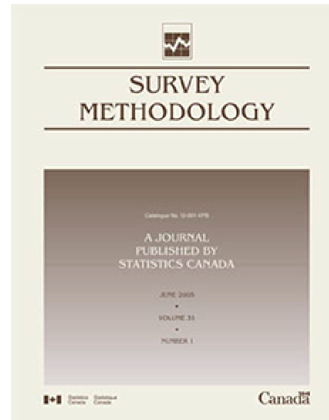
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology 41-1

One step or two? Calibration weighting from a complete list frame with nonresponse



by Phillip S. Kott and Dan Liao

Release date: June 29, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2015
Vol. 41, No. 1, pp. 165-181
Statistics Canada, Catalogue No. 12-001-X

One step or two? Calibration weighting from a complete list frame with nonresponse

Phillip S. Kott and Dan Liao[1]

Abstract

When a random sample drawn from a complete list frame suffers from unit nonresponse, calibration weighting to population totals can be used to remove nonresponse bias under either an assumed response (selection) or an assumed prediction (outcome) model. Calibration weighting in this way can not only provide double protection against nonresponse bias, it can also decrease variance. By employing a simple trick one can estimate the variance under the assumed prediction model and the mean squared error under the combination of an assumed response model and the probability-sampling mechanism simultaneously. Unfortunately, there is a practical limitation on what response model can be assumed when design weights are calibrated to population totals in a single step. In particular, the choice for the response function cannot always be logistic. That limitation does not hinder calibration weighting when performed in two steps: from the respondent sample to the full sample to remove the response bias and then from the full sample to the population to decrease variance. There are potential efficiency advantages from using the two-step approach as well even when the calibration variables employed in each step is a subset of the calibration variables in the single step. Simultaneous mean-squared-error estimation using linearization is possible, but more complicated than when calibrating in a single step.

Key Words: Probability sampling; Response model; Prediction model; Double protection; Simultaneous variance estimation.

1 Introduction

Survey sampling is a tool used primarily for estimating the parameters of a finite population based on a randomly drawn sample of its members. Probability samples come with design (sampling) weights, which are often the inverses of the individual member selection probabilities. As long as each population element has a positive selection probability, it is a simple matter to produce an estimator for the population total of a survey variable that is unbiased with respect to the probability-sampling mechanism. The ratio of two unbiased estimators of totals or any other smooth function of estimated totals, while not necessarily unbiased, is asymptotically unbiased and often consistent since its relative variance, like its relative bias, tends to zero as the sample size grows arbitrarily large.

Deville and Särndal (1992) introduced calibration weighting as a tool for adjusting design weights in such a way that the weighted sums of certain "calibration" variables equal their known (or better-estimated) population totals. As a consequence of these *calibration equations* holding, the standard error of an estimated total for a variable without a known population total is often reduced while remaining nearly (i.e., asymptotically) unbiased under the probability sampling mechanism.

Although originally developed to reduce standard errors, calibration weighting has also been used to remove selection biases resulting from unit nonresponse under certain assumptions (e.g., Folsom 1991; Fuller, Loughin and Baker 1994; Lundström and Särndal 1999; Folsom and Singh 2000). To this end, whether (or not) an element selected for the sample responds to a survey is treated as an additional phase of Poisson random sampling with unknown, but positive, selection probabilities. Calibration weighting estimates these Poisson selection probabilities implicitly and produces estimated totals that are nearly

1. Phillip S. Kott, Senior Research Statistician, RTI International, Rockville, Maryland 20852, U.S.A. E-mail: pkott@rti.org; Dan Liao, Research Statistician, RTI International, Rockville, Maryland 20852, U.S.A.

unbiased under the combined sample- and response-selection mechanisms, which is often called the "quasi-sampling design". See Oh and Scheuren (1983).

An important *caveat* is that although the sample-selection mechanism is fully under the control of the statistician, the response-selection mechanism is unknown. The response mechanism is assumed to have a particular form, and the failure of this assumption can result in biased estimators.

An alternative justification for calibration weighting involves a different type of modeling. It is easy to show that calibration weighting produces an estimator that is unbiased under a linear prediction (outcome) model if the expected value of the survey variable under the prediction model is a linear function of the calibration variables so long as the sampling and response mechanisms are ignorable, that is to say, the same prediction model applies whether or not the population element is sampled or whether it responds when sampled.

Unlike the selection model governing the response mechanism, it is possible for the linear prediction model to hold for one survey variable and not another. That is why most survey samplers prefer to assume a *selection model* when adjusting for unit nonresponse. Nevertheless, it is reassuring to know that if *either* model is correct, then the estimated total is nearly unbiased (i.e., has a relative bias that vanishes asymptotically), a property Kim and Park (2006) called "double protection" against nonresponse bias.

It is possible to simultaneously remove the selection bias and decrease standard error under the probability-sample mechanism in a single step by adjusting the design weights of unit respondents so that the estimated totals for a set of calibration variables equal their known population totals. Nevertheless, there are reasons for preferring the use of two calibration-weighting steps even when the sets of calibration variables used in both steps are the same or a subset of the calibration variables in a single step: the first step from the respondent sample to the original sample to remove selection bias and the second from the original sample to the population to decrease the variances of the resulting estimators.

Although Folsom and Singh (2000) and others have pointed out that calibration weighting can also be used to remove the selection bias due to under- or over-coverage of the sampling frame, we will direct our attention here on a single-stage sample drawn from a complete list frame without duplication. That is to say, we will assume that the sampling frame is identical to the target population (i.e., each population unit is listed on the frame).

The paper is structured as follows. Section 2 reviews some background theory on calibration weighting. Section 3 introduces a slightly new variance estimator that, like the variance estimator in Kott (2006), can be used to measure both the mean squared error of a calibration-weighted estimator under the quasi-sampling design and the variance under either the prediction model or the combination of the prediction model and original sampling mechanism, thus making the double protection against nonresponse bias arguably more useful for inference. The variance estimator in Kott applies only when calibrating to the population. Here we follow Folsom and Singh (2000) and allow the possibility that calibration is to the original sample.

Section 4 discusses the limitations of calibrating weighting in a single step and develops some theory for a two-step approach. Although our main purpose here is to argue the benefits of using two steps even when similar sets of calibration variables are employed in both steps, the calibration estimator we treat in this section is broader. Section 5 describes the results of some simulation experiments, while Section 6 offers a few concluding remarks.

2 One-step calibration weighting

2.1 Calibration weighting and unit nonresponse

In the absence of nonresponse (or frame errors), calibration weighting is a sampling-weight-adjustment method that creates a set of weights $\{w_k; k \in S\}$, asymptotically close to the original design weights, $d_k = 1/\pi_k$, that satisfy a set of *calibration equations* (one for each component of $\mathbf{z}_k$) :

$$\sum_S w_k \mathbf{z}_k = \sum_U \mathbf{z}_k,$$

where S denotes the sample, π_k the sample-selection probability of unit k, U the population of size $N, \mathbf{z}_k$ a vector with P components each having a known population total, and $\sum_A$ means $\sum_{k \in A}$.

Kott (2009) describes a conservative set of mild conditions under which $t_y = \sum_S w_k y_k$ is a nearly unbiased estimator for the population total $T_y = \sum_U y_k$ (i.e., the relative bias of t_y is asymptotically zero). Most importantly, each $\pi_k N/n$ is assumed to be bounded from below by a positive value as N and the (expected) sample size, n, grow arbitrarily large (we add the parenthetical "expected" in case the sample size is random).

In addition, the first four central population moments of each component of $\mathbf{z}_k$ is assumed to be bounded from above, while $N^{-1} \sum_U \mathbf{z}_k \mathbf{z}_k^T$ converges to a positive definite matrix.

Using calibration-weighting will tend to reduce mean squared error relative to the expansion estimator, $t_y^E = \sum_S d_k y_k$, when y_k is correlated with some components of $\mathbf{z}_k$. One should keep in mind, however, that most surveys have many y_k's.

A simple way to compute calibration weights is linearly with the following formula:

$$\begin{aligned} w_k &= d_k \left[1 + \left(\sum_U \mathbf{z}_j - \sum_S d_j \mathbf{z}_j\right)^T \left(\sum_S d_j \mathbf{z}_j \mathbf{z}_j^T\right)^{-1} \mathbf{z}_k\right] \\ &= d_k \left[1 + \mathbf{g}^T \mathbf{z}_k\right]. \end{aligned}$$

Fuller et al. (1994) and later Lundström and Särndal (1999) argued that this linear calibration can also be used to handle unit nonresponse. The sample S is replaced by the respondent sample R, while

$$\mathbf{g} = \left[(1-\theta)\left(\sum_U \mathbf{z}_j - \sum_R d_j \mathbf{z}_j\right)^T + \theta\left(\sum_S d_j \mathbf{z}_j - \sum_R d_j \mathbf{z}_j\right)^T\right]\left(\sum_R d_j \mathbf{z}_j \mathbf{z}_j^T\right)^{-1},$$

depending on whether the respondent sample is *calibrated to the population* $(\theta = 0)$ or *calibrated to the original sample* $(\theta = 1)$. Either way, the estimate is nearly unbiased under the quasi-sample-design that treats response as a second phase of random sampling so long as each unit's probability of response has the form:

$$p_k = 1/\left(1 + \boldsymbol{\gamma}^T \mathbf{z}_k\right), \tag{2.1}$$

and $\mathbf{g}$ is a consistent estimator for the unknown parameter vector $\boldsymbol{\gamma}$ in equation (2.1).

The problem with the response function in equation (2.1) is that the implicit estimator for p_k, $\hat{p}_k = 1/(1 + \mathbf{g}^T\mathbf{z}_k)$ can be negative. A nonlinear form of calibration weighting avoiding this possibility was suggested by Kott and Liao (2012) based on the generalized exponential form of Folsom and Singh (2000). It uses Newton's method (iterative Taylor-series approximations) to find a $\mathbf{g}$ such that the calibration equation (from here on, we refer to the vector of component calibration equations as the calibration equation):

$$\sum_R w_k \mathbf{z}_k = \sum_R d_k \alpha(\mathbf{g}^T\mathbf{z}_k)\mathbf{z}_k = (1-\theta)\sum_U \mathbf{z}_k + \theta \sum_S d_k \mathbf{z}_k \tag{2.2}$$

holds, where $\theta = 0$ or 1,

$$\alpha(\mathbf{g}^T\mathbf{z}_k) = \frac{\ell + \exp(\mathbf{g}^T\mathbf{z}_k)}{1 + \exp(\mathbf{g}^T\mathbf{z}_k)/u}, \tag{2.3}$$

ℓ, the lower bound of $\alpha(\cdot)$, is nonnegative (so that calibration weights are likewise nonnegative), and the upper bound of $\alpha(\cdot), u > \ell$, can be either finite or infinite.

Although there are other reasonable forms the *weight-adjustment function* $\alpha(\mathbf{g}^T\mathbf{z}_k)$ can take, we will restrict our attention to functions in the form in equation (2.3). This is a generalization of both raking where $\ell = 0, u = \infty$, and the implicit estimation of a logistic response model, where $\ell = 1, u = \infty$. In Deming and Stephan's original (1940) iterative-proportional-fitting algorithm for raking, the components of $\mathbf{z}_k$ were restricted to indicator functions. We use "raking" more broadly here to mean calibration weighting with a weight-adjustment function of the form $\alpha(\mathbf{g}^T\mathbf{z}_k) = \exp(\mathbf{g}^T\mathbf{z}_k)$.

When $\ell < 1$, equation (2.3) becomes the generalized-raking adjustment introduced in Deville and Särndal (1992) and discussed further in Deville, Särndal and Sautory (1993). Generalized raking not only lets the components of $\mathbf{z}_k$ be continuous but also allows the range of the $\alpha(\mathbf{g}^T\mathbf{z}_k)$ to be constrained between a positive ℓ and a (possibly) finite u.

Deville and Särndal (1992) required $\alpha(0) = \alpha'(0) = 1$. Since the authors were not treating samples with nonresponse (or incorrect frames), $\mathbf{g}^T\mathbf{z}_k$ needed to converge to 0 and $\alpha(\mathbf{g}^T\mathbf{z}_k)$ to 1 as the (expected) sample size grew arbitrarily large. When adjusting design weights for nonresponse, however, setting $\ell \geq 1$ is a more sensible strategy, so that the implicit estimated probability of response does not exceed 1.

Although the original definition of calibration weighting in Deville and Särndal (1992) involved minimizing the differences between the w_k and d_k in R as measured by some loss function, later formulations (e.g., Estevao and Särndal 2000) removed the loss function from the definition. Forcing w_k and d_k to be close makes little sense when calibration weighting is used to adjust for unit nonresponse since if a sampled k has a relatively small probability of response, then the difference between w_k and d_k *should* be relatively large.

Rather than assuming a response model with a particular functional form, an alternative justification for using calibration weighting as a mean of removing unit-nonresponse bias assumes a prediction model in which the survey variable y_k is itself a random variable such that $\mathrm{E}(y_k | \mathbf{z}_k) = \mathbf{z}_k^T\boldsymbol{\beta}$ for some unknown $\boldsymbol{\beta}$ whether or not k is sampled or whether it responds when sampled. Kott (2006) and others have observed the calibration-weighted estimator for $T_y = \sum_U y_k$ will be nearly unbiased under the prediction model when calibration is done to the population (when $\theta = 0$ in equation (2.2)) and under the combination of the prediction model and the original sample-selection mechanism when calibration is done to the original sample (when $\theta = 1$).

The property that a calibration-weighted estimator is nearly unbiased in some sense when *either* an assumed response model *or* an assumed prediction model holds has been called “double protection against nonresponse bias” by Kim and Park (2006). It is known as “double robustness” in the biostatics literature (Bang and Robins 2005) and attributed to Robins, Rotnitzky and Zhao (1994), which dealt with item rather than unit nonresponse.

The distribution of $y_k | \mathbf{z}_k$ under the prediction model is often assumed to be the same for sampled and nonsampled population members. That is to say, the sampling mechanism is assumed to be *ignorable*. In addition, the distribution of $y_k | \mathbf{z}_k$ is often assumed to be the same whether or not a population member responds when sampled, that is, that the response mechanism is also assumed to be ignorable (Little and Rubin 2002). Here, we make weaker analogous assumptions under the prediction model, namely, that $\mathrm{E}(y_k | \mathbf{z}_k)$ does not depend on whether k is sampled or when sampled responds. Let us say that the sampling and response mechanisms are assumed to be “first-moment ignorable”.

2.2 Instrumental variables

Deville (2000) observed that instrumental-variable calibration can be used to adjust for potential nonresponse bias by assuming a response model that depended on $\mathbf{x}_k$,

$$p_k = \left[\alpha\left(\boldsymbol{\gamma}^T \mathbf{x}_k\right)\right]^{-1} = \frac{1 + \exp\left(\boldsymbol{\gamma}^T \mathbf{x}_k\right)/u}{\ell + \exp\left(\boldsymbol{\gamma}^T \mathbf{x}_k\right)}, \tag{2.4}$$

but fitting calibration equations with $\mathbf{z}_k$:

$$\sum\nolimits_R w_k \mathbf{z}_k = \sum\nolimits_R d_k \alpha\left(\mathbf{g}^T \mathbf{x}_k\right) \mathbf{z}_k = (1 - \theta) \sum\nolimits_U \mathbf{z}_k + \theta \sum\nolimits_S d_k \mathbf{z}_k, \tag{2.5}$$

where the $\mathbf{g}$ satisfying equation (2.5) with $\theta = 0$ or 1 a consistent estimator of unknown parameter vector $\boldsymbol{\gamma}$ in equation (2.4). Some mild conditions are needed for this. Sufficient are the following: $N^{-1} \sum_R d_k \alpha\left(\boldsymbol{\gamma}^T \mathbf{x}_k\right) \mathbf{z}_k$ is a consistent and bounded estimator for $N^{-1}\left[(1 - \theta) \sum_U \mathbf{z}_k + \theta \sum_S d_k \mathbf{z}_k\right]$, $\alpha(\phi)$ is everywhere twice differentiable, and $N^{-1} \sum_R d_k \alpha'(\phi) \mathbf{z}_k \mathbf{x}_k^T$ is always invertible and bounded as the sample grows arbitrarily large.

Let $R_k = 1$ when $k \in R, 0$ otherwise. It is not hard to show that

$$\begin{aligned} \mathbf{g} - \boldsymbol{\gamma} &= -\left(\sum\nolimits_S d_k R_k \alpha'(c_k) \mathbf{z}_k \mathbf{x}_k^T\right)^{-1} \left\{\sum\nolimits_S d_k R_k \alpha\left(\boldsymbol{\gamma}^T \mathbf{x}_k\right) \mathbf{z}_k - \left[(1-\theta) \sum\nolimits_U \mathbf{z}_k + \theta \sum\nolimits_S d_k \mathbf{z}_k\right]\right\} \\ &\quad -\left(N^{-1} \sum\nolimits_S d_k R_k \alpha'(c_k) \mathbf{z}_k \mathbf{x}_k^T\right)^{-1} \left\{N^{-1} \sum\nolimits_S d_k R_k \alpha\left(\boldsymbol{\gamma}^T \mathbf{x}_k\right) \mathbf{z}_k - N^{-1}\left[(1-\theta) \sum\nolimits_U \mathbf{z}_k + \theta \sum\nolimits_S d_k \mathbf{z}_k\right]\right\} \end{aligned}$$

for some c_k between $\mathbf{g}^T \mathbf{x}_k$ and $\boldsymbol{\gamma}^T \mathbf{x}_k$, as Kott and Liao (2012) demonstrated when $\mathbf{x}_k = \mathbf{z}_k$.

Deville also noted that it is possible for components of the $\mathbf{x}_k$ to be survey variables with values known only for respondents. Chang and Kott (2008) extended the notion of calibration weighting to allow the dimension of the $\mathbf{z}_k$-vector to be greater than that of the $\mathbf{x}_k$-vector. We will *not* treat either possibility in the following sections.

Kim and Shao (2013) in treating nonignorable nonresponse call the components of $\mathbf{z}_k$ not wholly functions of the components of $\mathbf{x}_k$ "instrumental variables". To limit future confusion, we will henceforth use to term "model variables" to refer to the components of $\mathbf{x}_k$.

3 Variance estimation for the one-step calibration estimator

In this section, we let

$$t_y = \sum\nolimits_R w_k y_k = \sum\nolimits_R d_k \alpha\left(\mathbf{g}^T \mathbf{x}_k\right) y_k$$

be the calibration-weighted estimator for T_y, where $w_k = d_k \alpha\left(\mathbf{g}^T \mathbf{x}_k\right)$ when $k \in R$ is the calibration weight, and w_k is conveniently defined to be 0 when $k \notin R$. The weight-adjustment function $\alpha(\cdot)$ is defined implicitly by equation (2.4), and $\mathbf{g}$ is again chosen so that the calibration equation (2.5) holds for either $\theta = 0$ or 1.

We propose the following estimator for the variance t_y :

$$v\left(t_y\right) = \sum_{k,\, j \in S} \left(1 - \frac{\pi_k \pi_j}{\pi_{kj}}\right) \left[d_k\left(\theta \mathbf{z}_k^T \mathbf{b} + \alpha_k e_k\right)\right]\left[d_j\left(\theta \mathbf{z}_j^T \mathbf{b} + \alpha_j e_j\right)\right] + \sum_{k \in R} d_k\left(\alpha_k^2 - \alpha_k\right) e_k^2, \quad (3.1)$$

where π_{kj} is the joint selection probability of k and j under the original sampling design, $\pi_{kk} = \pi_k = 1/d_k, \pi_k = \alpha\left(\mathbf{g}^T \mathbf{x}_k\right)$ when $k \in R$ and 0 otherwise,

$$\mathbf{b} = \left[\sum\nolimits_R d_k \alpha'\left(\mathbf{g}^T \mathbf{x}_k\right) \mathbf{x}_k \mathbf{z}_k^T\right]^{-1} \sum\nolimits_R d_k \alpha'\left(\mathbf{g}^T \mathbf{x}_k\right) \mathbf{x}_k y_k, \quad (3.2)$$

and $e_k = y_k - \mathbf{z}_k^T \mathbf{b}$. We will show that $v\left(t_y\right)$ in equation (3.1) can be nearly unbiased in some sense if *either* a response model (Section 3.1) *or* prediction model holds (Section 3.2).

The variance estimator in equation (5.2) of Kott (2006) is identical to $v\left(t_y\right)$ in equation (3.1) when $\theta = 0$. The variance estimator in Kim and Haziza (2014) is also similar. Their prediction model is more general than the linear prediction model considered here.

This variance estimator $v\left(t_y\right)$ presupposes that the original sampling design is such that each element can only be drawn once. In Section 3.1, we see that when the probabilities of response are independent (Poisson), then under mild assumptions, $v\left(t_y\right)$ is a nearly unbiased estimator of the mean squared error of t_y under the quasi-sampling design whether or not the prediction model, $\mathrm{E}\left(y_k \mid \mathbf{x}_k, \mathbf{z}_k\right) = \mathbf{z}_k^T \boldsymbol{\beta},$ holds.

In Section 3.2, $v(t_y)$ is shown to be a nearly unbiased estimator for the combined prediction-model and original-sampling-design variance of t_y as an estimator for T_y whether or not the response model in equation (2.4) holds. Thus, $v(t_y)$ can be called a "simultaneous variance estimator".

3.1 Variance estimation under the response model

For ease of exposition we will assume that the response model in equation (2.4) with a finite u holds. Sufficient conditions for $v(t_y)$ to be a nearly unbiased estimator for the mean squared error of t_y (by which the bias converges to 0 as the sample size grows arbitrary large) are

$$\pi_{kj} \geq B_0 > 0 \tag{3.3}$$

$$\sum_{j=1}^{N} \left| \frac{\pi_{kj}}{\pi_k \pi_j} - 1 \right| \leq B_1 < \infty \text{ for every } k, \tag{3.4}$$

$$\frac{\sum_{j=1}^{N} \psi_j^r}{N} \leq B_2 < \infty \text{ where } \psi_j \text{ is } y_j \text{ or any component of } \mathbf{x}_j \text{ or } \mathbf{z}_j, \text{while } r = 1 \text{ or } 2, \tag{3.5}$$

and $N^{-1} \sum_R d_k \alpha' (\mathbf{g}^T \mathbf{x}_k) \mathbf{z}_k \mathbf{x}_k^T$ is of full rank and is bounded in probability as the sample size grows arbitrarily large.

From these, $\alpha'(\phi) = (1 - \alpha(\phi)/u) \exp(\phi)/[(1 + \exp(\phi)/u)]$ being bounded when u is finite, and the Cauchy-Schwarz inequality $((\sum a_k b_k)^2 \leq \sum a_k^2 \sum b_k^2)$, it is not hard to see not only that $\mathbf{g}$ is a consistent estimator for $\boldsymbol{\gamma}$, but also that $\mathbf{b}$ in equation (3.2) (which can be rendered $\mathbf{b} = [N^{-1} \sum_R d_k \alpha' (\mathbf{g}^T \mathbf{x}_k) \mathbf{x}_k \mathbf{z}_k^T]^{-1} N^{-1} \sum_R d_k \alpha' (\mathbf{g}^T \mathbf{x}_k) \mathbf{x}_k y_k)$ has a probability limit, call it $\mathbf{b}^*$, whether or not the prediction model holds. Moreover, both $\mathbf{b} - \mathbf{b}^*$ and $\mathbf{g} - \boldsymbol{\gamma}$ are $\mathbf{O}_p(1/\sqrt{n})$.

Observe that

$$\begin{aligned} (t_y - T_y)/N &= \theta\left(\sum_S d_k \mathbf{z}_k^T \mathbf{b}^* - \sum_U \mathbf{z}_k^T \mathbf{b}^*\right)/N \\ &+ \left[\sum_R d_k \alpha(\mathbf{g}^T \mathbf{x}_k) e_k^* - \sum_R d_k \alpha(\boldsymbol{\gamma}^T \mathbf{x}_k) e_k^*\right]/N \\ &+ \left[\sum_R d_k \alpha(\boldsymbol{\gamma}^T \mathbf{x}_k) e_k^* - \sum_U e_k^*\right]/N, \end{aligned}$$

where $e_k^* = y_k - \mathbf{z}_k^T \mathbf{b}^*$. The insertion of the $\alpha'(\cdot)$ into the "regression coefficient" $\mathbf{b}$ allows us to ignore the contribution to quasi-design mean squared error of the second term in this sum, $Q = \sum_R d_k [\alpha(\mathbf{g}^T \mathbf{x}_k) - \alpha(\boldsymbol{\gamma}^T \mathbf{x}_k)] e_k^*/N$. That is because $\sum_R d_k \alpha' (\boldsymbol{\gamma}^T \mathbf{x}_k) \mathbf{x}_k e_k = 0$ is true by definition, which implies $\sum_R d_k \alpha' (\boldsymbol{\gamma}^T \mathbf{x}_k) \mathbf{x}_k e_k^*$ is $\mathrm{O}_p(1/\sqrt{n})$ under our assumptions. Moreover, since $\alpha(\mathbf{g}^T \mathbf{x}_k) - \alpha(\boldsymbol{\gamma}^T \mathbf{x}_k) = \alpha'(c_k)(\mathbf{g} - \boldsymbol{\gamma})^T \mathbf{x}_k$ is also $\mathrm{O}_p(1/\sqrt{n})$, $Q = (\mathbf{g} - \boldsymbol{\gamma})^T \sum_R d_k \alpha'(c_k) \mathbf{x}_k e_k^*$ is $\mathrm{O}_p(1/n)$, which is asymptotically ignorable relative to the two $\mathrm{O}_p(1/\sqrt{n})$ components of $(t_y - T_y)/N$.

With the contribution of Q eliminated from consideration, an idealized, but not calculable, nearly unbiased estimator for the quasi-design mean squared error of t_y is

$$v_{I1}(t_y) = \sum_{k,j \in S} \left(1 - \frac{\pi_k \pi_j}{\pi_{kj}}\right) \left[d_k \left(\theta \mathbf{z}_k^T \mathbf{b}^* + e_k^*\right)\right] \left[d_j \left(\theta \mathbf{z}_j^T \mathbf{b}^* + e_j^*\right)\right] + \sum_{k \in R} \left(\frac{d_k e_k^*}{p_k}\right)^2 (1 - p_k), \quad (3.6)$$

where the first term on the right estimates the mean squared error before nonresponse (if any) and the second the added variance due to nonresponse.

An alternative nearly unbiased idealized mean squared error estimator, closer to being calculable, is

$$v_{I2}(t_y) = \sum_{k,j \in S} \left(1 - \frac{\pi_k \pi_j}{\pi_{kj}}\right) \left[d_k \left(\theta \mathbf{z}_k^T \mathbf{b}^* + \frac{R_k}{p_k} e_k^*\right)\right] \left[d_j \left(\theta \mathbf{z}_j^T \mathbf{b}^* + \frac{R_j}{p_j} e_j^*\right)\right] + \sum_{k \in R} d_k \left(\frac{e_k^*}{p_k}\right)^2 (1 - p_k), \quad (3.7)$$

where again $R_k = 1$ when $k \in R, 0$ otherwise. Since the $(R_k / p_k) e_k^*$ are independent under the response model with mean e_k^* and variance $\left(e_k^* / p_k\right)^2 p_k (1 - p_k)$, $\mathrm{E}\left[(R_k / p_k) e_k^* (R_j / p_j) e_j^*\right] = e_k^* e_j^*$ when $k \neq j$. By contrast, the following holds when $k = j$:

$$(1 - \pi_k) E\left[\left(d_k \frac{R_k}{p_k} e_k^*\right)^2\right] = (1 - \pi_k)\left[\left(d_k e_k^*\right)^2 + \left(\frac{d_k e_k^*}{p_k}\right)^2 p_k (1 - p_k)\right]$$

$$= (1 - \pi_k)\left(d_k e_k^*\right)^2 + \left(\frac{d_k e_k^*}{p_k}\right)^2 p_k (1 - p_k) - d_k \left(\frac{e_k^*}{p_k}\right)^2 p_k (1 - p_k).$$

The first summation on the right-hand side of equation (3.7) has terms where $k \neq j$ and terms where $k = j$, the latter of which causes the second summation in (3.7) to differ from the second summation on the right-hand side of equation (3.6). Note that the expectation under the response model of $\sum_R d_k \left(e_k^* / p_k\right)^2 (1 - p_k)$ in the second summation on the right-hand side of (3.7) is $\sum_S d_k \left(e_k^* / p_k\right)^2 p_k (1 - p_k)$.

Finally, $v_{I2}(t_y)$ can be replaced by the asymptotically identical, but computable, $v(t_y)$ in equation (3.1) since $\sum_{j \in S} \left(1 - \pi_k \pi_j / \pi_{kj}\right)$ is bounded for all k under assumptions (3.3) and (3.4), allowing e_k and α_k to be substituted for the unknown e_k^* and $1/p_k$, respectively (because $e_k^* - e_k$ and $\alpha_k - 1/p_k$ are $\mathrm{O}_p\left(1/\sqrt{n}\right)$ for all k).

3.2 Variance estimation under the prediction model

Matters are a bit simpler when we assume a prediction model holds but not necessarily the response model in equation (2.4). Suppose $\mathrm{E}\left(y_k \mid \mathbf{x}_k, \mathbf{z}_k\right) = \mathbf{z}_k^T \boldsymbol{\beta}$, whether or not k is sampled or responds when sampled, and the $\varepsilon_k = y_k - \mathbf{z}_k^T \boldsymbol{\beta}$ are uncorrelated random variables with variances equal to $\sigma_k^2 = \mathbf{z}_k^T \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ need not be specified other than having finite components.

The mean squared error of t_y as an estimator for T_y under that prediction model is the sum of the prediction variance of t_y as an estimator for $T_y, \sum_R (w_k^2 - w_k)\sigma_k^2$ (see, for example, Kott 2009, page 69), and the squared bias, $(\sum_S \mathbf{x}_k^T \boldsymbol{\beta} - \sum_U \mathbf{x}_k^T \boldsymbol{\beta})^2$, the latter being zero when $\theta = 0$. The combined variance of t_y as an estimator for T_y under the prediction model and original sample design is

$$V_C = \theta \mathrm{Var}_D \left(\sum_S \mathbf{x}_k^T \boldsymbol{\beta}\right) + \mathrm{E}_D \left[\sum_S (w_k^2 - w_k)\sigma_k^2\right],$$

where the subscript D denotes that the operation (variance or expectation) is with respect to the original sampling design. Recall $w_k = 0$ for $k \neq R$.

To see that $v(t_y)$ in equation (3.1) provides a nearly unbiased estimator for V_C, observe first that

$$e_k = y_k - \mathbf{z}_k^T \mathbf{b} = \varepsilon_k - \mathbf{z}_k^T \left[N^{-1} \sum_R d_j \alpha'(\mathbf{g}^T \mathbf{x}_j) \mathbf{x}_j \mathbf{z}_j^T\right]^{-1} N^{-1} \sum_R d_j \alpha'(\mathbf{g}^T \mathbf{x}_j) \mathbf{x}_j \varepsilon_j.$$

Let $\delta_{kj} = 1$ when $k = j$ and 0 otherwise. Because the ε_k are uncorrelated, and $\mathrm{E}(\varepsilon_k^2) = \sigma_k = \mathbf{z}_k^T \boldsymbol{\eta}$, it is now not hard to show that $\mathrm{E}(e_k e_j) = \delta_{kj}\sigma_k^2 + \mathrm{O}(1/n)$ for almost every k, j pair under the prediction model when $N^{-1} \sum_R d_k \alpha'(\mathbf{g}^T \mathbf{x}_k) \mathbf{z}_k \mathbf{x}_k^T$ converges to an invertible matrix, and assumptions (3.3), (3.4), and

$$\frac{\sum_{j=1}^{N} \psi_j^r}{N} \leq B_2 < \infty \text{ where } \psi_j \text{ is any component of } \mathbf{x}_j \text{ or } \mathbf{z}_j, \text{ and } r = 1, 2, 3, \text{ or } 4, \tag{3.8}$$

hold. Observe that the change from the assumptions in (3.5) to (3.8) makes the relative bias of $v(t_y)$ as an estimator for V_C (or $\sum_R (w_k^2 - w_k)\sigma_k^2$ when $\theta = 0$) $\mathrm{O}(1/n)$ rather than $\mathrm{O}(1/\sqrt{n})$.

4 Two-step calibration weighting

4.1 Calibration weighting in two steps

In practice, the components of $\mathbf{x}_k$ are often 0/1 group-membership identifiers, and the groups are mutually exclusive and exhaustive. In that situation, $\mathbf{g}^T \mathbf{x}_k$ can only take on P values. *Almost* any weight-adjustment function, $\alpha(\mathbf{g}^T \mathbf{x}_k)$, will yield equivalent results. An example is the linear function, $\alpha(\mathbf{g}^T \mathbf{x}_k) = 1 + \mathbf{g}^T \mathbf{x}_k$, of Lundström and Särndal (1999).

One popular weight-adjustment function that sometimes *cannot* be used (note the italicized "almost" in the previous paragraph) is $\alpha(\mathbf{g}^T \mathbf{x}_k) = 1 + \exp(\mathbf{g}^T \mathbf{x}_k)$, which assumes response is a logistic function of $\mathbf{x}_k$. The problem is that this weight-adjustment function cannot return values less than unity. We noted in the previous section, that sometimes one may need α_k to be less than 1. A routine that tries to use $\alpha(\mathbf{g}^T \mathbf{x}_k) = 1 + \exp(\mathbf{g}^T \mathbf{x}_k)$ and fit the calibration equations will fail.

This can be a particular problem when assuming a logistic response model and trying to calibrate to the population in a single step. There may be a component of $\mathbf{z}_k$, say z_{ka}, that is always nonnegative, but the original sample and response set are such that $\sum_R d_k z_{ka} > \sum_U z_{ka}$ even though $\sum_R d_k z_{ka}$ cannot exceed $\sum_S d_k z_{ka}$. Thus, calibrating to the population will always fail because no α_k can be less than 1.

Calibrating to the original sample, by contrast, need not fail, since $\sum_R d_k z_{ka} \le \sum_S d_k z_{ka}$. This suggest that one calibrates first to the original sample, which removes the response bias if the assumed response model holds, and then to the population, which removes the response bias if the prediction model holds. Estevao and Särndal (2002) discuss a variety of ways to calibrate in steps, but we focus on a single method here.

A second advantage of calibration weighting in two steps can be realized even when the calibration variables used in both steps are the same or a subset of those used in the single step. This happens when the response model holds, and the linear prediction model is only roughly true. Some version or "optimal" estimation can then be used in the second calibration-weighting step to increase efficiency. Rao (1994) introduced the notion of the optimal regression estimator. It was put into calibration-weighting form and discussed further in Bankier (2002) and Kott (2009, Section 4.2). Detail and how this can be done are provided in Sections 4.2 and 5.

4.2 Estimation and variance estimation when calibrating in two steps

In this subsection, we start with a fairly general two-step calibration estimator for a total and then address estimating its variance. The first calibration-weighting step, which is to the original sample, employs $\mathbf{x}_{1k}$ as the vector of response-model variables and $\mathbf{z}_{1k}$ as the calibration vector. Each has P_1 components. The weight-adjustment function has the form described in equation (2.4) with $\mathbf{g}_1$ now replacing $\mathbf{g}$. The calibration equation is $\sum_R d_k \alpha\left(\mathbf{g}_1^T \mathbf{x}_{1k}\right) \mathbf{z}_{1k} = \sum_S d_k \mathbf{z}_{1k}$.

The second calibration-weighting step, which is to the population, employs $\mathbf{x}_{2k}$ and $\mathbf{z}_{2k}$, each with P_2 components. The nonresponse bias under the response model is removed in the first step. For the weight-adjustment function for the second step, we propose using

$$h_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right) = \frac{\ell_k + \exp\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right)}{1 + \exp(\mathbf{g}_2^T \mathbf{x}_{2k})/u_k}, \tag{4.1}$$

where $u_k > \ell_k > 0$ may be set almost at whim (but see below). The right-hand side of equation (4.1) can vary across the k (and so can depend on d_k and α_k), yet $h_k(0) = h'_k(0) = 1$, making it asymptotically indistinguishable from the linear function: $1 + \mathbf{g}_2^T \mathbf{x}_{2k}$. For simplicity, we will call $h_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right)$ and $h'_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right), h_k$ and h'_k respectively. From a quasi-sampling-design viewpoint, both are asymptotically identical to unity. The second calibration equation is $\sum_S d_k h_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right) \mathbf{z}_{2k} = \sum_U \mathbf{z}_{2k}$. Because this equation must hold, there are limits on the available choices for u_k and ℓ_k in equation (4.1).

A good simultaneous variance estimator for $t_y = \sum_R w_k y_k = \sum_R d_k \alpha\left(\mathbf{g}_1^T \mathbf{x}_{1k}\right) h_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right) y_k$ is (as we shall see)

$$v\left(t_y\right) = \sum_{k,j \in S}\left(1 - \frac{\pi_k \pi_j}{\pi_{kj}}\right)\left[d_k\left(\mathbf{z}_{1k}^T \mathbf{b}_1 + \alpha_k h_k e_{1k}\right)\right]\left[d_j\left(\mathbf{z}_{1j}^T \mathbf{b}_1 + \alpha_j h_j e_{1j}\right)\right] + \sum_{k \in R} d_k\left(h_k^2 \alpha_k^2 - h_k \alpha_k\right) e_{1k}^2, \tag{4.2}$$

where

$$e_{2k} = y_k - \mathbf{z}_{2k}^T\left(\sum_S d_j \alpha_j h_j' \mathbf{x}_{2j} \mathbf{z}_{2j}^T\right)^{-1} \sum_S d_j \alpha_j h_j' \mathbf{x}_{2j} y_j, \tag{4.3}$$

$$\mathbf{b}_1 = \left(\sum_S d_f \alpha_f' \mathbf{x}_{1f} \mathbf{z}_{1f}^T\right)^{-1} \sum_S d_f \alpha_f' h_f \mathbf{x}_{1f} e_{2f}, \tag{4.4}$$

and

$$e_{1k} = e_{2k} - \mathbf{x}_{1k}^T \mathbf{b}_1. \tag{4.5}$$

Let $\mathbf{x}_k$ now be the vector composed of the non-duplicated components of $\mathbf{x}_{1k}$ and $\mathbf{x}_{2k}$ and define $\mathbf{z}_k$ analogously. Sufficient conditions for (4.2) to be a simultaneous variance estimator include the corresponding components of equation (4.1) depending on whether either the response model in equation (2.4) holds with $\mathbf{x}_{1k}$ replacing $\mathbf{x}_k$ or the prediction model is $\mathrm{E}\left(y_k \mid \mathbf{x}_k, \mathbf{z}_k\right) = \mathbf{z}_{2k}^T \boldsymbol{\beta}_2$, whether or not k is sampled or responds if sampled, and the $\varepsilon_{2k} = y_k - \mathbf{z}_{2k}^T \boldsymbol{\beta}_2$ are uncorrelated random variables with variances equal to $\sigma_{2k}^2 = \mathbf{z}_{2k}^T \boldsymbol{\eta}_2$, where $\boldsymbol{\eta}_2$ need not be specified other than having finite components. Now, both $N^{-1} \sum_R d_k \alpha'\left(\mathbf{g}_1^T \mathbf{x}_{1k}\right) \mathbf{z}_{1k} \mathbf{x}_{1k}^T$ and $N^{-1} \sum_R d_k h_k'\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right) \mathbf{z}_{2k} \mathbf{x}_{2k}^T$ are assumed to be of full rank and bounded as the sample size grows arbitrarily large.

The variance estimator in equation (4.2) is almost the same as the estimator in (3.1): $\mathbf{x}_k$ has been replaced with $\mathbf{x}_{1k}$ and $\mathbf{z}_k$ with $\mathbf{z}_{1k}$, while $h_k e_{2k}$ substitutes for y_k (we will get to a small difference shortly). Observe that e_{2k} is effectively an expression of the "residual" from the second calibration-weighting step. This residual is multiplied by the weight-adjustment factor h_k, which is asymptotically unity from the quasi-sampling-design-based perspective and a constant from the prediction-model viewpoint. The product is then used to create the first-step "regression-coefficient" $\mathbf{b}_1$ in equation (4.4) and its accompanying "residual" e_{1k} in equation (4.5). We do the second step regression first because $t_y - T_y = \sum_R w_k y_k - \sum_U y_k = \sum_R w_k e_{2k} - \sum_U e_{2k}$.

It is for estimating the prediction model of t_y as an estimator of $T_y, \sum_S\left(w_k^2 - w_k\right) \sigma_{2k}^2$, that the last appearance of h_k on the right-hand side of equation (4.2) is not squared, as it would be if $h_k e_{2k}$ substituted for y_k everywhere. From a quasi-design viewpoint, h_k is asymptotically identical to unity, so whether or not it is squared makes no asymptotic difference.

Observe that the h_j' have been inserted in equation (4.3) for the same reason as α' was inserted into $\mathbf{b}$ in equation (3.1). Since the h_j' are asymptotically unity, however, they are not really needed (and serve no function whatever from a prediction-model viewpoint). A similar argument applies to the h_f in equation (4.4): they are asymptotically unity from the quasi-sampling-design viewpoint (and part of an estimate of 0 from a prediction-model viewpoint).

5 Some simulations

Paralleling Kott and Liao (2012), we generated a synthetic population, U, of hospitals from the 2008 DAWN public-use file. After creating U, we independently drew 3,600 stratified simple random samples of size 400 from U using the strata definitions on the public-use file. These definitions incorporate information on location and hospital ownership (public or private) not directly provided on the file.

We set the stratum sample sizes roughly proportional to a size measure q_k, but never less than four. For q_k we used annual drug-related emergency-room visits, which was always positive. The DAWN actually has a size variable attached to every hospital in the frame: total emergency-room visits in a previous year according to the American Hospital Association. Unfortunately, it was not included on the public-use file. Design weights in our simulations varied between 4.375 and 48, which allowed us to treat the finite population correction factors as ignorable in variance estimation.

As in our original paper, we generated a respondent sample R for each simulated sample based on Bernoulli draw from the logistic function:

$$p_k = (1 + \exp(3.735 - 0.4 \log(q_k)))^{-1}, \tag{5.1}$$

We also created alternative respondent samples using

$$p_k = \left(1 + \exp\left(0.597 - 0.005 q_k^{1/2}\right)\right)^{-1}. \tag{5.2}$$

Both response models produce unweighted overall response rates of around 54%, which is similar to actual DAWN experience, where response is also a mildly increasing function of the size variable. Notice that $\alpha_k = 1/p_k$ is bounded even if neither probability can be expressed by equation (2.4) with a finite u.

As in the previous study, we focused on estimating population totals for three survey variables. Annual drug-related emergency-room visits with adverse pharmaceutical reaction and those resulting in deaths came from the public-use file. Since both these variables were roughly linear in our size measure, the third "survey" variable was artificially constructed. It was the size measure (annual drug-related emergency-room visits) raised to the 1.3 power.

We investigated eight estimators and estimates of their variance. These are summarized in Table 5.1. The first two featured calibration to the original sample only (equation (2.5) with $\theta = 1$), with response assumed to be logistic in the log of the size measure. That is to say, equation (2.3) was employed with $\mathbf{x}_k = (1 \;\log(q_k))^T$. The first estimator used $\mathbf{z}_k = (1 \;\log(q_k))^T$ as the calibration vector while the second used $\mathbf{z}_k = (1 \; q_k)^T$, which was more consistent with a reasonable prediction model, at least for adverse reactions and deaths.

Our third and fourth estimator featured calibration to the sample and population in a single step (equation (2.5) with $\theta = 1$ and then $\theta = 0$) using $\mathbf{x}_k = \mathbf{z}_k = (1 \;\log(q_k)\; q_k)^T$. They were designed to be nearly unbiased if either the logistic response model in $(1 \;\log(q_k))^T$ or the linear prediction model in $(1 \; q_k)^T$ held.

Table 5.1
Summary of simulation exercise (all results in percentages %)

Estimator	t_{y1}	t_{y2}	t_{y3}	t_{y4}	t_{y5}	t_{y6}	t_{y7}	t_{y8}
Calibration to Sample								
response-model variables: $\mathbf{x}_{1k}$	$(1\ \log(q_k))^T$	$(1\ \log(q_k))^T$	$(1\ \log(q_k)\,q_k)^T$	-	$(1\ \log(q_k))^T$	$(1\ \log(q_k))^T$	$(1\ \log(q_k))^T$	$(1\ \log(q_k))^T$
calibration variables: $\mathbf{z}_{1k}$	$(1\ \log(q_k))^T$	$(1\ q_k)^T$	$(1\ \log(q_k)\,q_k)^T$	-	$(1\ \log(q_k))$	$(1\ q_k)^T$	$(1\ \log(q_k))^T$	$(1\ q_k)^T$
Calibration to Population								
response-model variables: $\mathbf{x}_{2k}$	-	-	-	$(1\ \log(q_k)\,q_k)^T$	$(1\ \log(q_k)\,q_k)^T$	$(1\ \log(q_k)\,q_k)^T$	$f_k\,(1\ \log(q_k)\,q_k)^T$	$f_k\,(1\ \log(q_k)\,q_k)^T$
calibration variables: $\mathbf{z}_{2k}$	-	-	-	$(1\ \log(q_k)\,q_k)^T$	$(1\ \log(q_k)\,q_k)^T$	$(1\ \log(q_k)\,q_k)^T$	$(1\ \log(q_k)\,q_k)^T$	$(1\ \log(q_k)\,q_k)^T$
True Response: $p_k = 1/\{1 + \exp[3.735 + 0.4\log(q_k)]\}$								
Adverse Reactions								
Relative Bias of t_y	-0.07	0.06	-0.11	-0.13	-0.02	-0.07	0.10	0.09
Relative RMSE of t_y	4.97	3.98	4.01	2.45	2.51	2.57	2.40	2.39
Relative Bias of $v(t_y)$	8.60	12.59	12.52	6.24	6.76	6.16	6.76	6.48
Deaths								
Relative Bias of t_y	-0.17	0.06	-0.20	-0.26	-0.20	-0.30	0.04	-0.07
Relative RMSE of t_y	11.75	11.39	11.56	11.07	11.28	11.36	10.91	10.91
Relative Bias of $v(t_y)$	-1.34	-0.48	-0.90	-0.76	-1.00	-0.60	-0.12	-0.28
$(\text{Size})^{1.3}$								
Relative Bias of t_y	-0.16	-0.05	0.08	0.09	0.04	0.06	-0.02	0.01
Relative RMSE of t_y	6.92	5.07	5.06	0.95	1.05	1.12	0.89	0.89
Relative Bias of $v(t_y)$	10.01	18.49	17.47	-2.26	-3.41	-3.32	0.51	-2.12
True Response: $p_k = 1/\{1 + \exp[0.597 + 0.005\, q_k^{1/2}]\}$								
Adverse Reactions								
Relative Bias of t_y	2.87	-0.26	0.08	0.04	0.48	0.53	0.15	0.07
Relative RMSE of t_y	5.90	3.97	4.00	2.35	2.43	2.45	2.33	2.35
Relative Bias of $v(t_y)$	-18.22	11.63	11.95	9.90	8.82	7.35	7.19	6.67
Deaths								
Relative Bias of t_y	1.24	-1.88	0.47	0.36	1.03	1.20	-0.58	-0.67
Relative RMSE of t_y	11.42	11.01	11.41	10.95	11.18	11.26	10.69	10.72
Relative Bias of $v(t_y)$	5.30	3.00	6.27	6.24	5.65	5.06	6.21	5.90
$(\text{Size})^{1.3}$								
Relative Bias of t_y	5.17	1.05	-0.07	-0.05	-0.31	-0.36	0.01	0.08
Relative RMSE of t_y	9.11	5.31	5.05	0.85	0.97	1.01	0.80	0.82
Relative Bias of $v(t_y)$	-26.83	11.70	17.09	8.23	0.29	-3.98	5.17	2.90

$f_k = d_k\alpha_k - 1 = (d_k/\hat{p}_k) - 1$

Not surprisingly, the (empirical) relative mean squared error of the fourth estimator is always lower than the third. The reason is fairly obvious looking at equation (3.1) and considering the consequence of θ being 0 (calibration to the population) rather than 1 (calibration to the sample).

The fifth through eighth estimators were calibrated in two steps. The fifth and seventh estimators employed the calibration weighting from the first estimator in its first step, while the sixth and eighth employed the calibration weighting from the second estimator. The fifth and sixth used $\mathbf{z}_{2k} = \mathbf{x}_{2k} = (1 \ \log(q_k) \, q_k)^T$ in their second step, while the seventh and eighth were nearly pseudo-optimal (Kott 2011) using $\mathbf{z}_{2k} = (1 \ \log(q_k) \, q_k)^T$ and $\mathbf{x}_{2k} = (d_k \alpha_k - 1)\,\mathbf{z}_{2k}$ in their second step. All four employed the individual weight-adjustment functions:

$$h_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right) = \frac{1}{d_k \alpha_k} + \left(1 - \frac{1}{d_k \alpha_k}\right) \exp\left[\frac{\mathbf{g}_2^T \mathbf{x}_{2k}}{1 - \dfrac{1}{d_k \alpha_k}}\right].$$

As Kott (2011) showed these $h_k\left(\mathbf{g}_2^T \mathbf{x}_{2k}\right)$ are asymptotically identical to the weight-adjustment function, $1 + \mathbf{g}_2^T \mathbf{x}_{2k}$, when $\mathbf{g}_2^T \mathbf{x}_{2k} = \mathrm{O}_P\left(1/\sqrt{n}\right)$ but prevent any w_k from falling below unity. Each is a version of equation (4.1) with $\ell_k = 1/(d_k \alpha_k), c = 1,$ and $u = \infty$.

Because the nonresponse rate was so large, we did not encounter a problem computing the third and fourth estimator using any of the simulated respondent samples. The relative mean squared error of the fourth estimator was always slightly higher than that of the seventh and eighth estimators, which incorporated nearly pseudo-optimal calibration in their second step. Interestingly, this was not the case when comparing the fourth estimator to the fifth and sixth estimators which, although employing two steps, did not incorporate nearly pseudo-optimal calibration.

Observe that although the second estimator always had a smaller relative mean squared error than the first, being more consistent with a reasonable prediction model (even for $q_k^{1.3}$, the survey variable appeared closer to being linear in q_k than in $\log(q_k)$), the other analogous pairs (fifth vs sixth and seventh vs eighth) exhibited no clear pattern of superiority. This is because it is the second-step residuals that are effectively modeled in equation (4.4) not the y- values.

Generating the nonresponse with equation (5.2) than (5.1) did not seem to have much of an impact on the results except for the relative biases of the first estimator. For both adverse reactions and $(\text{size})^{1.3}$, the relative bias of this estimator is over 40% of the relative mean squared error. That is likely because both models that could be used to justify this estimator (response is logistic in the log of the size measure and the survey variable is linear in the log of the size measure) fail. Not surprisingly, since the relative bias is such a large part of the relative mean squared error in these two situations, $v(t_k)$ underestimates mean squared error badly. Nowhere else is the relative bias of $v(t_k)$ greater than 15%.

It seems that even our artificial variable, $(\text{size})^{1.3}$, was close enough to being linear in the size measure that bias was never an issue for any estimator other than the first. The first estimator itself had a negligible relative bias when response was a logistic model of the log of the size measure, as assumed.

6 Concluding remarks

In Section 4, we noted two reasons to prefer calibration weighting in two steps: to make implicitly fitting a logistic response model easier and to incorporate nearly quasi-optimal calibration. A side benefit of two-step calibration is more efficient estimation of the response model in step one since there is no sampling error to confound the estimation. This is useful when one wants to analyze the causes of unit nonresponse for its own sake.

We must concede, however, that the reduction in mean squared error using two steps was modest in our simulation experiments in Section 5. Moreover, the practical appeal of the simplicity of calibrating in a single step cannot be denied.

When calibration-weighting is used to adjust for nonresponse that is not missing at random as described in Chang and Kott (2008) and Kott and Chang (2010), the efficiency gains from a second step involving only calibration variables and functions of calibration variables model variables is likely to be sizeable.

When the finite population correction factors can be ignored, replication offers a much simpler approach to variance estimation than equation (3.7) even though the second summation on the right-hand side can be dropped in this situation. A different attractive alternative is the “collapsed” version of equation (4.2) that ignores the impact of the first calibration step:

$$\tilde{v}(t_y) = \sum_{k,j \in S} \left(1 - \frac{\pi_k \pi_j}{\pi_{kj}}\right) [w_k e_{2k}] [w_j e_{2j}] + \sum_{k \in R} d_k \left(h_k^2 \alpha_k^2 - h_k \alpha_k\right) e_{2k}^2.$$

This estimator clearly estimates the prediction-model variance if that model holds. A version of it – with the second summation removed – fared well in our simulation experiments (not shown). Some caution is needed before one draws too strong a conclusion from that result since the linear model was never too far from holding in our investigations.

Finally, a number of assumptions were made to simplify the exposition. The interested reader can extend the results to unbounded d_k, more general and not-necessarily-bounded weight-adjustment functions, or to allow the prediction-model errors to be correlated within primary sampling units. When N grows faster than n, the assumption that $\sigma_k^2 = \mathbf{z}_k^T \boldsymbol{\eta}$ can sometimes be dropped. See, for example, Kott (2009, page 69).

Acknowledgements

This paper was prepared for the Symposium on the Analysis of Survey Data and Small Area Estimation, in honour of the 75th Birthday of Professor J.N.K. Rao sponsored by the Fields Institute for Research in Mathematical Sciences. The authors would like to thank the organizers of the conference for the invitation to present this paper and the Institute for its generous funding of the conference without which this paper would never have been written. They would also like to thank a number of editors and referees for their helpful comments.

References

Bang, H., and Robins, J.M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61, 962-972.

Bankier, M. (2002). Regression estimators for the 2001 Canadian Census. Presented at the International Conference in Recent Advances in Survey Sampling.

Chang, T., and Kott, P.S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika,* 95, 557-571.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sample frequency table when the expected marginal total are known. *Annals of Mathematical Statistics,* 11, 427-444.

Deville, J.-C. (2000). Generalized calibration and application to weighting for non-response. In *COMPSTAT: Proceedings in Computational Statistics, 14th Symposium, Utrecht, The Netherlands*, (Eds., J.G. Bethlehem and P.G.M. Van der Heidjen), Heidelberg: Physica Verlag, 65-76.

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, 1013-1020.

Estevao, V.M., and Särndal, C.-E. (2002). The ten cases of auxiliary information for calibration in two-phase sampling. *Journal of Official Statistics*, 18, 233-255.

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

Folsom, R.E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *Proceedings of the American Statistical Association, Social Statistics Section*, 197-202.

Folsom, R.E., and Singh, A.C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Proceedings of the American Statistical Association, Survey Research Methods Section,* available online at http://www.amstat.org/sections/srms/Proceedings/, 598-603.

Fuller, W.A., Loughin, M.M. and Baker, H.D. (1994). Regression weighting in the presence of nonresponse with application to the 1987-88 Nationwide Food Consumption Survey. *Survey Methodology*, 20, 1, 75-85.

Kim, J.K., and Haziza, D. (2014). Doubly robust inference with missing survey data. *Statistica Sinica,* 24, 375-394.

Kim, J.K., and Park, H. (2006). Imputation using response probability. *Canadian Journal of Statistics,* 34, 1-12.

Kim, J.K., and Shao, J. (2013). *Statistical Methods for Handling Incomplete Data,* London: Chapman and Hall/CRC.

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

Kott, P.S. (2009). Calibration weighting: Combining probability samples and linear prediction models. In *Handbook of Statistics 29B: Sample Surveys: Inference and Analysis,* (Eds., D. Pfeffermann and C.R. Rao), New York: Elsevier.

Kott, P.S. (2011). A nearly pseudo-optimal method for keeping calibration weights from falling below unity in the absence of nonresponse or frame errors. *Pakistan Journal of Statistics*, 27, 391-396.

Kott, P.S., and Chang, T.C. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association,* 105, 1265-1275.

Kott, P.S., and Liao, D. (2012). Comparing weighting methods when adjusting for logistic unit Nonresponse. Presented at Federal Committee on Survey Methodology Research Conference, available online at http://www.fcsm.sites.usa.gov/files/2014/05/Kott_2012FCSM_III-B.pdf..

Little, R.J., and Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd Ed.), New York: John Wiley & Sons, Inc.

Lundström, S., and Särndal, C.-E. (1999). Calibration as a standard method for the treatment of nonresponse. *Journal of Official Statistics,* 15, 305-327.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustment for unit nonresponse. In *Incomplete Data in Sample Surveys,* (Eds., W.G. Madow, I. Olkin and D.B. Rubin), New York: Academic Press, 2.

Rao, J.N.K. (1994). Estimation of totals and distributing functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Robins J.M., Rotnitzky A. and Zhao L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, p. 846-866.