Journal of Official Statistics, Vol. 20, No. 1, 2004, pp. 1–18

The Effect of Multiple Weighting Steps on Variance Estimation

Richard Valliant[1]

Multiple weight adjustments are common in surveys to account for ineligible units on a frame, nonresponse by some units, and the use of auxiliary data in estimation. A practical question is whether all of these steps need to be accounted for when estimating variances. Linearization variance estimators and related estimators in commercial software packages that use squared residuals usually account only for the last step in estimation, which is the incorporation of auxiliary data through poststratification, regression estimation, or similar methods. Replication variance estimators can explicitly account for all of the steps in estimation by repeating each adjustment separately for each replicate subsample. Through simulation, this article studies the difference in these methods for some specific sample designs, estimators of totals, and rates of ineligibility and nonresponse. In the simulations reported here, the linearization variance estimators are negatively biased and produce confidence intervals for a population total that cover at less than the nominal rate, especially at smaller sample sizes. The jackknife replication estimator generally yields confidence intervals that cover at or above the nominal rate but do so at the expense of considerably overestimating empirical mean squared errors. A leverage-adjusted variance estimator, which is related to the jackknife estimator, has small positive bias and nearly nominal coverage. The leverage-adjusted estimator is less computationally burdensome than the jackknife but works well in the situations studied here where multiple weighting steps are used.

Key words: Ineligibility; jackknife; leverage; nonresponse; replication variance estimate.

1. Introduction

Multiple steps in weighting are common in survey estimation. Each step usually introduces a source of variability in an estimator that it may be important to reflect when estimating variances. A typical sequence of weighting steps in a probability sample is this:

1. Compute base weights.
2. Adjust weights to account for units with unknown eligibility.
3. Adjust weights for nonresponse.
4. Use auxiliary data.

Base weights in Step (1) are usually inverses of selection probabilities. In some surveys the eligibility of all initial sample units cannot be determined. For example, in a telephone survey of residential households, noncontacts may be businesses, nonworking numbers, or eligible residences, but their actual status is undetermined. In Step (2) the weight of such unknown cases is distributed among the cases whose eligibility is known.

[1] University of Michigan, Joint Program for Survey Methodology, 1218 Lefrak Hall, College Park MD 20742, U.S.A. Email: rvalliant@survey.umd.edu

A nonresponse adjustment in Step (3) consists of spreading the weight of the eligible nonrespondents over eligible respondents. The method studied here is to assign units to classes that have different response rates and to make a ratio weight adjustment to each unit within a class. Poststratification, raking, regression estimation, and more general calibration estimation are examples of the use of auxiliary data in Step (4).

The variance of an estimator is affected by the population structure of the variables being estimated, the complexity of the design used to collect data, and the form of the estimator itself. Intuition may lead us to believe that a variance estimator that somehow incorporates all of these complications is better than one that does not. However, literature that directly addresses this question is limited.

The two major competitors in finite population variance estimation are replication and linearization. There is a wealth of literature studying the model-based and design-based properties of these alternatives. Generally, the situations covered are ones in which different simplifying assumptions are made – e.g., sampling is with replacement, there is no nonresponse, or the estimator is of a certain type. One reason for simplifying is that theory is quite difficult to develop when all four weighting steps are used and the sample design is complex.

For replication variance estimators there is evidence in particular cases that it is necessary to repeat each step of estimation separately for each replicate subsample in order to produce a consistent or approximately unbiased variance estimate. Empirical results, however, are not uniform. Lemeshow (1979) and Ernst and Williams (1987) found that recalculation of weights is necessary for the BRR method. Rust (1987) reports results for a survey of student grant records in which a nonresponse adjustment and ratio estimation were used along with BRR. For the estimates in that study, it made little difference whether the nonresponse adjustments and ratio estimation were repeated for each half-sample or not.

Valliant (1993) showed theoretically and empirically that poststratification factors must be recomputed for every replicate in order for the BRR or jackknife estimators to be consistent in two-stage sampling. Yung and Rao (1996; 2000) obtained similar results for the jackknife in stratified, multistage sampling both with and without nonresponse and poststratification.

There are a number of articles that cover some, but not all, of the four steps when applying Taylor series variance estimators. Lundström and Särndal (1999) adapted a linearization estimator for two-phase sampling to the situation where nonresponse is the mechanism leading to the second phase. Rao (1996) derived a modified linearization variance estimator that accounted for mean imputation. Rao's estimator uses a standard linearization variance formula but the deviates needed for the formula are specialized for the case of mean imputation for the nonrespondents (see Rao 1996, expression (21)).

Expedient methods that ignore some complexities are often used in practice. For example, in a survey that uses an unknown eligibility adjustment, a nonresponse adjustment, and poststratification, a linearization estimator may be used that accounts for design features like stratification and clustering but not for the other complexities of the estimator. An improvement would be to account for stratification, clustering, and the last step of poststratification while ignoring the unknown eligibility and nonresponse adjustments. When such expedients give acceptable performance is an open question.

An extreme case where a shortcut would clearly be incorrect would be an estimate of the total number of units in a poststratum. The variance of this estimate would be zero, but a linearization estimator that ignores poststratification would be nonzero.

Lago et al. (1987) compared linearization variance estimators that ignored poststratification to replication estimators that properly accounted for it. For the Hispanic Health and Nutrition Examination Survey, they found that the shortcut linearization estimator was similar to a balanced repeated replication (BRR) estimator for most statistics. However, for mean height, weight, and blood cholesterol level, which were highly correlated with the age/sex poststratification variables, the linearization variance was a severe overestimate. Smith et al. (2000) compared jackknife variance estimates that reflected stratification and clustering plus nine separate weighting steps with Taylor series variance estimates that accounted only for stratification and clustering. Although the linearization estimates tended to be smaller than the jackknife, the authors did not assess which was closer to the truth.

Shortcut implementations of linearization estimators are fairly common in practice for at least two reasons. First, linearizing complex estimators is difficult and commercial software packages limit how faithfully a user can reflect the complexities of a design and an estimator. Few studies report direct comparisons of shortcut linearization estimators and more elaborate replication estimators. This article attempts to fill that gap by empirically comparing some alternative variance estimators systematically for various combinations of eligibility rate, response rate, type of estimator, and sample size.

Section 2 defines notation. Section 3 introduces some alternative variance estimates. Several simulation studies employing multiple weighting steps are reported in Section 4. The last section is a brief conclusion.

2. Notation and an Estimator of a Total

For the illustrations in this article we consider only stratified and unstratified, single-stage sampling but include all four of the weighting steps listed in Section 1. Suppose that the strata are numbered $h = 1, \ldots, H$, the frame size in stratum h is N_h, the number of initial sample units is n_h, and the set of initial sample units is s_h. Denote the base weight for sample unit hi as w_{hi}. Define the following sets of sample cases:

s_{ER} = set of eligible sample respondents
s_{ENR} = set of eligible sample nonrespondents
s_{IN} = set of sample units known to be ineligible
s_{UNK} = set of sample units whose eligibility status is unknown

The full sample s is the union of these four sets. The set of units whose eligibility status is known is $s_{KN} = s_{ER} \cup s_{ENR} \cup s_{IN}$.

Suppose that the sample is also divided into classes, $c = 1, \ldots, C$, that are used for the unknown eligibility adjustment. Another set of classes, $d = 1, \ldots, D$, is used for the nonresponse adjustment. Both of these sets of classes may cut across strata. In practice, the eligibility adjustment and nonresponse adjustment classes may often be the same. Let s_c denote the set of sample units in class c and $s_{c,KN} = s_c \cap s_{KN}$ the set with known eligibility in class c. Then, the unknown eligibility adjustment for sample units in class c is

$$a_{1c} = \frac{\sum_{(hi)\in s_c} w_{hi}}{\sum_{(hi)\in s_{c,KN}} w_{hi}} \tag{1}$$

and the eligibility-adjusted weight for a unit with known eligibility in class c is $w_{1hi} = w_{hi}a_{1c}$, $(hi) \in s_{c,KN}$. The summations over $(hi) \in A$ for some set A means to sum over all strata and the units within each stratum that are members of the set. After this step, the units with unknown eligibility are eliminated.

Next, denote the set of cases in class d as s_d, those that are known to be eligible in class d as $s_{d,E} = s_d \cap (s_{ER} \cup s_{ENR})$ and the set of eligible respondents in class d to be $s_{d,ER} = s_d \cap s_{ER}$. The nonresponse adjustment for units in class d is

$$a_{2d} = \frac{\sum_{(hi)\in s_{d,E}} w_{1hi}}{\sum_{(hi)\in s_{d,ER}} w_{1hi}} \tag{2}$$

The nonresponse-adjusted weight is then

$$w_{2hi} = \begin{cases} w_{1hi}a_{2d} & (hi) \in s_{d,E} \\ w_{1hi} & (hi) \in s_{IN} \end{cases} \tag{3}$$

i.e., the weights for eligible respondents are adjusted while the weights for known ineligibles remain the same as they were after the unknown eligibility adjustment. The nonrespondents, s_{ENR}, are eliminated. After this step, the units with nonzero weight, which are used in estimation, are s_{ER} and s_{IN}. The known ineligibles are retained on the grounds that their presence in the sample is a reflection of other nonsample ineligibles in the frame.

To illustrate the use of auxiliary data, we take the case of the general regression (GREG) estimator (see e.g., Särndal, Swensson, and Wretman 1992) in single-stage sampling. The GREG is motivated by a linear model in which the Y's are independent random variables with $E_M(Y_{hi}) = \mathbf{x}'_{hi}\boldsymbol{\beta}$ and $\mathrm{var}_M(Y_{hi}) = v_{hi}$, where $\mathbf{x}_{hi}$ is a p-vector of auxiliaries for unit hi and $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector. The g-weight for unit hi is $g_{hi} = 1 + (\mathbf{T}_x - \hat{\mathbf{T}}_x)'\mathbf{A}^{-1}\mathbf{x}_{hi}/v_{hi}$, where $\mathbf{T}_x$ is the vector of population totals of the auxiliaries, $\hat{\mathbf{T}}_x = \sum_{(hi)\in(s_{ER}\cup s_{IN})} w_{2hi}\mathbf{x}_{hi}$ is the estimator of the x-totals using the weights after eligibility adjustment and nonresponse adjustment, and $\mathbf{A} = \sum_{(hi)\in(s_{ER}\cup s_{IN})} w_{2hi}\mathbf{x}_{hi}\mathbf{x}'_{hi}/v_{hi}$.

In the current context, we compute the g-weights using the eligible respondents, s_{ER}, and the known ineligibles, s_{IN}. Using $s_{ER} \cup s_{IN}$ presumes that the population control totals $\mathbf{T}_x$ include some units that are actually ineligible but cannot be separated out. This can occur if the population counts are made from a frame that is somewhat out-of-date. If the population controls include only eligibles, then the g-weight would be computed on the basis of only the eligible respondents, s_{ER}. After the g-weight adjustment, the weight for sample unit i is $w_{3hi} = w_{2hi}g_{hi}$, $(hi) \in s_{ER} \cup s_{IN}$.

The final weights after the three stages of adjustment would, thus, be defined by

$$w^*_{hi} = \begin{cases} a_{1c}a_{2d}g_{hi}w_{hi} & (hi) \in s_{c,KN} \cap s_{d,ER} \\ a_{1c}g_{hi}w_{hi} & (hi) \in s_{c,KN} \cap s_{IN} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Note that the final weight differs depending on the combination of eligibility and nonresponse adjustment classes to which a unit belongs and the unit's g-weight value, in addition to which of the four eligibility classes (s_{ER}, s_{ENR}, s_{IN}, or s_{UNK}) the sample unit belongs to. For estimation the eligibles are treated as a domain, and the ineligibles are assigned data values of zero both for point estimation and for variance estimation (see, e.g., Cochran 1977, Sections 2.13, 5A.14). After these sequential adjustments, even an estimated total of the form $\hat{T} = \sum_{(hi) \in s_{ER} \cup s_{IN}} w^*_{hi} Y_{hi}$ is nonlinear in the design-based sense because the weights involve various sample-dependent ratio adjustments.

We will cover two specific cases of the GREG in the simulations described later. The linear regression estimator with a single auxiliary x in an unstratified design is

$$\hat{T}_{LR} = \sum\nolimits_{i \in s_{ER} \cup s_{IN}} w_{2i} g_i Y_i \qquad (5)$$

where $g_i = 1 + (\mathbf{T}_x - \hat{\mathbf{T}}_x)' \mathbf{A}^{-1} \mathbf{x}_i$, $\hat{\mathbf{T}}_x = \sum_{s_{ER \cup IN}} w_{2i} \mathbf{x}_i$, $\mathbf{x}'_i = (1\, x_i)$, and $\mathbf{A} = \sum_{s_{ER} \cup s_{IN}} w_{2i} \mathbf{x}_i \mathbf{x}'_i$. The second is the poststratified estimator denoted by $\hat{T}_{PS}$. If we let $k = 1, \ldots, K$ index the poststrata and $s_{PS,k}$ be the set of population units in poststratum k, then the g-weight for a unit is

$$g_{hi} = \begin{cases} N_k / \hat{N}_k & (hi) \in s_{PS,k} \cap (s_{ER} \cup s_{IN}) \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

where N_k is the population count in poststratum k (which may include some ineligibles) and $\hat{N}_k = \sum_{(hi) \in s_{PS,k} \cap (s_{ER} \cup s_{IN})} w_{2hi}$, i.e., the estimate of the poststratum count based on eligible responding sample units and the sample units that are known to be ineligible.

3. Variance Estimators

We will study several variance estimators that, in varying degrees, account for the complexity of the design and the estimator of the total. The variation of the jackknife studied here is to divide the units within a stratum into random groups and delete one group at a time. If the initial sample is divided into G_h random groups within each stratum, then the delete-one-group jackknife is defined as

$$v_J = \sum_h \frac{G_h - 1}{G_h} \sum_{g=1}^{G_h} \left[\hat{T}_{(hg)} - \hat{T} \right]^2 \qquad (7)$$

where $\hat{T}_{(hg)}$ is the estimated total based on deleting all initial sample units in group (hg) and then repeating all weighting steps – base weight calculation, adjustment for unknown eligibility, nonresponse adjustment, and use of auxiliary data. The total number of groups is $G = \sum_h G_h$. If $G_h = n_h$ and the groups are disjoint, then (7) is just the standard delete-one jackknife. Many variants of the grouped jackknife may be used in practice (see, e.g., Rust and Rao 1996). In the simulations, we will consider only the case of disjoint random groups formed within each stratum with $G_h = \bar{G}$, i.e., an equal number of groups per stratum. The union of the groups in a stratum is the initial stratum sample.

The subsamples created by deleting one group at a time are usually referred to as replicates. In implementing the jackknife, each group is deleted one at a time to create the replicates. This is done without regard to the disposition of the initial unit as a respondent, a nonrespondent, an unknown, or an ineligible. This procedure is available in WesVar® (Westat 2000) and SUDAAN® (Shah et al., 1996).

Several versions of linearization and related variance estimators might be used when the basic design is single-stage stratified sampling. Use of some of the choices would simply be a mistake but could be selected by a naïve user of some software packages. Other choices might be reasonable if some steps in weighting make small contributions to the variance. The simplest variance estimator is one that would be appropriate for the Horvitz-Thompson estimator in a stratified sample selected with varying probabilities and with replacement (Särndal, Swensson, and Wretman 1992, Expression 2.9.9). If we interpret the weight w_{hi}^* as the inverse of an adjusted selection probability and add an *ad hoc* finite population correction (*fpc*) factor, this variance estimator is

$$
\begin{aligned}
v_{naive1} = {} & \sum\nolimits_h (1 - f_{h,ER \cup IN}) \\
& \times \frac{n_{h,ER \cup IN}}{n_{h,ER \cup IN} - 1} \sum\nolimits_{i \in s_{h,ER \cup IN}} \left(w_{hi}^* Y_{hi} - \frac{1}{n_{h,ER \cup IN}} \sum\nolimits_{i \in s_{h,ER \cup IN}} w_{hi}^* Y_{hi} \right)^2
\end{aligned}
\tag{8}
$$

where

$s_{h,ER \cup IN}$ = set of sample eligible respondents and known ineligibles in stratum h,
$n_{h,ER \cup IN}$ = sample size in stratum h of eligible respondents and known ineligibles, and
$f_{h,ER \cup IN} = n_{h,ER \cup IN}/N_h$.

This estimator is available in SUDAAN using the option `DESIGN = STWOR` in a procedure statement, in STATA® (Stata Corporation 2001) using the procedure `svytotal`, and in SAS® `PROC SURVEYMEANS` (SAS Institute 2001). Units that are in s_{IN} (known ineligibles) have their Y values set to zero so that the eligibles are appropriately treated as a domain. We label this variance estimator "naïve" because it treats the resulting sample of eligible respondents and ineligibles as a with-replacement sample (but adds an ad hoc *fpc*) and ignores the adjustments for unknown eligibility and nonresponse along with the poststratification step (or other use of auxiliary data).

Another variant would be to exclude the s_{IN} cases and treat s_{ER} as a stratified without-replacement (*stwor*) sample. This estimator would be

$$
v_{naive2} = \sum\nolimits_h (1 - f_{h,ER}) \frac{n_{h,ER}}{n_{h,ER} - 1} \sum\nolimits_{i \in s_{h,ER}} \left(w_{hi}^* Y_{hi} - \frac{1}{n_{h,ER}} \sum\nolimits_{i \in s_{h,ER}} w_{hi}^* Y_{hi} \right)^2 \tag{9}
$$

where $s_{h,ER}$, $n_{h,ER}$, and $f_{h,ER}$ are defined in terms of the set of eligible respondents in stratum h. This estimator will typically be smaller than v_{naive1} since the ineligibles do not enter the calculation as zeroes. It is possible to compute this estimator in SAS, STATA, and SUDAAN by restricting the dataset to the eligible respondents only. Arguments can sometimes be made to condition on achieved sample sizes rather than averaging over all sizes that could be obtained under a design, in which case v_{naive2} might be appropriate. In this case, we assume that the distribution of the number of sample eligibles is unknown and, thus, not an ancillary statistic that can be conditioned on.

The variance estimator that is usually referred to as the linearization estimator is

$$v_L = \sum\nolimits_h (1 - f_{h,ER\cup IN}) \times \frac{n_{h,ER\cup IN}}{n_{h,ER\cup IN} - 1} \sum\nolimits_{i \in s_{h,ER\cup IN}} \left(w_{2hi} r_{hi} - \frac{1}{n_{h,ER\cup IN}} \sum\nolimits_{i \in s_{h,ER\cup IN}} w_{2hi} r_{hi} \right)^2 \tag{10}$$

where $r_{hi} = Y_{hi} - \hat{Y}_{hi}$ with $\hat{Y}_{hi} = \mathbf{x}'_{hi}\hat{\mathbf{B}}$, $\hat{\mathbf{B}} = \mathbf{A}^{-1}\sum_{(hi)\in s_{ER\cup IN}} w_{2hi}\mathbf{x}_{hi}Y_{hi}/v_{hi}$, and w_{2hi} is the weight for unit i after adjustment for unknown eligibility and nonresponse. Although v_L uses a residual, r_{hi}, appropriate to the GREG, it can have poor conditional properties since the w_{2hi} weights are used rather than w^*_{hi} (see e.g., Valliant 1993). That is, in samples where $\hat{\mathbf{T}}_x$ is not near $\mathbf{T}_x$, v_L has a conditional bias.

The jackknife linearization estimator is very similar to (10), but uses the w^*_{hi} weights:

$$v_{JL} = \sum\nolimits_h (1 - f_{h,ER\cup IN}) \times \frac{n_{h,ER\cup IN}}{n_{h,ER\cup IN} - 1} \sum\nolimits_{i \in s_{h,ER\cup IN}} \left(w^*_{hi} r_{hi} - \frac{1}{n_{h,ER\cup IN}} \sum\nolimits_{i \in s_{h,ER\cup IN}} w^*_{hi} r_{hi} \right)^2 \tag{11}$$

Use of w^*_{hi} leads to better conditional performance when the g_{hi} are not near 1. This estimator approximates a jackknife in which the set $s_{ER\cup IN}$ is treated as the initial sample, which is assumed to be selected with replacement. As in (9), an ad hoc *fpc* is added in (11). The jackknife linearization estimator does not make separate adjustments for unknown eligibility and nonresponse for each replicate.

The special case of v_{JL} that is appropriate for poststratification is available in SUDAAN, using the POSTWGT and POSTVAR options of some procedures. For poststratification, the difference between the linearization estimator v_L and the jackknife linearization estimator is that v_{JL} includes a factor $(N_k/\hat{N}_k)^2$ for each unit that is in poststratum k. This inclusion imparts better conditional properties to v_{JL} in samples where $N_k/\hat{N}_k$ is not near 1. Accounting for poststratification does not appear to be possible in STATA v.7 or SAS v.8 unless the user writes his or her own code.

Another, related approximation to the jackknife was derived by Valliant (2002). This estimator adjusts each weighted residual using a leverage, $\Delta_{hi} = w_{2hi}\mathbf{x}'_{hi}\mathbf{A}^{-1}\mathbf{x}_{hi}/v_{hi}$, associated with sample unit (hi):

$$v^*_J = \sum\nolimits_h (1 - f_{h,ER\cup IN}) \sum\nolimits_{s_{h,ER\cup IN}} \left(\frac{w^*_{hi} r_{hi}}{1 - \Delta_{hi}} - \frac{1}{n_{h,ER\cup IN}} \sum\nolimits_{s_{h,ER\cup IN}} \frac{w^*_{hi} r_{hi}}{1 - \Delta_{hi}} \right)^2 \tag{12}$$

Since $\Delta_{hi} < 1$, v^*_J will be larger than v_{JL}, which will typically give higher confidence interval coverage rates. As the number of eligible respondents and known ineligibles increases, $\Delta_{hi} \to 0$ so that the difference between v^*_J and v_{JL} will diminish.

Särndal, Swensson, and Wretman (1992, Expressions 7.2.11 and 7.9.8) give an estimator of the variance of the GREG based on weighted squared residuals. Adapted to the situation here their estimator is

$$v_{SSW} = \sum\nolimits_h (1 - f_{h,ER\cup IN}) \frac{n_{h,ER\cup IN}}{n_{h,ER\cup IN} - 1} \sum\nolimits_{s_{h,ER\cup IN}} \left(w^*_{hi} r_{hi} \right)^2 \tag{13}$$

If the sampling fraction is small, the sample is large, and the sample is selected with replacement, then v_{SSW} is about the same as the estimator $\hat{V}_T$ in Särndal (1996). Since the stratum mean of the $w^*_{hi}r_{hi}$ will be near zero under some reasonable conditions (see e.g., Valliant 2002), v_{SSW} and the jackknife linearization estimator v_{JL} will also be very similar. In the linear regression literature, estimators like v_{SSW} are known as sandwich estimators and date from Horn, Horn, and Duncan (1975) and White (1982).

A variation on v_{SSW} is had by setting all g-weights to 1, leading to

$$v_\pi = \sum\nolimits_h (1 - f_{h,ER\cup IN}) \frac{n_{h,ER\cup IN}}{n_{h,ER\cup IN} - 1} \sum\nolimits_{s_{h,ER\cup IN}} (w_{2hi} r_{hi})^2 \tag{14}$$

(see Särndal et al., 1992, Expression 7.9.9). This estimator is very close to the linearization estimator, v_L, in (10) since the stratum mean of $w_{2hi}r_{hi}$ will be near zero. Thus, v_π has the same poor conditional properties as v_L.

Table 1 summarizes the design and estimation steps accounted for by the different variance estimators. Only the jackknife v_J explicitly accounts for stratification, unknown eligibility adjustment, nonresponse adjustment, and use of auxiliary data in estimation. The linearization estimator, v_L, and the estimator v_π are shown as partially accounting for the use of auxiliary data because each uses the appropriate residual but uses the weights w_{2hi} rather than w^*_{hi}. The only estimator that does not treat eligibles as a domain is v_{naive2}.

4. Empirical Evaluation

To compare the different variance estimators, we conducted simulation studies using two populations. The first is a poststratified population similar to ones found in human populations in which groups of units have different means. The second has a target variable whose mean depends on a single auxiliary variable, as might be the case in some business populations.

4.1. Poststratified population

A stratified population, with specifications shown in Table 2, was generated in which poststratification was appropriate. The population has five design strata and five classes

Table 1. Design and estimation steps accounted for by different variance estimators

Variance estimator	Stratification	Unknown eligibility adjustment	Nonresponse adjustment	Use of auxiliary data (e.g., poststratification, regression estimation)	Treatment of eligibles as a domain
v_J	✓	✓	✓	✓	✓
v_{naive1}	✓				✓
v_{naive2}	✓				
v_L	✓			partial	✓
v_{JL}	✓			✓	✓
v^*_J	✓			✓	✓
v_{SSW}	✓			✓	✓
v_π	✓			partial	✓

Table 2. Specifications for the poststratified population

	Design stratum or poststratum				
	1	2	3	4	5
N_h	1,000	1,000	1,000	1,000	1,000
P_h	0.247	0.259	0.256	0.248	0.243
N_k/N	0.30	0.24	0.18	0.16	0.12
η_k	0.1	0.2	0.3	0.4	0.5

that are used as poststrata. The poststrata cut across the strata. The variable Y used in estimation is a 0-1 Bernoulli variable with means, η_k, ranging from 0.1 to 0.5 across the poststrata. The proportion of the population in each poststratum ranges from 0.30 to 0.12. For each unit in the population, a poststratum indicator was generated independently of design stratum membership. Each design stratum has $N_h = 1{,}000$ units and the stratum means, P_h, of Y in the realized population range from 0.243 to 0.259. (The expected value of Y in each stratum is $\sum_{k=1}^{5} \eta_k N_k/N = 0.256$. The range of P_h is due to random variation.) In other words, there is little difference among the design strata in the means of the estimation variable while there is considerable difference in the poststrata. Consequently, the use of the poststratified estimator will be effective in reducing variances.

Parameters in the simulation, shown in Table 3, were the proportion by design-stratum whose status was known, the proportion eligible among those with known status, and the proportion responding among those with known status that were eligible. For each unit in the population Bernoulli random variables were generated with probabilities given in Table 3 to determine whether a unit had a known status, was eligible, and was a respondent. This procedure was repeated for every sample that was selected. In addition to the response probabilities shown in the last line of Table 3, we ran simulations with rates of 0.40, 0.45, 0.50, 0.55, and 0.60 across the design strata. Results were qualitatively similar to the ones we report here.

Stratified simple random samples (*stsrs*) of size $n = 100$, 250, and 500 were selected without replacement. An equal number of sample units was allocated to each of the five design strata. Four versions of the grouped jackknife were computed: $G = 10, 25, 50, 100$. In each case, the initial sample within each stratum was divided into $G/5$ random groups. Note that the combination ($n = 100$, $G = 100$) corresponds to the standard delete-one jackknife. For the variance estimators, other than the jackknife, 4,000 samples were

Table 3. Simulation parameters for the proportions known, eligible, and responding in the poststratified population

	Design stratum				
	1	2	3	4	5
Known status	0.70	0.75	0.80	0.85	0.90
Eligible	0.70	0.75	0.80	0.85	0.90
Responding	0.60	0.65	0.70	0.75	0.80

selected for each sample size. Because substantially more computing was needed for the jackknife, 1,000 *stsrs*'s were selected for that variance estimator.

The adjustment for unknown eligibility, defined by (1), and the nonresponse adjustment, defined by (2), were made within each design-stratum. The eligible responding units and the known ineligible units were then poststratified as shown in (6).

4.2. *HMT population*

Hansen, Madow, and Tepping (1983) used a population in which Y depended on a single auxiliary x. In their population $E(Y|x) = 0.4 + 0.25x$ and $\text{var}(Y|x) = 0.0625x^{3/2}$ with both x and Y having particular gamma distributions. We generated an HMT population of size $N = 5{,}000$; details of the construction are in HMT (1983). The probability of response was modeled as logistic. The following function was developed through some experimentation and led to an average response rate of about 60% in simple random samples and 77% in probability proportional to x samples:

$$p(x) = \{1 + \exp[-(-2.1972 + 0.3081x)]\}^{-1}$$

This logistic function with these parameters leads to an increasing probability of response as x increases. In an establishment survey, for example, this corresponds to units having more employees being more likely to respond. No ineligibles were used for the simulation using the HMT population.

The estimator of the population total was the linear regression estimator defined in (5). Two types of samples were selected: simple random samples selected without replacement (*srswor*) and probability proportional to size (*pps*) samples where the size was x. We selected the *pps* samples by first randomizing the order of the population and then selecting a systematic *pps* sample with a random starting point. Samples of size $n = 100$, 200, and 500 were selected using both *srswor* and *pps*. Four versions of the grouped jackknife were computed: $G = 10$, 25, 50, 100. In each case, the initial sample was randomly divided into G groups with each group having n/G units. As for the poststratified population, 1,000 samples were selected for each sample size for the jackknife. For the other variance estimators, 4,000 samples were selected.

The nonresponse adjustment, defined by (2), was made within groups formed by sorting the initial sample based on x. Groups were formed that had about 10 respondents each for $n = 100$, 20 respondents when $n = 200$, and 50 when $n = 500$.

4.3. *Simulation results – ignorable nonresponse*

Figure 1 summarizes results for the poststratified population for ignorable nonresponse. The figure gives columns for the relative bias (relbias) of a variance estimator, coverage of 95% confidence intervals (CI's), mean half-width of the confidence intervals, and the standard error of the half-widths. The relbias of a variance estimator v is computed as $100(\bar{v} - mse(\hat{T}))/mse(\hat{T})$ where $mse(\hat{T}) = \sum_{s=1}^{S}(\hat{T}_s - T_s)^2/S$. S is the total number of samples, $\hat{T}_s$ is the estimated total from sample s, and T_s is the total of Y for the eligible units across the whole population. Note that we take the empirical *mse* as the target for variance estimation, rather than the empirical variance of $\hat{T}$. The population total T_s varies from one

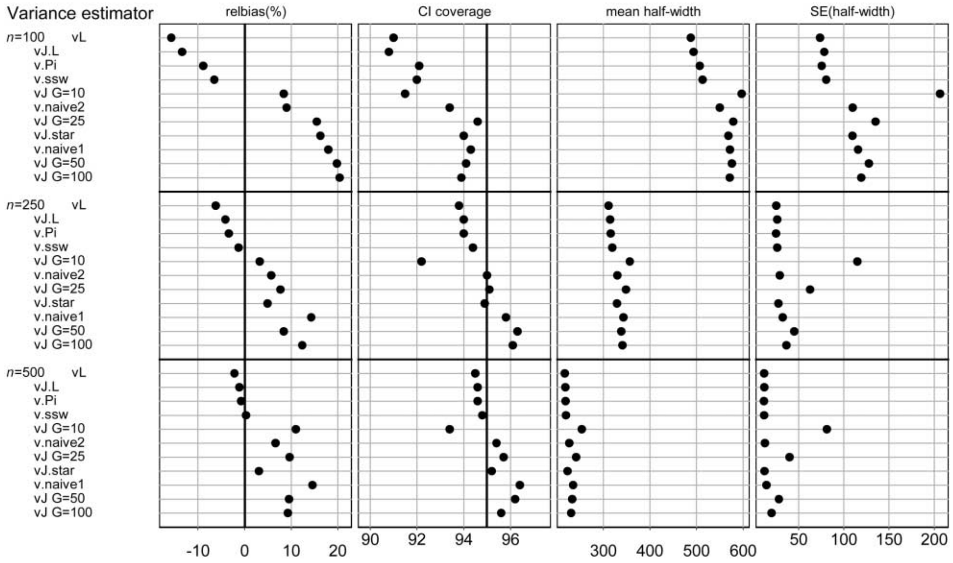

Fig. 1. Comparisons of the relbias, coverage rates and mean half-widths of 95% confidence intervals, and standard error of half-widths of different estimators of variance for the post stratified estimator of a total. Key to row labels: $vL = v_L$; $vJ.L = v_{JL}$; $v.Pi = v_\pi$; $v.ssw = v_{SSW}$; $vJ(G = nnn) = v_J$ with $G = nnn$ groups; $v.naive2 = v_{naive2}$; $vJ.star = v_J^*$; $v.naive1 = v_{naive1}$

sample to another because the population indicator variables for whether a unit is eligible were regenerated independently for each sample.

In all the simulations using ignorable nonresponse, the estimates of the totals are approximately unbiased after adjustment for ineligibility, nonresponse, and use of auxiliary data through poststratification or linear regression estimation. Thus, the relbiases and confidence interval coverages reported in this section are unaffected by any bias in the $\hat{T}$'s.

The 95% confidence interval using a variance estimate v was calculated as $\hat{T}_s \pm t_{DF}\sqrt{v}$ where t_{DF} is a multiplier from the t-distribution with DF degrees of freedom, and DF is the degrees of freedom associated with v. The half-width of an interval is $t_{DF}\sqrt{v}$. For the jackknife variance estimates, we used $DF = \sum_h (G_h - 1) = G - H$. For v_{naive2}, we used $DF = \sum_h (n_{h,ER} - 1) = n_{ER} - H$. For the other variance estimators, we used $DF = \sum_h (n_{h,ER \cup IN} - 1) = n_{ER \cup IN} - H$. Use of multipliers from the t-distribution, rather than 1.96 from the standard normal, makes a considerable difference in the width of intervals for the jackknife with $G = 10$ or 25, but is less important for the other variance estimators.

Table 4. Sample sizes used in estimation in poststratified population

n	Number of eligible respondents + known ineligibles, $n_{ER \cup IN}$	
	Mean	Range
100	61.5	(42, 77)
250	153.6	(126, 181)
500	307.2	(270, 345)

Table 4 lists the mean and range of the number of eligible respondents plus known ineligibles across the samples. There is a substantial reduction from the initial sample size because of cases that had unknown status or were nonrespondents. Thus, a major source of variation is the number of sample units used in evaluating both the estimate of the total and the variance of that estimate.

The variance estimators are sorted in Figure 1 by the relbias obtained for samples of size $n = 100$. For $n = 100$ v_L, v_{JL}, v_π, and v_{SSW} all have negative biases with the linearization variance estimator being the worst at -15.7%. The relbiases for the other estimators range from 8.4% for v_J $(G = 10)$ to 20.4% for v_J $(G = 100)$. The full delete-one jackknife, thus, has the largest relbias when $n = 100$. The biases diminish for $n = 250$ and 500 although the pattern persists of negative biases for v_L, v_{JL}, v_π, and v_{SSW} and positive biases for the other variance estimators.

Underestimation by v_L, v_{JL}, v_π, and v_{SSW} leads to CI's that cover at less than the nominal rate. When $n = 100$, the coverage rate with these choices is at most 92%. For $n = 500$ coverage for these estimators is near 95%. For the jackknife choices, overestimation of the *mse* does not necessarily lead to overcoverage by the CI's. When $n = 100$, 15–20% relbias produces coverage rates of 93.9% to 94.6% for v_J $(G = 25, 50, 100)$. The least biased of the grouped jackknife choices, v_J $(G = 10)$, has the worst CI coverage at 91.5% for $n = 100$. For the two larger sample sizes, v_J $(G = 25, 50, 100)$ all have at least 95% coverage. The approximate jackknife, v_J^*, performs well, having relbias less than 5% for $n = 250$ and 500 and having coverage rates of 94.0, 94.9, and 95.2 at the three sample sizes.

The average half-widths and standard errors of the half-widths show some differences between the variance estimates. Average lengths are somewhat longer for the jackknife estimates and the related estimate v_J^*, especially for $n = 100$. Longer intervals are due to the variance estimates and the multipliers from the *t*-distribution being larger for the jackknife estimates. The estimate that stands out for its high variability is v_J $(G = 10)$. The stability of the grouped jackknife increases as the number of groups increases – a phenomenon that is well known among practitioners (see also Wolter 1985, Section 4.2.5).

The estimators, v_{naive1} and v_{naive2}, are theoretically incorrect for the poststratified estimator but are included here since users of some software packages might select them. Both are overestimates since each uses the wrong residual for the poststratified estimator. The relbias of v_{naive1} ranges from about 18% at $n = 100$ to 14.6% at $n = 250$. Note that there is no decrease in relbias when moving from $n = 250$ to $n = 500$. v_{naive2} is smaller than v_{naive1} and is actually less biased because it ignores the fact that the eligibles are a domain. The positive bias of v_{naive1} leads to overcoverage by the confidence intervals, although the problem is not severe. At $n = 250$ and 500, for example, the empirical coverage rate using v_{naive1} is 96.4%.

As a point of comparison, we ran the simulations with no unknowns, no ineligibles, and no nonresponse. Results are shown in Table 5 for $n = 100$. The outcomes for $n = 250$ and 500 have similar patterns but the differences among the variance estimators are less pronounced. There is a difference between the relbiases with and without the complications of unknowns, ineligibles, and nonrespondents. With these complications the relbiases generally become more extreme in both the positive and negative directions, but this effect is larger for the jackknife estimators. For example, the relbias of v_J $(G = 100)$ is 6.25% without the complications but is 20.45% with them. Of course, much of the difference

Table 5. Comparison of relbiases for samples of n = 100 from the poststratified population with and without unknowns, ineligibles, and nonresponse. Results for the grouped jackknife are based on 1,000 samples; results for the other estimators use 4,000 samples

Variance estimator	Relbias (%)	
	Without unknowns, ineligibles, nonresponse	With unknowns, ineligibles, nonresponse
v_L	-10.16	-15.68
v_{JL}	-6.27	-13.32
v_π	-6.42	-8.85
v_{SSW}	-2.36	-6.44
$v_J\ (G = 10)$	4.68	8.46
v_{naive2}	–	9.06
$v_J\ (G = 25)$	5.27	15.54
v_J^*	5.92	16.31
v_{naive1}	12.21	17.99
$v_J\ (G = 50)$	5.79	19.86
$v_J\ (G = 100)$	6.25	20.45

between the two columns in Table 5 is due to the fact that the samples without unknowns, ineligibles, and nonresponse are the full size of $n = 100$ while the samples with these complications average 61.5 respondents plus known ineligibles (see Table 4).

Figure 2 summarizes the results for the HMT population using the linear regression estimator, defined by (5), along with simple random samples. The same summary statistics are graphed as in Figure 1-relbias, coverage of 95% coverage, mean half-width of CIs, and

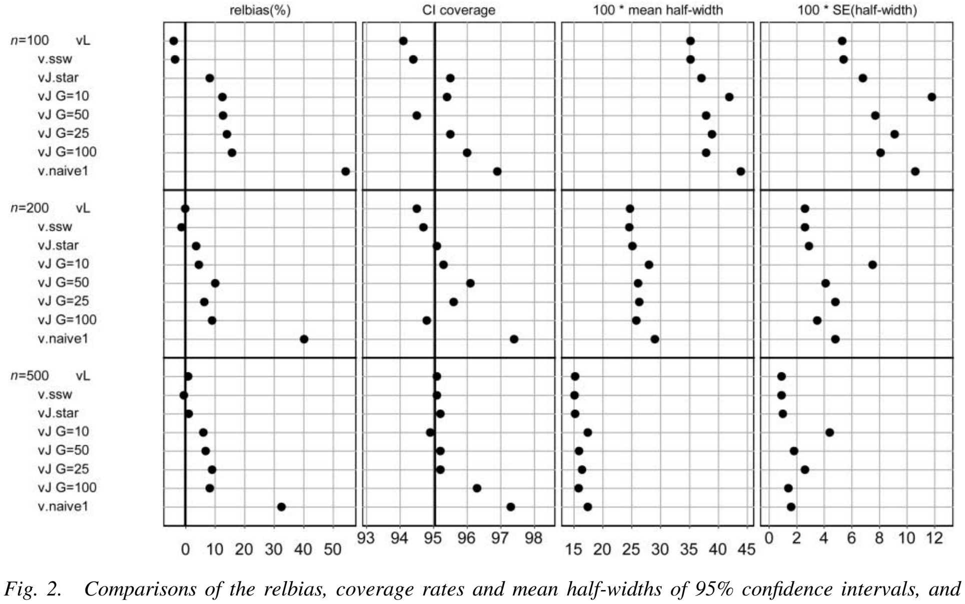

Fig. 2. Comparisons of the relbias, coverage rates and mean half-widths of 95% confidence intervals, and standard error of half-widths of different estimators of variance for the linear regression estimator of a total in the HMT population. Simple random samples. Key to row labels: $vL = v_L$; $v.ssw = v_{SSW}$; $vJ.star = v_J^*$; $vJ\ (G = nnn) = v_J$ with $G = nnn$ groups; $v.naive1 = v_{naive1}$

standard error of CI half-widths. For the linear regression estimator, $v_L = v_\pi$ since $\sum_{s_{ER}} w_{2i} r_i = 0$, and $v_{SSW} = v_{JL}$ since $\sum_{s_{ER}} w_i^* r_i = 0$. The degrees of freedom used for the CIs were $G - 1$ for the grouped jackknife and $n_{ER} - 1$ for the other estimators. Table 6 lists the mean and range of respondents for the three sample sizes. As in the simulations using the poststratified population, there is considerable variation in the size of the simulated samples.

The estimators $v_L(= v_\pi)$ and $v_{SSW}(= v_{JL})$ are negatively biased at $n = 100$ but the biases are less than 5%. At $n = 200$ and 500 these estimators are nearly unbiased. The grouped jackknife estimators and v_J^* are overestimates at all sample sizes, with v_J^* having the smallest positive bias at all three sample sizes. The estimator v_{naive1} was again included because it is easy to select in some survey software packages. It is extremely biased at all three samples (32.4 to 54.2%) and should not be used in this situation.

Confidence interval coverage is near the nominal 95% for $v_L(= v_\pi)$ and $v_{SSW}(= v_{JL})$. The coverage rates for the grouped jackknife vary depending on the number of groups used and range from 94.5 to 96.3% across the three sample sizes. Although v_{naive1} is a severe overestimate, this results in overcoverage by CIs of only two to three percent. For $n = 200$, for example, the relbias is 40.1% but the confidence interval coverage is 97.4%. Differences in average half-widths are not particularly remarkable. The standard errors of half-widths are clearly smaller for $v_L(= v_\pi)$, $v_{SSW}(= v_{JL})$, and v_J^*. The leverage-adjusted estimator v_J^* is again one of the better choices since it combines small positive bias, good confidence interval coverage, and CI widths that are not too variable.

Figure 3 shows the results for the *pps* samples. When $n = 100$, v_L and v_{SSW} have negative biases of -17.1 and -13.7%; their bias remains negative but approaches zero for the larger sample sizes. The grouped jackknife estimators all have positive biases that diminish with increasing sample size. When $n = 100$, the biases range from 33.2 for v_J $(G = 50)$ to 52.2% for v_J $(G = 100)$, and the range is 13.1–14.9% when $n = 500$. Notice that at the smallest sample size, the full delete-one jackknife, v_J $(G = 100)$, has the largest bias, which was also true for the $n = 100$ *srswor* samples from HMT and the $n = 100$ *stsrs* samples from the poststratified population. As in the other simulations, v_J^* is conservative but its positive bias is not so large as for the grouped jackknife estimators.

Underestimation by v_L and v_{SSW} leads to undercoverage by the confidence intervals, especially at $n = 100$. The grouped jackknife estimators have better CI coverage for $n = 100$ and 200, which is obtained at the cost of overestimation. The empirical coverage of v_J^* is again near the nominal 95% at all sample sizes. As might be expected, v_L and v_{SSW}

Table 6. Sample sizes used in estimation in the HMT population

	Number of respondents —*srswor* samples		Number of respondents —*pps* samples	
n	Mean	Range	Mean	Range
100	59.1	(43, 76)	77.1	(62, 90)
200	118.4	(92, 142)	154.2	(131, 175)
500	296.0	(258, 335)	385.8	(348, 417)

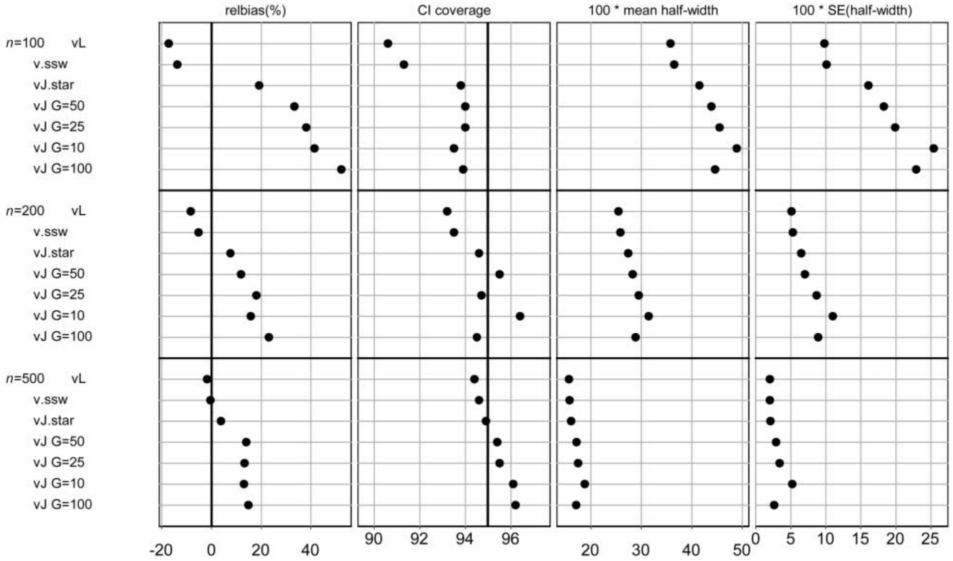

Fig. 3. Comparisons of the relbias, coverage rates and mean half-widths of 95% confidence intervals, and standard error of half-widths of different estimators of variance for the linear regression estimator of a total in the HMT population. Probability proportional to size samples. Key to row labels: vL = v_L; v.ssw = v_{SSW}; vJ.star = v_J^*; vJ (G = nnn) = v_J with G = nnn groups.

give the shortest and most stable interval lengths at all sample sizes while v_J^* is intermediate in width and stability.

We also ran the *pps* simulations with complete response. Results are shown in Table 7 for $n = 100$. The biases of the estimators are affected substantially by whether there is nonresponse. As in the poststratified population, the relbiases with nonresponse generally become more extreme in both the positive and negative directions. For example, v_L is almost unbiased with full response (relbias $= -0.49\%$) but its relbias is -17.11% with nonresponse. The relbias of v_J $(G = 100)$ is 8.44% with full response, but is 52.30% with

Table 7. Comparison of relbiases for pps samples of n = 100 from the HMT population with and without nonresponse. Results for the grouped jackknife are based on 1,000 samples; results for the other estimators use 4,000 samples

	Relbias (%)	
Variance estimator	With full response	With nonresponse
v_L	−0.49	−17.11
v_{SSW}	2.71	−13.67
v_J^*	7.41	19.16
v_J $(G = 50)$	8.83	33.34
v_J $(G = 10)$	5.43	41.46
v_J $(G = 25)$	10.46	38.14
v_J $(G = 100)$	8.44	52.30
v_{naive1}	23.93	199.00

nonresponse. The changes are much more extreme than those in Table 5 for the (*stsrs*, $n = 100$) samples.

The estimator v_{naive1} is extremely biased at all sample sizes for the linear regression estimator, regardless of whether there is nonresponse, since the residual it uses is incorrect. For example, the relbiases for $n = 100$ in Table 7 are 23.93% with full response and 199.00% with nonresponse. Results for v_{naive1} were omitted from Figure 3 since its biases at all sample sizes were so large.

We also conducted simulations using the poststratified population where the nonresponse was not ignorable. Parameters in Table 3 were used for units with $Y = 0$, while units with $Y = 1$ had response probabilities of 0.9 times the values on the third line of the table. Other features of the simulations were the same as described earlier. This led to $\hat{T}_{PS}$ having a bias of -8.82%. Because none of the variance estimators accounts for the bias of $\hat{T}_{PS}$, the relbiases were all shifted in the negative direction as compared to those in the third column of Table 5. Consequently, confidence interval coverage deteriorated as compared to the simulations with ignorable nonresponse. The grouped jackknife fared better in this regard since it tends to overestimate the variance. For example, for $n = 100$ v_J $(G = 100)$ covered 92.4% of the time while the coverage rate for v_L was 87.7%. Nonignorable nonresponse led to degraded performance by all of the variance estimators and none of them can be strongly recommended.

5. Conclusion

Two general types of variance estimators used in survey sampling are ones based on squared residuals, like linearization variance estimators, and replication variance estimators. The former are computationally less demanding since explicit variance formulas are evaluated. However, the linearization and related estimators require a separate derivation for each type of statistic. Special case theory and programming may be needed if existing software packages do not meet users' needs.

The replication estimators are more intensive computationally since they involve repeated calculation of an estimate based on subsamples of the full sample. However, the computational algorithm is fairly simple in the full-response case: compute the estimate, no matter how complicated, separately for the full sample and each replicate, and combine the results using a variance formula appropriate to the replication method. For many types of (differentiable) estimators no special case theoretical formulas are required.

Theory for the two types of estimators shows that asymptotically there is little difference in large samples with full response. The basic design-based or model-based theory does involve some strong assumptions, e.g., the first-stage is selected with replacement or the first-stage units are independent. More importantly, the possibilities that there are ineligible units and that some units will not respond are often not considered when comparing the variance estimators. For the replication variance estimators, there is literature showing that adjusted data values can be used to create consistent variance estimators. However, these adjustments may vary depending on the form of the basic estimator (total, mean, ratio, etc.) and are not included in the commercial software packages now available.

In the simulations presented here, the linearization estimators and several other variance estimators based on squared residuals are underestimates. This problem is considerably worse when there are ineligible units and ignorable nonresponse, as opposed to full response, and leads to confidence intervals that cover at less than the nominal rate. The one exception among the squared-residual estimators is a leverage-adjusted estimator that approximates the jackknife and tends to be somewhat of an overestimate. The leverage adjustment does successfully compensate for the weight adjustments for ineligibility and nonresponse without being excessively conservative.

The contention that replicate estimators are superior to linearization estimators because they facilitate accounting for various stages of weight adjustment is true but only in a limited sense. The grouped jackknife estimator, which recomputes weight adjustments for every replicate, tends to be an overestimate and the degree of overestimation can be substantial with smaller sample sizes. This overestimation is accompanied by some overcoverage by confidence intervals, although the excess above the nominal level is small. Thus, nominal coverage is obtained but at the expense of potentially large overestimation.

In summary, when there is ignorable nonresponse, the only estimator in this study that combines reasonably small positive bias with near-nominal confidence interval coverage is the leverage-adjusted estimator v_J^*. Although the linearization and related estimators and the jackknife estimator may converge to the same value in large samples, the convergence can be slow and is adversely affected by ineligibility and nonresponse even when the nonresponse mechanism is ignorable. Plus, the two types of estimators converge to the desired value from opposite sides, so that neither is ideal. The leverage-adjusted estimator is only slightly more complicated to calculate than a linearization estimator and is much less computationally burdensome than a replication estimator. Of course, no simulation study can cover all possibilities. However, the populations and steps in weighting adjustment used here are realistic, and we expect these findings will be germane to a variety of surveys.

6. References

Cochran, W.G. (1977). Sampling Techniques. New York: John Wiley and Sons.

Ernst, L.R. and Williams, T. (1987). Some Aspects of Estimating Variances by Half-sample Replication in CPS. Proceedings of the American Statistical Association, Section on Survey Research Methods, 480–485.

Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983). An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys. Journal of the American Statistical Association, 78, 776–793.

Horn, S.D., Horn, R.A., and Duncan, D.B. (1975). Estimating Heteroscedastic Variances in Linear Models. Journal of the American Statistical Association, 70, 380–385.

Lago, J., Massey, J., Ezzati, T., Johnson, C., and Fulwood, R. (1987). Evaluation of the Design Effects for the Hispanic Health and Nutrition Examination Survey. Proceedings of the American Statistical Association, Section on Survey Research Methods, 595–600.

Lemeshow, S. (1979). The Use of Unique Statistical Weights for Estimating Variances with the Balanced Half-Sample Technique. Journal of Statistical Planning and Inference, 3, 315–323.

Lundström, S. and Särndal, C.-E. (1999). Calibration as a Standard Method for Treatment of Nonresponse. Journal of Official Statistics, 15, 305–327.

Rao, J.N.K. (1996). On Variance Estimation with Imputed Survey Data. Journal of the American Statistical Association, 91, 499–506.

Rust, K. (1987). Practical Problems in Sampling Error Estimation. Bulletin of the International Statistical Institute, Invited paper 10.3, 1–18.

Rust, K. and Rao, J.N.K. (1996). Variance Estimation for Complex Estimators in Sample Surveys. Statistics in Medicine, 5, 381–397.

Särndal, C.-E. (1996). Efficient Estimators with Simple Variance in Unequal Probability Sampling. Journal of the American Statistical Association, 91, 1289–1300.

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

SAS Institute (2001). SAS Stat Version 8. Cary NC: SAS Institute.

Shah, B.V., Barnwell, B., Bieler, G., Boyle, K., Folsom, R., Lavange, L., Wheeless, S., and Williams, R. (1996). Technical Manual: Statistical Methods and Algorithms Used in SUDAAN. Research Triangle Park, NC: Research Triangle Institute.

Smith, P.J., Srinath, K.P., Battaglia, M.P., Graubard, B.I., Barker, L., Hoaglin, D.C., Frankel, M., and Khare, M. (2000). Issues Relating to the Use of Jackknife Methods in the National Immunization Survey. Proceedings of the American Statistical Association, Section on Survey Research Methods, 709–714.

Stata Corporation (2001). STATA Statistical Software, Version 7. College Station TX: Stata Corporation.

Valliant, R. (1993). Poststratification and Conditional Variance Estimation. Journal of the American Statistical Association, 88, 89–96.

Valliant, R. (2002). Variance Estimation for the General Regression Estimator. Survey Methodology, 28, 103–114.

Westat (2000). WesVar 4.0 User's Guide. Rockville MD: Westat.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. Econometrica, 50, 1–25.

Wolter, K. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.

Yung, W. and Rao, J.N.K. (1996). Jackknife Linearization Variance Estimators under Stratified Multi-stage Sampling. Survey Methodology, 22, 23–31.

Yung, W. and Rao, J.N.K. (2000). Jackknife Variance Estimation under Imputation for Estimators Using Poststratification Information. Journal of the American Statistical Association, 95, 903–915.

Received June 2002
Revised April 2003