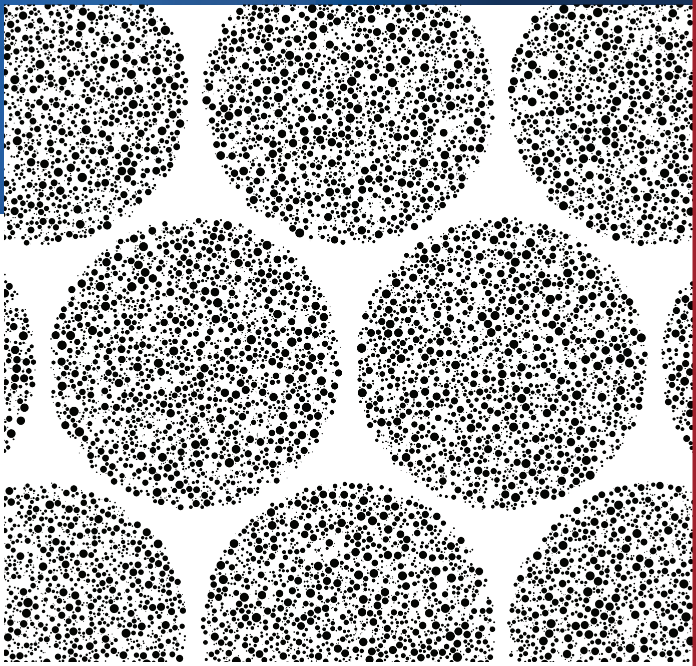
RTI Press®

Methods Report

ISSN 2378-7813

May 2024



Calibration Weighting With a Blended (Probability and Nonprobability) Sample: Mean and Variance Estimation When Errors Can Come from Both Samples

Phillip S. Kott, and Jamie Ridenhour

RTI INTERNATIONAL

RTI Press publication MR-0053-2405

RTI International is an independent, nonprofit research organization dedicated to improving the human condition. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

Suggested Citation

Kott, P. S., and Ridenhour, J. (2024). *Calibration Weighting With a Blended (Probability and Nonprobability) Sample: Mean and Variance Estimation When Errors Can Come from Both Samples*. RTI Press Publication No. MR-0053-2405. Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2024.mr.0053.2405

This publication is part of the RTI Press Methods Report series.

RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
E-mail: rtipress@rti.org
Website: www.rti.org

Contents

About the Authors i

Abstract ii

Introduction 1

Solving a Calibration Equation 2

Calibration Weighting a Blended Sample 3

Estimation from the Nonprobability Sample 5

An Example 5

Some Concluding Remarks 8

References 9

Appendix. SAS-Callable SUDAAN Code Used in Tables 1 and 2 11

About the Authors

Phillip S. Kott, PhD, is a senior research statistician in Official Statistics at RTI International. https://orcid.org/0000-0002-9801-7171

Jamie Ridenhour, PhD, is Director, Statistics, in Official Statistics at RTI International. https://orcid.org/0000-0001-6216-9512

RTI Press Associate Editor

Cynthia Bland

Abstract

We show how calibration weighting can be employed to combine a probability and a nonprobability sample of the same population in a statistically defensible manner. This is done by assuming that the probability of a population element being included in the nonprobability sample can be modeled as a logit function of variables known for all members of both samples. Estimating these probabilities for the members of the nonprobability sample with a calibration equation and treating their inverses as quasi-probability weights is a key to creating composite weights for the blended sample. We use the WTADJX procedure in SUDAAN® to generate those weights and then measure the standard errors of the resulting estimated means and totals as well as assess the potential for bias in those estimates. The appendix contains the SAS-callable code for the SUDAAN procedures used in this paper.

Introduction

Nonprobability survey samples—whose members do not have known probabilities of sample inclusion—are everywhere and have considerable potential for bias (see Baker et al., 2013, and the references therein). It has become popular to attempt to remove the bias of an estimate derived from a nonprobability sample by first combining that sample, denoted here by S_0, with a probability sample S_1 that covers the same population U but shares no members with it. After that, one estimates the probability γ_k that a population unit k in the blended sample $S = S_0 \cup S_1$ was originally from the nonprobability sample as a function of a vector of covariates $\mathbf{z}_k$ available for members from both samples (when used here, "sample" always refers to a respondent sample).

Valliant and Dever (2011) suggest that the inverse of this estimated probability—which they, following much of the literature, call a "propensity"—can be used as a quasi-probability sampling weight either directly, $w_k = 1/\hat{\gamma}_k$, or indirectly after some form or poststratification. For example, Lee (2006) proposes sorting the blended sample by $\hat{\gamma}_k$ values, then breaking the sample into cells of nearly equal size, and finally assigning the weight $w_k = \widehat{N}_{1c}/n_{0c}$, where $\widehat{N}_{1c}$ is the estimated-from-the-probability-sample population size of cell c containing nonprobability-sampled unit k, and n_{0c} is the number of members of S_0 in c.

Although estimating $\gamma_k = \Pr(k \in S_0 \mid k \in S ; \mathbf{z}_k)$ by fitting a logistic regression on the unweighted blended sample is often treated as an estimate for $\Pr(k \in S_0 \mid \mathbf{z}_k)$, Robbins and colleagues (2021) argue that a better estimate for the quasi-probability that k is in the nonprobability sample when $S_0 \cap S_1 = \emptyset$ is

$$p_{0k} = \Pr(k \in S_0 \mid \mathbf{z}_k) = \pi_k \hat{\gamma}_k / (1 - \hat{\gamma}_k),$$

where π_k is the probability that k is chosen for the probability sample (which can include an adjustment for unit nonresponse when it is needed). It is assumed that π_k is known for members in the population that are not in S_1.

To see how $p_{0k} = \Pr(k \in S_0 \mid \mathbf{z}_k)$ is derived, start with

$$\pi_k = \Pr(k \in S_1 \mid k \in S ; \mathbf{z}_k) \times \Pr(k \in S \mid \mathbf{z}_k), \text{ and}$$

$$\Pr(k \in S_0 \mid \mathbf{x}_k) = \Pr(k \in S_0 \mid k \in S ; \mathbf{x}_k) \times \Pr(k \in S \mid \mathbf{x}_k),$$

then solve for $\Pr(k \in S_0 \mid \mathbf{z}_k)$. Elliott and Valliant (2017) make a similar point, suggesting a more sophisticated method could be used to estimate γ_k.

Robbins and colleagues (2021) offer methods for weighting the blended sample, but for now, we assume there are survey items of interest collected in the nonprobability sample but not in the probability sample so that quasi-probability weights for those items are only needed for the members of S_0.

There are two critical assumptions underlying the use of p_{0k}. One is that the probability and nonprobability sample have no member in common. This can be assured by removing any member of S_1 from the nonprobability sample. The other is that $\Pr(k \in S_0 \mid k \in S ; \mathbf{z}_k)$ can be modeled, whether by a logistic function (as in Robbins et al., 2021) or some other functional form (as suggested by Elliott & Valliant, 2017). We believe that it is more reasonable to assume that $\Pr(k \in S_0 \mid \mathbf{z}_k)$ itself can be modeled by a logistic (or some other) function whether or not $S_0 \cap S_1 = \emptyset$.

We first describe in general terms how that assumption can be used to generate quasi-probability weights for a nonprobability sample given either population totals for the component of $\mathbf{z}_k$ or their probability-sample–estimated analogues. In our setup, when the population total is known for a component $\mathbf{z}_k$, it does not need to be collected on the probability sample (only the nonprobability sample).

Using the WTADJX procedure in SUDAAN 11, we then show how to estimate population means for variables collected from the nonprobability sample and for variables of interest collected from a blended sample comprising a nonprobability and a probability sample drawn from the same population when there are auxiliary variables collected on both samples that can be used to reduce or remove the biases of the estimates for the variables of interest.

This methodology was applied to a stratified simple random probability sample blended with a nonprobability sample drawn from the same population. The methodology can be used to assess the potential for bias in estimates based on the blended sample.

Solving a Calibration Equation

The model

$$\Pr(k \in S_0 | \mathbf{z}_k) = [1 + \exp(\mathbf{z}_k^T \mathbf{g})]^{-1} \quad (1)$$

is a selection model. If correctly specified, it models the probability that $k \in U$ is included in the nonprobability sample S_0 (which can involve self-selection and response) as a logistic function of the vector $\mathbf{z}_k$ with unknown parameter-vector $\mathbf{g}$. Kott (2019) points out that this selection model can often be estimated by solving a calibration equation when each component of the population total $\mathbf{T_z} = \sum_{k \epsilon U} \mathbf{z}_k$ is either known or consistently estimated from a probability sample, which itself can have been weighted to account for unit nonresponse (here "a consistent estimate" computed from a probability sample converges into the population parameter it estimates as the probability sample size and population sizes grow arbitrarily large).

A calibration equation that can be used to estimate $\mathbf{g}$ in equation (1) is

$$\sum_{k \epsilon S_0} [1 + \exp(\mathbf{z}_k^T \hat{\mathbf{g}})] \mathbf{z}_k = \widehat{\mathbf{T}}_\mathbf{z} , \quad (2)$$

where each component of $\widehat{\mathbf{T}}_\mathbf{z}$ is either assumed to be a known population quantity or a consistent estimate from a probability sample. The solution for $\hat{\mathbf{g}}$ in equation (2), when it exists, can usually be found using Newton's method as Deville and colleagues (1993) show. That approach has been programmed into the SUDAAN 11® routines WTADJUST and WTADJX (RTI International, 2012), the R routines 'calib' and 'gencalib' in 'Sampling' (Tille & Matei, 2023), and elsewhere. We will describe how to use SUDAAN's WTADJX for our purposes later. Other software packages could be used in a similar manner.

When a solution to equation (2) exists—and we will assume here it does—$\hat{\mathbf{g}}$ is a consistent estimator for $\mathbf{g}$ under mild conditions we assume to hold. The *quasi-probability weight* for $k \in S_0$ is then

$$w_k = [1 + \exp(\mathbf{z}_k^T \hat{\mathbf{g}})]. \quad (3)$$

The theory supporting this use of an assumed selection model like that in equation (1) together with a calibration equation like (2) first to estimate consistently the parameters of the selection model and then with an equation like (3) to use those estimates in generating quasi-probability weights (asymptotically equal to the inverses of the sample members' probabilities of inclusion into the nonprobability sample) is analogous to the quasi-random theory supporting the use and calibration-equation fitting of a selection model for the response/nonresponse mechanism in a probability sample. See, for example, Kott and Liao (2012).

In the nonresponse-adjustment setting of a probability sample, $S_1 \subset S_1^*$, the selection (response) model $\mathbf{Pr}(k \in S_1 | \mathbf{z}_k; S_1^*) = [1 + \exp(\mathbf{z}_k^T \mathbf{g})]^{-1}$, where S_1^* is the probability sample before unit nonresponse, replaces equation (1), and $\boldsymbol{w}_k = \boldsymbol{\pi}_k^{-1}[1 + \mathbf{exp}(\mathbf{z}_k^T \hat{\mathbf{g}})]$ replaces equation (3), where $\boldsymbol{\pi}_k$ is the probability that k has been chosen for S_1^*.

The assumption that every member of the population has a probability of selection into the nonprobability sample equal to $\mathbf{Pr}(k \in S_0 | \mathbf{z}_k) = [1 + \mathbf{exp}(\mathbf{z}_k^T \mathbf{g})]^{-1}$ or any other monotonic differentiable function is a strong one.

An alternative justification for using equations (1) through (3) in creating weights in estimating a total like $Y = \sum_{k \in U} y_k$ from a nonprobability sample of y-values follows. Suppose each $y_k \in U$ behaves like random variables with mean $\mathbf{z}_k^T \boldsymbol{\beta}$ for some unknown parameter $\boldsymbol{\beta}$ whether (or not) k is in the nonprobability sample. That is, selection is ignorable in expectation with respect to this *prediction* model, so called because the model predicts a value for y_k (Royall, 1970). Given $\widehat{\mathbf{T}}_\mathbf{z}$ and assuming equation (2) has a solution, if *either* this prediction model or the selection model holds among the members of the population, estimating Y with $\widehat{Y} = \sum_{k \in S_0} \boldsymbol{w}_k \boldsymbol{y}_k$ will be nearly unbiased in some sense (technically, $\widehat{Y}$ is a predictor, not an estimator, for the random variable Y under the prediction model). See Kott and Liao (2012) for a proof of this assertion. A similarly "doubly robust" approach can be found in in Chen and colleagues (2020).

When one of more components of $\widehat{\mathbf{T}}_\mathbf{z}$ is consistently estimated from a probability sample, the near unbiasedness of $\widehat{Y}$ requires the combination of probability-sampling inference (for $\widehat{\mathbf{T}}_\mathbf{z}$) and either the selection model or prediction model (for $\widehat{Y} | \widehat{\mathbf{T}}_\mathbf{z}$).

Nevertheless, we call the former the selection-model framework and the latter the prediction-model framework.

Observe that $\sum_{k\in S_0} w_k y_k / \sum_{k\in S_0} w_k$ is a nearly unbiased predictor for the population mean $\bar{y} = \sum_{k\in U} y_k / \sum_{k\in U} 1$, when each $y_k \in U$ behaves like a random variable with mean $\mathbf{z}_k^T \boldsymbol{\beta}$, and 1 is either a component of $\mathbf{z}_k$ or the linear combination of the components of $\mathbf{z}_k$.

As mentioned previously, when selected members of a probability sample S_1 have design weights $\{d_k\}$ before unit nonresponse, where $d_k = \pi_k^{-1}$, then we can weight the unit respondents with

$$w_k = d_k[1 + \exp(\mathbf{z}_k^T \hat{\mathbf{g}})], \tag{4}$$

when response is a logistic function of $\mathbf{z}_k$ (which need not be the same as the vector in equation [1]), and $\hat{\mathbf{g}}$ (which likewise need not be the vector in equation [1]) satisfies the calibration equation $\sum_{k\in S_1} d_k [1 + \exp(\mathbf{z}_k^T \hat{\mathbf{g}})]\, \mathbf{z}_k = \hat{\mathbf{T}}_{\mathbf{z}}$.

Calibration weighting was originally proposed to reduce the standard error of an estimated total in the absence of nonresponse. It works when y_k can be approximated by a linear function of the components of $\mathbf{z}_k$, and the weight-adjustment function within the square brackets of equation (4) is replaced by exp $(\mathbf{z}_k^T \hat{\mathbf{g}})$, where $\hat{\mathbf{g}}$ converges to $\mathbf{0}$ and consequently the w_k converges to d_k as the probability sample grows arbitrarily large.

Both weight adjustments are special cases of the following more general weight-adjustment function:

$$\alpha(\theta) = \frac{L + \exp(\theta)}{1 + \exp(\theta)/U}, \tag{5}$$

where $[L, U]$ is the range of $\alpha(\theta)$, and $0 \le L < U \le \infty$. Software packages that do calibration weighting via what has been called "the logit transformation" in equation (5) allow the user to set L and U. Some packages (like 'calib' in Tille & Matei, 2023) are only designed for probability samples with full response and restrict L to a value less than 1. When used to adjust for unit nonresponse or nonprobability selection, however, the range for the implicitly estimated probability of unit response or selection is $[1/U, 1/L]$. Consequently, it is sensible to set L at either 1 or a value greater than 1.

Calibration Weighting a Blended Sample

Suppose we have a probability sample and a nonprobability sample both chosen from the same population. We denote them by S_1 and S_0, respectively. At first, suppose both collect a variable y_k with the intention of estimating its population mean. The probability sample is a stratified multistage sample, which may suffer from some unit nonresponse. If used by itself, a vector $\mathbf{z}_{1k}$ of variables including one with known population totals can be employed to generate calibration weights for the probability sample rendering estimates for the population mean using those weights both nearly unbiased with respect to the selection model (probability sampling is a type of selection model) and with respect to the linear prediction model: $\mathrm{E}(y_k) = \mathbf{z}_{1k}{}^T \boldsymbol{\beta}_1$. If there is any unit nonresponse, the selection model assumes that the probability of unit response is correctly specified by the inverse of a weight-adjustment function of the components of $\mathbf{z}_{1k}$, while the linear prediction model assumes unit nonresponse is ignorable in expectation.

Similarly, a vector $\mathbf{z}_{0k}$ of variables including one with known population totals can be used to generate calibration weights for the nonprobability sample rendering estimates for the population mean using those weights both nearly unbiased with respect to the selection model when the probability of selection into the nonprobability sample is correctly specified by the inverse of a weight-adjustment function of the components of $\mathbf{z}_{0k}$ and with respect to the linear outcome model: $\mathrm{E}(y_k) = \mathbf{z}_{0k}{}^T \boldsymbol{\beta}_0$, assuming selection into S_0 is ignorable in expectation.

Many of the components of $\mathbf{z}_{1k}$ and $\mathbf{z}_{0k}$ may coincide. We do not require that the two samples be disjoint but they must be selected independently.

The WTADJUST procedure in SUDAAN can be used to create weights and estimate the population means as described previously as long as the weight-adjustment function in equation (5) is used for both samples. WTADJUST allows L and U to differ across the members of a sample. Here, one can set values L_1 and U_1 for every member of S_1 and values L_0 and U_0 for every member of S_0. When U_f, $f = 0$ or 1, is unspecified, it is treated as virtually infinite (10^{20}), and a finite centering parameter C_f needs to be added to WTADJUST for the members of S_f; say, max{1, $2L_f$}, but the choice (as long as it is finite) has no impact on the results.

WTADJUST will also estimate standard errors that are nearly unbiased under the selection-model framework. Moreover, any linear combination of the two estimates is also a nearly unbiased estimator of the population mean and has a standard error that can be estimated (under the selection-model framework) using WTADJUST.

To this end, let S be the union of S_1 and S_0. A sample member may be in S twice, with each such member treated as two separate members of the blended sample S. We treat the H design strata of S_1 and the entire nonprobability sample as the $H + 1$ design stratum of the blended sample S. For any k in S, let $\mathbf{z}_k{}^T = (\mathbf{z}_{1k}{}^T\ \mathbf{z}_{0k}{}^T)$, where all the components of $\mathbf{z}_{1k}$ are 0 when k is in S_0 and all the components of $\mathbf{z}_{0k}$ are 0 when k is in S_1. The L and U parameters are the same for each member of S_1, and they are the same for each member of S_0, but the former and latter pairs may differ.

Consider the following calibration equation, which can be used to create quasi-probability weights for any positive λ:

$$\sum_{k\epsilon S} d_k \alpha_{(k)}\left(\mathbf{z}_k^T \hat{\mathbf{g}}\right)\mathbf{z}_k = \begin{pmatrix} \lambda \sum_{k\epsilon S_1} \pi_k^{-1} \alpha_1\left(\mathbf{z}_{1k}^T \hat{\mathbf{g}}_1\right)\mathbf{z}_{1k} \\ \sum_{k\epsilon S_0} \alpha_0\left(\mathbf{z}_{0k}^T \hat{\mathbf{g}}_0\right)\mathbf{z}_{0k} \end{pmatrix} = \begin{pmatrix} \lambda \mathbf{T}_{\mathbf{z}_1} \\ \mathbf{T}_{\mathbf{z}_0} \end{pmatrix}$$
$$= \mathbf{T}_{\mathbf{z}}(\lambda), \qquad (6)$$

where $\alpha_{(k)}\left(\mathbf{z}_k^T\hat{\mathbf{g}}\right) = \alpha_f\left(\mathbf{z}_{fk}^T\hat{\mathbf{g}}_f\right)$ for $k \in S_f$, $\alpha_f\left(\mathbf{z}_{fk}^T\hat{\mathbf{g}}_f\right)$ is weight-adjustment function for S_f ($f = 0$ or 1), $\hat{\mathbf{g}}^T = \left(\hat{\mathbf{g}}_1^T\ \ \hat{\mathbf{g}}_2^T\right)$, and

$$d_k = \delta_k \lambda \pi_k{}^{-1} + (1 - \delta_k), \qquad (7)$$

$\delta_k = 1$ when k was originally from S_1 and 0 otherwise (we labeled the weight on the left-hand side of this equation d_k for convenience, although it depends on the choice of λ; a more formal label would be $d_k{}^{(\lambda)}$). Observe that the relative contribution of the probability sample when estimating $\bar{y}$ is $\lambda/(1+\lambda)$.

WTADJUST estimates both $\bar{y}$ with the weights implied by equation (6) and the standard error of that estimate.

WTADJUST has one glaring limitation, however. It cannot be used to estimate standard errors when the probability of selection into the nonprobability sample includes variables with unknown population totals that need to be estimated by the probability sample. For that, one needs WTADJX (or something like it; Chen et al., 2020, discuss another approach).

For our purposes, the equation for the quasi-weights in S using WTADJX is

$$w_k = d_k \frac{L_k + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)}{1 + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)/U_k}, \qquad (8)$$

where $L_k = L_1\delta_k + L_0(1-\delta_k)$ and $U_k = U_1\delta_k + U_0(1-\delta_k)$, and the model variable $\mathbf{x}_k$ is a vector with the same number of components as the vector of values on which we are calibrating, such as the $\mathbf{z}_k$.in equation (1). (When $\mathbf{x}_k = \mathbf{z}_k$, WTADJUST can be used in place of WTADJX.)

Let $\mathbf{q}_k$ denote a vector of additional variables included in the nonprobability sample's selection model,

$$\Pr(k \in S_0 | \mathbf{z}_{0k}, \mathbf{q}_k) = \frac{1 + \exp\left(\mathbf{z}_{0k}^T\mathbf{g}_0 + \mathbf{q}_k^T\mathbf{g}_{\mathbf{q}}\right)/\mathrm{U}_0}{L_0 + \exp\left(\mathbf{z}_{0k}^T\mathbf{g}_0 + \mathbf{q}_k^T\mathbf{g}_{\mathbf{q}}\right)},$$

but with unknown population totals that need be estimated by the probability sample. With

$$\mathbf{x}_k{}^T = (\mathbf{z}_{1k}{}^T\ \mathbf{z}_{0k}{}^T\ [1-\delta_k\ ]\mathbf{q}_k{}^T), \text{ and}$$
$$\mathbf{z}_k{}^T = (\mathbf{z}_{1k}{}^T\ \mathbf{z}_{0k}{}^T\ \{1 - \delta_k[1 + 1/\lambda]\}\mathbf{q}_k{}^T), \qquad (9)$$

WTADJX can be used to estimate the population mean $\bar{y}$ by finding the $\hat{\mathbf{g}}$ satisfying:

$$\sum_{k\epsilon S} d_k \frac{L_k + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)}{1 + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)/U_k}\mathbf{z}_k = \begin{pmatrix} \lambda\sum_{k\epsilon U}\mathbf{z}_{1k} \\ \sum_{k\epsilon U}\mathbf{z}_{0k} \\ \mathbf{0} \end{pmatrix}, \qquad (10)$$

where $\mathbf{0}$ has as many components as $\mathbf{q}_k$ (so that $\sum_{k\epsilon S_0} w_k \mathbf{q}_k = \sum_{k\epsilon S_1} w_k \mathbf{q}_k$) . To estimate the population total $Y = \sum_{k\in U} y_k$, the quasi-weights in equation (8) need to be divided by 1+λ.

One can vary the choice of λ to find the optimal value that minimizes the standard error of the estimated mean. Recall that every choice for λ results in nearly unbiased estimation. Moreover, observe that given how the d_k in equation (7) and $\mathbf{z}_k$ in equation (9) are defined, the choice of λ has no impact on the $\hat{\mathbf{g}}$ satisfying equation (10).

After choosing λ, the components of the vector on the right-hand side of equation (10) are known before sampling. Because of this, WTADJX can estimate the standard error of an estimated total or mean (under the selection-model framework), assuming the indicators of selection into the nonprobability sample are independent of each other. When there is unit nonresponse in the probability sample, an analogous assumption is made about the indicators' unit response.

Ignoring finite population correction (as we will), the key to nearly unbiased variance estimation via linearization is the near (i.e., asymptotic) equality of $\sum_S w_k e_k = \sum_S d_k \alpha(\mathbf{x}_k^T \hat{\mathbf{g}}) e_k$ and $\sum_S d_k \alpha(\mathbf{x}_k^T \mathbf{g}) e_k$, where

$$e_k = y_k - \mathbf{z}_k^T \left[\sum_S d_j \alpha'\left(\mathbf{x}_j^T \hat{\mathbf{g}}\right) \mathbf{x}_j \mathbf{z}_j^T \right]^{-1} \sum_S d_j \alpha'\left(\mathbf{x}_j^T \hat{\mathbf{g}}\right) \mathbf{x}_j y_j. \quad (11)$$

The inclusion of the $\alpha'\left(\mathbf{x}_j^T \hat{\mathbf{g}}\right)$ terms in e_k allows us to avoid directly accounting for $\hat{\mathbf{g}}$ itself being an estimate in large-sample variance estimation. For more theoretical details on variance estimation for a calibrated estimator when $\mathbf{x}_k \neq \mathbf{z}_k$, see Kott and Liao (2015).

When there are many variables for which one needs to estimate a population mean from the blended sample, the optimal λ will likely vary across the variables. Consequently, a compromise will be needed if one desires a single λ to be used for all variables.

Estimation from the Nonprobability Sample

Often y-values are collected from the nonprobability sample but not the probability sample. We can treat whether (or not) an element of the blended sample was originally a member of the nonprobability sample as a class variable in WTADJX and then estimate $\bar{y}$ and the standard error of that estimate with WTADJX. The estimates of $\bar{y}$ based on the blended sample and for the class defined as the probability sample will be missing, while the estimate for the class defined by the nonprobability sample will be a nearly unbiased estimate for $\bar{y}$. A nearly unbiased estimate for its standard error will accompany it. The selection of λ in equation (1) does not matter. Setting λ = 1 for simplicity is a straightforward choice.

An Example

Benoit-Bryan and Mulrow (2021) describe a simulated population of 113,549 (N) individuals created from the *Culture and Community in a Time of Crisis* survey of behavior and attitudes before and during the Covid-19 crisis. Both 10,000 stratified simple random probability samples of 1,000 persons and 10,000 vaguely described nonprobability samples of 4,000 persons were drawn from the population. There are no missing values in the data set as it now stands (in Mulrow, 2022). This allows for pure analyses of blending methodologies.

To demonstrate our methodology, we focus on a single probability and a single nonprobability sample. Our goal is to estimate population means for 14 survey variables of interest (that were chosen by Benoit-Bryan and Mulrow for a competition) using information from 32 not-of-interest (NOI) survey variables (chosen by us) as well as variables for which the population means were known. The last group includes 9 region indicators, 3 levels of urbanization, an Hispanicity indicator, 6 race categories, 4 age categories, and 7 education-level categories.

Most of the survey variables of interest and NOI variables were yes/no (1/0). Two of the survey variables of interest were originally on a five-point Likert scale. Thus, we had for analytical purposes 20 variables of interest whose proportion of 1s we were trying to estimate.

The survey variables are described and given variable names (e.g., q7_22) below:

Variables of Interest

q7_22	Attended classical music in 2019
q10_1	Missed experiencing artwork, performances
q_11	Offered online exhibitions or galleries
q25_11	Will see a play or musical when able in short term
q1_15	Participated in a live interactive event in past 30 days

 https://doi.org/10.3768/rtipress.2024.mr.0053.2405

q6_9 Want more fun in life

q11_4 Offered online materials or activities for kids

q10_3 Miss celebrating cultural heritage

q7_14 Attended community festival in 2019

q6_1 Want more hope in life

q1_6 Watch movie or tv series in past 30 days

q25_13 Will take art, music, or dance class when able in short term

The two five-level original variables of interest (and their replacements):

q17 During Covid, how important are arts and cultural organizations

q18 Before Covid, how important were arts and cultural organizations

_2 Slightly (e.g., q17 = 2 becomes q17_2 = 1)

_3 Moderately

_4 Important

_5 Very

NOI Variables

In 2019, did you attend or participate in . . .

q7_1 Art museum

q7_2 Children's museum

q7_3 Art gallery/fair

q7_4 Botanical garden

q7_5 Zoo or aquarium

q7_6 Science or technology museum

q7_7 Natural history museum

q7_8 Public park

q7_9 Architectural tour

q7_10 Public/street art

q7_11 Film festival

q7_12 Music festival

q7_13 Performing arts festival

q7_15 Craft or design fair

q7_16 Read books/literature

q7_17 Food and drink experience

q7_18 Nonmusical play

q7_19 Musical

q7_20 Variety or comedy show

q7_21 Popular music

q7_23 Jazz music

q7_24 Opera

q7_25 World music

q7_26 Contemporary dance

q7_27 Ballet

q7_28 Regional dance

q7_29 Historic attraction/museum

q7_30 Television program

q7_31 Movies/film

q7_32 Library

q7_33 Cultural center

q7_34 Video games/online gaming

There were 18 strata in the stratified simple random probability sample. We assigned each member of that sample to Domain 1 and the members of the nonprobability sample to Domain 2 and to Stratum 19. Implicitly assuming $\lambda = 1$, both domains were at first calibrated separately to 30 population variables defined by region, urbanization, race, Hispanicity, age, and education level using WTADJUST with the weight-adjustment-function parameters in equation (5) set at $L = 0$ and $U = 10^{20}$ (virtually infinity).

An attempt at setting an initial weight of 1 with $L = 1$ for the nonprobability sample failed for technical reasons related to how the SUDAAN program runs rather than to the underlying theory. We set the initial weight for each member of the nonprobability sample at 25 (slightly less than N/n_0 = 113,549/4,000 = 28.38725, which is what we set as the initial weight for each member) and L at 1/25 = 0.04. That is mathematically equivalent to assuming the probability on inclusion is a logistic function, because $1 + \exp(\mathbf{x}_j^T \mathbf{g}) = 25\,(1/25 + \exp(\mathbf{x}_j^T \mathbf{g}^*))$, so that only the coefficients of the intercepts for $\mathbf{g}$ and $\mathbf{g}^*$, or their equivalents, differ.

Table 1 displays differences in the estimated means (the proportion of 1s expressed as a percent) for the variables of interest computed from the probability and nonprobability samples separately using

Table 1. Differences in estimated domain means of variables of interest (in percentage point) using WTADJUST

Variable	Difference	*t*-value	*p*-value
q7_22	-5.87071	-3.09817	0.00196
q17_5	-5.67015	-2.77586	0.00553
q17_4	5.19282	2.75408	0.00591
q6_9	5.42981	2.65408	0.00798
q1_15	-4.96485	-2.44432	0.01455
q10_1	-4.51152	-2.40899	0.01603
q18_5	-4.76654	-2.38859	0.01695
q18_3	2.11118	1.71476	0.08645
q18_4	2.19730	1.18603	0.23567
q25_13	-1.26753	-1.09874	0.27194
q1_6	1.21429	1.07647	0.28177
q11_1	-2.11664	-1.05061	0.29349
q18_2	0.49995	0.97542	0.32940
q17_2	0.61545	0.78579	0.43203
q10_3	-0.67250	-0.78239	0.43402
q25_11	-1.15565	-0.61314	0.53981
q11_4	-0.70367	-0.36134	0.71786
q6_1	0.63118	0.31783	0.75063
q17_3	-0.31811	-0.22483	0.82212
q7_14	-0.03445	-0.01706	0.98639

WTADJUST for both estimates (the code is in the Appendix). These differences are sorted by their ascending *p*-values.

Because there are so many differences being measured (20), it is advisable to use a Bonferroni correction when assessing whether estimated differences are significant. This correction divides the *p*-values for the variables of interest by 20. If the difference with the smallest *p*-value among the variables of interest remains significant at some level, then the hypothesis of no difference between probability-sample and nonprobability-sample estimates for the variables of interest fails. The Bonferroni correction is known to be conservative, so it may be prudent to assesses significance at the .1 level rather than at the conventional .05 level. With this in mind, the hypothesis of no differences between the estimates fails among the variables of interest (.00196 < .1/20 = .005). It would fail at the Bonferroni-corrected .05 level as well.

Table 2 is analogous to Table 1 displaying the differences among the estimates for the NOI variables with the 10 smallest *p*-values (the code is in the Appendix). We took the four NOI variables with significant differences at the Bonferroni-corrected .1 level (*p*-value < .1/32) and added them to the selection model for the nonprobability samples and recomputed the estimates for the nonprobability sample using WTAJDX. The revised differences between the probability-sample estimates and nonprobability-sample estimates for the variables of interest are displayed in Table 3. The *t*- and *p*-values of these differences were computed, without loss of generality, setting $\lambda = 1$.

Using a Bonferroni correction, the null hypothesis of no significant difference between the probability-sample estimates and revised nonprobability-sample estimates among the variables of interest no longer fails at either the .1 level (.1/20 < .00594) or at the .05 level. This gives us some confidence that the revised nonprobability-sample estimates for the variables of interest, unlike the original ones, are nearly unbiased. That confidence is mitigated somewhat by the observation that 6 out of the 20 differences

Table 2. Ten largest differences in estimated domain means of NOI variables using WTADJUST (in terms of their *p*-values)

Variable	Difference	*t*-value	*p*-value
Q7_28	-3.15237	-3.58028	0.00035
Q7_4	-6.97418	-3.45283	0.00056
Q7_5	-6.05481	-3.22138	0.00128
Q7_32	-6.15135	-3.07216	0.00214
Q7_24	-4.34578	-2.67472	0.00750
Q7_1	-4.42137	-2.51749	0.01185
Q7_27	-3.69502	-2.33730	0.01946
Q7_11	-3.42707	-2.27782	0.02278
Q7_13	-4.50724	-2.20741	0.02733
Q7_6	-3.90890	-2.09830	0.03593

in Table 3 have *p*-values less than .05. Even though the differences are not independent, we would have hoped to see only one difference to have a *p*-value less than or equal to .05.

Some Concluding Remarks

In practice, users will not be able to compare their estimates with full-population statistics as could have been done here. Nevertheless, users will be able to assess whether calibrated estimates from their nonprobability sample are significantly different from analogous probability estimates, as Tables 1

Table 3. Revised differences in estimated domain means of variables of interest after using WTADJX

Variable	Difference	*t*-value	*p*-value
q6_9	5.64438	2.75222	0.00594
q7_22	-4.92017	-2.59090	0.00960
q17_4	4.82479	2.54737	0.01088
q10_1	-4.12273	-2.18439	0.02898
q17_5	-4.40463	-2.14937	0.03165
q1_15	-4.23211	-2.09599	0.03613
q18_5	-3.68356	-1.85614	0.06349
q1_6	1.56700	1.37592	0.16891
q7_14	2.56980	1.31451	0.18874
q18_3	1.53855	1.23678	0.21623
q18_4	1.85522	1.00026	0.31723
q25_11	-1.67931	-0.88706	0.37509
q17_3	-1.04803	-0.72978	0.46556
q18_2	0.36094	0.69342	0.48808
q17_2	0.46723	0.59242	0.55360
q11_4	1.09904	0.57548	0.56499
q10_3	-0.44952	-0.52322	0.60084
q25_13	-0.60311	-0.52262	0.60126
q6_1	0.89665	0.45045	0.65240
q11_1	-0.78112	-0.38946	0.69695

and 3 show. This remains true even if the method used here for choosing which NOI variables needed to be included in the calibration—beginning with a pared list of potential NOI variables without any interactions followed by the use of Bonferroni-adjusted *p*-values to reduce those variables further—proves not to be generally successful.

One thing that we did not consider is variance estimation under the linear-prediction-model framework (which is dubious for binary variables like those in our example). When there are no estimated variable totals from the probability sample used in the calibration equation for the nonprobability sample, and one assumes the errors in the linear models for the probability and nonprobability samples are independent across sample members, the near independence of the residuals in equation (11) suggest the variance estimator developed in the test is nearly unbiased under the prediction-model framework *as long as the probability and nonprobability samples are distinct*. This is an unnecessary assumption under a selection-model framework. When it fails, a delete-a-group jackknife variance estimator (Kott, 2001) may be used with group *membership* of any repeated sample member for both its appearances. When it does not fail, however, replacing $\alpha\left(\mathbf{x}_j^T\hat{\mathbf{g}}\right)$ with its derivative $\alpha'\left(\mathbf{x}_j^T\hat{\mathbf{g}}\right)$ in equation (11) is unnecessary because $\hat{\mathbf{g}}$ is no longer an estimator for a selection-model parameter. Although the selection-model–based WTADJUST and WTADJX make this replacement when estimating variances, to our knowledge, analogous routines in R, such those found in 'Survey' (Lumley, 2023), do not.

We can show that the test being nearly unbiased under the prediction-model framework remains true when there are estimated variable totals from the probability sample used in the calibration equation for the nonprobability sample, although the revised prediction model assumes $\mathrm{E}(y_k)$ is a linear function of the components of $\mathbf{x}_k$ rather than $\mathbf{z}_k$ as in Kott and Chang (2010) (and the expected value of each component of $\mathbf{z}_k$ is a linear function of the components of $\mathbf{x}_k$). The proof of this assertion is beyond the scope of this endeavor.

References

Baker, R., Brick, M., Bates, N., Battaglia, M., Couper, M., Dever, J., Gile, K., & Tourangeau, R. (2013). *Report of the AAPOR task force on non-probability sampling*. American Association for Public Opinion Research. https://doi.org/10.1093/jssam/smt008

Benoit-Bryan, J. & Mulrow, E. (2021). Exploring nonprobability methods with simulations from a common data source: Culture and community in a time of crisis. *JSM Proceedings, Survey Research Methods Section*. American Statistical Association.

Chen, Y., Li, P., & Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, *115*(532), 2011–2021. https://doi.org/10.1080/01621459.2019.1677241

Deville, J., Särndal, C., & Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, *88*(423), 1013–1020. https://doi.org/10.1080/01621459.1993.10476369

Elliott, M., & Valliant, R. (2017). Inference for non-probability samples. *Statistical Science*, *32*(2), 249–264. https://doi.org/10.1214/16-STS598

Kott, P. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, *17*, 521–526.

Kott, P. (2019). A partially successful attempt to integrate a web-recruited cohort into an address-based sample. *Survey Research Methods*, *13*, 95–101.

Kott, P., & Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*, *105*(491), 1265–1275. https://doi.org/10.1198/jasa.2010.tm09016

Kott, P., & Liao, D. (2012). Providing double protection for unit nonresponse with a nonlinear calibration weighting routine. *Survey Research Methods*, *6*, 105–111.

Kott, P., & Liao, D. (2015). One step or two? Calibrating weighting from a complete list frame with nonresponse. *Survey Methodology*, *41*, 165–181.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics*, *22*, 329–349.

Lumley, R. (2023). *Package 'survey'* [Software routine]. https://cran.r-project.org/web/packages/survey/survey.pdf

Mulrow, E. (2022). *Exploring methods for combining probability and nonprobability samples with simulations from a common data source*. https://osf.io/ygpzm/

Robbins, M., Ghosh-Dastidar, B., & Ramchand, R. (2021). Blending probability and nonprobability samples with applications to a survey of military caregivers. *Journal of Survey Statistics and Methodology*, *9*(5), 1114–1145. https://doi.org/10.1093/jssam/smaa037

Royall, R. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, *57*(2), 377–387. https://doi.org/10.1093/biomet/57.2.377

RTI International. (2012). *SUDAAN language manual, release 11.0*. RTI International.

Tille, Y. & Matei, A. (2023). *Package 'sampling'* [Software routine]. http://cran.r-project.org/web/packages/sampling/sampling.pdf (procedure: *calib*, and procedure: *gencalib*).

Valliant, R., & Dever, J. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods & Research*, *40*(1), 105–137. https://doi.org/10.1177/0049124110392533

Appendix. SAS-Callable SUDAAN Code Used in Tables 1 and 2

Table 1: WTADJUST code

```
*The user needs to specify <inputdat>;
*This can be put inside a macro to do a separate run
    for each &var.;
*<numstrat> is a numeric stratification variable from
    the design information;
*<sudcaseid> is a numeric case ID;
*<dwt> is the design weight;
*The values in the postwgt statement are the
    control totals for each variable on the
    class statement repeated for each level of
    domain;
*DOMAIN takes the value 1 for the observations in
    the probability data and the value 2 for the
    observations in the nonprobability data;
*The output variables populate what is in Table 1;

PROC WTADJUST DESIGN = STRWR ADJUST =
    POST DATA = <inputdat> NOTSORTED
    MAXITER=500;
  NEST <numstrat>;
  IDVAR <sudcaseid>;
  WEIGHT <dwt>;
  POSTWEIGHT
   5290 106063 2196
   5290 106063 2196
   157 2077 9480 4771 39074 38429 18643 918
   157 2077 9480 4771 39074 38429 18643 918
   11810 21452 29406 46048 4833
   11810 21452 29406 46048 4833
   3852 3169 416 94984 2592 2254 6282
   3852 3169 416 94984 2592 2254 6282
   1541 108431 3577
   1541 108431 3577
   8391 19824 16859 8235 22560 2999 9180 7100
    18401
   8391 19824 16859 8235 22560 2999 9180 7100
    18401
  ;
  CLASS DOMAIN NUM_Q38 NUM_Q40
    NUM_AGECAT NUM_CENSUSRACE
    METROMICRORECODE REGION9 &var.
    / NOFREQS;
  VAR &var.;
  MODEL _ONE_ = NUM_Q38 NUM_Q40
    NUM_AGECAT NUM_CENSUSRACE
    METROMICRORECODE REGION9 /
    NOINT;
  LOWERBD LOWER_BOUND;
  CENTER CENTER_VAL;
  VDIFFVAR DOMAIN =(1 2);
  OUTPUT MEAN SE_MEAN T_MEAN P_MEAN
    / FILENAME=VDIFF_&var. REPLACE;
RUN;
```

Table 2: WTADJX Code

```
*The code can also be run within a macro across
    multiple values of &var.;
*Same notes as above with regard to <numstrat>,
    <sudcaseid>, <dwt> and DOMAIN;
*The Mi and Ci (i = 1,2,3,4) variables, which
    correspond respectively to the components
    of xk and zk in equation (9), are for the four
    NOI variables with significant differences at
    the Bonferroni-corrected .1 level (p-value
    < .1/32);

PROC WTADJX DESIGN = STRWR ADJUST =
    POST DATA = <inputdat> NOTSORTED
    MAXITER=500;
  NEST <numstrat>;
  IDVAR <sudcaseid>;
  WEIGHT <dwt>;
  LOWERBD lower_bound;
  CENTER center_val;
```

```
    CLASS DOMAIN NUM_Q38 NUM_Q40
          NUM_AGECAT NUM_CENSUSRACE
          METROMICRORECODE REGION9
          &VAR. / NOFREQS;
    VAR &var.;
MODEL _ONE_ = NUM_Q38 NUM_Q40
          NUM_AGECAT NUM_CENSUSRACE
          METROMICRORECODE REGION9 M1
          M2 M3 M4 / NOINT;
    CALVARS NUM_Q38 NUM_Q40 NUM_
          AGECAT NUM_CENSUSRACE
          METROMICRORECODE REGION9 C1
          C2 C3 C4 / NOINT;
    POSTWGT
    5290 106063 2196
    5290 106063 2196
    157 2077 9480 4771 39074 38429 18643 918
    157 2077 9480 4771 39074 38429 18643 918
    11810 21452 29406 46048 4833
    11810 21452 29406 46048 4833
    3852 3169 416 94984 2592 2254 6282
    3852 3169 416 94984 2592 2254 6282
    1541 108431 3577
    1541 108431 3577
    8391 19824 16859 8235 22560 2999 9180 7100
          18401
    8391 19824 16859 8235 22560 2999 9180 7100
          18401
    0 0 0 0
    ;
    VDIFFVAR DOMAIN =(1 2);
    OUTPUT MEAN SE_MEAN T_MEAN P_
          MEAN / FILENAME=POSTADJX_&var.
          REPLACE;
RUN;
```

www.rti.org/rtipress

RTI Press publication MR-0053-2405