A Comparison of Variances for Poststratification to Estimated Control Totals

Jill A. Dever[1] and Richard Valliant[1,2]
[1]Joint Program in Survey Methodology, University of Maryland
[2]Institute for Social Research, University of Michigan

Key Words: estimated-control calibration, sampling frame coverage bias, survey-estimated control totals

1. Introduction

Calibration estimators, a label first used by Deville and Särndal (1992), identify a class of estimators that borrow strength from auxiliary information to improve the efficiency of survey estimates over more traditional weighting methods. The $g\,(g \geq 1)$ auxiliary variables are assumed to be (linearly) related with the set of key survey variables. This association is directly related to the gains in efficiency. These estimators are used in all types of surveys throughout the research world.

The poststratified estimator, a well-known and widely applied calibration estimator, is sometimes used to correct for sampling frame problems such as undercoverage (e.g., Kott 2006). Undercoverage occurs when the sampling frame fails to contain all units for the population under study. For example, estimates for the Behavioral Risk Factor Surveillance System (BRFSS), a nationwide random-digit-dial (RDD) telephone survey conducted by the Centers for Disease Control and Prevention (CDC), are poststratified (i.e., calibrated) to population counts that include households with and without land-line telephone service (CDC 2006).

A primary assumption with calibration is that the control totals, to which the auxiliary sample estimates are calibrated, are either true population values known without error, or are taken from an independent, highly precise survey that is much larger than the survey requiring calibration. In some cases, however, these controls are estimates obtained from other surveys which possess a non-negligible sampling variance. For example, there are efforts to calibrate Web panel surveys to separate, higher-quality reference surveys that are not much larger than the panel surveys themselves (e.g., Krotki 2007; Terhanian, et al. 2000).

Calibration variance estimators have been developed for *traditional* or fixed-control calibration. Many researchers apply these formulae even though the controls are estimated. The tacit assumption is that any additional variance (and bias) associated with these controls is negligible and can be ignored. Currently, the validity of this assumption can not be checked. We label the methodology which properly accounts for the estimated controls as estimated-control (EC) calibration.

The goal of our research is to develop and evaluate estimators for complex sampling designs under EC calibration. In this paper, we focus specifically on the estimated-control poststratified (ECPS) estimator of a population total for data collected from a two-stage design where n_h first-stage sampling units are selected with replacement from within H design strata. Through theoretical development (*section 3*) and a simulation study, we compare the properties for variance estimators developed for the ECPS with variance estimators chosen under the naïve "population control total" assumption. Both linearization and replication variance estimators are examined. Illustrations are given of the effects on variances of different levels of precision in the estimated controls. The specifications for the simulation study are discussed in *section 4*, followed by a summary of the results (*section 5*). We conclude the paper with an overview of future research in this area. We begin in *section 2* with a brief summary of the extensive literature related to sample weight calibration, and a definition of the ECPS.

2. Estimated-Control Poststratified Estimator

The general form of a calibration estimator is best described as an expansion or linear weighting estimator (Estevao and Särndal 2000). An estimated population total of a variable y is $\hat{t}_y = \sum_{k \in s} w_k y_k$, where the calibrated analysis weight (w_k) for the k^{th} unit in the sample s is a function of the design weight π_k^{-1} and a calibration adjustment factor g_k, also known as a g-weight (Särndal, Swensson, and Wretman 1992). G-weights are traditionally calculated by minimizing a specified function that measures the distance between the design and calibrated weights. A model which defined the relationship between the outcome variable (y_k) and the auxiliary variables ($\mathbf{x}_k'$), referred to in the literature as either the calibration or prediction model, is also required to completely specify the resulting point estimator. For example, the generalized least squares (or chi-square) distance function $\sum_{k \in g} \left(w_k - \pi_k^{-1} \right)^2 \Big/ c_k \pi_k^{-1}$ and a linear prediction model such that $E_M \left(y_k \right) = \mathbf{x}_k' \mathbf{B}$ and $Var_M \left(y_k \right) = c_k \sigma^2$ generates a closed-form solution to the calibration equations called a generalized regression estimator (GREG) for $c_k = 1$ (Deville and Särndal 1992). The distance function is minimized subject to a set of calibration constraints (or calibration equations) defined as:

$$\mathbf{t}_x = \hat{\mathbf{t}}_x \tag{2.1}$$

where $\hat{\mathbf{t}}_x = \sum_{k \in s} w_k \mathbf{x}_k$, $\mathbf{t}_x = \sum_{k \in U} \mathbf{x}_k$, $\mathbf{x}_k$ is a vector of length G containing auxiliary or benchmark variable values for element *k*, $\mathbf{t}_x$ is the vector of population controls (counts) corresponding to the g auxiliary variables, and $\hat{\mathbf{t}}_x$ is the corresponding vector of estimated controls.

Poststratification, a well known and widely used calibration technique, has been shown to reduce errors associated with sampling, nonresponse, and coverage (e.g., Särndal, Swensson, and Wretman 1992; Kott 2006). The formula for a poststratified estimator of a population total for a survey with 100 percent response is

$$\hat{t}_{yPS} = \sum_{g=1}^{G} N_g \frac{\hat{t}_{yg}}{\hat{N}_g} = \sum_{g=1}^{G} N_g \frac{\sum_{k \in s_g} \pi_k^{-1} y_k}{\sum_{k \in s_g} \pi_k^{-1}} \tag{2.2}$$

where N_g is the number population units in the g^{th} poststratum ($g = 1, \ldots, G$), $\hat{N}_g$ is estimated number of population units in poststratum g, and s_g represents the set of sample units belonging to poststratum *g*. Though a simplified notation is used, $\hat{t}_{yg}$ is calculated under the particular sampling design of the sample survey, hereafter referred to as the *analytic survey*. Relating the poststratified estimator with the calibration equations (2.1), $\hat{\mathbf{t}}_x = \sum_{k \in s_g} w_k \mathbf{x}_k = \sum_{k \in s_g} \pi_k^{-1} = \hat{N}_{g\pi}$, where $\mathbf{x}_k = x_k = 1$ if $k \in s_g$ (zero otherwise) and $\mathbf{t}_x$ corresponds to the vector of N_g's. The poststratified estimator is a type of GREG generated under a group-mean prediction model with one covariate, i.e., the model expectation and variance are defined as $E_M \left(y_k \right) = \beta_g$ and $Var_M \left(y_k \right) = \sigma_g^2$ respectively for $k \in U_g$, the population units within poststratum g.

To facilitate our discussion of EC calibration, we label the survey requiring calibration as the *analytic survey* and the source of the control totals as the *benchmark survey*. Note that more than one benchmark survey may be tapped for the control totals. However, for our discussion, we will assume only one benchmark survey.

The estimated-control poststratified estimator (ECPS) of a population total under the assumption of complete response is defined by a slight modification to the traditional poststratified estimator (2.2):

$$\begin{aligned} \hat{t}_{yECPS} &= \sum_{g=1}^{G} \hat{N}_{Bg} \frac{\hat{t}_{Ayg}}{\hat{N}_{Ag}} \\ &= \sum_{g=1}^{G} \left(\sum_{k \in s_{Bg}} w_{Bk} \right) \frac{\sum_{k \in s_{Ag}} \pi_{Ak}^{-1} y_k}{\sum_{k \in s_{Ag}} \pi_{Ak}^{-1}} . \end{aligned} \tag{2.3}$$

The subscripts A and B are used to denote estimation from either the analytic or benchmark surveys. The population values (N_g) in (2.2) are replaced with estimates from the benchmark survey ($\hat{N}_{Bg}$) and are calculated as the sum of the benchmark survey weights (w_{Bk}) within poststratum g. The calculation of the analytic survey estimates, $\hat{t}_{Ayg}$ and $\hat{N}_{Ag}$, remains the same as defined previously (i.e., $\hat{t}_{Ayg} = \hat{t}_{yg}$ and $\hat{N}_{Ag} = \hat{N}_g$); we apply the A subscript to ensure clarity of the estimation source.

3. Variance Estimators for the ECPS

The usual set of variance estimators have been developed for traditional calibration and are available in software designed to analyze survey data (e.g., R 2005; SAS 2004; Stata 2004; and SUDAAN 2004). Taylor series linearization, a variance technique available for any real-valued function with continuous first- and second-order partial derivatives, is discussed in, for example, Wolter (1985) and Binder (1995). Särndal, Swensson, and Wretman (1989, 1992) developed an approximate linearization population variance for the GREG of a population total as a function of the population residuals from the specified prediction model and the *g*-weights. Stukel, Hidiroglou, and Särndal (1996) discuss a *g*-weighted GREG variance formula developed by Hidiroglou, Fuller, and Hickman (1980) for a stratified multi-stage design using the linear substitute (or ultimate cluster) method (Kalton 1979). Replication methods, such as balanced repeated replication and jackknife, have been discussed for a variety of calibration estimators in sources such as Valliant 1993, Canty and Davison 1999, and Demnati and Rao 2004. However, limited work has been completed on variance estimation for EC calibration. Our research will contribute to this very area.

Four estimated-control variance estimators for the ECPS that account for the variance in the control totals were compared for this study. They include two newly developed linearization estimators, and two delete-one jackknife variance estimators. With the delete-one jackknife, replicates are created by deleting one primary sampling unit (PSU) and adjusting the weights for the remaining PSUs within the corresponding design stratum. This results in a total of $R=\sum_{h=1}^{H} n_h$ replicates calculated by summing the number of PSUs per stratum (n_h) across the strata ($h=1,...,H$). We additionally compare the properties of these estimators with a linearization variance estimator and a delete-one jackknife under the naïve assumption that the estimated controls are population values known without error (see, e.g., section 6.6 in Särndal, Swensson, and Wretman 1992).

An effective variance estimator will reproduce the corresponding population sampling variance in expectation. The approximate (or asymptotic) population sampling variance for the ECPS is adapted from similar work by Fuller (1998) and has the form

$$\begin{aligned} AV\left(\hat{t}_{yECPS}\right) &= \mathbf{N}_B'\mathbf{V}_A\mathbf{N}_B + tr\left[\mathbf{V}_A\mathbf{V}_B\right] + \overline{\mathbf{Y}}_A'\mathbf{V}_B\overline{\mathbf{Y}}_A \\ &\cong \mathbf{N}_B'\mathbf{V}_A\mathbf{N}_B + \overline{\mathbf{Y}}_A'\mathbf{V}_B\overline{\mathbf{Y}}_A \end{aligned} \tag{3.1}$$

where $\mathbf{N}_B$ is a vector of population counts within the G poststrata; $\overline{\mathbf{Y}}_A$ is a G-length vector with population components of the form $\bar{y}_g = t_{yg}/N_g$; $\mathbf{V}_A$ is the (variance-) covariance matrix of the estimated components of the vector $\overline{\mathbf{Y}}_A$ (i.e., $\hat{\bar{y}}_g$); and $\mathbf{V}_B$ is the covariance matrix of the G benchmark control estimates $\left(\hat{N}_{B1},\ldots,\hat{N}_{BG}\right)$. The first component, $\mathbf{N}_B'\mathbf{V}_A\mathbf{N}_B$, is the approximate variance for the traditional poststratified estimator ($\hat{t}_{yPS}$), i.e., the benchmark estimates are treated as fixed. The later component, $\overline{\mathbf{Y}}_A'\mathbf{V}_B\overline{\mathbf{Y}}_A$, is variance associated with the benchmark estimates with the analytic estimates are treated as fixed. The term $tr\left[\mathbf{V}_A\mathbf{V}_B\right]$ is assumed to be of lower order, effectively $O\left(n_A^{-1}\right)O\left(n_B^{-1}\right)$.

We discuss the estimators for the population sampling variance below. The sample estimators are calculated by substituting sample estimates for the corresponding variance components.

3.1 Taylor Series Linearization (ECTS)

We derived a first-order linearization variance estimator for the $\hat{t}_{yECPS}$ of the following form:

$$var\left(\hat{t}_{yECPS}\right) = \hat{\mathbf{N}}_B'\hat{\mathbf{V}}_A\hat{\mathbf{N}}_B + tr\left[\hat{\mathbf{V}}_A\hat{\mathbf{V}}_B\right] + \hat{\overline{\mathbf{Y}}}_A'\hat{\mathbf{V}}_B\hat{\overline{\mathbf{Y}}}_A \tag{3.2}$$

where $\hat{\mathbf{N}}_B$ is the estimated vector of the G benchmark controls, and $\hat{\mathbf{V}}_A$ is the estimated covariance matrix of the estimates $\hat{\overline{\mathbf{Y}}}_A = \left(\hat{t}_{Ay1}/\hat{N}_{A1},\ldots,\hat{t}_{AyG}/\hat{N}_{AG}\right)$. Retaining an estimate of the positive term $tr\left[\hat{\mathbf{V}}_A\hat{\mathbf{V}}_B\right]$ in (3.2) means that this estimator should be viewed as a conservative approximation of (3.1). We leave this term in the formula for comparative purposes.

3.2 Residual Linearization Method (ECRL)

Another approximate variance estimator is

$$var\left(\hat{t}_{yECPS}\right) = var_{res}\left(\hat{t}_{yPS}\right) + \hat{\overline{\mathbf{Y}}}_A'\hat{\mathbf{V}}_B\hat{\overline{\mathbf{Y}}}_A \tag{3.3}$$

which is derived by substituting the first component in (3.1) with the variance estimator for the $\hat{t}_{yPS}$ (2.2) that is a function of the prediction model residuals and the g-weights, i.e., $var_{res}\left(\hat{t}_{yPS}\right)$. The second component remains as specified previously.

3.3 Fuller Two-Phase Jackknife Method (ECF2)

Isaki, Tsay, and Fuller (2004) applied a two-phase delete-one jackknife variance estimator developed by Fuller (1998) to an EC calibration situation. The premise behind Fuller's methodology is to take a spectral (eigenvalue) decomposition of the benchmark covariance matrix ($\mathbf{V}_B$), develop benchmark adjustments that are a function of the resulting eigenvalues and eigenvectors, and add the adjustments to the controls to create a set of replicate controls. A randomly chosen subset of the R replicates is calibrated to the G constructed replicate controls. Specifically, the benchmark control total for the r^{th} variance replicate within the h^{th} design stratum is defined as

$$\hat{\mathbf{N}}_{B(hr)} = \hat{\mathbf{N}}_B + c_h\delta_{(hr)}\sum_{g=1}^{G}\delta_{g|(hr)}\hat{\mathbf{z}}_g' \tag{3.4}$$

where $\hat{\mathbf{N}}_B$ is the G-length vector of estimated controls; c_h, in general, is a constant related to the chosen replication variance method ($c_h = \sqrt{n_h/(n_h-1)}$ specifically for the delete-one jackknife); $\delta_{(hr)}$ is a zero/one indicator that identifies the G (out of R) randomly chosen replicates to receive an adjustment ($\hat{\mathbf{z}}_g$) for the estimated controls, $(hr)=1,\ldots,R$; $\delta_{g|(hr)}=1$ if the g^{th} component ($g=1,...,G$) of the benchmark covariance decomposition is randomly chosen for the assignment; and $\hat{\mathbf{z}}_g = \hat{\mathbf{q}}_g\sqrt{\hat{\lambda}_g}$, a function of an eigenvector ($\hat{\mathbf{q}}_g$) and the associated eigenvalue ($\hat{\lambda}_g$) from the decomposition of the estimated covariance matrix, $\hat{\mathbf{V}}_B = \sum_{g=1}^{G}\hat{\mathbf{z}}_g\hat{\mathbf{z}}'_g$. Thus, given that $\delta_{(hr)}=1$ for a particular replicate (hr), one indicator $\delta_{g|(hr)}$ must also equal one; however, if $\delta_{(hr)}=0$, then *all* indicators $\delta_{g|(hr)}$ equal zero.

Fuller (1998) demonstrated that the (delete-one) jackknife sample variance of the replicate controls reproduces the estimated benchmark covariance matrix ($\hat{\mathbf{V}}_B$). He further showed that the design-expectation of the components of the resulting jackknife variance estimator are asymptotically equivalent to (3.1) in addition to a cross-product term that converges to zero with $O\left(n_A^{-1}\right)$.

3.4 Nadimpalli-Judkins-Chu Jackknife Method (ECNJ)

Nadimpalli, Judkins, and Chu (2004) developed a jackknife variance estimator by randomly perturbing all instead of a subsample of the replicate controls in the following way:

$$\hat{\mathbf{N}}_{B(hr)} = \hat{\mathbf{N}}_B + c_h\left[R_h\varepsilon_{(hr)} \times diag\left[se\left(\hat{\mathbf{N}}_B\right)\right]\right] \quad (3.5)$$

where $R_h = 1/\sqrt{Hn_h}$ is related to the estimated number of replicates within stratum h; $diag\left[se\left(\hat{\mathbf{N}}_B\right)\right]$ is the diagonal matrix of dimension G containing the estimated standard errors of the benchmark controls ($\hat{\mathbf{N}}_B$); and $\varepsilon_{(hr)}$ is a G-length vector of values randomly generated from the standard normal distribution for each replicate. The remaining terms are the same as specified for the Fuller jackknife method above (*section 3.3*). Unlike the Fuller methods, the sample variance of the NJC replicate controls does not reproduce the covariance matrix ($\hat{\mathbf{V}}_B$); however, in expectation the variance of the replicate controls equals $\mathbf{V}_B = E\left(\hat{\mathbf{V}}_B\right)$.

Use of the NJC jackknife method would be plausible in two cases – the complete benchmark covariance matrix for the controls is unavailable (e.g., estimates taken from a previous report), or that the covariance terms would reduce the size of the variance estimates, and therefore the resulting values defined by (3.5) would be conservative estimates. The diagonal covariance matrix for $\hat{\mathbf{N}}_B$ in NJC would be correct if the estimated poststratum counts were actually uncorrelated. However this is unlikely because of the multinomial structure of $\hat{\mathbf{N}}_B$. Given the setup for the NJC method, the expectation of the variance estimator will *not* approximate (3.1); the bias term is related to the difference between the design-expectation of $diag\left[se\left(\hat{\mathbf{N}}_B\right)\right]$ and $\mathbf{V}_B$.

4. Simulation Study

4.1 Simulation Parameters

We complement the theoretical evaluation of the six variance estimators presented in the previous section with an empirical evaluation through a simulation study. The simulation population is a random subset of the 2003 National Health Interview Survey (NHIS) public-use file containing records for 21,664 adults. These records were divided into 25 strata, each containing six PSUs. We selected 1,000 samples of size 1,000 to estimate the population totals and associated variances for two variables from the NHIS: (*i*) Health insurance coverage (NOTCOV)—whether a person was without any type of health insurance within the last 12 months; and (*ii*) Medical care delayed (PDMED12M)—whether a person delayed medical care or not because of cost in last 12 months. Samples were selected in two stages – a probability proportional to size (with replacement) sample of two PSUs per stratum and a simple random sample of 20 persons within each sampled PSU. We excluded nonresponse from consideration in our current simulation study to minimize extraneous factors that might affect our comparisons.

Poststratification may reduce variances slightly but in household surveys is mainly used to correct for sampling frame undercoverage, as well as other problems inherent with surveys. The 1,000 simulation samples were selected to mimic a sampling frame that suffers from differential undercoverage, i.e., telephone survey frames. The 16 poststratification cells were defined by an eight-level age variable crossed with gender. The coverage rates for the 16 cells by analysis variable are provided in *Table 1*. A coverage rate equal to 1.0 would indicate full coverage. Before each sample was selected, the frame was designated as a stratified random subsample of the full population of 21,664. For example, 90% of the male population less than five years of age (age <5, male) were randomly selected to be in the frame for NOTCOV. This process of subsetting the population to the frame was independently implemented for each sample.

Table 1. Coverage Rates for 16 Poststratification Cells by Analysis Variable

	Not covered by health insurance (NOTCOV)		Delayed medical care in last 12mo (PDMED12M)	
Age	**Male**	**Female**	**Male**	**Female**
<5	0.9	0.9	0.9	0.9
5–17	0.8	0.8	0.8	0.8
18–24	0.5	0.5	0.6	0.5
25–44	0.5	0.5	0.6	0.5
45–64	0.8	0.8	0.6	0.5
65–69	0.9	0.9	0.9	0.5
70–74	0.9	0.9	0.9	0.7
75 +	0.9	0.9	0.9	0.8

We suspect that the decision for researchers to use either a traditional or an EC calibration variance estimator depends on the precision of the control totals. We calculated the EC covariance matrix ($\hat{\mathbf{N}}_B$ using the notation from section 3) from the complete NHIS public-use data file (N=92,148) and adjusted the values to reflect a sample of size comparable with our simulation population (N=21,664). A few example correlations for the covariance matrix $\hat{\mathbf{N}}_B$ are provided in *Table 2*; the off-diagonal values range from -0.05 to 0.75 with a mean value of 0.22. From this revised matrix we calculated four covariance matrices for the simulation by dividing the original matrix by the adjustment factors 1.0, 3.6, 18, and 72. The adjustments reflect approximate effective sample sizes of 92,000, 25,500 (=92,000/3.6), 5,100, and 1,280, respectively.

Table 2. Estimated Control Total Correlations for Males in Age Groups Ranging from 18 to 69

	18-24	**25-44**	**45-64**	**65-69**
18–24	1.00	0.37	0.29	0.01
25–44	0.37	1.00	0.31	0.10
45-64	0.29	0.31	1.00	0.19
65-69	0.01	0.10	0.19	1.00

The simulation was conducted in **R** (Lumley 2005, R Development Core Team 2005) because of its extensive capabilities for analyzing survey data and efficiency with simulated analyses. Code was developed to calculate the linearization and replicate variance estimates for the EC poststratified estimator discussed above because such relevant code did not exist.

4.2 Evaluation Criteria

The empirical results for the six variance estimators were compared using three measures across the 1,000 simulation samples: the estimated percent relative bias of the variance estimator $\left(\left[\frac{1}{1000}\sum_s var\left(\hat{\theta}\right)\Big/\left(rmse\right)^2\right]-1\right)$; the 95% confidence interval coverage rate $\left(P\left(|\,\hat{Z}\,|\leq z_{1-\alpha/2}\right)\right.$ where $\left.\hat{Z}=\left[\hat{\theta}-\theta\right]\Big/\sqrt{var(\hat{\theta})}\right)$; and the standard deviation of the estimated standard errors $\left(\frac{1}{(1000-1)}\sum_s\left[\sqrt{var\left(\hat{\theta}_s\right)}-\sum_s\sqrt{var\left(\hat{\theta}_s\right)}\Big/1000\right]^2\right)$.

The relative bias and the root mean square error (*rmse*) of our point estimators are calculated as $(1000)^{-1}\sum_s\left(\hat{\theta}_s-\theta\right)\Big/\theta$ and $\sqrt{\sum_s\left(\hat{\theta}_s-\theta\right)^2\Big/1000}$, respectively. The term θ refers to total calculated from the full population, not total calculated from the "covered" frame.

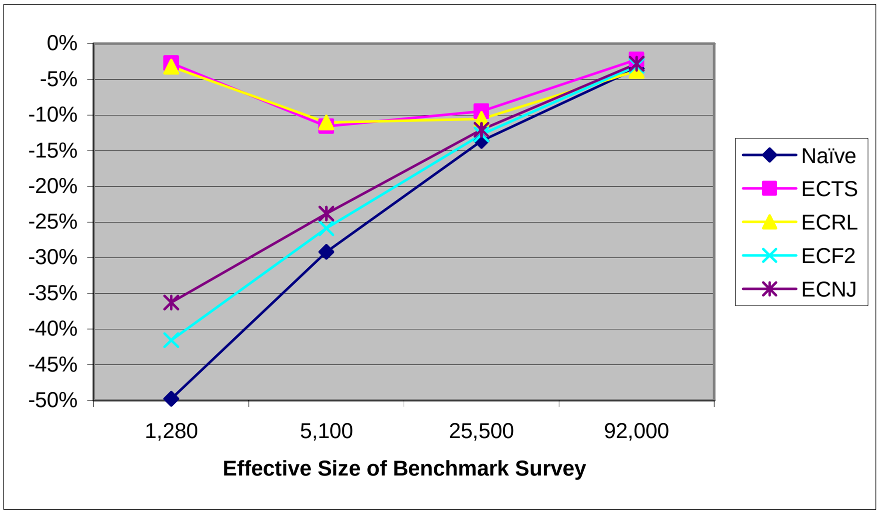
5. Simulation Results

We first examine the results of our point estimators to justify the need for calibration, and move on to a comparison of empirical results for our set of variance estimators.

5.1 Point Estimator

To justify the need for calibration, we initially evaluated the Horvitz-Thompson (HT) estimates ($\sum_s \pi_k^{-1} y_k$) for our analysis variables which are known to be design-unbiased under pristine conditions . The average percent relative bias for the HT estimator was approximately 38 percent for NOTCOV and slightly higher for PDMED12M at 41 percent. These large values also show that some correction is needed to adjust for the non-negligible levels of bias. The percent relative bias for the poststratified estimator was much lower – values range from 1.3 to 2.6 percent for both analysis variables. Minor fluctuations in the levels were detected as the efficiency of the benchmark covariance changed for both point estimators.

Figure 1. Percent Relative Bias of Five Variance Estimators by Efficiency of Benchmark Survey Estimates for Total Number Not Covered by Health Insurance in Last 12 Months.



5.2 Variance Estimators

Adding to the discussion in Section 3, the empirical results for the ideal variance estimator should show a *percent relative bias* either near zero or somewhat positive for a conservative measure. Figure 1 shows the percent relative bias (y axis) for the five of the six variance estimators using the NOTCOV variable and benchmark estimates with increasing levels of efficiency (left to right on the x axis). A similar pattern is shown for the variable PDMED12M. We focus on the general pattern in the data as shown in Figure 1 instead of displaying pages of tabular results.

The traditional poststratified variance estimator (Naïve) assumes that the benchmark estimates are true population values and is represented by a blue diamond in Figure 1. We also included the delete-one jackknife in the simulation, again assuming that $\hat{\mathbf{N}}_B$ was constant. The empirical results for the two naïve variance estimators (linearization and jackknife) were so similar that they are represented by a single line. When the efficiency of the benchmark estimates is low (represented as an effective sample size of 1,280), the benchmark estimates make a substantial contribution to the variance, and the estimator severely underestimates the true variance by as much as 50 percent. The relative bias improves with the increased efficiency of the benchmark survey estimates because they make up a smaller proportion of the variance of $\hat{\theta}$.

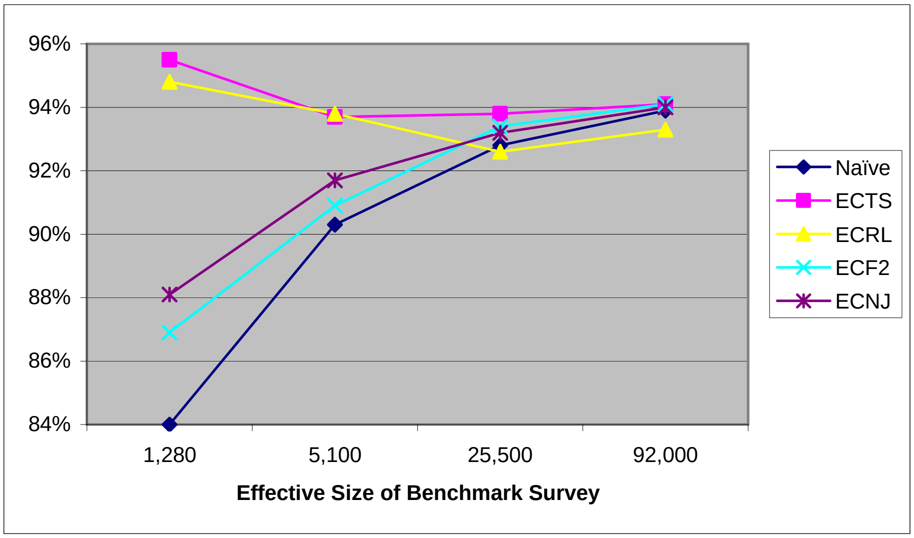
A pattern similar to the Naïve estimators is seen with the Nadimpalli-Judkins-Chu jackknife method (ECNJ). However, the ECNJ is a noticeable improvement over the Naïve estimators by reducing the underestimate in some cases by as much as 25 percent.

The Taylor Series Linearization (ECTS), the Residual Linearization Method (ECRL), and the Fuller Two-Phase Jackknife Method (ECF2) are all similar and produced the lowest levels of relative bias. Any differences resulting from simulation variation were negligible.

When the precision of the control totals is high (represented here as an effective sample size of 92,000), the difference between the methods is negligible. One could argue by a visual examination of Figure 1 that a slight decrease in the precision (effective size of 25,000) is also negligible. This suggests that a relative threshold could be identified to determine when the new EC calibration techniques are required instead of using traditional methods already available.

The patterns exhibited for the percent relative bias are reflected in the *coverage rates for the 95 percent confidence interval* for the estimated totals (Figure 2). The coverage rates for the Naïve and ECNJ estimators are much lower than the desired rate (95 percent), approaching 80 and 86 percent, respectively as the effective size of the benchmark survey decreases. These levels again improve as the efficiency of the benchmark estimates improves.

Figure 2. Coverage Rates for 95 Percent Confidence Intervals of Five Variance Estimators by Efficiency of Benchmark Survey Estimates for Total Number Not Covered by Health Insurance in Last 12 Months.



However, the coverage rates for the remaining variance estimators which account for the complete benchmark covariance matrix are much higher. The coverage rates for the ECTS, ECRL, and ECF2 methods range from levels slightly below 93 to 96 percent with the lower levels occurring when the contribution of the benchmark estimates to the overall variance is much lower than from the analytic survey.

Again, we see that the coverage rates by variance estimator are comparable when the precision of the benchmark controls is relatively high. Differences may be detected if the point estimates exhibited higher levels of bias than shown in our simulation study.

The discussion so far indicates that there is minimal difference between the ECTS, ECRL, and ECF2 methods. We finally look to the *standard deviation of the estimated standard errors* in an attempt to distinguish the estimators. *Table 3* contains the percent relative difference of the standard deviations for the ECRL, ECF2, and ECNJ from the ECTS value for the NOTCOV variable. Again, a similar pattern is reflected for the PDMED12M variable. The variation in the Fuller two-phase (ECF2) estimates is larger than in the estimates from the linearization methods (ECTS and ECRL), and is slightly lower than the variation in the ECNJ estimates. In our simulation study, we made one random adjustment to the benchmark controls per replicate as shown in (3.4). One way to stabilize ECF2 may be to use multiple random assignments for each sample, compute ECF2 from each assignment, and average those values. This proved too computationally demanding for our simulation study but may warrant further examination.

Table 3. Percent Relative Difference of the Standard Deviations for the ECRL, ECF2, and ECNJ Estimates from the ECTS Value for NOTCOV

Variance Estimators	Effective Size of Benchmark Survey			
	1,280	**5,100**	**25,500**	**92,000**
ECRL	-1.1%	0.3%	-1.1%	-3.1%
ECF2	14.3%	7.4%	2.0%	0.5%
ECNJ	15.1%	7.6%	2.3%	0.6%

6. Conclusions and Future Work

The theoretical and analytical work discussed in this paper support the need for a new methodology to address calibration using estimated control totals, i.e., estimated-control (EC) calibration. Traditional variance estimators severely underestimate the population sampling variance resulting in, for example, incorrect decisions for hypothesis tests and sub-optimal sample allocations when the design is optimized in the future.

Currently, the linearization variance estimators developed under our research shows the most promise for EC calibration. However, the applicability of replication methods is cited for public-use files released without sampling design information to further protect data confidentiality and respondent privacy. Hence, the Fuller two-phase method presented here is widely applicable.

Our future research will include an examination of bias in the point estimators, a generalization to linear calibration, and adaptation to other statistics including a ratio-estimated mean. We additionally are investigating whether threshold values are identifiable which determine (1) when there is negligible difference between traditional and estimated-control variance estimation to suggest when the extra effort is not needed, and (2) when the benchmark controls are too imprecise to use in a calibration model. We also plan to investigate the theoretical implications for measurement errors in the analytic as well as the benchmark survey, and methods to improve the benchmark estimates which includes for example, collapsing cells to create an "optimal" set of poststratification cells.

References

Binder, D.A. (1995), "Linearization Methods for Single Phase and Two-Phase Samples: A Cookbook Approach," *Survey Methodology*, **22**(1), pp. 17-22.

Canty, A.J. and Davison, A.C., (1999), "Resampling-based Variance Estimation for Labour Force Surveys," *The Statistician*, **48**, pp. 379-391.

Centers for Disease Control and Prevention (2006), Technical Information and Data for the Behavioral Risk Factor Surveillance System (BRFSS) – BRFSS Weighting Formula, Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, September 11, 2006.

Demnati, A. and Rao, J.N.K. (2004), "Linearization Variance Estimators for Survey Data," *Survey Methodology*, **30**(1), pp. 17-26.

Deville, J-C. and Särndal, C-E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, **87**(418), pp. 376-382.

Estevao, V.M. and Särndal, C-E. (2000), "A Functional Form Approach to Calibration," *Journal of Official Statistics*, **16**(4), pp. 379-399.

Fuller, W.A. (1998), "Replication Variance Estimation for the Two-Phase Samples," *Statistica Sinica*, **8**, pp. 1153-1164.

Hidiroglou, M.A., Fuller, W.A., and Hickman, R.D. (1980), *SUPERCARP*, Department of Statistics, Iowa State University, Ames, Iowa.

Isaki, C.T., Tsay, J.H., and Fuller, W.A. (2004), "Weighting Sample Data Subject to Independent Controls," *Survey Methodology*, **30**(1), pp. 35-44.

Kalton, G. (1979), "Ultimate Cluster Sampling," *Journal of the Royal Statistical Society*, Series A (General), **142**(2), pp. 210-222.

Kott, P.S. (2006), "Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors," *Survey Methodology*, **32**(2), pp. 133-142.

Krotki, K. (2007), "Combining RDD and Web Panel Surveys," *Proceedings of the Survey Research Methods Section*, American Statistical Association (in print).

Lumley, T. (2005). Survey: analysis of complex survey samples. R package version 3.0-1. University of Washington: Seattle.

Nadimpalli, V., Judkins, D., and Chu, A. (2004). "Survey Calibration to CPS Household Statistics." *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 4090-4094.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.R-project.org.

Research Triangle Institute (2004), *SUDAAN Language Manual*, Release 9.0, Research Triangle Park, NC: Research Triangle Institute.

Särndal, C-E., Swensson, B., and Wretman, J. (1989), "The Weighted Residual Technique for Estimating the Variance of the General Regression Estimator of the Finite Population Total," *Biometrika*, **76**(3), pp. 527-537.

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

SAS Institute Inc. (2004), *SAS/STAT(r) 9.1 User's Guide*, Cary, NC: SAS Institute Inc.

StataCorp (2005), *Stata Statistical Software: Release 9*, Survey Data, College Station, TX: StataCorp LP.

Stukel, D.M., Hidiroglou, M.A., and Särndal, C-E. (1996), "Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization," *Survey Methodology*, **22**(2), pp. 117-125.

Terhanian., G., Bremer, J., Smith, R., and Thomas, R. (2000). *Correcting Data from Online Survey for the Effects of Nonrandom Selection and Nonrandom Assignment*. Research Paper: Harris Interactive.

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

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