview Survey, 1995-2004. *National Center for Health Statistics*, *Vital Health Statistics* 2000; 2, 130.
6. Petroni R, King KE. Evaluation of Survey of Income and Program Participation's cross-sectional noninterview adjustment. *Proceedings of the Survey Research Methods Section*, *American Statistical Association 1988*; 342-347.
7. Little RJ. Survey nonresponse adjustments. *International Statistical Review*, 1986; 54: 139-157.

Table 3. Estimators of Mean of *Y*

Response Weight	Assumed Model	Estimator	Weight	Response Rate
1. wrr(x)		$\dfrac{\sum_x \sum_z w^*_{xz} r_{xz} \bar{y}_{xz}}{\sum_x \sum_z w^*_{xz} r_{xz}}$	$w^*_{xz} = \dfrac{(N_z/n_z)(r/N)}{\hat{\phi}_x^*}$	$\hat{\phi}_x^* = \dfrac{\sum_z (r_{xz}/\pi_z)}{\sum_z (n_{xz}/\pi_z)}$
2. urr(x)		$\dfrac{\sum_x \sum_z w_{xz} r_{xz} \bar{y}_{xz}}{\sum_x \sum_z w_{xz} r_{xz}}$	$w_{xz} = \dfrac{(N_z/n_z)(r/N)}{\hat{\phi}_x}$	$\hat{\phi}_x = \dfrac{r_{x+}}{n_{x+}}$
3. ml(xz)/ urr(xz)	$[XZ]^Y$	$\dfrac{\sum_x \sum_z w'_{xz} r_{xz} \bar{y}_{xz}}{\sum_x \sum_z w'_{xz} r_{xz}}$	$w'_{xz} = \dfrac{(N_z/n_z)(r/N)}{\hat{\phi}_{xz}}$	$\hat{\phi}_{xz} = \dfrac{r_{xz}}{n_{xz}}$
4. ml(x)	$[X]^Y$	$\dfrac{\sum_x w_x^* \bar{y}_x}{\sum_x w_x^*}$	$w^*_x = \sum_z w'_{xz} r_{xz}$	
5. ml(z)	$[Z]^Y$	$\sum_z \dfrac{N_z}{N} \bar{y}_z$		
6. ml(null)	$[\phi]^Y$	$\sum_z \dfrac{r_z}{r} \bar{y}_z$		
7. ml(x+z)	$[X+Z]^Y$	$\dfrac{\sum_x \sum_z w'_{xz} r_{xz} (\hat{\mu} + \hat{\alpha}_{1x} + \hat{\alpha}_{2z})}{\sum_x \sum_z w'_{xz} r_{xz}}$		
8. urr(x+z)	$[XZ]^Y$	$\dfrac{\sum_x \sum_z w_{xz}^{(u)} r_{xz} \bar{y}_{xz}}{\sum_x \sum_z w_{xz}^{(u)} r_{xz}}$	$w_{xz}^{(u)} = \dfrac{(N_z/n_z)(r/N)}{\hat{\phi}_{xz}^{(u)}}$	$\hat{\phi}_{xz}^{(u)}$ from unweighted additive logistic model
9. wrr(x+z)	$[XZ]^Y$	$\dfrac{\sum_x \sum_z w_{xz}^{(w)} r_{xz} \bar{y}_{xz}}{\sum_x \sum_z w_{xz}^{(w)} r_{xz}}$	$w_{xz}^{(w)} = \dfrac{(N_z/n_z)(r/N)}{\hat{\phi}_{xz}^{(w)}}$	$\hat{\phi}_{xz}^{(w)}$ from weighted additive logistic model

Table 4. 10000*RMSE of 1000 Replicate Samples (n=312).

	Generated Model for Y and R $[\ ]^Y$	$[\ ]^R$	urr(xz) $[XZ]^Y$	wrr(x)	urr(x)	ml(x) $[X]^Y$	ml(z) $[Z]^Y$	ml(null) $[\phi]^Y$	ml(x+z) $[X+Z]^Y$	wrr(x+z)	urr(x+z)
			Estimator, and Assumed Model where Applicable								
1	XZ	XZ	398*	513	494	438	655	526	398	399	391
2	XZ	X+Z	406*	582	534	410	704	618	410	406	405
3	XZ	X	419*	421	419	528	730	462	412	418	419
4	XZ	Z	364*	527	511	384	376	385	366	363	366
5	XZ	ϕ	361*	365	365	502	364	430	361	361	362
6	X+Z	XZ	381	562	536	441	696	531	382*	400	381
7	X+Z	X+Z	387	770	706	403	838	715	384*	390	387
8	X+Z	X	392	406	399	646	750	437	388*	393	391
9	X+Z	Z	367	608	588	404	378	402	367*	368	366
10	X+Z	ϕ	334	346	343	612	349	513	334*	334	333
11	X	XZ	397	421	434	391*	779	875	395	398	395
12	X	X+Z	397	419	474	390*	838	956	391	398	396
13	X	X	425	423	424	394*	842	986	422	424	425
14	X	Z	346	368	376	346*	369	369	346	348	346
15	X	ϕ	364	362	363	343*	380	381	363	363	364
16	Z	XZ	360	558	577	437	338*	515	358	368	362
17	Z	X+Z	381	696	762	385	344*	466	376	382	381
18	Z	X	375	390	393	674	350*	757	372	372	372
19	Z	Z	348	617	628	393	346*	398	347	348	348
20	Z	ϕ	338	353	353	640	337*	658	337	337	337
21	ϕ	XZ	409	433	439	406	390	383*	408	412	408
22	ϕ	X+Z	408	437	455	403	372	364*	404	409	408
23	ϕ	X	421	415	415	383	401	354*	416	417	417
24	ϕ	Z	388	415	416	386	388	386*	388	388	388
25	ϕ	ϕ	374	371	371	341	372	341*	372	372	372
		Mean	382	471	471	443	507	528	380	383	381

*ML estimate of $\bar{Y} = N^{-1}\sum_j \sum_k N_{jk}\bar{Y}_{jk}$.

Lowest RMSE Shaded in Grey.

Table 5. 10000*(Average Bias) of 1000 Replicate Samples (n=312)

	Generated Model for Y and R $[\ ]^Y$	$[\ ]^R$	urr(xz) $[XZ]^Y$	wrr(x)	urr(x)	ml(x) $[X]^Y$	ml(z) $[Z]^Y$	ml(null) $[\phi]^Y$	ml(x+z) $[X+Z]^Y$	wrr(x+z)	urr(x+z)
			Estimator, and Assumed Model where Applicable								
1	XZ	XZ	-11*	288	246	-163	539	366	-7	-56	-6
2	XZ	X+Z	2*	392	288	-34	595	495	-56	1	2
3	XZ	X	1*	-1	-1	-358	630	308	52	-1	0
4	XZ	Z	-1*	365	335	-108	-1	-78	-1	-3	-1
5	XZ	ϕ	8*	6	7	-362	7	-254	7	5	6
6	X+Z	XZ	-18	393	354	-207	597	386	-18*	-116	-54
7	X+Z	X+Z	0	656	568	57	759	619	-2*	-5	-1
8	X+Z	X	-7	-10	-7	-525	663	269	-5*	-12	-9
9	X+Z	Z	-5	473	446	-151	-10	-121	-5*	-7	-5
10	X+Z	ϕ	5	1	3	-516	7	-381	4*	2	4
11	X	XZ	17	16	-53	17*	684	799	17	-21	31
12	X	X+Z	3	4	-178	2*	761	893	-1	1	3
13	X	X	-3	-4	-6	-3*	748	927	-1	-5	-4
14	X	Z	-12	-9	-58	-13*	-5	39	-12	-11	-12
15	X	ϕ	0	0	2	-8*	0	146	0	0	1
16	Z	XZ	1	423	444	-213	0*	-368	2	-33	-22
17	Z	X+Z	1	592	662	-20	1*	-291	0	0	1
18	Z	X	-9	-25	-23	-545	-14*	-665	-10	-11	10
19	Z	Z	-3	514	527	-157	-3*	-167	-3	-4	-3
20	Z	ϕ	1	-3	-1	-531	0*	-552	0	-2	0
21	ϕ	XZ	-5	-9	-10	-4	-3	0*	-6	-6	-6
22	ϕ	X+Z	-25	-21	-22	-28	-20	-23*	-28	-27	-26
23	ϕ	X	-8	-6	-6	-2	-4	0*	-4	-8	-8
24	ϕ	Z	14	11	11	17	14	16*	14	14	14
25	ϕ	ϕ	0	0	0	2	0	2*	0	0	0
		Mean	-2	162	141	-154	238	95	-3	-12	-3
		Mean of Absolute Average Bias	6	169	170	162	243	327	10	14	9

*ML estimate of $\bar{Y} = N^{-1}\sum_j \sum_k N_{jk}\bar{Y}_{jk}$.

Smallest Absolute Average Bias Shaded in Grey.