# Creating Statistically-Defensible Calibrated Weights for a Blended Sample and Measuring the Accuracy of the Resulting Estimates

Phillip S. Kott[1], Jamie Ridenhaur[2]
[1,2] RTI International, 3040 East Cornwallis Road, P.O. Box 12194,
Research Triangle Park, NC 27709

**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 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. The WTADJX procedure in SUDAAN® is employed to generate those weights and then measure the standard errors or resulting estimated means and totals.

**Key Words:** Selection model, outcome model, double protection, logit function, WTADJX.

## 1. Introduction

Nonprobability surveys samples – one 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 nonprobability sample by first combining that sample, denoted here by $S_0$, with a probability sample $S_1$ covering the same population $U$ but with which it shares no members. After that, one estimates the probability $\gamma_k$ that a population unit $k$ in the blended sample $S = S_0 \cup S_1$ was originally from the nonprobability sample given 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 their $\hat{\gamma}_k$ values, then breaking the sample into cells of nearly equal size, and finally assigning the weight $w_k = \hat{N}_{1c}/n_{0c}$, where $\hat{N}_{1c}$ 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 *et* *al.* (2020) argues 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 $\pi_k$ 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 $\pi_k$ is known for members in the population that is not in $S_1$.

To see how $p_{0k} = \Pr(k \in S_0 \mid \mathbf{z}_k)$ is derived, start with $\pi_k = \Pr(k \in S_1 \mid k \in S ; \mathbf{z}_k) \times \Pr(k \in S \mid \mathbf{z}_k)$, 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 $\gamma_k$.

Robbins *et al.* goes on to 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 the probability sample so that a 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.*) or some other functional form (as suggested by Elliott and Valliant). 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 $S_0 \cap S_1 = \emptyset$.

In Section 2, we 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 need not be collected on the probability sample (only the nonprobability sample).

In Section 3, using the WTADJX procedure in SUDAAN 11 we show how to estimate population means for variables collected from the nonprobability sample and from a blended sample composed of a nonprobability and a probability sample drawn from the same population.

Section 4 explores an example of estimated means based on blending a stratified simple random probability sample and a nonprobability sample drawn from the same population. Section 5 provides a brief discussion of unaddressed issues.

## 2. Solving a Calibration Equation

The model

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

is a selection model. If correctly specified, it models the probability that $k \in U$ is included in the 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) point out that this selection model can often be estimated by solving a calibration equation when each component of the population total $\mathbf{T}_\mathbf{z} = \sum_{k \epsilon U} \mathbf{z}_k$ 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

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

where each component of $\hat{\mathbf{T}}_\mathbf{z}$ 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. That algorithm has been programmed into the SUDAAN 11(R) routines WTADJUST and WTADJX (RTI 2012), the R routines ‘calib’ and ‘gencalib’ in ‘Sampling’ (Tille and Mattei, 2021), and elsewhere. In Section 3, we will describe how to use SUDAAN’s WTADJX for our purposes. Other software packages could be employed 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

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

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 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) with an equation like (3) 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 $\Pr(k \in S_1 | \mathbf{z}_k;\, S_1^*) = \left[1 + \exp\left(\mathbf{z}_k^T \mathbf{g}\right)\right]^{-1}$, where $S_1^*$ is the probability sample before unit nonresponse, replaces equation (1), and $w_k = \pi_k^{-1}[1 + \exp\,(\mathbf{z}_k^T \hat{\mathbf{g}})]$, replaces equation (3), where $\pi_k$ 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 $\Pr(k \in S_0 |\, \mathbf{z}_k) = [1 + \exp\,(\mathbf{z}_k^T \mathbf{g})]^{-1}$ 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 $\mathbf{z}_k \boldsymbol{\beta}$ 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 $\hat{\mathbf{T}}_\mathbf{z}$ 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 $\hat{Y} = \sum_{k \in S_0} w_k\, y_k$ will be nearly unbiased in some sense (technically, $\hat{Y}$ is a predictor, not an estimator, for the random variable $Y$ under the prediction model). The reader is again directed to Kott and Liao for a proof of this assertion. A similarly "doubly robust" approach can be found in in Chen et al. (2020).

When one of more components of $\hat{\mathbf{T}}_{\mathbf{z}}$ is consistently estimated from a probability sample, the near unbiasedness of $\hat{Y}$ requires the combination of probability-sampling inference (for $\hat{\mathbf{T}}_{\mathbf{z}}$) and either the selection model or prediction model (for $\hat{Y}|\hat{\mathbf{T}}_{\mathbf{z}}$). 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 $\bar{y} = \sum_{k\in U} y_k / _{k\in U}\, 1$, when each $y_k \in U$ behaves like a random variables with mean $\mathbf{z}_k\boldsymbol{\beta}$, and 1 is either a component of $\mathbf{z}_k$ or the linear combination of the components of $\mathbf{z}_k$.

As alluded to above, 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

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

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 $\sum_{k\epsilon S_1} d_k [1 + \exp\,(\mathbf{z}_k^T \hat{\mathbf{g}})]\, \mathbf{z}_{\boldsymbol{k}} = \hat{\mathbf{T}}_{\mathbf{z}}$.

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 $\exp\left(\mathbf{z}_k^T \hat{\mathbf{g}}\right)$, where $\hat{\mathbf{g}}$ converges to $\mathbf{0}$ and consequently the $w_k$ 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 and Matei 2013) 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 value greater than 1.

## 3. Calibration Weighting a Blended Sample

Suppose we have a probability and a nonprobability sample both chosen from the same population. We again 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 1 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: $E(y_k) = \mathbf{z}_{1k}^T\boldsymbol{\beta}_1$. 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 expection.

Similarly, a vector $\mathbf{z}_{0k}$ of variables including 1 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: $E(y_k) = \mathbf{z}_{0k}^T\boldsymbol{\beta}_0$ assuming selection into $S_0$ is ignorable in expection.

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 above so 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 $S_1$. 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 (so 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 (variance estimation under the prediction-model framework is discussed in the final section). Moreover, any linear combination of the two estimates is also a nearly unbiased estimator of the population mean and has a standard error 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 strata of the blended sample $S$. Let $\mathbf{z}_k^T = (\mathbf{z}_{1k}^T \ \ \mathbf{z}_{0k}^T)$, where the components of $\mathbf{z}_{1k}^T$ are $\mathbf{z}_{1k}$ for members originally from $S_1$ and 0 for the members orginally $\mathbf{z}_{0k}$. The components of $\mathbf{z}_{0k}^T$ are defined conformally. The $L$ and $U$ parameters are the same for all members of $S_1$ and they are the same for all member of $S_0$, but the former and latter pairs may differ.

Consider the following calibration equation, which can be used to crease quasi-probability weights:

$$\sum\nolimits_{k\epsilon S_1} \pi_k^{-1}\alpha_1\left(\mathbf{z}_{1k}^T\hat{\mathbf{g}}_1\right)\mathbf{z}_{1k}^T + \sum\nolimits_{k\epsilon S_0} \alpha_0\left(\mathbf{z}_{0k}^T\hat{\mathbf{g}}_0\right)\mathbf{z}_{0k}^T = \lambda\mathbf{T}_{\mathbf{z_1}}^T + \mathbf{T}_{\mathbf{z_0}}^T$$

or

$$\sum\nolimits_{k\in S} d_k\, \alpha(\mathbf{z}_k^T\hat{\mathbf{g}})\mathbf{z}_k = \mathbf{T}_{\mathbf{z}(\lambda)} \qquad (6)$$

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

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

$\delta_k$ = 1 when $k$ was originally from $S_1$ and 0 otherwise, and $\mathbf{z}_k^{(\lambda)} = \lambda\mathbf{z}_{1k} + \mathbf{z}_{0k}$ . 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 can not 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

$$w_k = d_k \frac{L_k + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)}{1 + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)/U_k}, \tag{7}$$

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 variables included in the nonprobability sample's selection model,

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

but have unknown population totals that need be estimated by the probability sample. With $\mathbf{x}_k^T = (\mathbf{z}_{1k}^T\ \ \mathbf{z}_{0k}^T\ \ [1-\delta_k]\mathbf{q}_k^T)$ and $\mathbf{z}_k^T = (\mathbf{z}_{1k}^T\ \ \mathbf{z}_{0k}^T\ \ \{1-\delta_k[1 + 1/\lambda]\}\mathbf{q}_k^T)$, WTADJX can be used to estimate the population mean $\bar{y}$ by finding the $\hat{\mathbf{g}}$ satisfying:

$$\sum_{k \in S} d_k \frac{L_k + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)}{1 + \exp\left(\mathbf{x}_k^T\hat{\mathbf{g}}\right)/U_k}\mathbf{z}_k = \begin{pmatrix} \lambda\sum_{k \in U}\mathbf{z}_{1k} \\ \sum_{k \in U}\mathbf{z}_{0k} \\ \mathbf{0} \end{pmatrix}, \tag{8}$$

where $\mathbf{0}$ has as many components as $\mathbf{q}_k$ (so that $\sum_{k \in S_0} w_k\mathbf{q}_k = \sum_{k \in S_1} w_k\mathbf{q}_k$) . One can vary the choice of $\lambda$ in an attempt to find the optimal value that minimizes the standard error of the estimated mean. (Recall that every chose for $\lambda$ results in nearly unbiased estimation.)

To estimate the population total $Y = \sum_{k \in U} y_k$ the quasi-weights described above need to be divided by $1+\lambda$.

After choosing $\lambda$, the components of the vector on the right-hand side of equation (8) 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 indictators of selection into the nonprobability sample are independent of each other. When there is unit nonresponse in the probability sample, an analogous assumption in 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

$$e_k = y_k - \mathbf{z}_k^T \left[\sum_S \alpha'(\mathbf{x}_j^T \hat{\mathbf{g}})\, \mathbf{x}_j \mathbf{z}_j^T\right]^{-1} \sum_S \alpha'(\mathbf{x}_j^T \hat{\mathbf{g}}) \mathbf{x}_j y_j, \tag{9}$$

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 one-step 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 $\lambda$ will likely vary across the variables. Consequently, a compromise will be needed if one desire a single $\lambda$ to be used for all variables.

**Estimation from the Nonprobability Sample**

Suppose *y*-values are only collected from the nonprobability 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 $\lambda$ no longer matters; setting $\lambda = 1$ is a straightforward choice.

## 4. 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 vaguely-described nonprobability samples 4,000 persons were drawn from the population. For our purposes, we will 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) using information from 32 not-of-interest (NOI) survey variables (chosen by us) as well as variables for which the population means are known. The last group includes 9 region indicators (with no missing values), 3 levels of urbanization, an Hispanicity indicator, 6 race categories, 4 age categories, and 7 education-level categories. For all categorical variables except region, a missing survey response is treated as an additional category.

Most of the survey variables of interest and NOI variables are yes/no (1/0) with a missing response treated as a "no" (0). Two of the survey variables of interest were originally on a five-point Likert scale. Missing responses are placed in the least-interested levels with the remaining four levels treated as four separate yes/no variables. Thus, we have for analytical purposes 20 variables of interest whose proportion of 1's we are 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 |
| q_10 | 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 |
| q6_9 | Want more fun in life |
| q11_4 | Offered online materials or activities for kids |
| q10_3 | Miss celebrating cultural heritage |
| q7_4 | Attended community festival in 2019 |
| q6_1 | Want more hope in life |
| q1_6 | Watch movie of tv series in past 30 days |
| q25_13 | Will take art, music, or dance class when able in short term |

<u>The two five-level original variables of interest (and their replacments)</u>

| | |
|---|---|
| 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 of technology museum |
| q7_7 | Natural history museum |
| q7_8 | Public park |
| q7_9 | Architectual tour |
| q7_10 | Public/street art |
| q7_11 | Film festival |
| q7_12 | Music festival |
| q7_13 | Perfomring 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 assign 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 are 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 an intial weight of 1 and set $L = 1$ with for the nonprobability sample failed for technical reasons related to how the SUDAAN program runs rather than the underlying theory. One thing to try (which we have not yet done) is to set the intial weight for each member of the nonprobability sample at 25 (which is slightly less than $N/n_0$ = 113,549/4,0000 = 28.38725, what we actually 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($\mathbf{x}_j^T\mathbf{g}^*$), where 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 1's expressed as a percent) for the variables of interest and the NOI variables. Using a conservative Bonferroni correction because there are so many differences being measured, only those in yellow are deemed significant at the 0.1 level. One variable of intereset has a significant difference as do four NOI variables. Adding those four NOI variables for the selection model for the nonprobability sample and using WTADJX as described in the previous section, produces the differences on the right-hand side of Table 2 for the variables of interest. None are signficant at the 0.1 level after Bonferroni correction.

We did not have time to investigate what $\lambda$ to choose. We did look at $\lambda$ = 1 and compared the estimated $t$-values ($(\hat{p} - p)/\hat{v}^{1/2}$) for the means for the 20 variables of interest computed from the WTADJX-calibrated blended sample with those computed from the WTADJUST-calibrated probability sample. These are displayed in Table 3. For both sets, only one 1 out of 20 of the estimated $t$-values is greater than 2 in absolute value, as we would expect. Table 3 reveals that there are four such when computed from the WTADJUST-calibrated blended sample.

If the 20 estimated means were unbiased and independent, then the sample mean and standard deviation of their $t$-values would be 0 and 1 respectively. They are not independent; still Table 3 computes those values. They are close to perfect for the WTADJUST-calibrated probability sample (0.052 and 1.01), which we know produces nearly unbiased estimates. They are not that good for the WTADJUST-calibrated blended sample (0.380 and 1.36), which we suspect somerimes produces biased estmates. They are much closer for the WTADJX-calibrated blended sample (0.164 and 1.13),

which we hope produces nearly unbiased estimates or, at least, estimates with relatively small biases.
The average effective sample sizes $(p(1-p)/\hat{v})$, assuming heroically that all three methods are unbiased, is approximately 739 for the WTADJUST-calibrated probability sample (which we know is unbiased), over three time higher, approximately 2,390, for the WTADJUST-calibrated blended sample (which evidence suggests is biased), and a bit lower than that, approximatley 2,481, for the WTADJX-calibrated blended sample (which we have reason to believe is either unbiased or has a greatly-reduced bias). This suggest there are gains to be made from incorporating probability and nonprobability samples using WTADJX as described in the previous section.

## 5. Some Concluding Remarks

There remains much to be done, such as choosing a λ for the samples under investigation, assessing how well the method described in the text works for the remaining 9,999 simulated probability and nonprobability samples in the data sets described in Benoit-Bryan and Mulrow (2021), exploring ways to improve the model for inclusion into the nonprobability sample (e.g., by including more NOI variables estimatable from the probability sample, adding interaction terms, or setting bounds on the logit function). There was no probability-sample nonresponse in the data set investigated but the theory develop herein describes how to handle that possibility.

One thing that is not yet discussed is variance estimation under the linear-prediction-model framework. When there is no estimated variable totals from the probability sample used in estimating the probability of inclusion into 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 (9) suggest the variance estimator developed in the test is nearly unbiased under the prediction-model framework so long as the probability and nonprobability samples are distinct. If not, a delete-a-group jackknife variance estimator (Kott, 2001) may be used with group membership of any repeatedly sample member the same for both its appearances. We can show that this remains true when there are estimated variable totals from the probability sample used in estimating the probability of inclusion into the nonprobability sample, although the revised prediction model assumes $E(y_k)$ 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 paper.

Table 1. Differences in Estimated Domain Means (as Percents) Using WTADJUST

| Variable of Interest | Difference | t-value | p-value |
|---|---|---|---|
| q7_22 | -5.87071 | -3.09817 | 0.00196 |
| q17_4 | 5.21195 | 2.76443 | 0.00572 |
| q17_5 | 5.54057 | 2.71230 | 0.00670 |
| q6_9 | 5.39458 | 2.63609 | 0.00841 |
| q10_1 | 4.67418 | 2.49856 | 0.01250 |
| q1_15 | 4.96485 | 2.44432 | 0.01455 |
| q18_5 | 4.74582 | 2.37813 | 0.01744 |
| q18_3 | 2.13906 | 1.73783 | 0.08230 |
| q18_4 | 2.14557 | 1.16103 | 0.24569 |
| q1_6 | 1.21429 | 1.07647 | 0.28177 |
| q11_1 | 2.11664 | 1.05061 | 0.29349 |
| q18_2 | 0.49995 | 0.97542 | 0.32940 |
| q10_3 | 0.83516 | 0.95075 | 0.34178 |
| q25_13 | 1.06669 | 0.90289 | 0.36663 |
| q17_2 | 0.64217 | 0.82035 | 0.41206 |
| q25_11 | 0.95481 | 0.50363 | 0.61455 |
| q11_4 | 0.70367 | 0.36134 | 0.71786 |
| q6_1 | 0.59595 | 0.30044 | 0.76385 |
| q17_3 | 0.29020 | 0.20513 | 0.83748 |
| q7_14 | 0.03445 | 0.01706 | 0.98639 |

| NOI Variables | 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 |
| Q7_19 | 3.93684 | 1.93262 | 0.05334 |
| Q7_16 | 3.23589 | 1.83943 | 0.06591 |
| Q7_33 | 3.43996 | 1.82965 | 0.06736 |
| Q7_7 | 3.48459 | 1.80583 | 0.07101 |
| Q7_2 | 2.40930 | 1.78957 | 0.07358 |
| Q7_10 | -2.82599 | -1.47158 | 0.14120 |
| Q7_12 | -2.36151 | -1.31812 | 0.18753 |
| Q7_23 | 2.25852 | 1.29604 | 0.19502 |
| Q7_18 | 2.48310 | 1.22351 | 0.22119 |
| Q7_25 | 1.50790 | 1.11109 | 0.26658 |
| Q7_15 | 2.06517 | 1.05442 | 0.29174 |
| Q7_3 | 2.01458 | 0.99738 | 0.31863 |
| Q7_29 | 1.87405 | 0.98767 | 0.32336 |
| Q7_20 | 1.39394 | 0.81699 | 0.41397 |
| Q7_31 | 1.29923 | 0.78177 | 0.43438 |
| Q7_8 | 1.22418 | 0.73358 | 0.46324 |
| Q7_30 | 1.20529 | 0.64647 | 0.51801 |
| Q7_34 | 0.68921 | 0.46450 | 0.64231 |
| Q7_17 | 0.53617 | 0.26842 | 0.78838 |
| Q7_9 | 0.25032 | 0.14476 | 0.88490 |
| Q7_21 | 0.09669 | 0.04947 | 0.96054 |
| Q7_26 | 0.05184 | 0.03653 | 0.97086 |

Table 2. Differences in Estimated Domain Means (as Percents) Among the Variables of Interest: Comparing Using WTADJUST to WTADJXT

| Variable of Interest (WTADJUST) | Difference | t-value | p-value | Variable of Interest (WTADJX) | Difference | t-value | p-value |
|---|---|---|---|---|---|---|---|
| q7_22 | -5.87071 | -3.09817 | 0.00196 | q6_9 | 5.57247 | 2.71589 | 0.00663 |
| q17_4 | 5.21195 | 2.76443 | 0.00572 | q7_22 | -4.92017 | -2.59090 | 0.00960 |
| q17_5 | 5.54057 | 2.71230 | 0.00670 | q17_4 | 4.85731 | 2.56490 | 0.01035 |
| q6_9 | 5.39458 | 2.63609 | 0.00841 | q10_1 | -4.34892 | -2.30739 | 0.02107 |
| q10_1 | 4.67418 | 2.49856 | 0.01250 | q1_15 | -4.23211 | -2.09599 | 0.03613 |
| q1_15 | 4.96485 | 2.44432 | 0.01455 | q17_5 | -4.28403 | -2.09036 | 0.03664 |
| q18_5 | 4.74582 | 2.37813 | 0.01744 | q18_5 | -3.65736 | -1.84277 | 0.06542 |
| q18_3 | 2.13906 | 1.73783 | 0.08230 | q1_6 | 1.56700 | 1.37591 | 0.16891 |
| q18_4 | 2.14557 | 1.16103 | 0.24569 | q7_14 | 2.56980 | 1.31450 | 0.18870 |
| q1_6 | 1.21429 | 1.07647 | 0.28177 | q18_3 | 1.55871 | 1.25309 | 0.21023 |
| q11_1 | 2.11664 | 1.05061 | 0.29349 | q18_4 | 1.81761 | 0.98224 | 0.32602 |
| q18_2 | 0.49995 | 0.97542 | 0.32940 | q10_3 | -0.67571 | -0.76722 | 0.44298 |
| q10_3 | 0.83516 | 0.95075 | 0.34178 | q25_11 | -1.45157 | -0.76291 | 0.44555 |
| q25_13 | 1.06669 | 0.90289 | 0.36663 | q17_3 | -1.0170 | -0.70833 | 0.47877 |
| q17_2 | 0.64217 | 0.82035 | 0.41206 | q18_2 | 0.36094 | 0.69342 | 0.48808 |
| q25_11 | 0.95481 | 0.50363 | 0.61455 | q17_2 | 0.49590 | 0.62915 | 0.52928 |
| q11_4 | 0.70367 | 0.36134 | 0.71786 | q11_4 | 1.09904 | 0.57548 | 0.56499 |
| q6_1 | 0.59595 | 0.30044 | 0.76385 | q6_1 | 0.82474 | 0.41475 | 0.67834 |
| q17_3 | 0.29020 | 0.20513 | 0.83748 | q11_1 | -0.78112 | -0.38946 | 0.69695 |
| q7_14 | 0.03445 | 0.01706 | 0.98639 | q25_13 | -0.37537 | -0.31749 | 0.75089 |

Table 3. *t*-values for the Estimated Means of the Variables of Interest Computed Three Ways

| Variable | Probability Sample WTADJUST | Blended Sample WTADJUST (λ=1) | Blended Sample WTADJX (λ=1) |
|---|---|---|---|
| q10_1 | -0.99065 | 0.67878 | 0.50094 |
| q11_1 | 1.00639 | 2.89611 | 2.16935 |
| q25_1 | 0.20710 | 0.88154 | 1.13134 |
| q1_15 | -1.19463 | 0.25967 | -0.09930 |
| q6_9 | 1.67386 | 0.43554 | 0.34664 |
| q11_4 | 1.02645 | 2.24001 | 1.25431 |
| q10_3 | -0.79613 | -0.48440 | -0.66086 |
| q7_14 | -0.08139 | -0.13219 | -1.32624 |
| q7_22 | -0.99647 | 1.28732 | 0.77016 |
| q6_1 | 0.95140 | 1.43630 | 1.31127 |
| q1_6 | 0.81299 | 0.38169 | 0.06789 |
| q25_13 | -0.50996 | -0.01880 | -0.60255 |
| q17_2 | -0.45649 | -1.66936 | -1.46063 |
| q17_3 | -1.34542 | -2.25072 | -1.71539 |
| q17_4 | 2.29803 | 1.48587 | 1.66113 |
| q17_5 | -0.90640 | 1.04816 | 0.42534 |
| q18_2 | 0.33557 | -0.34277 | -0.07007 |
| q18_3 | -0.22834 | -2.16198 | -1.65571 |
| q18_4 | 0.58373 | -0.08872 | 0.08769 |
| q18_5 | -0.35780 | 1.72005 | 1.14356 |
| Mean | 0.052 | 0.380 | 0.164 |
| Standard Deviation | 1.01 | 1.13 | 1.36 |

## References

Baker, R., J, Brick, N. Bates, M. Battaglia, M. Couper, J. Dever, K. Gile, and R. Tourangeau. 2013. *Report of the AAPOR Task Force on Non-Probability Sampling*. Deerfield, IL: AAPOR.

Benoit-Bryan, J. and E. Mulrow. 2021. "Exploring Nonprobability Methods with Simulations from a Common Data Source: Culture and Community in a Time of Crisis." *In JSM Proceedings,* Survey Research Methods Section. Alexandra, VA: American Statistical Association.

Chen, Y., P. Li, and C. Wu. 2020. "Doubly Robust Inference with Nonprobability Survey Samples." *Journal of the American Statistical Association*, 115, 2011-2021.

Elliott, M. and R. Valliant. 2017. "Inference for Non-Probability Samples. *Statistical Science*, 32, 249–264."

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. 2001. "The Delete-a-Group Jackknife." *Journal of Official Statistics,* 17, 521-526.

Kott, P. and T. Chang. 2010. "Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse." *Journal of the American Statistical Association, 105*, 1265–1275.

Kott, P., & Liao, D. 2015. "One Step or Two? Calibrating Weighting from a Complete List Frame with Nonresponse." *Survey Methodology,* 41, 165–181.

Kott, P. and D. Liao. 2012. "Providing Double Protection for Unit Nonresponse with a Nonlinear Calibration Weighting Routine." *Survey Research Methods,* 6, 105–111.

Lee, S. 2006. "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys." *Journal of Official Statistics*, 22, 329-49.

Robbins, M., Ghosh-Dastidar B., and Ramchand, R. 2020. "Blending Probability and Nonprobability Samples with Applications to a Survey of Military Caregivers." *Journal of Survey Statistics and Methodology*, forthcoming. Available at https://doi.org/10.1093/jssam/smaa037.

Royall, R. 1970. "On Finite Population Sampling Theory Under Certain Linear Regression Models." *Biometrika*, 57, 377–387.

RTI International 2012. *SUDAAN Language Manual, Release 11.0.* Research Triangle Park, NC: RTI International.

Tille, Y. and A. Matei. 2013. *Package* '*Sampling.*' A software routine available at http://cran.r-project.org/web/packages/sampling/sampling.pdf (procedure: *calib*, and procedure: *gencalib*).

Valliant, R. and J. Dever. 2011. "Estimating Propensity Adjustments for Volunteer Web Surveys." *Sociology, Methods and Research*, 40, 105–137.