Estimated-Control Poststratified Variance Estimators for Proportions

Jill A. Dever[1] and Richard Valliant[2]
[1]RTI International, 701 13th Street NW, Suite 750, Washington, DC 20005
[2]Joint Program in Survey Methodology, University of Maryland and Institute for Social Research, University of Michigan, 1218 LeFrak Hall College Park, MD 20742

Abstract
Calibration estimators, such as a poststratified estimate of a population proportion, use auxiliary data to improve the efficiency of survey estimates. Traditionally, the control totals used in the poststratification are assumed to be population values with no sampling variance. Often, however, estimates from other surveys are used because the population controls either do not exist or are not readily accessible. In this situation, many researchers apply traditional variance estimators to cases where the control totals are estimated, thus assuming that any additional sampling variance associated with these controls is negligible. We compare the mean square error for linearization and replication variance estimators of proportions when the uncertainty in the control totals is either addressed or ignored. Illustrations are given of the effects of different levels of variability in the estimated controls on the overall variance estimates. Comparisons are also made to previous work conducted in this area by the authors on estimated population totals.

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

1. Introduction

Poststratified estimators, a specific type of calibration estimator (Deville & Särndal 1992), are used in a variety of surveys as a method to reduce variances or to correct for frame deficiencies. With this method, sampling weights are adjusted (i.e., benchmarked) so that they sum to a set of G ($G \geq 1$) poststratum totals. Gains in efficiency are best when the variables used to define the poststrata are associated with the set of key variables collected in the survey.

Frame deficiencies such as undercoverage are of particular interest to the research presented here. Undercoverage occurs when the sampling frame fails to contain all units for the population under study (e.g., Särndal, Swensson, & Wretman 1992; Kott 2006). For example, data collected from January to December 2007 through the U.S. National Health Interview Survey (NHIS) indicate that coverage rates for a landline telephone survey vary greatly by state, as well as region of the country. Approximately five percent of Vermont households were estimated to have only wireless-phone usage in comparison to 26.2 percent of households in Oklahoma (Blumberg et al. 2009).

A primary assumption with poststratification is that the control totals used in the weight adjustments are either true population values known without error, or are estimated from an independent, highly precise survey that is much larger than the survey requiring poststratification. 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., Terhanian, et al. 2000).

Variance estimators have been developed for poststratification under the population control total assumption. Many researchers apply this type of formula even though the controls are estimated. The tacit assumption is that any additional variance associated with these controls is negligible and can be ignored. Currently, this assumption can not be validated as reasonable. We label the methodology which properly accounts for the estimated controls as estimated-control (EC) poststratification.

The goal of our research is to develop and evaluate estimators for complex sampling designs under EC poststratification. In this paper, we focus specifically on the estimated-control poststratified (ECPS) estimator of a population mean calculated as the ratio of two estimated totals for a prototypical complex survey design. The data for the survey requiring poststratification is assumed to be collected from a two-stage design where m_{Ah} first-stage sampling units are selected with replacement from within H design strata. Only a random sampling design is assumed for the source of the estimated poststratum totals. We begin in *section 2* with a brief summary of the research to date on poststratification and provide an explicit definition of the ECPS mean. 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 control totals. The set-up for the simulation study is provided in *section 4*, followed by a summary of the results (*section 5*). We compare the results presented here with those derived from prior research on ECPS totals. The paper is concluded with a summary of future research in this area.

2. Estimated-Control Poststratified Estimator

The general form of a poststratification estimator can be described as a linear weighting estimator (Estevao & Särndal 2000). An estimated population total of a variable y is $\hat{t}_y = \sum_{k \in s} w_k y_k$, where the poststratified weight (w_k) for the k^{th} unit in the random sample s is a function of the design weight π_k^{-1} and a poststratification-adjustment factor a_k, also known as a g-weight (Särndal, Swensson, & Wretman 1992). Suppose that the finite population U from the sample s is randomly drawn can be divided into g =1, ..., G mutually exclusive and exhaustive poststrata. Then, the a_k's are calculated to satisfy the following set of constraints:

$$\mathbf{N} = \hat{\mathbf{N}} \tag{1}$$

where $\hat{\mathbf{N}} = \left(\hat{N}_1, \ldots, \hat{N}_G\right)'$, a vector of survey estimated poststratum counts; $\hat{N}_g = \sum_{k \in s_g} a_k \, \pi_k^{-1}$, the estimated count for poststratum g defined as the sum of the poststratified weights for members of that poststratum, i.e., $k \in s_g$; and $\mathbf{N} = \left(N_1, \ldots, N_G\right)'$, a vector of population poststratum counts.

The formula for a poststratified estimator of a population total of a variable *y*, under the potentially naïve population control total assumption, is defined as

$$\hat{t}_{yNP} = \mathbf{N}'\hat{\mathbf{N}}_A^{-1}\hat{\mathbf{t}}_{Ay} \tag{2}$$

where $\hat{\mathbf{t}}_{Ay} = \left(\hat{t}_{Ay1}, \ldots, \hat{t}_{AyG}\right)'$, a *G*-length column vector of estimated totals for *y* by poststratum such that $\hat{t}_{Ayg} = \sum_{k \in s_{Ag}} \pi_k^{-1} y_k$, and $\hat{\mathbf{N}}_A = diag\left(\hat{N}_{A1}, \ldots, \hat{N}_{AG}\right)$, a G-dimension diagonal matrix with $\hat{N}_{Ag} = \sum_{k \in s_{Ag}} \pi_k^{-1}$. The subscript *A* identifies estimates calculated from the *analytic survey*, a term we reserve for the survey requiring poststratification. The estimator of the population mean is thus defined as

$$\hat{\bar{y}}_{NP} = \frac{\hat{t}_{yNP}}{N} \tag{3}$$

where $N = \sum_g N_g$, the size of the finite population *U*.

Sometimes, however, the population counts are unknown and must be estimated from a *benchmark survey*. For these situations, we define the estimated-control poststratified (ECPS) estimator of a population total as

$$\hat{t}_{yP} = \hat{\mathbf{N}}_B'\hat{\mathbf{N}}_A^{-1}\hat{\mathbf{t}}_{Ay}, \tag{4}$$

by substituting $\mathbf{N}$ in expression (2) with $\hat{\mathbf{N}}_B = \left(\hat{N}_{B1}, \ldots, \hat{N}_{BG}\right)'$, a vector of poststratum totals estimated from the benchmark survey, i.e., $\hat{N}_{Bg} = \sum_{l \in s_{Bg}} w_l^{-1}$. Note that we may also express $\hat{\mathbf{N}}_A^{-1}\hat{\mathbf{t}}_{Ay}$ as $\hat{\bar{\mathbf{Y}}}_A = \left(\hat{t}_{Ay1} \big/ \hat{N}_{A1}, \ldots, \hat{t}_{AyG} \big/ \hat{N}_{AG}\right)'$ so that $\hat{t}_{yP} = \hat{\mathbf{N}}_B'\hat{\bar{\mathbf{Y}}}_A$. The subscript *B* in expression (4) distinguishes the benchmark survey estimates from the analytic survey estimates (e.g., $\hat{\mathbf{t}}_{Ay}$). The corresponding ECPS estimator of a population mean, the interest of our research presented here, is defined as

$$\hat{\bar{y}}_P = \frac{\hat{t}_{yP}}{\hat{N}_B}. \tag{5}$$

Note that the denominator in (5) is calculated by using expression (4) and y_k=1 and reduces to $\hat{N}_B = \hat{\mathbf{N}}_B'\mathbf{1}_G$, a function of the estimated benchmark poststratum counts and a *G*-length vector of ones ($\mathbf{1}_G$).

3. Variance Estimators for the ECPS

Variance estimators have been developed for poststratification estimators that use population control totals. The estimated standard errors are calculated using survey analysis packages including R® (R Development Core Team 2005); SAS® (2004), Stata® (2004), and SUDAAN® (2008). 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 (2007) and Binder (1995).

A poststratified estimator, a type of generalized regression estimator (GREG), can be expressed in terms of a group-mean regression model with model expectation equal to $E_M\left(y_k\right)=\bar{y}_g$, the mean for population units in poststratum g, and a poststratum-specific variance $Var_M\left(y_k\right)=\sigma_g^2$ (see., e.g., Särndal, Swensson, & Wretman 1992). Särndal, Swensson, and Wretman (1989, 1992) developed an approximate linearization population variance for a population total as a function of the model residuals and the g-weights. Stukel, Hidiroglou, and Särndal (1996) discuss a g-weighted 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 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 poststratification.

Four ECPS variance estimators that incorporate the variability in the estimated control totals were compared for this study. They include one linearization estimator and three 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 $m_A=\sum_{h=1}^{H} m_{Ah}$ replicates calculated by summing the number of PSUs per stratum (m_h) across the H design strata. We additionally compare the properties of these estimators with the "usual" linearization variance estimator that does not account for the estimated control totals (see, e.g., section 6.6 in Särndal, Swensson, & Wretman 1992).

An effective variance estimator will reproduce the corresponding population sampling variance in expectation. To derive the asymptotic population sampling variance for the estimated mean $\hat{\bar{y}}_P$ defined in (5), we first derive the variance for $\hat{t}_{yP}=\hat{\mathbf{N}}_B'\hat{\bar{\mathbf{Y}}}_A$ by applying the unconditional variance formula given in, e.g., Casella & Berger (2002, Theorem 4.4.7):

$$AV\left(\hat{t}_{yP}\right)=\mathbf{N}_B'\mathbf{V}_A\mathbf{N}_B+\bar{\mathbf{Y}}_A'\mathbf{V}_B\bar{\mathbf{Y}}_A \tag{6}$$

where $\mathbf{N}_B$ is a vector of population counts within the G poststrata defined by the benchmark survey target population such that $E_B\left(\hat{\mathbf{N}}_B\right)=\mathbf{N}_B$, the expectation with respect to the benchmark survey design; $\bar{\mathbf{Y}}_A=\left(t_{Ay1}/N_{A1},\ldots,t_{AyG}/N_{AG}\right)'$ with $E_A\left(\hat{t}_{Ayg}\right)=t_{Ayg}$ and $E_A\left(\hat{N}_{Ag}\right)=N_{Ag}$, the expectation with respect to the analytic survey design; $\mathbf{V}_A$ is the covariance matrix of the estimated components of the vector $\bar{\mathbf{Y}}_A$; and $\mathbf{V}_B$ is the covariance matrix of the G poststratum benchmark estimates given in $\hat{\mathbf{N}}_B$. The first term in expression (6) is the approximate variance for the standard poststratified estimator where the benchmark estimates are treated as fixed, i.e., $AV\left(\hat{t}_{yNP}\right)=\mathbf{N}_B'\mathbf{V}_A\mathbf{N}_B$. The second term in expression (6), $\bar{\mathbf{Y}}_A'\mathbf{V}_B\bar{\mathbf{Y}}_A$, is variance component associated only with the variation in the benchmark poststratum estimates, i.e., the analytic estimates are treated as fixed.

Using equation (6) and $\hat{N}_B = \hat{\mathbf{N}}'_B \mathbf{1}_G$, the asymptotic variance of $\hat{\bar{y}}_P$ (5) is defined as:

$$AV\left(\hat{\bar{y}}_P\right) = \left(\frac{1}{\hat{N}_B}\right)^2 \left[AV\left(\hat{t}_{yP}\right) + \hat{\bar{y}}_P^2 \; AV\left(\hat{N}_B\right) - 2\hat{\bar{y}}_P \; ACov\left(\hat{t}_{yP}, \hat{N}_B\right)\right] \quad (7)$$

where $AV\left(\hat{N}_B\right) = \mathbf{1}'_\mathbf{G} \mathbf{V}_B \mathbf{1}_\mathbf{G}$, the asymptotic variance of the benchmark estimated total, and $ACov\left(\hat{t}_{yP}, \hat{N}_B\right) = \left(\bar{\mathbf{Y}}_A - \bar{y}_P \mathbf{1}_G\right)' \mathbf{V}_B \left(\bar{\mathbf{Y}}_A - \bar{y}_P \mathbf{1}_G\right)$, the asymptotic covariance of two estimators used to define $\hat{\bar{y}}_P$.

Having defined the population sampling variance for the estimated mean, we next define the set sample variance estimators included in our research. We begin with two first-order Taylor linearization variance estimators followed by three formulae for a jackknife variance estimator.

3.1 Linearization Variance Estimation for Standard Poststratification Variance Estimator (Naïve)

The sample variance estimator defined for a standard poststratification adjustment is found is many sampling textbooks and is expressed in matrix form as

$$var_{Naive}\left(\hat{\bar{y}}_P\right) = \left(\frac{1}{\hat{N}_B}\right)^2 var_{Naive}\left(\hat{t}_{yP}\right) \quad (8)$$

with $var_{Naive}\left(\hat{t}_{yP}\right) = \hat{\mathbf{N}}'_B \hat{\mathbf{V}}_A \hat{\mathbf{N}}_B$. Note that the second and third components of expression (7) are not estimated. If (*i*) the benchmark estimates are precise so that the estimated covariance matrix $\hat{\mathbf{V}}_B = var\left(\hat{\mathbf{N}}_B\right)$ is negligible and (*ii*) the estimates in $\hat{\bar{y}}_P$ are positively correlated, then expression (8) could overestimate the population sampling variance. Other conditions may suggest that estimates obtained with this formula will not be noticeably different from those generated with the ECPS variance estimators described next. We make this comparison with our simulation study described in Section 4.

3.2 Taylor Series Linearization (ECTS)

The form of the first-order Taylor linearization sample variance that accounts for the variation in the benchmark poststratification totals can be described as a method-of-moments estimator:

$$var_{ECTS}\left(\hat{\bar{y}}_P\right) = \left(\frac{1}{\hat{N}_B}\right)^2 \left[var_{EC}\left(\hat{t}_{yP}\right) + \hat{\bar{y}}_P^2 \; var_{EC}\left(\hat{N}_B\right) - 2\hat{\bar{y}}_P \; cov_{EC}\left(\hat{t}_{yP}, \hat{N}_B\right)\right] \quad (9)$$

where $var_{EC}\left(\hat{t}_{yP}\right) = \hat{\mathbf{N}}'_B \hat{\mathbf{V}}_A \hat{\mathbf{N}}_B + \hat{\bar{\mathbf{Y}}}'_A \hat{\mathbf{V}}_B \hat{\bar{\mathbf{Y}}}_A$ with $\hat{\mathbf{V}}_A = var\left(\hat{\mathbf{N}}_A\right)$, the estimated covariance of the poststratum totals estimated from the analytic survey; $var_{EC}\left(\hat{N}_B\right) = \mathbf{1}'_G \hat{\mathbf{V}}_B \mathbf{1}_G$ for $\mathbf{1}_G$ defined as a *G*-length vector of ones; and $cov_{EC}\left(\hat{t}_{yP}, \hat{N}_B\right) = \left(\hat{\bar{\mathbf{Y}}}_A - \hat{\bar{y}}_P \mathbf{1}_G\right)' \hat{\mathbf{V}}_B \left(\hat{\bar{\mathbf{Y}}}_A - \hat{\bar{y}}_P \mathbf{1}_G\right)$. Expression (9) can be recast as a standard poststratification sample variance estimator plus terms that account for the variation in the benchmark poststratum totals.

3.3 Fuller Two-Phase Jackknife Method (ECF2m)

Isaki, Tsay, and Fuller (2004) applied a two-phase delete-one jackknife variance estimator developed by Fuller (1998) to a survey with estimated control totals. The jackknife sample variance estimator takes the form

$$var_{ECF2}\left(\hat{\bar{y}}_P\right) = \sum_{h=1}^{H}\sum_{r=1}^{m_{Ah}} c_h^{-2}\left(\ddot{\bar{y}}_{P(r)} - \hat{\bar{y}}_P\right)^2 \tag{10}$$

The r^{th} replicate estimator for the population mean is defined as:

$$\ddot{\bar{y}}_{P(r)} = \frac{\ddot{t}_{yP(r)}}{\ddot{N}_{B(r)}} = \frac{\ddot{\mathbf{N}}'_{B(r)}\hat{\mathbf{N}}_{A(r)}^{-1}\hat{\mathbf{t}}_{Ay(r)}}{\ddot{\mathbf{N}}'_{B(r)}\mathbf{1}_G} \tag{11}$$

where $\hat{\mathbf{N}}_{A(r)}^{-1}$ and $\hat{\mathbf{t}}_{Ay(r)}$ are calculated with the formula given for expression (2) after the r^{th} PSU has been removed from the sample. The Fuller's approach relies on a spectral (eigenvalue) decomposition of the benchmark covariance matrix ($\mathbf{V}_B$) to provide the vector of estimators $\ddot{\mathbf{N}}_{B(r)}$. Adjustments, defined as functions of the resulting eigenvalues and eigenvectors, are added to the estimated controls $\hat{\mathbf{N}}_B$ to create a set of replicate controls. Namely,

$$\ddot{\mathbf{N}}_{B(r)} = \hat{\mathbf{N}}_B + c_h\hat{\mathbf{z}}_{B(r)} \tag{12}$$

where $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$, a constant associated with the delete-one jackknife; $\hat{\mathbf{z}}_{(r)} = \delta_{(r)}\sum_{g=1}^{G}\delta_{g|(r)}\hat{\mathbf{z}}_g$; $\delta_{(r)}$ is a zero/one indicator that identifies the G out of m_A randomly chosen replicates to receive an adjustment; $\delta_{g|(r)} = 1$ if the g^{th} component of the benchmark covariance decomposition is randomly chosen for the assignment given that replicate r is selected for adjustment; 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 $\hat{\lambda}_g$ such that $\hat{\mathbf{V}}_B = \sum_{g=1}^{G}\hat{\mathbf{z}}_g\hat{\mathbf{z}}'_g$. Thus, given $\delta_{(r)} = 1$ for a particular replicate, a single indicator $\delta_{g|(r)}$ must also equal one; however, if $\delta_{(r)} = 0$, then *all* indicators $\delta_{g|(r)}$ equal zero. Details to construct the computer code for $\ddot{\mathbf{N}}_{B(r)}$ are provided in Dever and Valliant (2007).

Fuller (1998) demonstrated that the delete-one jackknife sample variance of the replicate controls reproduces the estimated benchmark covariance matrix, i.e., $var_{ECF2}\left(\hat{\mathbf{N}}_B\right) = \hat{\mathbf{V}}_B$ with $\hat{\mathbf{N}}_B$ defined by solving equation (12), for every sample. Using this and a geometric approximation, it can be shown that the design expectation of the resulting jackknife variance estimator is asymptotically equivalent to $AV\left(\hat{\bar{y}}_P\right)$ in (7) only if the respective components are calculated with values from design-consistent estimators.

The two additional jackknife variance estimators included in our research follow the same pattern as the ECF2m where the vector of benchmark poststratification totals $\hat{\mathbf{N}}_B$ are adjusted for the replicate estimates. We continue with another variance estimator also presented in 2004.

3.4 Nadimpalli-Judkins-Chu Jackknife Method (ECNJCm)

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:

$$\ddot{\mathbf{N}}_{B(r)} = \hat{\mathbf{N}}_B + c_h\, R_h\, \hat{\mathbf{S}}_B\, \boldsymbol{\eta}_{(r)} \tag{13}$$

where $R_h = 1\big/\sqrt{Hm_{Ah}}$ is a function of the number of population PSUs estimated from stratum h; $\hat{\mathbf{S}}_B = \sqrt{diag\left(\hat{\mathbf{V}}_B\right)}$, the diagonal matrix of dimension G containing the estimated standard errors from the benchmark covariance matrix; and $\eta_{(r)}$ is a G-length vector of standard normal values independently generated for each replicate. The remaining terms are specified for the ECF2m following expression (12).

Unlike the ECF2m, the sample variance of the ECNJCm replicate controls given in (13) reproduces the benchmark covariance matrix $\mathbf{V}_B$ *in expectation* only if the off-diagonal terms are truly zero. In most cases, $\mathbf{V}_B$ will not be diagonal which suggests that var_{ECNJC} may not be an effective estimator for the population sampling variance. The magnitude of the over- or under-estimation is related to the sign of the missing off-diagonal terms in $\mathbf{V}_B$, as well as the association between the *y* variable and the variables used to define the poststrata.

3.5 Multivariate Normal Jackknife Method (ECMV)

The multivariate normal method (ECMV) is a generalization of the ECNJCm and to our knowledge is first developed for our research. The ECMV incorporates the complete covariance matrix $\hat{\mathbf{V}}_B$ and relies on large-sample theory so that the poststratum adjustments may be modeled as coming from a G-dimensional multivariate normal distribution with a mean vector of zeros and estimated covariance matrix $\hat{\mathbf{V}}_B$. The replicate controls for the ECMV take the form

$$\ddot{\mathbf{N}}_{B(r)} = \hat{\mathbf{N}}_B + c_h\, R_h\, \hat{\boldsymbol{\varepsilon}}_{(r)} \tag{14}$$

where $\hat{\boldsymbol{\varepsilon}}_{(r)}$ is a G-length vector of values such that $\hat{\boldsymbol{\varepsilon}}_{(r)} \sim MVN_G\left(\mathbf{0}, \hat{\mathbf{V}}_B\right)$. The remaining terms were defined previously for the ECF2m with expression (12) and for the ECNJCm with expression (13).

Unlike the Fuller method, $var_{ECMV}\left(\hat{\mathbf{N}}_B\right) \neq \hat{\mathbf{V}}_B$. The ECMV methodology relies on the design-and model-based properties of the estimator to show that $E\left[var_{ECMV}\left(\hat{\mathbf{N}}_B\right)\right] = E\left(\hat{\mathbf{V}}_B\right)$. Thus, in expectation, the ECF2m and ECMV variance estimators are asymptotically equivalent and both should perform reasonably in the empirical study detailed in the next section. However, because $var_{ECMV}\left(\hat{\mathbf{N}}_B\right) \neq \hat{\mathbf{V}}_B$, we hypothesized that ECMV variance estimates will less stable than the ECF2m as examined through the variability of the estimates across the simulation samples.

4. Simulation Study

4.1 Simulation Parameters

We complement the theoretical evaluation of the five 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 "population" records were divided into H=25 strata, each containing six PSUs. We selected 4,000 samples of size 2,000 to estimate the population totals and associated variances for two NHIS analysis variables: NOTCOV=1 indicates that an adult *did not* have health insurance coverage in the 12 months prior to the NHIS interview (approximately 17 percent of the population); and PDMED12M=1 indicates that an adult *delayed* medical care because of cost in the 12 months prior to the interview (approximately 7 percent of the population). For brevity and because the results are similar, we only present the results for NOTCOV in the subsequent discussion. Samples were selected in two stages – a *with-replacement* probability proportional to size sample of two PSUs per stratum and a simple random sample of 40 persons within each sampled PSU. Nonresponse is not included in the simulation study presented here but will be addressed in future research.

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 4,000 simulation samples were selected to mimic a sampling frame that suffers from differential undercoverage, i.e., telephone survey frames. Sixteen (G=16) poststratification cells were defined by an eight-level age variable crossed with gender. The coverage rates for the 16 cells ranged, for example, from 50 percent for females 18-24 years of age to 90 percent for males ages 65-69. No cells were defined as having complete coverage, i.e., 100 percent. Before each sample was selected, the frame was randomly generated by selected a stratified random subsample from the full population of 21,664 using the associated coverage rates. For example, 90 percent of the male population 65-69 years of age was randomly sampled to be in the sampling frame for NOTCOV. The sampling frames were randomly generated prior to selecting the 2,000 analytic survey sample units.

The decision for researchers to use either the standard or the estimated-control (EC) poststratification variance estimator should depend on the precision of the poststratum control totals. The vector of benchmark totals $\hat{\mathbf{N}}_B$ and the associated covariance matrix $\hat{\mathbf{V}}_B$ were estimated from the complete NHIS public-use data file (92,148 records) and ratio adjusted to reflect a sample of size comparable with our simulation population (N=21,664). To address varying levels of precision in the benchmark estimates, we calculated four covariance matrices. The first corresponds to an effective survey size of 21,644. The remaining three were generated by dividing this first matrix by 3.6, 18, and 72 to reflect approximate effective sample sizes of 6,000 (≈ 21,700/3.6), 1,200, and less than 500, respectively.

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 to date the relevant code did not exist.

4.2 Evaluation Criteria

The empirical results for the five variance estimators were compared using three measures across the 4,000 simulation samples:

(*i*) the estimated percent relative bias of the variance estimator,

$$\left(\tfrac{1}{4000}\sum_j var\left(\hat{\bar{y}}_{P(j)}\right) - mse\right) \Big/ mse \qquad (15)$$

where $var\left(\hat{\bar{y}}_{P(j)}\right)$ is one of the five variance estimates evaluated for sample *j* and *mse* is the mean square error of $\hat{\bar{y}}_{P(j)}$ defined below;

(*ii*) the 95% confidence interval coverage rate,

$$\tfrac{1}{4000}\sum_j I\left(|\hat{z}_j| \le z_{1-\alpha/2}\right) \qquad (16)$$

where $\hat{z}_j = \left(\hat{\bar{y}}_{P(j)} - \bar{y}\right) \Big/ \sqrt{var\left(\hat{\bar{y}}_{P(j)}\right)}$; and,

(*iii*) the standard deviation of the estimated standard errors, calculated as the square root of

$$\tfrac{1}{3999}\sum_j \left(\sqrt{var\left(\hat{\bar{y}}_{P(j)}\right)} - \tfrac{1}{4000}\sum_j \sqrt{var\left(\hat{\bar{y}}_{P(j)}\right)}\right)^2 . \qquad (17)$$

The relative bias and empirical mean square error of our point estimators are calculated as $\tfrac{1}{4000}\sum_j \left(\hat{\bar{y}}_{P(j)} - \hat{\bar{y}}\right) \Big/ \hat{\bar{y}}$ and $\tfrac{1}{4000}\sum_s \left(\hat{\bar{y}}_{P(j)} - \bar{y}\right)^2$, respectively.

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 poststratification, we initially compared the percent relative bias for the two estimated means as a function of only the design weights, i.e., $\hat{\bar{y}}_{HT} = \hat{\mathbf{N}}_A^{-1}\hat{\mathbf{t}}_{Ay}$, against those that included the poststratification adjustment, i.e., $\hat{\bar{y}}_P$ given in expression (5). Relative biases of zero are ideal; however, values near zero are more reasonable with simulation studies. The unadjusted weights resulted in point estimates that underestimate the population value by as much as 9 percent. The ECPS adjustment corrects for undercoverage resulting in a slight overestimate of the population means by approximately one percent. Therefore, poststratification with estimated control totals is justified over the use of only the design weights. This same conclusion was drawn from a similar analysis using estimated totals in our previous research.

5.2 Relative Bias in Variance Estimators

Adding to the brief theoretical discussion in Section 3, the empirical results for a desirable variance estimator should show a *percent relative bias* (15) 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). The horizontal line represents zero bias. The vertical line represents studies for which the

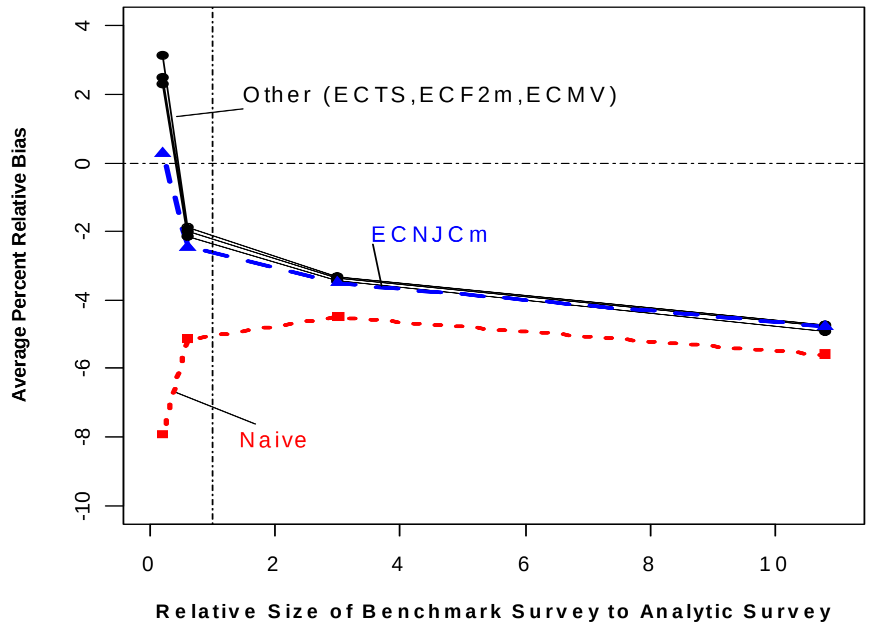

Figure 1: Percent Bias Relative to Empirical MSE for Five Variance Estimators by the Relative Size of the Benchmark Survey to the Analytic Survey

analytic and benchmark surveys are equal in effective size. Again, a similar pattern was displayed for the variable PDMED12M.

The standard poststratified variance estimator (naïve), shown in red in Figure 1, assumes that the benchmark estimates are true population values. The average bias indicates that this estimator is most negatively biased among the variance estimators examined in our research. This finding is consistent with the theoretical evaluation briefly discussed in Section 3. The relative bias is smallest when the benchmark survey is approximately three times larger than the analytic survey. The bias increases dramatically when the benchmark controls are estimated from a much smaller survey. This general pattern was also displayed for estimated totals in the authors previous research (Dever and Valliant 2007), though the levels of bias were much more pronounced for the previous point estimator.

The levels of bias for the EC variance estimators are similar and all show an improvement over the naïve estimator. The relative contribution of the benchmark estimates to the estimated variance is negatively related to the relative size of the benchmark survey to the analytic survey. As the benchmark contribution increases, the bias in the EC variance estimators is reduced to levels that are somewhat conservative (3 percent overestimate). The ECNJCm variance estimator has slightly higher levels of bias than the other EC variance estimators under our simulation study. This is a marked improvement over the comparative levels of bias than was shown for totals (Dever and Valliant 2007).

Table 1: Percent Bias Relative to Empirical MSE for Five Variance Estimators by the Relative Size of the Benchmark Survey and Type of Point Estimator

Point Estimator	Variance Estimator	Relative Size of Benchmark to Analytic Survey			
		0.2	0.6	3.0	10.8
Mean	Naïve	-7.9	-5.1	-4.5	-5.6
	ECTS	2.3	-2.1	-3.4	-4.9
	ECF2m	3.1	-1.9	-3.3	-4.7
	ECNJCm	0.3	-2.4	-3.5	-4.8
	ECMV	2.5	-2.0	-3.4	-4.7
Total	Naïve	-56.0	-31.0	-14.2	-12.2
	ECTS	-0.2	-8.4	-8.2	-10.1
	ECF2m	0.1	-8.2	-8.3	-10.1
	ECNJCm	-40.0	-24.2	-11.9	-11.1
	ECMV	-0.2	-8.1	-8.1	-10.0

Table 1 provides a comparison of the percent relative biases for the estimated mean of NOTCOV displayed in Figure 1 against the levels for the estimated total used in the numerator of the mean. The levels of bias are noticeably larger for totals than means with the naïve and ECTS variance estimators. The difference is similar though less stark with the remaining EC variance estimators.

5.3 Confidence Interval Coverage

The next criterion used to compare the variance estimators was the *empirical coverage rates* (16) for the 95 percent confidence intervals associated with the two outcome variables. Coverage rates for the estimated means under all simulation conditions were fairly stable and had at least a 94 percent coverage rate. We additionally did not detect a linear trend with the increasing size of the benchmark survey. The confidence intervals for the naïve variance estimator were almost twice as wide (average width was approximately 1.2) as the EC confidence intervals and showed a negative shift reflecting the underestimation displayed in Figure 1.

The confidence interval coverage rates for estimated totals are similar for the ECTS, ECF2m, and ECMV variance estimators. However, the naïve and ECNJCm coverage rates fall below the 90 percent when the benchmark survey is small relative to the analytic survey.

5.4 Stability of the Variance Estimators

The results presented above suggest that there are minimal theoretical, as well as empirical, differences between the ECTS, ECF2m, and ECMV methods. A comparison of the variation in the variance estimates, calculated with expression (17), suggests that the ECTS variance estimator is most stable among those examined though the relative increase for the other estimators was less than five percent. This corresponds with the discussion given in Krewski & Rao (1981). The difference in the stability of the ECF2 and ECMV methods is less noticeable with estimated means than with estimated totals.

6. Conclusions and Future Work

The theoretical and analytical work discussed in this paper addresses poststratification using estimated control totals, i.e., estimated-control (EC) poststratification. Traditional variance estimators can 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. The level of underestimation is related to the precision of the benchmark control totals. This underestimation, however, is less severe with means, a function of two estimated totals, in comparison with estimated totals.

The ECNJCm method (Nadimpalli, Judkins, & Chu 2004) can also produce variance estimates that are too small. Our simulation studies suggest that the bias in the ECNJCm variance estimates is much less with the means than with totals such that the differences between this and the other EC estimators examined is not acute. However, additional theory is needed to support this claim under general conditions.

Our recommendation for a variance estimator for an ECPS estimator of a mean falls to one of the three remaining EC formula. Theoretically, the linearization variance estimator (ECTS), the Fuller two-phase jackknife estimator (ECF2m), and the multivariate normal jackknife estimator (ECMV) are asymptotically equivalent. The empirical results suggest that the differences among the three methods in practice are negligible. Choosing between the ECTS and one of the jackknife methods must be based on the type of analysis or public-use file desired for the study.

More work is needed to reduce the negative relative biases seen in the empirical study (Table 1) because slightly conservative variance estimates are generally more desirable. Our future research will also include a generalization to linear calibration (e.g., GREG) and to other statistics. We additionally plan to investigate whether threshold values can be identified that determine (*i*) when there will be a negligible difference between variance estimates using the standard and ECPS formulae, and (2) when the benchmark controls are too imprecise to use in poststratification. We also plan to investigate the theoretical implications for measurement errors in the analytic and benchmark surveys, and methods to improve the benchmark estimates which includes for example, collapsing cells to create an "optimal" set of poststratification cells.

Acknowledgements

This work was competed as part of the first author's doctoral dissertation at the Joint Program in Survey Methodology, University of Maryland. She would like to thank the members of her committee, Richard Valliant, Phillip Kott, Frauke Kreuter, Stephen Miller, and Paul Smith for their guidance.

References

Binder, D.A. (1995). Linearization methods for single phase and two-phase samples: a cookbook approach. *Survey Methodology*, **22**(1), 17-22.

Blumberg, S.J., Luke, J.V., Davidson, G., Davern, M.E.,Yu, T-C., & Soderberg, K. 2009. Wireless substitution: state-level estimates from the National Health Interview Survey, January–December 2007. National Health Statistics Reports, 14, March 11, 2009.

Canty, A.J. & Davison, A.C. (1999), Resampling-based variance estimation for labour force surveys, *The Statistician*, **48**, 379-391.

Casella, G. & Berger, R. L. (2002). *Statistical Inference*. California: Duxbury Press, 2nd ed.

Demnati, A. & Rao, J.N.K. (2004). Linearization variance estimators for survey data. *Survey Methodology*, **30**(1), 17-26.

Dever, J.A. & Valliant, R. (2007). A comparison of variances for poststratification to estimated control totals. In *JSM Proceedings*, Section on Survey Research Methods. Alexandria, VA: American Statistical Association. 3465–3472.

Deville, J-C. & Särndal, C-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**(418), 376-382.

Estevao, V.M. & Särndal, C-E. (2000). A functional form approach to calibration. *Journal of Official Statistics*, **16**(4), 379-399.

Fuller, W.A. (1998). Replication variance estimation for the two-phase samples. *Statistica Sinica*, **8**, 1153-1164.

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

Isaki, C.T., Tsay, J.H., & Fuller, W.A. (2004). Weighting sample data subject to independent controls. *Survey Methodology*, **30**(1), 35-44.

Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society*, Series A (General), **142**(2), 210-222.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, **32**(2), 133-142.

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

Nadimpalli, V., Judkins, D., & Chu, A. (2004). Survey calibration to CPS household statistics. In *JSM Proceedings*, Section on Survey Research Methods. Alexandria, VA: American Statistical Association. 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 (2008), *SUDAAN Language Manual*, Release 10.0, Research Triangle Park, NC: Research Triangle Institute.

Särndal, C-E., Swensson, B., & Wretman, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, **76**(3), 527-537.

Särndal, C-E., Swensson, B., & 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., & Särndal, C-E. (1996). Variance estimation for calibration estimators: a comparison of jackknifing versus Taylor linearization. *Survey Methodology*, **22**(2), 117-125.

Terhanian., G., Bremer, J., Smith, R., & 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**, 89-96.

Wolter, K.M. (2007), *Introduction to Variance Estimation*, New York: Springer-Verlag Inc.