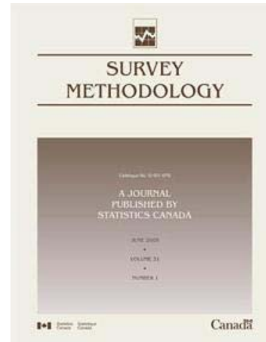
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

A comparison of variance estimators for poststratification to estimated control totals



by Jill A. Dever and Richard Valliant

June 2010

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2010
Vol. 36, No. 1, pp. 45-56
Statistics Canada, Catalogue No. 12-001-X

A comparison of variance estimators for poststratification to estimated control totals

Jill A. Dever and Richard Valliant [1]

Abstract

Calibration techniques, such as poststratification, use auxiliary information to improve the efficiency of survey estimates. The control totals, to which sample weights are poststratified (or calibrated), are assumed to be population values. Often, however, the controls are estimated from other surveys. Many researchers apply traditional poststratification variance estimators to situations where the control totals are estimated, thus assuming that any additional sampling variance associated with these controls is negligible. The goal of the research presented here is to evaluate variance estimators for stratified, multi-stage designs under estimated-control (EC) poststratification using design-unbiased controls. We compare the theoretical and empirical properties of linearization and jackknife variance estimators for a poststratified estimator of a population total. Illustrations are given of the effects on variances from different levels of precision in the estimated controls. Our research suggests (i) traditional variance estimators can seriously underestimate the theoretical variance, and (ii) two EC poststratification variance estimators can mitigate the negative bias.

Key Words: Estimated-control poststratification; Sampling frame coverage bias; Survey-estimated control totals.

1. Introduction

Poststratified estimators, and other calibration estimators, are used in many types of surveys to reduce variances or to correct for frame deficiencies. Specific examples include large U.S. government surveys, such as the Consumer Expenditure Survey (see, *e.g.*, Jayasuriya and Valliant 1996); surveys of specialized populations, such as the U.S. Department of Defense Survey of Health Related Behaviors among Military Personnel (Bray, Hourani, Rae, Dever, Brown, Vincus, Pemberton, Marsden, Faulkner and Vandermaas-Peeler 2003); and a myriad of surveys outside the U.S. including the Canadian Retail Trade Survey (see, *e.g.*, Hidiroglou and Patak 2006), the Swedish Labour Force Survey (Mirza and Hörngren 2002), and the British Household Panel Survey (Taylor, Brice, Buck and Prentice-Lane 2007).

Calibration estimators, such as those generated under poststratification, are used to minimize errors associated with incomplete sampling frames (*i.e.*, undercoverage) and with sampling and nonresponse (see, *e.g.*, Särndal, Swensson and Wretman 1992; Lessler and Kalsbeek 1992; Kott 2006). For example, estimates from the Behavioral Risk Factor Surveillance System (BRFSS), a nationwide random-digit-dial (RDD) telephone survey conducted by the U.S. Centers for Disease Control and Prevention (CDC), are poststratified to counts that include households with and without landline telephone service (Centers for Disease Control and Prevention 2006). The decrease in the errors is linked to the association of the population control totals with the frame undercoverage, patterns of non-ignorable nonresponse, and the variable of interest (Kim, Li and Valliant 2007).

When relevant population controls do not exist, many researchers use survey-estimated control totals, and apply traditional variance formulae as if the controls were known without error. For example, Nadimpalli, Judkins and Chu (2004) adjusted weights for the *2003 National Survey of Parents and Youth* to the number of U.S. households with children ages 9-18 estimated from the *Current Population Survey* (CPS) using a ratio-raking algorithm (www.census.gov/cps). Estimates of how people in the U.S. spend their time can be calculated from *The American Time Use Survey* using weights that have been poststratified to projected estimates from the U.S. decennial Census (Killion 2006). More recently, researchers at the Pew Research Centers calibrated weights for a set of 2008 U.S. presidential pre-election surveys to population estimates from the March 2007 CPS, as well as to estimates on telephone usage patterns from the July-December 2007 *National Health Interview Survey* (Keeter, Dimock and Christian 2008).

The goal of our research is to develop and evaluate variance estimators for point estimates with weights that contain a poststratification adjustment to a set of survey-estimated control totals. We label the methodology which properly accounts for the estimated controls as *estimated-control* (*EC*) *poststratification*. In this paper, we focus specifically on the EC poststratified (ECPS) estimator of a population total for data collected from a stratified, multi-stage design, where the first-stage sampling units are selected *with replacement*. The remainder of this section gives a brief review of weight calibration and poststratification. Section 2

1. Jill A. Dever, RTI International. Email: jdever@rti.org; Richard Valliant, Survey Research Center, University of Michigan and Joint Program in Survey Methodology, University of Maryland. Email: rvalliant@survey.umd.edu.

contains an explicit definition of the ECPS estimator under study, followed in Section 3 by an evaluation of the bias properties. Through a theoretical evaluation (Section 4) and a simulation study, we compare variance estimators developed for the ECPS estimator with a variance estimator chosen under the naïve "population control total" assumption. Both linearization and replication variance estimators are examined in our research. We provide illustrations on the effects of different levels of precision in the estimated controls on the variance estimates. The specifications for the simulation study are detailed in Section 5, followed by a summary of the results (Section 6). We conclude the paper with a brief summary and an overview of future research in this area.

Calibration estimators (Deville and Särndal 1992), such as a poststratified estimator of a population total, borrow strength from auxiliary information to improve the efficiency of survey estimates over simpler weighting methods. When the auxiliary variables are (linearly) related to the set of key survey variables, calibration estimators can be very efficient.

The general form of a *traditional* or *fixed-control* calibration estimator is best described as an expansion estimator or "linear weighting" estimator as discussed in Estevao and Särndal (2000). Define s to be the set of sample elements from a probability sample, and $d_k = 1/\pi_k$ to be the design weight for element k such that $\pi_k = \Pr(k \in s)$. An estimated population total of a variable y is $\hat{t}_y = \sum_{k \in s} w_k y_k$, where the calibration weight $(w_k = a_k d_k)$ for the k^{th} element defined as a function of the design weight, d_k, and a calibration-adjustment factor, a_k, also known as a g-weight (Särndal *et al.* 1992). The calibration weights are calculated by minimizing a specified function that measures the distance between the design and calibration weights subject to a set of constraints defined as:

$$\mathbf{t}_{Ux} = \hat{\mathbf{t}}_{Ax} \qquad (1)$$

where $\mathbf{t}_{Ux} = \sum_{k \in U} \mathbf{x}_k$, the vector of population controls (counts) corresponding to the $G(G \geq 1)$ auxiliary variables; $\hat{\mathbf{t}}_x = \sum_{k \in s} w_k \mathbf{x}_k$, the estimated population controls corresponding to the components of $\mathbf{t}_{Ux}$; and $\mathbf{x}_k$ is a vector of length G containing auxiliary or benchmark variable values for element k. Note that $\mathbf{x}_k$ may contain ones and zeros to indicate the presence or absence of a certain characteristic (*e.g.*, age 18-25), or larger values (*e.g.*, number of children). An example of such a calibration system is the generalized least squares (or chi-square) distance function $\sum_{k \in s_A} (w_k - d_k)^2 / c_k d_k$ that is minimized subject to the constraints in (1). This system generates a closed-form solution called the generalized regression estimator (GREG) for $c_k = 1$ (Deville and Särndal 1992). The poststratified estimator is a special case of the GREG.

Variance estimation techniques for the poststratified estimator, and more generally for the GREG, have been widely studied. Binder (1995) demonstrates techniques used to calculate a *Taylor linearization* variance estimator for the GREG. Additional references for the linearization variance estimator under poststratification (and calibration more generally) include Deville, Särndal and Sautory (1993), Demnati and Rao (2004), and Hidiroglou and Patak (2006). Särndal, Swensson and Wretman (1989) developed an approximate linearization variance for the GREG of a population total as a function of the population residuals from a specified model and the design weights (d_k). Valliant (1993) and Yung and Rao (1996) modified the residual-based variance estimator by multiplying the sample residuals by the calibration weights $w_k (= a_k d_k)$. They demonstrated that this revised estimator, created by linearizing the associated jackknife, reduced the bias associated with the original formula. This variance estimator is also discussed in Särndal *et al.* (1992), Stukel, Hidiroglou and Särndal (1996), and in Chapter 11 of Särndal and Lundström (2005). Properties of replication variance estimators (*i.e.*, jackknife and BRR) have been examined in, for example, Valliant (1993), Rust and Rao (1996), Canty and Davison (1999), Théberge (1999), Rao and Shao (1999), Yung and Rao (1996; 2000), and Kott (2006).

An assumption in the articles above is that the control totals, to which the auxiliary sample estimates are adjusted, 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 estimated from other surveys with non-negligible sampling variances. 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, Bremer, Smith and Thomas 2000).

Many researchers apply formulae developed for traditional poststratification even though the controls have been estimated. The tacit assumption is that any additional error (variance and bias) associated with these controls is negligible and can be ignored. Currently, the validity of this assumption can not be checked until a complete picture of EC poststratification has been developed.

2. The estimated-control poststratified estimator

To facilitate our discussion of the estimated-control poststratified estimator, we label the survey requiring poststratification as the *analytic survey* and the source of the control totals as the *benchmark survey*. In practice, more than one benchmark survey may be tapped for the control totals. However, we will assume only one benchmark survey for the theoretical development so that control total variances and covariances are estimable.

Let U represent the finite target population containing N elements and $t_y = \sum_{k \in U} y_k$ represent the population total of interest for a variable y. Let s_A represent a random sample of size n_A from the frame U_A for the analytic survey. A random sample s_B of size n_B is selected for the benchmark survey from the corresponding sampling frame U_B. We allow the possibility that each of the frames, U_A and U_B, do not completely cover the target population U. However, coverage is treated as a random event so that all elements in the target population have a positive probability of being covered by either the analytic or the benchmark survey frame.

As a convention throughout the paper, an "A" subscript signifies an association with the analytic survey such as a sample design parameter or an estimate. A "B" subscript identifies the benchmark survey quantities. These subscripts are absent from the parameters associated with the population of interest, *i.e.*, t_y.

For the stratified, multi-stage design assumed for the analytic survey, $m_{Ah} (m_{Ah} \geq 2)$ primary sampling units (PSUs), indexed by i, are selected *with replacement* from a total of M_{Ah} PSUs in the h^{th} design stratum ($h = 1, \ldots, H$ with $H \geq 2$). We assume that n_{Ahi} elements, each indexed by k, are selected from N_{Ahi} in PSU hi in such a way that an unbiased estimate of the PSU total can be made. The design weight, d_k, is calculated as the inverse of the unconditional inclusion probability for $k \in s_{Ahi}$, the set of analytic survey elements within the hi^{th} PSU. Thus, n_A, the size of the analytic survey sample, is calculated as $n_A = \sum_{h=1}^{H} \sum_{i=1}^{m_{Ah}} n_{Ahi}$. Elements for the benchmark survey are randomly drawn from the corresponding sampling frame; no explicit specifications are made for the random sampling method.

Poststratification can be used to correct for sampling and coverage errors. Therefore, we allow undercoverage in the analytic-survey, as well as, the benchmark-survey sampling frames. Additionally, we do not consider the effects of nonresponse.

Suppose that the population U can be divided into $g = 1, ..., G$ mutually exclusive and exhaustive poststrata. When the population count of elements, N_g, is known for each poststratum, the traditional poststratified estimator of a total for y is defined as

$$\hat{t}_{yPS} = \sum_{g=1}^{G} N_g \frac{\hat{t}_{Ayg}}{\hat{N}_{Ag}}, \tag{2}$$

where y_k is the value of the analysis variable y for element k; $\hat{t}_{Ayg} = \sum_{k \in s_A} \delta_{gk} d_k y_k$, the total of y in poststratum g estimated from the analytic survey data; $\hat{N}_{Ag} = \sum_{k \in s_A} \delta_{gk} d_k$, the analytic survey estimated total in poststratum g; and $\delta_{gk} = 1$ indicates membership in the g^{th} poststratum and zero otherwise. Note that $\hat{t}_{Ayg}$ may also be expressed as $\hat{t}_{Ayg} = \sum_{k \in s_{Ag}} d_k y_k$, where s_{Ag} indicates the set of analytic survey elements in poststratum g. The "hat" notation in the expression above is used to distinguish a population estimator (*e.g.*, $\hat{N}_{Ag}$) from the known population parameter (*e.g.*, N_g). If the count of elements in poststratum g is estimated by setting $y_k = 1$ in the formula for $\hat{t}_{Ayg}$, then $\hat{t}_{yPS}$ equals N_g. In this sense, $\hat{t}_{yPS}$ is poststratified to the population counts $N_1, \ldots, N_G$.

In certain situations, however, the population counts are not available and must be estimated from a benchmark survey. Define the ECPS estimator of a population total of a variable y as

$$\hat{t}_{yP} = \sum_{g=1}^{G} \hat{N}_{Bg} \frac{\hat{t}_{Ayg}}{\hat{N}_{Ag}}. \tag{3}$$

The number of population elements in the g^{th} poststratum $(g = 1, \ldots, G)$ estimated from the benchmark survey is denoted as $\hat{N}_{Bg} = \sum_{l \in s_{Bg}} w_l$, where s_{Bg} is the set of sample elements in poststratum g from the benchmark survey and w_l is the weight associated with the l^{th} element. The calibration-adjustment factors applied to the analytic survey design weights for $\hat{t}_{yP}$ are calculated as $a_k = \hat{N}_{Bg} / \hat{N}_{Ag}$ for $k \in s_{Ag}$.

Relating the poststratified estimators to the calibration system discussed in the previous section, $\hat{\mathbf{t}}_{Ax}$ is a G-length vector of estimated population counts for each poststratum such that $\hat{\mathbf{t}}_{Ax} = (\hat{t}_{Ax1}, \ldots, \hat{t}_{AxG})'$, where $\hat{t}_{Axg} \equiv \hat{N}_{Ag} = \sum_{k \in s_A} d_k \delta_{gk}$ and $x_k \equiv \delta_{gk} = 1$ if the element k is a member of the g^{th} poststratum and 0 otherwise. The vector $\mathbf{t}_{Ux}$ corresponds either to $\mathbf{N} = (N_1, \ldots, N_G)'$ for the $\hat{t}_{yPS}$ estimator given in (2), or to $\hat{\mathbf{N}}_B = (\hat{N}_{B1}, \ldots, \hat{N}_{BG})'$, a $G \times 1$ vector of benchmark control estimates, for the $\hat{t}_{yP}$ estimator given in (3).

The estimator $\hat{t}_{yP}$ can be expressed in matrix notation as $\hat{t}_{yP} = \hat{\mathbf{N}}'_B \hat{\bar{\mathbf{Y}}}_A$ where $\hat{\bar{\mathbf{Y}}}_A = (\hat{\mathbf{N}}_A)^{-1} \hat{\mathbf{t}}_{Ay}$, a $G \times 1$ vector of analytic survey estimates of the form $\hat{\bar{\mathbf{Y}}}_A = [\hat{t}_{A1} / \hat{N}_{A1}, \ldots, \hat{t}_{AG} / \hat{N}_{AG}]'$; $\hat{\mathbf{N}}_A = \text{diag}(\hat{N}_{A1}, \ldots, \hat{N}_{AG})$, a diagonal matrix of poststratum totals estimated from the analytic survey; and $\hat{\mathbf{t}}_{Ay} = [\hat{t}_{A1}, \ldots, \hat{t}_{AG}]'$ is a $G \times 1$ vector of poststratum totals for the outcome variable estimated from the analytic survey. The remaining variables associated with the matrix notation were defined previously.

An effective poststratification adjustment can reduce the bias in the resulting point estimates and will either reduce or minimally inflate the variance in comparison to the unadjusted weight. This effect is well known for traditional poststratification; we provide the comparative evaluation under an estimated-control setting in the next sections.

3. Bias in the ECPS of a population total

Traditional poststratification is known for reducing the bias associated with an incomplete sampling frame. This reduction is most successful when poststrata are formed such that the within-poststratum correlation of y_k with the probability of the k^{th} element being included on the sampling frame is very near zero (Kim, Li and Valliant 2007).

To evaluate the (unconditional) design-based bias for $\hat{t}_{yP}$, we must account for the random property of four components – the analytic and benchmark sample designs and the population coverage propensities for the corresponding sampling frames. Following the work of Kim, Li and Valliant (2007, equation 2), the approximate design bias of $\hat{t}_{yP}$ as an estimator of the population total $t_y = \sum_{k \in U} y_k$ is calculated as

$$\begin{aligned} \text{Bias}(\hat{t}_{yP}) &= E(\hat{t}_{yP}) - t_y \\ &\cong \sum_{g=1}^{G} \left[t_{yg} \left\{ \frac{N_{Bg}}{N_g} - 1 \right\} + N_{Bg} \text{Cov}(y_g, \phi_{Ag}) \bar{\phi}_{Ag}^{-1} \right] \end{aligned} \tag{4}$$

where N_g is the population size for the set of elements U_g within poststratum g; $N_{Bg} = E(\hat{N}_{Bg})$, the expected value of the poststratum estimates under the benchmark survey design; $\text{Cov}(y_g, \phi_{Ag}) = N_g^{-1} \sum_{k \in U_g} (y_k - \bar{y}_g)(\phi_{Ak} - \bar{\phi}_{Ag})$, the population covariance between the outcome variable (y_k) and the coverage propensities (ϕ_{Ak}) within poststratum g; $\bar{y}_g = t_{yg}/N_g$, the g^{th} poststratum mean of y; $t_{yg} = \sum_{k \in U_g} y_k$, the population total of y within poststratum g; and $\bar{\phi}_{Ag} = N_{Ag}/N_g$, the average coverage propensity within the poststratum under the analytic survey design with $N_{Ag} = E(\hat{N}_{Ag})$. Note that the population total may also be expressed as $t_y = \sum_g t_{yg}$.

Components of the bias are zero only under certain conditions. (*i*) If $N_{Bg} = N_g$ for all g (*i.e.*, no coverage errors in the benchmark sampling frame), then the bias is dependent only on the association between the outcome variable and the coverage propensities, $\text{Cov}(y_g, \phi_{Ag})$. The value of $\text{Bias}(\hat{t}_{yP})$ then reduces to the formula provided in Kim, Li and Valliant (2007, equation 2) for the traditional poststratified estimator, $\hat{t}_{yPS}$. (*ii*) If the coverage probabilities are constant within each poststratum (*i.e.*, $\phi_{Ak} = \bar{\phi}_{Ag}$, $k \in U_g$ for all g), then the second bias component is zero. Only if *both* conditions are satisfied can we say that $\hat{t}_{yP}$ is approximately unbiased. Some may argue that a "perfect" combination of poststrata could be formed such that the positive and negative components cancel; however, we believe this likelihood to be so rare as to be virtually impossible.

Having examined bias, we present an evaluation of the variance of $\hat{t}_{yP}$. For some estimators, the contribution of the bias (squared) to the total mean square error (MSE) is small relative to the variance.

4. Variance estimation for the ECPS

Variance estimators have been developed for traditional poststratification and are available in software designed to analyze survey data, *e.g.*, R® (R Development Core Team 2009), SAS® (SAS Institute Inc. 2009), Stata® (StataCorp 2010), and SUDAAN® (Research Triangle Institute 2008). However, limited work has been completed on variance estimation for EC poststratification.

Four EC variance estimators for $\hat{t}_{yP}$ that account for the variance in the control totals are presented in the following subsections after defining the population sampling variance. They include one newly developed linearization variance estimator, and three delete-one-PSU (delete-one) jackknife variance estimators. With the delete-one jackknife, replicates are created by sequentially deleting one 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 analytic-survey PSUs per stratum (m_{Ah}) across the H strata $(h = 1, ..., H)$.

An effective variance estimator will reproduce the corresponding population sampling variance in expectation. The approximate (or asymptotic) population sampling variance of $\hat{t}_{yP} = \hat{\mathbf{N}}_B' \hat{\bar{\mathbf{Y}}}_A$ has the following form:

$$\begin{aligned} \text{AV}(\hat{t}_{yP}) &= \mathbf{N}_B' \mathbf{V}_A \mathbf{N}_B + 2\bar{\mathbf{Y}}_A' \text{Cov}(\hat{\mathbf{N}}_B, \hat{\bar{\mathbf{Y}}}_A) \mathbf{N}_B + \bar{\mathbf{Y}}_A' \mathbf{V}_B \bar{\mathbf{Y}}_A \\ &= \mathbf{N}_B' \mathbf{V}_A \mathbf{N}_B + \bar{\mathbf{Y}}_A' \mathbf{V}_B \bar{\mathbf{Y}}_A \end{aligned} \tag{5}$$

where $\mathbf{N}_B = E(\hat{\mathbf{N}}_B)$, a vector of expected values for the benchmark poststratum counts within the G poststrata; $\hat{\mathbf{N}}_B = (\hat{N}_{B1}, \ldots, \hat{N}_{BG})'$ is a G-length vector of control totals estimated from the benchmark survey; $\bar{\mathbf{Y}}_A$ is a G-length vector with population components of the form $\bar{y}_{Ag} = t_{Ayg}/N_{Ag}$; $\mathbf{V}_A$ is the population (variance-)covariance matrix of the estimated components of the vector $\bar{\mathbf{Y}}_A$; and $\mathbf{V}_B$ is the covariance matrix of the G benchmark control estimates $\hat{\mathbf{N}}_B$. 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 component, $\bar{\mathbf{Y}}_A' \mathbf{V}_B \bar{\mathbf{Y}}_A$, is the variance associated with the benchmark estimates conditioned on the analytic survey sample; this is the EC poststratification variance component. Because we assume that the analytic and benchmark surveys are independent, the covariance of estimates from the two surveys is, by definition, zero. Hence, the component $\text{Cov}(\hat{\mathbf{N}}_B, \hat{\bar{\mathbf{Y}}}_A)$ above is eliminated from the expression.

Krewski and Rao (1981), Rao and Wu (1985), and others demonstrated the asymptotic consistency of the linearization and jackknife variance estimators for nonlinear functions. However, this examination needs to be extended to the EC poststratification. We discuss the set of EC variance estimators for the population sampling variance below identified or developed for our research. The sample estimators were calculated by substituting sample estimates for the corresponding variance parameters. We begin with an evaluation of a traditional or naïve poststratified variance estimator that does not account for the variation in the estimated controls.

4.1 A traditional variance estimator for EC poststratification (Naïve)

A variety of variance estimators have been developed for poststratification estimators. With all of the methods, the controls are assumed to be fixed and known without error. Therefore, $\bar{\mathbf{Y}}'_A \mathbf{V}_B \bar{\mathbf{Y}}_A$, the second (positive) component in expression (5), is zero because $\mathbf{V}_B = \mathbf{0}$ by assumption. The linearization variance estimator has the form

$$\text{var}_{\text{Naïve}}(\hat{t}_{yP}) = \hat{\mathbf{N}}'_B \hat{\mathbf{V}}_A \hat{\mathbf{N}}_B \quad (6)$$

where $\hat{\mathbf{N}}_B$ is the vector of the G benchmark control total estimates, and $\hat{\mathbf{V}}_A$ is the estimated covariance matrix of the estimates $\hat{\bar{\mathbf{Y}}}_A = (\hat{t}_{Ay1}/\hat{N}_{A1}, \ldots, \hat{t}_{AyG}/\hat{N}_{AG})$. Because the second component in the second line of (5) is not estimated, any variance formula developed for traditional post-stratification will by definition underestimate the population sampling variance. However, highly precise benchmark estimates may contribute a negligible EC-poststratification variance component to the overall estimate. Thus, the difference between the estimates for traditional and EC poststratification will for these situations also be negligible.

4.2 Taylor series linearization (ECTS)

A linearization variance estimator for the $\hat{t}_{yP}$ has the form:

$$\text{var}_{\text{ECTS}}(\hat{t}_{yP}) = \hat{\mathbf{N}}'_B \hat{\mathbf{V}}_A \hat{\mathbf{N}}_B + \hat{\bar{\mathbf{Y}}}'_A \hat{\mathbf{V}}_B \hat{\bar{\mathbf{Y}}}_A \quad (7)$$

where $\hat{\mathbf{V}}_B$ is the estimated benchmark covariance matrix for the set of G control totals. The remaining terms are defined for expression (6). The ECTS formula is a function of the variance under traditional poststratification and an additive inflation term associated with the variation in the benchmark controls, *i.e.*, $\text{var}_{\text{ECTS}}(\hat{t}_{yP}) = \text{var}_{\text{Naïve}}(\hat{t}_{yP}) + \hat{\bar{\mathbf{Y}}}'_A \hat{\mathbf{V}}_B \hat{\bar{\mathbf{Y}}}_A$.

Ideally, the benchmark survey analysis file would be available to calculate the values for $\hat{\mathbf{V}}_B$. However, researchers may have to rely on published estimates for only the marginal control totals, *i.e.*, point and variance estimates by one characteristic instead of the counts and covariance estimates for a set of characteristics. The implications of having limited information are discussed further in Section 4.4.

4.3 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 poststratification situation. The premise behind Fuller's methodology (ECF2) is to take a spectral (eigenvalue) decomposition of the benchmark covariance matrix $(\hat{\mathbf{V}}_B)$, develop benchmark adjustments that are a function of the resulting eigenvalues and eigenvectors, and add the adjustments to the vector of benchmark controls $(\hat{\mathbf{N}}_B)$ to create a set of replicate controls. A randomly chosen subset of the m_A replicates is poststratified to the G constructed replicate controls where the total number of PSUs must equal or exceed the number of poststrata, *i.e.*, $m_A \geq G$. Specifically, the benchmark control total for the r^{th} replicate is defined as

$$\hat{\mathbf{N}}_{B(r)} = \hat{\mathbf{N}}_B + c_h \hat{\mathbf{z}}'_{(r)} \quad (8)$$

where $\hat{\mathbf{z}}'_{(r)} = \delta_{(r)} \sum_{g=1}^{G} \delta_{g|(r)} \hat{\mathbf{z}}'_g$; $c_h = \sqrt{m_{Ah}/(m_{Ah} - 1)}$, a constant related to the delete-one jackknife variance method; $\delta_{(r)}$ 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 $(\hat{\lambda}_g)$ where $\hat{\mathbf{V}}_B = \sum_{g=1}^{G} \hat{\mathbf{z}}_g \hat{\mathbf{z}}'_g$, by definition. Thus, given that $\delta_{(r)} = 1$ 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.

The delete-one jackknife can take multiple forms depending on the centering value. We chose the somewhat conservative variance estimator centered about the full-sample estimate for our research (v_4 in Wolter 2007, section 4.5). The delete-one jackknife variance estimator, $\text{var}_{\text{ECF2}}(\hat{t}_{yP})$, is calculated as follows under the Fuller method for a stratified, multi-stage design.

$$\begin{aligned}\text{var}_{\text{ECF2}}(\hat{t}_{yP}) &= \sum_{h=1}^{H} \frac{(m_{Ah} - 1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{t}_{yP(r)} - \hat{t}_{yP})^2 \\ &= \sum_{h=1}^{H} \frac{(m_{Ah} - 1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\hat{t}_{yP(r)} - \hat{t}_{yP} + c_h \hat{\mathbf{z}}'_{(r)} \hat{\mathbf{B}}_{A(r)})^2 \quad (9)\end{aligned}$$

where the terms in (9) are defined below. Note that the association of the r^{th} replicate to a particular design stratum is defined through the stratum membership of the eliminated PSU. The replicate estimates in (9) are defined as $\hat{t}_{Ayg(r)} = \sum_h \sum_{i \in s_{Ah}} d_{i(r)} \sum_{k \in s_{Ahi}} \delta_{gk}\, d_k\, y_k$ and $\hat{N}_{Ag(r)} = \sum_h \sum_{i \in s_{Ah}} d_{i(r)} \sum_{k \in s_{Ahi}} \delta_{gk}\, d_k$, where the PSU-subsampling weights are calculated as

$$d_{i(r)} = \begin{cases} 0 & \text{if } r = i,\ i \in s_{Ah} \\ 1 & \text{if } h \neq h' \text{ for } r \in s_{Ah} \text{ and } i \in s_{Ah'} \\ m_{Ah}/(m_{Ah} - 1) & \text{if } r \neq i \text{ but } h = h'. \end{cases} \quad (10)$$

The remaining terms in (9) are $\hat{\mathbf{B}}_{A(r)} = \hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)}$, the estimated mean of the outcome variable within poststratum g and replicate r;

$$\ddot{t}_{yP(r)} = \sum_{g=1}^{G} \hat{N}_{Bg(r)} (\hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)}), \quad (11)$$

a function of replicate estimates with $\hat{N}_{Bg(r)}$ defined as the g^{th} component in expression (8); $\hat{t}_{yP(r)}$ is the replicate estimate under traditional poststratification, namely $\sum_{g=1}^{G} \hat{N}_{Bg}(\hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)})$; and $\hat{t}_{yP}$ is the estimated total given in expression (3) calculated from the complete sample file. Squaring the terms in (9) results in a variance component conditioned on the benchmark controls, a component due to the benchmark control variability, and a cross-term of lower order that is approximately equal to zero in expectation. The design-expectation of the resulting jackknife variance estimator is asymptotically equivalent to $\text{AV}(\hat{t}_{yP})$ in (5) only if the respective components are calculated with values from design-consistent estimators. Fuller (1998) also demonstrated that the jackknife variance of the replicate controls, $\text{var}_{\text{ECF2}}(\hat{\mathbf{N}}_B)$, reproduces the estimated benchmark covariance matrix $\hat{\mathbf{V}}_B$ for every sample.

Currently no software exists to calculate the ECF2. The six steps needed to calculate $\text{var}_{\text{ECF2}}(\hat{t}_{yP})$ using any appropriate programmable package are as follows:

1. Calculate the full-sample estimate $\hat{t}_{yP}$ using expression (3).
2. Determine the G eigenvalues $\hat{\lambda}_g$ and eigenvectors $\hat{\mathbf{q}}_g$ for $\hat{\mathbf{V}}_B$, and calculate the replicate adjustments $\hat{\mathbf{z}}_g = \hat{\mathbf{q}}_g \sqrt{\hat{\lambda}_g}$. Concatenate the $G \times G$ matrix of $\hat{\mathbf{z}}_g$'s with a $G \times (m_A - G)$ matrix of zeros, and randomly sort the columns. Call this new $G \times m_A$ matrix $\hat{\mathbf{Z}}$.
3. Calculate a vector of length m_A with values equal to $c_h = \sqrt{m_{Ah}/(m_{Ah} - 1)}$ ordering from $h = 1$ to H. Populate each row of a $G \times m_A$ matrix, called $\mathbf{C}$, with this vector, *i.e.*, the row values are repeated . The m_A-length vector of jackknife stratum weights, $\mathbf{W}_R$, is created with components equal to $(m_{Ah} - 1)/m_{Ah}$ where the deleted PSU is extracted from stratum h.
4. Calculate the Hadamard (or element-wise) product (Searle 1982, page 49) of $\hat{\mathbf{Z}}$ and $\mathbf{C}$ denoted as $\hat{\mathbf{Z}} \bullet \mathbf{C}$. Replicate the vector $\hat{\mathbf{N}}_B$ into the columns of a $G \times m_A$ matrix and add to $\hat{\mathbf{Z}} \bullet \mathbf{C}$. This new $G \times m_A$ matrix, called $\hat{\mathbf{N}}_{BR}$, contains the replicate benchmark controls discussed in expression (8) for all m_A replicates.
5. Calculate the replicate estimates $\hat{\bar{y}}_{Ag(r)} = \hat{t}_{Ayg(r)}/\hat{N}_{Ag(r)}$ by removing in-turn one PSU from the analytic survey sample file, adjusting the weights for the remaining PSUs ($\mathbf{W}_R$ values), and summing the weighted values for the numerator and denominator within poststratum g. Call the resulting $G \times m_A$ matrix $\hat{\bar{\mathbf{Y}}}_R$.
6. Calculate the m_A replicate estimates, $\ddot{t}_{yP(r)}$, by first multiplying the elements $\hat{\mathbf{N}}_{BR}$ by $\hat{\bar{\mathbf{Y}}}_R$ and summing down the rows within a column. Next, subtract $\hat{t}_{yP}$ from each of the m_A values and square the terms, multiply by the PSU-subsampling weight adjustments specified in (10), and sum across the m_A estimates. The resulting value is the estimated variance using the Fuller method, $\text{var}_{\text{ECF2}}(\hat{t}_{yP})$.

4.4 Nadimpalli-Judkins-Chu jackknife method (ECNJC)

Nadimpalli *et al.* (2004) developed a delete-one jackknife variance estimator that randomly perturbs the control totals for the complete set of replicates instead of adjusting only a subsample of replicates as discussed for the ECF2. The benchmark survey replicate control totals have the following form:

$$\hat{\mathbf{N}}_{B(r)} = \hat{\mathbf{N}}_B + c_h R_h \hat{\mathbf{S}}_B \boldsymbol{\eta}_{(r)} \quad (12)$$

where $c_h = \sqrt{m_{Ah}/(m_{Ah} - 1)}$, as with the ECF2; $R_h = \sqrt{1/(H m_{Ah})}$, a function of the total number of analytic-survey strata (H) and PSUs (m_{Ah}); $\hat{\mathbf{S}}_B$ is a *diagonal* matrix of estimated standard errors for the benchmark controls; and $\boldsymbol{\eta}_{(r)}$ is a G-length vector of values randomly generated for each replicate from the standard normal distribution. The remaining terms are specified for the ECF2 following expression (8). Note that the covariance estimates included in the ECF2, *i.e.*, the off-diagonal values of $\hat{\mathbf{V}}_B$, are set to zero for the ECNJC.

The corresponding delete-one jackknife variance estimator of the poststratified total is calculated as follows:

$$\begin{aligned} \text{var}_{\text{ECNJC}}(\hat{t}_{yP}) &= \sum_{h=1}^{H} \frac{(m_{Ah} - 1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{t}_{yP(r)} - \hat{t}_{yP})^2 \\ &= \sum_{h=1}^{H} \frac{(m_{Ah} - 1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\hat{t}_{yP(r)} - \hat{t}_{yP} \\ &\qquad + c_h R_h \boldsymbol{\eta}'_{(r)} \hat{\mathbf{S}}_B \hat{\mathbf{B}}_{A(r)})^2, \end{aligned} \quad (13)$$

where $\ddot{t}_{yP(r)}$ is computed as described for the ECF2 in (11) but with $\hat{N}_{Bg(r)}$ defined by the g^{th} component in (12). Unlike the ECF2, the sample variance of the ECNJC replicate controls given in (12) reproduces the benchmark covariance matrix $\mathbf{V}_B$ *in expectation* only if the covariance terms are truly zero (see Appendix A for details). If $\mathbf{V}_B$ is not diagonal, $\text{var}_{\text{ECNJC}}$ fails this test.

Use of the ECNJC would be plausible in two cases: (*i*) the complete benchmark covariance matrix for the controls is unavailable (*e.g.*, estimates taken from a previous report), or (*ii*) the covariance terms are negative so that the resulting values defined by (12) would lead to conservative variance estimates. The diagonal matrix for $\hat{\mathbf{S}}_B$ 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 ECNJC, the expectation of the variance estimator will *not* approximate $\text{AV}(\hat{t}_{yP})$ in (5); the bias term is related to the difference between the design expectation of $\hat{\mathbf{S}}_B^2$ and $\mathbf{V}_B$.

4.5 Multivariate normal jackknife method (ECMV)

The multivariate normal method (ECMV) is a generalization of the ECNJC and to our knowledge is first discussed in this paper. The ECMV uses the complete covariance matrix $\hat{\mathbf{V}}_B$ and relies on large-sample theory so that the control total adjustments may be modeled as coming from a G-dimensional multivariate normal (MVN) distribution. The replicate controls for the ECMV have the form

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

where $\hat{\boldsymbol{\varepsilon}}_{(r)}$ is a G-length vector of random variables such that $\hat{\boldsymbol{\varepsilon}}_{(r)} \overset{\text{i.i.d.}}{\sim} \text{MVN}_G(\mathbf{0}, \hat{\mathbf{V}}_B)$; $c_h = \sqrt{m_{Ah}/(m_{Ah}-1)}$; and $R_h = \sqrt{1/(H\, m_{Ah})}$.

The delete-one jackknife variance estimator for the ECMV is calculated as

$$\begin{aligned}\text{var}_{\text{ECMV}}(\hat{t}_{yP}) &= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\ddot{t}_{yP(r)} - \hat{t}_{yP})^2 \\ &= \sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\hat{t}_{yP(r)} - \hat{t}_{yP} \\ &\qquad + c_h R_h \hat{\boldsymbol{\varepsilon}}'_{(r)} \hat{\mathbf{B}}_{A(r)})^2, \end{aligned} \qquad (15)$$

where $\ddot{t}_{yP(r)}$ is computed as described for the ECF2 in (11) but with $\hat{N}_{Bg(r)}$ defined by the g^{th} component in (14). Unlike the Fuller method, $\text{var}_{\text{ECMV}}(\hat{\mathbf{N}}_B) \neq \hat{\mathbf{V}}_B$; instead, the ECMV must rely on the design-based properties of the estimator. The design expectation of this estimator is evaluated with respect to the MVN distribution conditioned on the benchmark estimates (E_ε), and then with respect to the benchmark survey design (E_B). As shown in Appendix B.1,

$$E_B[E_\varepsilon(\text{var}_{\text{ECMV}}(\hat{\mathbf{N}}_B)|B)] = E_B(\hat{\mathbf{V}}_B). \qquad (16)$$

If $\hat{\mathbf{V}}_B$ is an approximately unbiased estimator of $\mathbf{V}_B$, then the population covariance matrix is reproduced with this method.

Under the Fuller two-phase method, $\text{Var}[\text{var}_{\text{ECF2}}(\hat{\mathbf{N}}_B)] = \text{Var}(\hat{\mathbf{V}}_B)$ because $\text{var}_{\text{ECF2}}(\hat{\mathbf{N}}_B) = \hat{\mathbf{V}}_B$. To compare ECF2 and ECMV further, note that if we define $y_k = 1$ in the analytic survey, then $\hat{t}_{yP} = \mathbf{1}'\hat{\mathbf{N}}_B$. As shown in Appendix B.2,

$$\text{Var}[\text{var}_{\text{ECMV}}(\mathbf{1}'\hat{\mathbf{N}}_B)] =$$

$$\text{Var}_B[\mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1}] + \frac{2}{H\bar{m}_A^*}[E_B(\mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1})^2] > \text{Var}_B[\mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1}] \quad (17)$$

where $\bar{m}_A^*$ is the harmonic mean of the PSU sample sizes per stratum in the analytic survey. This suggests that the var_{ECF2} and the var_{ECMV} have similar large sample expectations, though in practice the ECMV is likely to be more variable than the ECF2. We examine this issue through a simulation study described in the next section.

5. Description of simulation study

We complement the theoretical evaluation of the five variance estimators discussed in the previous section with an analysis of simulation results.

5.1 Simulation parameters

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. Samples were selected from this "population" using a two-stage design. Two PSUs were selected *with replacement* using probabilities proportional to the total number of adults (PPS) within the PSU. From within each sample PSU, we selected simple random samples of (n_{Ahi} =) 20 and 40 persons *without replacement* giving total sample sizes of 1,000 and 2,000, respectively. Two within-PSU sample sizes were considered for this study to evaluate the effects of smaller analytic survey variance components, calculated by increasing n_A, on the variance of $\hat{t}_{yP}$. For each combination of PSU and person-level samples (*i.e.*, 50 PSUs and either 1,000 or 2,000 persons), we selected 4,000 simulation samples. We calculated the estimated population totals and associated variances for two binary NHIS 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).

We exclude nonresponse from consideration in our current simulation study to minimize factors that might affect our comparisons. (Note: The interview questions for these variables can be found in the family core instrument at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2003/qfamilyx.pdf. Responses from questions FHI.070 and FAU.010/FAU.020 were used to generate the variables NOTCOV and PDMED12M, respectively).

Poststratification may reduce variances slightly. However, in household surveys, this technique is mainly used to correct for sampling frame undercoverage, as well as other problems inherent with surveys. Each of the 4,000 simulation samples was selected to mimic a sampling frame for the analytic survey that suffers from differential undercoverage, such as those used for many telephone surveys. Sixteen ($G = 16$) poststratification cells were defined by an eight-level age variable crossed with gender. The coverage rates for the 16 cells were created based on the population means for each age group by gender and range in value from 0.5 to 0.9. 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 percent of the male population 65-69 years of age was randomly selected to be in the sampling frame for the NOTCOV simulations. This process of subsetting the population to the frame was independently implemented for each sample and for each outcome variable.

We suspect that the decision for researchers to use either a traditional or an EC poststratification variance estimator depends on the precision of the control totals. We calculated the benchmark covariance matrix ($\hat{\mathbf{V}}_B$) from the complete NHIS public-use data file (92,148 records) and ratio adjusted the values to reflect a sample size comparable with our simulation population ($N = 21{,}664$). The off-diagonal values of $\hat{\mathbf{V}}_B$ range from -0.05 to 0.75 with a mean value of 0.22. From this 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 benchmark surveys with an approximate effective sample size of 21,700, 6,000 (≈ 21,700/3.6), 1,200, and less than 500, respectively.

The simulation was conducted in R® (Lumley 2009; R Development Core Team 2009) 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 the relevant code does not currently exist.

5.2 Evaluation criteria

The empirical results for the five variance estimators discussed in the previous section (Naïve, ECTS, ECF2, ECNJC, and ECMV) are compared using three measures across the $j = 1, \ldots, 4{,}000$ simulation samples, and the two outcome variables (NOTCOV and PDMED12M). The measures include: (*i*) the estimated percent relative bias of the variance estimator, $(1/4{,}000\sum_j \operatorname{var}(\hat{t}_{yP_j}) - \text{mse})/\text{mse}$ where $\operatorname{var}(\hat{t}_{yP_j})$ is one of the five variance estimates evaluated for sample *j* and *mse* is the mean square error of $\hat{t}_{yP}$ defined below; (*ii*) the 95% confidence interval coverage rate, $1/4{,}000\sum_j I(|\hat{z}_j| \le z_{1-\alpha/2})$ where $\hat{z}_j = (\hat{t}_{yP_j} - t_y)/\sqrt{\operatorname{var}(\hat{t}_{yP_j})}$; and, (*iii*) the standard deviation of the estimated standard errors, calculated as the square root of $1/(4{,}000-1)\sum_j(\sqrt{\operatorname{var}(\hat{t}_{yP_j})} - 1/4{,}000\sum_j\sqrt{\operatorname{var}(\hat{t}_{yP_j})})^2$. The relative bias and the root mean square error of our point estimators are calculated as $1/4{,}000\sum_j(\hat{t}_{yP_j} - t_y)/t_y$ and $\sqrt{\text{mse}} = \sqrt{1/4{,}000\sum_s(\hat{t}_{yP_j} - t_y)^2}$, respectively.

6. Simulation study results

6.1 Point estimator

To justify the need for poststratification, we initially evaluated the Horvitz-Thompson estimate ($\sum_{s_A} d_k y_k$) for the two outcome variables. This estimator is known to be design-unbiased under pristine conditions. The percent relative bias indicates that the HT estimator is negatively biased, underestimating the population total by 38 percent for NOTCOV and 41 percent for PDMED12M. These large values show that some correction is needed to adjust for the non-negligible levels of bias. The percent relative bias for the poststratified estimator $\hat{t}_{yP}$ was much lower – the $\hat{t}_{yP}$ is positively biased by no more than two percent for both outcome variables.

6.2 Variance estimators

Adding to the theoretical evaluation discussed in Section 4, the empirical results for an effective variance estimator should possess a *percent relative bias* either near zero or somewhat positive for a conservative measure (see Section 5.2 for the formula of the percent relative bias).

The percent relative biases generated from our simulation study are provided in Table 1. Bias estimates for the Naïve and ECNJC variance estimators are larger than for the other EC estimators for all our simulations. Estimates for the ECTS are somewhat smaller than the values calculated for the ECF2 and ECMV estimators for relatively small benchmark surveys. However, the differences are negligible as the size of the benchmark survey increases.

Table 1
Percent relative bias estimates for five variance estimators by outcome variable and relative size of the benchmark survey to the analytic survey

Outcome Variable	Variance Estimator	Relative Size (n_A = 1,000) 0.3	1.2	6.0	21.7	Relative Size (n_A = 2,000) 0.2	0.6	3.0	10.8
NOTCOV	Naïve	-50.3	-23	-10.7	-9.2	-56.0	-31	-14.2	-12.2
	ECTS	-4.5	-4.5	-6.1	-7.7	-0.2	-8.4	-8.2	-10.1
	ECF2	-4.7	-4.6	-5.8	-7.5	0.1	-8.2	-8.3	-10.1
	ECNJC	-36.7	-17.1	-8.9	-8.2	-40	-24.2	-11.9	-11.1
	ECMV	-4.3	-4.1	-6.0	-7.5	-0.2	-8.1	-8.1	-10.0
PDMED12M	Naïve	-34.4	-14.5	-5.7	-3.9	-48.1	-23.4	-10	-10.1
	ECTS	-3.3	-3.7	-2.7	-2.6	-4.7	-6.4	-5.1	-7.8
	ECF2	-3.5	-3.5	-2.4	-2.3	-4.6	-6.8	-5.2	-7.8
	ECNJC	-24.5	-10.5	-4.0	-2.7	-35.1	-17.6	-7.6	-8.4
	ECMV	-3.0	-3.3	-2.4	-2.2	-4.3	-6.3	-5.0	-7.7

The traditional poststratified estimator (Naïve) was most negatively biased among those compared as expected. When the benchmark survey is smaller than the analytic survey (and therefore produces estimates less precise than the analytic survey), the Naïve estimator is negatively biased by as much as 56 percent. The level of bias improved as the relative size of the benchmark survey increased; however, the Naïve estimator still resulted in, at best, a four percent underestimate. The ECNJC estimator fared slightly better than the Naïve estimator though the bias (-2.7 to -40 percent) is still larger than the other EC variance estimators, which range between -10.1 and 0.1 percent.

For a small benchmark survey relative to the size of the analytic survey (*i.e.*, relative size less than one), the levels of (absolute) bias dramatically increased for the Naïve and ECNJC estimators. The opposite effect is noted for the other EC variance estimators. The variance component associated with the benchmark survey, *e.g.*, $\hat{\hat{\mathbf{Y}}}'_A \hat{\mathbf{V}}_B \hat{\hat{\mathbf{Y}}}_A$ shown for var_{ECTS} in (7), becomes the dominate term within the EC variance estimators as the precision of the benchmark survey estimates decreases. Thus the benchmark variance component somewhat corrects for the underestimation associated with the analytic variance component. Additional research is needed to determine if a threshold exists for when such a counterbalance of bias can occur. The overall negative bias of our estimates is similar to the bias of linearization variance estimators as shown in another context by Rao and Wu (1985, section 4) and Wu (1985). However, further research is also needed to determine how to minimize the underestimation.

Note that the relative sizes of 21.7 when $n_A = 1{,}000$ and 10.8 when $n_A = 2{,}000$ both imply benchmark survey sample sizes of about 21,600. Thus the $O(M^2/m_B)$ component of the variance, $\bar{\mathbf{Y}}'_A \mathbf{V}_B \bar{\mathbf{Y}}_A$, is more prominent for the estimates in Table 1 based on $n_A = 2{,}000$. This leads to larger relative biases in these estimates, relative to those produced under $n_A = 1{,}000$, even though the analytic survey sample size is larger.

The patterns exhibited for the percent relative bias are reflected in the *coverage rates for the 95 percent confidence intervals* for the estimated totals but are not provided for sake of brevity. The Naïve and ECNJC estimators are more likely to experience confidence intervals coverage rates below 95 percent. These rates approach the appropriate level as the precision of the benchmark survey estimates improves. However, the remaining EC variance estimators had coverage rates near acceptable levels regardless of the relative size of the surveys and therefore are more robust.

The discussion so far suggests that there are minimal theoretical, as well as empirical, differences between the ECTS, ECF2, and ECMV methods. We finally look to the *standard deviation of the estimated standard errors* (SEs) in an attempt to distinguish the estimators. An examination of this variability can provide insight on the (empirical) stability of the variance estimators, *i.e.*, an unstable variance estimator could generate a poor variance estimate based on the nuances of a particular sample. Table 2 contains the percent relative increase in the standard deviations for the ECF2 and the ECMV both in comparison to the ECTS.

The variation in the ECMV variance estimates was noticeably larger than for ECF2 but only for relatively small benchmark surveys. The difference increased as the size of the analytic survey increased. This suggests that the ECF2 may be preferred over the ECMV due to increased stability in the variance estimates. However, further research is being conducted on the threshold for when the instability can affect the estimates.

7. Conclusions and future work

The theoretical and analytical work discussed in this paper support the need for a new methodology to address post-stratification 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 implemented in the future.

Table 2
Percent increase in instability of variance estimates relative to the ects by outcome variable and relative size of the benchmark survey

Outcome Variable	Variance Estimator	Relative Size (n_A = 1,000)				Relative Size (n_A = 2,000)			
		0.3	**1.2**	**6.0**	**21.7**	**0.2**	**0.6**	**3.0**	**10.8**
NOTCOV	ECF2	12.0	5.5	2.3	0.2	15.1	8.4	2.1	0.6
	ECMV	21.2	7.4	1.8	0.3	30.8	8.5	2.4	0.7
PDMED12M	ECF2	7.7	3.8	1.1	0.4	12.0	6.3	2.1	0.7
	ECMV	11.5	4.0	0.9	0.5	22.6	7.6	2.2	1.1

The EC linearization variance estimator var_{ECTS} in expression (7) shows promise for EC poststratification. This estimator is especially effective at reducing the percent relative bias experienced with the Naïve variance estimator in (6) when the benchmark survey is small relative to the analytic survey. The replication variance estimator var_{ECF2} given in (9) is recommended specifically for studies requiring replicate weights such as when public-use analysis files are released without sampling design information to further protect data confidentiality and respondent privacy. The alternative replication estimator var_{ECMV} also performed well and is somewhat easier to implement than var_{ECF2}.

Implementation of the recommended variance estimators requires specialized computer programs because the capabilities are currently not available in standard software. The linearization estimator may be more approachable because implementation involves a modification to available variance estimates, *e.g.*, $\text{var}_{\text{ECTS}}(\hat{t}_{y\text{ECPS}}) = \text{var}_{\text{Naïve}}(\hat{t}_{y\text{ECPS}}) + \hat{\hat{\mathbf{Y}}}'_A \hat{\mathbf{V}}_B \hat{\hat{\mathbf{Y}}}_A$. We provide a step-by-step discussion of the procedures required for the var_{ECF2} (see Section 4.3) to facilitate the creation of the computer program.

Extensions to this research to be presented at a later date include a generalization to linear calibration, to other statistics including a ratio-estimated mean, and to domain estimation. We additionally are investigating whether threshold values are identifiable which determine (*i*) when there are negligible differences between traditional and EC variance estimation, and (*ii*) when the benchmark controls are too imprecise to use for calibration. We also plan to investigate the theoretical implications of measurement errors in the analytic as well as the benchmark surveys.

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 thanks the members of her committee, Richard Valliant, Phillip Kott, Frauke Kreuter, Stephen Miller and Paul Smith for their guidance. The authors also thank the associate editor and referees for their constructive comments which clarified the presentation.

Appendix A

Derivation of $\text{var}_{\text{ECNJC}}(\hat{\mathbf{N}}_B)$

For the following derivations, let E_ε represent the expectation with respect to a standard normal distribution. All other terms are defined in the body of the paper.

$$\text{var}_{\text{ECNJC}}(\hat{\mathbf{N}}_B) = \sum_{h=1}^{H} \frac{m_{Ah}-1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\hat{\mathbf{N}}_{B(r)} - \hat{\mathbf{N}}_B)(\hat{\mathbf{N}}_{B(r)} - \hat{\mathbf{N}}_B)'$$

$$= \frac{1}{H} \hat{\mathbf{S}}_B \left(\sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} \boldsymbol{K}_{(r)} \right) \hat{\mathbf{S}}_B$$

where $\boldsymbol{K}_{(r)} = \boldsymbol{\eta}_{(r)} \boldsymbol{\eta}'_{(r)}$, a $G \times G$ cross-product matrix of standard normal values; and $\hat{\mathbf{S}}_B^2 = \text{diag}(\hat{\mathbf{V}}_B)$. Because $E_\varepsilon(\boldsymbol{K}_{(r)}) = \mathbf{I}_G$, a G-dimension identity matrix, we have $E_\varepsilon[\text{var}_{\text{ECNJC}}(\hat{\mathbf{N}}_B)] = \text{diag}(\hat{\mathbf{V}}_B)$. Therefore, $\text{var}_{\text{ECNJC}}(\hat{\mathbf{N}}_B)$ does not reproduce $\hat{\mathbf{V}}_B$ in expectation.

Appendix B

Evaluation of the ECMV

For the following derivations, let E_B and Var_B represent the expectation and variance with respect to the benchmark survey sampling design. Also, let E_ε and Var_ε represent the expectation and variance with respect to the G-dimensional multivariate normal distribution, $\text{MVN}_G(\mathbf{0}, \hat{\mathbf{V}}_B)$. All other terms are defined in the body of the paper.

B.1: Derivation of $E[\text{var}_{\text{ECMV}}(\hat{\mathbf{N}}_B)]$ given in (15)

Using expression (14) and $c_h^2 = m_{Ah}/(m_{Ah} - 1)$,

$$E[\text{var}_{\text{ECMV}}(\hat{\mathbf{N}}_B)] = E_B \left[E_\varepsilon \left(\sum_{h=1}^{H} \frac{(m_{Ah}-1)}{m_{Ah}} \sum_{r=1}^{m_{Ah}} (\hat{\mathbf{N}}_{B(r)} - \hat{\mathbf{N}}_B)(\hat{\mathbf{N}}_{B(r)} - \hat{\mathbf{N}}_B)' \middle| B \right) \right],$$

$$= \frac{1}{H} E_B \left[\sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} E_\varepsilon(\hat{\boldsymbol{\varepsilon}}_{(r)} \hat{\boldsymbol{\varepsilon}}'_{(r)} | B) \right]$$

$$= \frac{1}{H} \sum_{h=1}^{H} \frac{1}{m_{Ah}} \sum_{r=1}^{m_{Ah}} E_B(\hat{\mathbf{V}}_B) = E_B(\hat{\mathbf{V}}_B).$$

B.2: Derivation of $\mathrm{Var}[\mathrm{var}_{\mathrm{ECMV}}(\hat{\mathbf{N}}_B)]$ given in (15)

When $y_k = 1$ so that $\hat{t}_{yP} = \mathbf{1}'\hat{\mathbf{N}}_B$, $\mathrm{var}_{\mathrm{ECMV}}(\mathbf{1}'\hat{\mathbf{N}}_B) = H^{-1}\sum_{h=1}^{H} m_{Ah}^{-1}\sum_{r=1}^{m_{Ah}} \mathbf{1}'\hat{\boldsymbol{\varepsilon}}_{(r)}\hat{\boldsymbol{\varepsilon}}'_{(r)}\mathbf{1}$. Using the formula for the variance of a quadratic form (Searle 1982, section 13.5), we have

$$\begin{aligned}
\mathrm{Var}[\mathrm{var}_{\mathrm{ECMV}}(\mathbf{1}'\hat{\mathbf{N}}_B)] \\
&= \mathrm{Var}_B\left[\frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}} E_{\varepsilon}(\mathbf{1}'\hat{\boldsymbol{\varepsilon}}_{(r)}\hat{\boldsymbol{\varepsilon}}'_{(r)}\mathbf{1}|B)\right] \\
&\quad + E_B\left[\frac{1}{H^2}\sum_{h=1}^{H}\frac{1}{m_{Ah}^2}\sum_{r=1}^{m_{Ah}} \mathrm{Var}_{\varepsilon}(\mathbf{1}'\hat{\boldsymbol{\varepsilon}}_{(r)}\hat{\boldsymbol{\varepsilon}}'_{(r)}\mathbf{1}|B)\right] \\
&= \mathrm{Var}_B\left[\frac{1}{H}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\sum_{r=1}^{m_{Ah}} \mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1}\right] \\
&\quad + E_B\left[\frac{1}{H^2}\sum_{h=1}^{H}\frac{1}{m_{Ah}}\{2tr(\mathbf{1}\mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1}\mathbf{1}'\hat{\mathbf{V}}_B)\}\right] \\
&= \mathrm{Var}_B[\mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1}] + \frac{2}{H\bar{m}_A^*}[E_B(\mathbf{1}'\hat{\mathbf{V}}_B\mathbf{1})^2],
\end{aligned}$$

where $\bar{m}_A^* = (H^{-1}\sum_{h=1}^{H} m_{Ah}^{-1})^{-1}$ is the harmonic mean of m_{Ah}.

References

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

Bray, R., Hourani, L., Rae, K., Dever, J., Brown, J., Vincus, A., Pemberton, M., Marsden, M., Faulkner, D. and Vandermaas-Peeler, R. (2003). 2002 Department of Defense Survey of Health Related Behaviors Among Military Personnel. Tech. Rep. RTI/7841/006-FR, U.S. Department of Defense prepared by RTI International. URL http://dodwws.rti.org/2002WWFinalReportComplete05-04.pdf.

Canty, A.J., and Davison, A.C. (1999). Resampling-based variance estimation for Labour Force Surveys. *The Statistician*, 48, 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, 17-26.

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

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, 88(423), 1013-1020.

Estevao, V.M., and 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., and Patak, Z. (2006). Raking ratio estimation: An application to the Canadian Retail Trade Survey. *Journal of Official Statistics*, 22(1), 71-80.

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

Jayasuriya, B.R., and Valliant, R. (1996). An application of regression and calibration estimation to post-stratification in a Household Survey. *Survey Methodology*, 22, 2, 127-137.

Keeter, S., Dimock, M. and Christian, L. (2008). Calling Cell Phones in '08 Pre-Election Polls. NEWS Release (December 18, 2008): Pew Research Center for the People & the Press. URL http://people-press.org/reports/pdf/cell-phone-commentary.pdf.

Killion, R.A. (2006). Weighting Specifications for The American Time Use Survey (ATUS) for 2006. U.S. Bureau of the Census, Internal Memo (Doc.#ATUS-16).

Kim, J.J., Li, J. and Valliant R. (2007). Cell collapsing in poststratification. *Survey Methodology*, 33, 2, 139-150.

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

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, 9(5), 1010-1019.

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

Lessler, J.T., and Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons, Inc.

Lumley, T. (2009). Survey: Analysis of complex survey samples. R package version 3.19. University of Washington: Seattle.

Mirza, H., and Hörngren, J. (2002). The Sampling and the Estimation Procedure in the Swedish Labour Force Survey. Technical report, Statistics Sweden, Stockholm: Sweden.

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

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.R-project.org.

Rao, J.N.K., and Shao, J. (1999). Modified balanced repeated replication for complex survey data. *Biometrika*, 86(2), 403-415.

Rao, J.N.K., and Wu, C.F.J. (1985). Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association*, 80(391), 620-630.

Research Triangle Institute (2008). *SUDAAN Language Manual*. Release 10.0, Research Triangle Park, NC: Research Triangle Institute.

Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 283-310.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. England: John Wiley & Sons, Inc.

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), 527-537.

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

SAS Institute Inc. (2009). *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Searle, S.R. (1982). *Matrix Algebra Useful for Statistics*. New York: John Wiley & Sons, Inc.

StataCorp (2010). *Stata Statistical Software: Release 11*. 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, 117-125.

Taylor, M.F., Brice, J., Buck, N. and Prentice-Lane, E. (2007). *British Household Panel Survey User Manual Volume A*: *Introduction*, *Technical Report and Appendices*. University of Essex, Colchester.

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.

Thěberge, A. (1999). Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94(446), 635-644.

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 Science+Business Media, LLC.

Wu, C.F.J. (1985). Variance estimation for the combined ratio and combined regression estimators. *Journal of the Royal Statistical Society*, Series B, 47(1), 147-154.

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(451), 903-915.