sciendo

Journal of Official Statistics, Vol. 39, No. 1, 2023, pp. 45–78, http://dx.doi.org/10.2478/JOS-2023-0003

Investigating an Alternative for Estimation from a Nonprobability Sample: Matching plus Calibration

Zhan Liu[1] *and Richard Valliant*[2]

Matching a nonprobability sample to a probability sample is one strategy both for selecting the nonprobability units and for weighting them. This approach has been employed in the past to select subsamples of persons from a large panel of volunteers. One method of weighting, introduced here, is to assign a unit in the nonprobability sample the weight from its matched case in the probability sample. The properties of resulting estimators depend on whether the probability sample weights are inverses of selection probabilities or are calibrated. In addition, imperfect matching can cause estimates from the matched sample to be biased so that its weights need to be adjusted, especially when the size of the volunteer panel is small. Calibration weighting combined with matching is one approach to correct bias and reduce variances. We explore the theoretical properties of the matched and matched, calibrated estimators with respect to a quasirandomization distribution that is assumed to describe how units in the nonprobability sample are observed, a superpopulation model for analysis variables collected in the nonprobability sample, and the randomization distribution for the probability sample. Numerical studies using simulated and real data from the 2015 US Behavioral Risk Factor Surveillance Survey are conducted to examine the performance of the alternative estimators.

Key words: Calibration adjustment; doubly robust estimation; nearest neighbour matching; sample matching; target sample; volunteer panels.

1. Introduction

Probability samples have been the standard for finite population estimation for many decades. However, probability samples can have many nonsampling problems like low contact and response rates or missing data for units that do respond. Response rates in US surveys, in particular, have been declining for at least two decades (Brick and Williams 2013). Since nonprobability samples can be faster and cheaper to administer and collect, some organizations are gravitating toward their use (Terhanian and Bremer 2012). Baker et al. (2013) review the reasons that nonprobability samples, like volunteer internet panels, may be used rather than a probability sample. Among them are lower costs and compressed data collection periods. Quick turnaround can be especially important to gauge public well-being in health crises like the COVID-19 pandemic of 2020.

There are a variety of problems with nonprobability samples, especially among persons in a panel that have been recruited to participate in future surveys (e.g., see Baker et al. 2013; Valliant and Dever 2011; Valliant et al. 2018). These include selection bias, coverage error, panel nonresponse, attrition, and measurement error. We concentrate on the use of matching and calibration to adjust for the first two of these−selection bias and

[1] School of Mathematics and Statistics, Hubei University, China. Email: eleen_20040109@163.com
[2] Universitiy of Michigan, Institute for Social Research, 4620 North Park Avenue Apt 1406W, Chevy Chase, Maryland, 20815, U.S.A. Email: valliant@umich.edu

© Statistics Sweden

coverage error. Selection bias occurs if the sample differs from the nonsample in such a way that the sample cannot be projected to the full population without some type of statistical adjustment. Coverage error can occur if, for example, a volunteer panel consists of only persons with access to the Internet, assuming that the entire population of a country is the target of the survey. Other, more subtle forms of coverage error can occur if certain demographic groups would rarely or never participate in the particular type of nonprobability survey being conducted.

Because the selection of a nonprobability sample is not controlled by a survey designer, estimation methods other than standard design-based approaches are needed. At least six alternatives can be considered for weighting and estimation with nonprobability samples:

(1) Naïve method where all units are assigned the same weight,
(2) Quasi-randomization where a pseudo-inclusion probability is estimated for each nonprobability unit,
(3) Superpopulation modeling of analysis variables (Y's),
(4) Doubly robust estimation where quasi-randomization and superpopulation modeling are combined,
(5) Mass imputation of Y's into a probability sample using values from a nonprobability sample to form an imputation model,
(6) Matching of a nonprobability sample to a probability sample whose units are used as donors of weights to the nonprobability sample.

The naïve method of equal weighting is rarely, if ever, appropriate because nonprobability samples are not generally distributed proportionally across demographic or other important groups in the population. Alternatives (2)–(4) were reviewed by Elliott and Valliant (2017) and Valliant (2020) and further studied by Chen et al. (2020). Wang et al. (2020) refined alternative (2) by kernel-smoothing the propensity weights. Alternative (5) was proposed by Kim et al. (2021) and involves fitting an imputation model using data from the nonprobability sample and imputing Y values to the units in the probability sample using that model. The probability sample with imputed values is provided to analysts but not the nonprobability sample. Mass imputation solves the weighting problem by using the weights associated with the probability sample. The dissertation of Wang (2020) studied a version of (6) in which a kernel-smoothing method was used to proportionally distribute the probability sample weights to units in the nonprobability sample. We study another, somewhat simpler version of alternative (6), and particularly address some problems with the method.

1.1. Notation and Models Used for Analysis

Both a probability sample, denoted by S_p and a nonprobability sample, denoted by S_{np} will be used in subsequent sections. The target population for which estimates are made is U and has N units. To examine properties of estimators, three distributions will be used. Expectations taken with respect to the sample design used to select the probability sample S_p will be denoted by a π subscript. The probability of selection of unit i in S_p is π_i. To analyze the nonprobability sample S_{np}, we assume that its units are selected by an unknown quasi-randomization distribution; expectations taken with respect to that

distribution will be specified by a subscript R. The probability that unit j is included in S_{np} is $R(\mathrm{x}_j)$ where x_j is a C-vector of covariates or auxiliaries associated with unit j. To simplify notation in later sections, we set $R(\mathrm{x}_j) \equiv R_j$. The analysis variable Y will also be assumed to be generated by a superpopulation model, ξ. Consider the linear model for Y_i defined by

$$Y_i = \mathbf{x}_i\beta + \varepsilon_i (i \in U), \tag{1}$$

where β is a $C \times 1$ parameter vector, and the ε_i are independent, random errors with mean zero and variance σ_i^2. Theory for nonlinear models can also be worked out, as in Chen et al. (2020), but a linear model is convenient for purposes here. Under model (1), the expected value of the population total, $Y_U = \sum_{i \in U} Y_i$, is $E_\xi(Y_U) = \mathbf{X}_U\beta$ where $\mathbf{X}_U = \sum_{i \in U} \mathrm{x}_i$.

The remainder of the article is organized as follows. Section 2 describes how matching can be applied to obtain basic weights for units in the nonprobability sample and reviews the methods of matching. Section 3 presents the theory for bias and variance in different situations. Section 4 investigates properties of matched estimators when the nonprobability sample is calibrated to population totals of covariates. In Section 5, the sample matching and the calibration adjustment are applied in a simulation study using artificial data. In Section 6, an application to a real population is conducted to evaluate the performance of the proposed estimates. The last section summarizes our findings.

2. Applications of Matching

Sample matching has been an option for estimating treatment effects in causal inference for some time (e.g., see Cochran 1953; Rubin 1973; Rosenbaum and Rubin 1983). Moreover, it has been widely applied in evaluation research, observational studies and epidemiological studies (Rothman et al. 2008). More recently, it also has been applied as a way of identifying a sample in market research, public opinion surveys (e.g., Vavreck and Rivers 2008; Terhanian and Bremer 2012), and other nonprobability sampling surveys, especially using volunteer panel surveys. Baker et al. (2013) review some of the applications of matching in survey sampling. Its purpose in non-probability sampling surveys is to reduce selection bias and to estimate population characteristics. Another application of statistical matching is to overcome the problem of missing data created when some persons do not consent to having their survey responses linked to administrative databases (Gessendorfer et al. 2018).

The basic idea of sample matching in survey sampling is that first a random, probability sample, S_p, is selected from the sampling frame of the target population. This probability sample is matched to a pool of nonprobability cases, for example, a volunteer panel of persons. The resulting matched sample from the nonprobability pool is denoted by S_{np}. The probability sample should have none of the coverage problems of the nonprobability sample. This probability sample is also called a reference sample (Lee 2006) and can be an existing survey (or subsample of one) rather than one specially conducted to serve as the reference. For example, in the US the ACS, American Community Survey (U.S. Census Bureau 2023) is one possibility for a large, well-conducted household, reference survey. The probability sample should be representative of the target population in the sense that it can be used to make unbiased and/or consistent estimates of population quantities.

We assume that S_p does not include the Y variables for which estimates are to be made; these are collected from the nonprobability sample.

The application of matching described by Rivers (2007) is one in which S_p is a simple random sample (*srs*). The nonprobability sample S_{np} is obtained by a one-to-one match of S_p to a much larger pool of nonprobability cases, yielding a set S_{np} of the same size as S_p. Since S_p was treated as an *srs*, every unit in S_{np} was given the same weight. When S_p is an *srs*, the distribution across various characteristics of S_{np} is expected to be the same as that of the population. However, in an evaluation of the nonprobability samples offered by nine commercial vendors, Kennedy et al. (2016) found that a nonprobability sample may still produce biased estimators even though it had the same demographic distribution as the population. In other words, matching to an *srs* S_p to obtain S_{np} can be inadequate without further weighting. Rivers and Bailey (2009) describe an election polling application where the sample was obtained by matching, as described above, but inverses of estimated propensities of being in the nonprobability sample were used as weights.

Sample matching in alternative (6), as applied in this article, fits into the quasi-randomization approach. Each unit in a probability sample is matched to a unit in the nonprobability sample based on a set of covariates. The logical extension of Rivers (2007) is for the probability sample unit to "donate" its weight to the matched, nonprobability sample unit. The intuitive argument to justify this is that if the nonprobability units match the probability units on an extensive list of covariates, then the S_{np} units are exchangeable for the S_p units, S_{np} constitutes the same sort of sample as S_p, and the units in S_{np} can be weighted in the same way. This approach has the advantage of straightforward retention of all analytic data collected in the nonprobability sample unlike alternative (5) which could require a separate imputation model for every Y variable.

The probability sample used for matching can be larger, smaller, or equal in size to the non-probability sample, although the method in which S_{np} is selected to have the same size as S_p has advantages. If a pool of nonprobability units is used that is much larger than the probability sample, finding a close match for each unit in the probability sample may be more feasible. This would be the case when a large panel of volunteers has been accumulated. For example, the method used by many panel vendors now is to stockpile a large set of persons who have agreed to participate in online surveys when requested. For example, (see https://www.ipsos.com/en-au/ipsos-online-access-panels). Toluna, the parent company of Harris Interactive, does something similar. Their promotional material says that it has a panel of over 40 million consumers in 70 markets (see Toluna 2023).

If the nonprobability sample is smaller than S_p, a unit in S_{np} may be matched to more than one unit in S_p, making it unclear how to weight the S_{np} cases. In this article, we assume that the resulting sample size of the matched, nonprobability sample, S_{np}, equals the sample size of the probability sample, S_p, that it is matched against. Denote this sample size by n.

2.1. *Methods of Matching*

Which matching algorithm to use is a question. There are various algorithms, including nearest neighbour matching, caliper and radius matching, stratification and interval matching, as well as kernel and local linear matching (Caliendo and Kopeinig 2008).

Among these matching algorithms, nearest neighbour matching based on Euclidean distance is most straightforward. It contains, as special cases, single nearest neighbour matching without replacement, single nearest neighbour matching with replacement and multiple nearest neighbour matching. In single nearest neighbour matching, for a unit in S_p only one unit from the nonprobability pool can be chosen as its matching unit based on the covariates present in both data sets. If single nearest neighbour matching is done without replacement, a unit in the nonprobability pool can be chosen only once as a match. Single nearest neighbour matching without replacement may, however, have poor performance when the target sample and the volunteer panel have very different covariate distributions (Dehejia and Wahba 2002). To overcome this problem, single nearest neighbour matching with replacement and multiple nearest neighbour matching were proposed to increase the average quality of matching and reduce the bias (Smith and Todd 2005). Other matching methods have been suggested that use more than one unit in the nonprobability pool as the matching unit for an individual in the probability sample. Caliper and radius matching use this approach.

Another issue in sample matching is that the matching process will become relatively more difficult as the number of relevant covariates increases. This is the curse of dimensionality noted by Rosenbaum and Rubin (1983). In order to solve this problem, they propose the propensity score, which is the conditional probability of receiving a treatment given the covariates X, denoted by $p(X) = P(D = 1|X)$, where D is the binary indicator taking either the value 1 (receiving treatment, e.g., participation in a volunteer panel) or 0 (not receiving treatment). Rosenbaum and Rubin (1983) have proved that matching on the propensity score $p(X)$ is also valid when it is valid to match on the covariates X. Compared with direct matching based on all covariates, propensity score matching can reduce multiple dimensions (many covariates) to a single dimension, greatly simplifying the matching process. Consequently, it has been widely used in medical and epidemiological studies, economics, market research and a host of other fields (Schonlau et al. 2009; Baker et al. 2013).

3. Estimation from Matched Sample

In this section, we introduce estimators of means and totals based on the matched sample S_{np} and derive their properties. An estimator of a population total is

$$\hat{Y}_M = \sum_{j \in S_{np}} \tilde{w}_j y_j, \tag{2}$$

where $\tilde{w}_j$ is the weight from the probability sample unit that is matched to unit j in S_{np} and y_j is the Y value observed for that unit. We assume that these weights are appropriately scaled for estimating population totals. In particular, $\hat{N}_M = \sum_{j \in S_{np}} \tilde{w}_j$ is an estimator of N, the size of the target population. The mean of Y is estimated by $\hat{\bar{Y}}_M = \hat{Y}_M / \hat{N}_M$. We will consider two cases of weighting of the probability sample S_p:

(1) The weight used for each unit in S_p is the inverse of the selection probability of that unit, that is, $\tilde{w}_j = \pi_j^{-1}$; the estimator of the total with this weight is denoted by $\hat{Y}_{M1}$ subsequently;

(2) The S_p weights are those for a general regression (GREG) estimator; the estimator with this weight is denoted by $\hat{Y}_{M2}$.

Case (1), in which the S_{np} weight is the inverse of the selection probability of its matched unit in S_p is the simplest baseline case for S_{np} weighting. If $\hat{Y}_{M1}$ is biased in this case, then calibration of S_{np} must correct that flaw. Note that the GREG in case (2) includes the commonly used poststratification estimator and requires that the population total $\mathbf{X}_U$ of the covariate vector be known. Whether the weights above are related to the pseudo-inclusion probabilities of the units in S_{np} largely determines whether $\hat{Y}_{M1}$ and $\hat{Y}_{M2}$ are biased or not as shown below. The arguments given are largely heuristic, although they can be formalized using technical conditions like those in Chen et al. (2020).

Properties of estimators can be calculated in several ways: with respect to the ξ-model only, with respect to the R-distribution only, with respect to the π-distribution, or with respect to a combination of the distributions. In subsequent sections, we compute biases and variances with respect to the combined $R\pi$-distribution. The $R\pi$ calculation is analogous to the design-based calculations used in much of sampling theory. In addition, bias and variance calculations are made with respect to the ξ-model and combined $R\pi\xi$-distributions. The calculations made with respect to the ξ-distribution are conditional on the S_{np} and S_p samples. In principle, ξ calculations are more reflective of the statistical properties for the particular sets of units in S_{np} and S_p.

3.1. Bias of the Matched Estimator for Case (1)

Taking the expectation of $\hat{Y}_{M1} - Y_U$ under case (1) with respect to the pseudo-randomization distribution only gives

$$E_R\left(\hat{Y}_{M1} - Y_U\right) = E_R\left(\sum_{j \in S_{np}} \pi_j^{-1} y_j\right) - Y_U$$
$$= \sum_{j \in U} \frac{R_j}{\pi_j} y_j - Y_U.$$

If $R_j = \pi_j$, then $\hat{Y}_M$ will be R-unbiased. However, this does not have to be true generally. For example, if $R_j = Pr(j \in S_{np} | \mathbf{x}_j)$ is a complicated logistic function of a set of covariates that were not used in determining $\tilde{w}_j$, the estimator is R-biased. In this case, some approaches can be employed to reduce bias. For example, some important covariates can first be selected by establishing the relationship models between the S_p weights and covariates, then sample matching can be conducted based on these important covariates. Another situation leading to R-bias would be when the pseudo-inclusion mechanism is *nonignorable*, that is, $Pr(j \in S_{np} | \mathbf{x}_j, y_j) \neq Pr(j \in S_{np} | \mathbf{x}_j)$. Since in a probability sample, the selection mechanism is always ignorable ($\tilde{w}_j \neq 1/Pr(j \in S_{np} | \mathbf{x}_j, y_j)$) when inclusion in the nonprobability sample depends on Y.

If the expectation is taken over the Y-model, the result is

$$E_\xi\left(\hat{Y}_{M1} - Y_U | S_{np}, S_p\right) = \left(\hat{\mathbf{X}}_{np}(\pi) - \mathbf{X}_U\right)\beta,$$

where $\hat{X}_{np}(\pi) = \sum_{S_{np}} \mathbf{x}_j / \pi_j$. The ξ-bias is non-zero unless $\hat{\mathbf{X}}_{np}(\pi) = \mathbf{X}_U$. If $\hat{\mathbf{X}}_{np}(\pi)$ is an unbiased estimator of $\mathbf{X}_U$ under the quasi-randomization R-distribution, $\hat{Y}_{M1}$ will be

unbiased when averaged over both the $R-$ and ξ-distributions (and, equivalently, over the R, π, and ξ distributions). But, as for $\hat{Y}_{M1}$, $\hat{\mathbf{X}}_{np}(\pi)$ will be biased if the correct R-model is not linked to the S_p weights, that is, if $\tilde{w}_j = \pi_j^{-1} \neq 1/R_j$.

3.2. Bias of the Matched Estimator for Case (2)

If the weights from the probability sample have been calibrated to population totals of some covariates $\mathbf{x}$, the bias calculation changes some what. Take the case of the general regression (GREG) estimator being used for S_p. That is, $\tilde{w}_j = g_j/\pi_j$ where

$$g_j = 1 + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_p^{-1} \mathbf{x}_j / \tilde{\sigma}_j^2 \tag{3}$$

with $\hat{\mathbf{X}}_p = \sum_{S_p} \mathbf{x}_j/\pi_j$ and $\tilde{\mathbf{A}}_p = \sum_{S_p} \mathbf{x}_j \mathbf{x}_j^T / \left(\pi_j \tilde{\sigma}_j^2\right)$. The values of $\tilde{\sigma}_j^2$ are often set to a constant in practice, but for completeness, we include $\tilde{\sigma}_j^2$ in subsequent formulas. If $\tilde{\sigma}_j^2$'s are used in estimators of totals, they will be generally assumed values of the model variances in Equation (1); but, we do not require that $\tilde{\sigma}_j^2 = \sigma_j^2$. Note also that the π_j's must be available separately for each unit in the probability sample in order to recover $\mathbf{A}_p$ separately from the $\tilde{w}_j$. In some public-use files, users may only be presented with the $\tilde{w}_j = g_j/\pi_j$ and not π_j.

The estimator of the total is then

$$\hat{Y}_{M2} = \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_p^{-1} \sum_{j \in S_{np}} \mathbf{x}_j y_j / \left(\pi_j \tilde{\sigma}_j^2\right), \tag{4}$$

where $\hat{Y}_{np}(\pi) = \sum_{j \in S_{np}} y_j/\pi_j$. This ξ-bias is

$$E_\xi\left(\hat{Y}_{M2} - Y_U | S_{np}, S_p\right) = \hat{\mathbf{X}}_{np}(\pi)\beta + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_p^{-1} \tilde{\mathbf{A}}_{np}(\pi)\beta - \mathbf{X}_U\beta \tag{5}$$

where $\tilde{\mathbf{A}}_{np}(\pi) = \sum_{S_{np}} \mathbf{x}_j \mathbf{x}_j^T / \left(\pi_j \tilde{\sigma}_j^2\right)$. Thus $\hat{Y}_{M2}$ is ξ-model biased even though the weights in S_p are calibrated on the x's. The R-expectation (which is also the $R\pi$-expectation) is

$$E_R\left(\hat{Y}_{M2}\right) = \sum_U \frac{R_j}{\pi_j} y_j + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_p^{-1} \sum_U \frac{R_j}{\pi_j} \frac{\mathbf{x}_j y_j}{\tilde{\sigma}_j^2}$$

which is also generally not equal to Y_U.

If $R_j = \pi_j$ and sampling for S_{np} and S_p is ignorable, reasonable assumptions are that $N\tilde{\mathbf{A}}_p^{-1}$ and $N^{-1}\tilde{\mathbf{A}}_{np}(\pi)$ converge in probability to $N^{-1}\tilde{\mathbf{A}}_U = N^{-1}\sum_U \mathbf{x}_j \mathbf{x}_j^T/\tilde{\sigma}_j^2$. (See assumption (v) in the Appendix). In that case, $\tilde{\mathbf{A}}_p^{-1}\tilde{\mathbf{A}}_{np}(\pi) \xrightarrow{p} \mathrm{I}_c$ with $\mathbf{I}_C$ being the $C \times C$ identity matrix, and $E_\xi\left(\hat{Y}_{M2} - Y_U | S_{np}, S_p\right) \to \left(\hat{\mathbf{X}}_{np}(\pi) - \hat{\mathbf{X}}_p\right)\beta$.

Taking the expectation of Equation (5) with respect to the R- and π-distributions shows that $\hat{Y}_{M2}$ is approximately $R\pi\xi$-unbiased, but this depends on $R_j = \pi_j$ for all units in S_{np}. Under the same conditions (i.e., $R_j = \pi_j$ and $N\tilde{\mathbf{A}}_p^{-1}$ and $N^{-1}\tilde{\mathbf{A}}_{np}(\pi)$ converging),

$$E_R\left(\hat{Y}_{M2}\right) \doteq Y_U + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)\tilde{\mathbf{B}}_U$$

where $\tilde{\mathbf{B}}_U = \tilde{\mathbf{A}}_U^{-1} \sum_U \mathbf{x}_j y_j / \tilde{\sigma}_j^2$. Consequently, $E_R E_\pi\left(\hat{Y}_{M2}\right) \doteq Y_U$, assuming that $\hat{\mathbf{X}}_p$ is π-unbiased. Similarly, $E_R E_\pi E_\xi\left(\hat{Y}_{M2} - Y_U\right) = 0$.

The results in Subsections 3.1 and 3.2 can be summarized as follows:

- Case (1), $\tilde{w}_j$ is the inverse of the selection probability for its matched unit in the probability sample, $S_{p,}$ that is, $\tilde{w}_j = \pi_j^{-1}$
  - $\hat{Y}_{M1}$ is ξ-biased when the linear model (1) holds;
  - $\hat{Y}_{M1}$ is $R\pi$-unbiased if $R_j = \pi_j$, that is, the probability of a unit's being in the nonprobability sample, S_{np}, equals its probability of being in the probability sample, S_p;
  - $\hat{Y}_{M1}$ is $R\pi\xi$-unbiased when the linear model (1) holds and $R_j = \pi_j$;
- Case (2), $\tilde{w}_j$ is the GREG weight for its matched unit in S_p
  - $\hat{Y}_{M2}$ is ξ-biased under Equation (1) even though S_p is calibrated on the x's in the model;
  - $\hat{Y}_{M2}$ is R-biased in general;
  - $\hat{Y}_{M2}$ is approximately $R\pi$-unbiased and $R\pi\xi$-unbiased if $R_j = \pi_j$;

The key requirement (in addition to ignorability) for unbiasedness of any type is that the observation probability of a unit in the nonprobability sample should be equal to the selection probability of its matched unit from the probability sample. This seems unlikely to be exactly true in most applications.

3.3. Variance of the Matched Estimator in Case (1)

Since a variance estimator is mainly useful in a situation where a point estimator is unbiased or consistent, we concentrate on the case where $R_j = \pi_j$ and $\hat{Y}_{M1}$ is R-unbiased. Calculation of the variance of $\hat{Y}_{M1}$ with respect to the pseudo-inclusion probability distribution depends on the joint distribution of the indicators, $\{\delta_j\}_{j\in U}$ where $\delta_j = 1$ if $j \in S_{np}$ and 0 if not. If the δ_j have the same joint distribution as that of the indicators for being in the probability sample, then S_{np} can be treated as having the same sample design as S_p. If so, $V_R\big(\hat{Y}_{M1}\big) = V_\pi\big(\hat{Y}_{M1}\big)$, and the variance estimator for $\hat{Y}_{M1}$ would be determined by the sample design for S_p. For example, if the probability sample was a stratified, cluster sample, then the variance estimator appropriate to that design would be used.

When $\tilde{w}_j = \pi_j^{-1}$, a more realistic assumption, given the way that nonprobability samples are often acquired, is to treat the $\{\delta_j\}_{j\in U}$ as being independent. With that assumption, the R-variance can be estimated with a formula appropriate for a Poisson sample. Another option is the formula for a sample selected with replacement and with probabilities equal to $R_j = \pi_j$:

$$v_{R\pi}\big(\hat{Y}_{M1}\big) = \frac{n}{n-1}\sum_{j\in S_{np}}\left(\tilde{w}_j y_j - \frac{1}{n}\sum_{j'\in S_{np}}\tilde{w}_{j'}y_{j'}\right)^2. \tag{6}$$

Because $\hat{Y}_{M1}$ does not depend on S_p, $V_{R\pi}(\hat{Y}_{M1}) = V_R(\hat{Y}_{M1})$ and Equation (6) can be interpreted as an estimator of either. Estimator (6) is convenient because it is the default in survey software packages like R survey, Stata, and SAS. However, as shown in Appendix (Subsection 8.1), $v_{R\pi}$ is a biased estimator of the model variance given below.

The ξ-variance under Equation (1) in case (1) is $V_\xi\big(\hat{Y}_{M1}\big) = \left(\sum_{S_{np}}\sigma_j^2/\pi_j^2\right)$ which can be estimated by

$$v_\xi\big(\hat{Y}_{M1}\big) = \sum_{S_{np}}\frac{e_j^2}{\pi_j^2}, \tag{7}$$

where $e_j^2 = (y_j - \mathbf{x}_j^T \hat{\tilde{\mathbf{B}}}_{np}(\pi))^2$ is an approximately ξ-unbiased estimator of σ_j^2 with

$$\hat{\tilde{\mathbf{B}}}_{np}(\pi) = \left(\sum_{S_{np}} \mathbf{x}_j \mathbf{x}_j^T / \left(\pi_j \tilde{\sigma}_j^2 \right) \right)^{-1} \sum_{S_{np}} \mathbf{x}_j y_j / \left(\pi_j \tilde{\sigma}_j^2 \right)$$

(MacKinnon and White 1985). Note that, because the Y's are not available in the probability sample, we must estimate β from the nonprobability sample.

The $R\pi\xi$-variance, in general, is equal to

$$\begin{aligned} V_{R\pi\xi}\left(\hat{Y}_{M1} | S_p, S_{np}\right) &= E_R\left\{ V_{\pi\xi}\left(\hat{Y}_{M1} | S_p, S_{np}\right) \right\} + V_R\left\{ E_{\pi\xi}\left(\hat{Y}_{M1} | S_p, S_{np}\right) \right\} \\ &= E_R\left\{ E_\pi V_\xi\left(\hat{Y}_{M1} | S_p, S_{np}\right) + V_\pi E_\xi\left(\hat{Y}_{M1} | S_p, S_{np}\right) \right\} + \\ &\quad V_R\left\{ E_\pi E_\xi\left(\hat{Y}_{M1} | S_p, S_{np}\right) \right\}. \end{aligned} \tag{8}$$

For case (1) $\hat{Y}_{M1}$ does not depend on S_p and Equation (8) reduces to $V_{R\pi\xi}\left(\hat{Y}_{M1} | S_p, S_{np}\right) = V_{R\xi}\left(\hat{Y}_{M1} | S_p, S_{np}\right) = E_R\left\{ V_\xi\left(\hat{Y}_{M1} | S_p, S_{np}\right) \right\} + V_R\left\{ E_\xi\left(\hat{Y}_{M1} | S_p, S_{np}\right) \right\}$.

As shown in Appendix (Subsection 8.1) for case (1) with $R_j = \pi_j$, this is

$$V_{R\pi\xi}\left(\hat{Y}_{M1}\right) = \sum_U \frac{\sigma_j^2}{\pi_j} + \beta^T V_R\left(\hat{\mathbf{X}}_{np}\right)\beta. \tag{9}$$

Notice that, even though $\hat{Y}_{M1}$ does not directly depend on x, the $R\pi\xi$-variance does after accounting for the ξ-model structure. Expression (9) can be estimated by

$$\upsilon_{R\pi\xi}\left(\hat{Y}_{M1}\right) = \sum_{S_{np}} \frac{e_j^2}{\pi_j^2} + \hat{\tilde{\mathbf{B}}}_{np}(\pi)^T \upsilon_R\left(\hat{\mathbf{X}}_{np}\right) \hat{\tilde{\mathbf{B}}}_{np}(\pi), \tag{10}$$

where $\upsilon_R\left(\hat{\mathbf{X}}_{np}\right)$ is, for example, a version of Equation (6) adapted to estimate a covariance matrix.

3.4. Variance of the Matched Estimator in Case (2)

The ξ-model variance is $V_\xi\left(\hat{Y}_{M2}\right) = \sum_{S_{np}} \left(g_j/\pi_j\right)^2 \sigma_j^2$, which can be estimated by

$$\upsilon_\xi\left(\hat{Y}_{M2}\right) = \sum_{S_{np}} \left(g_j/\pi_j\right)^2 e_j^2. \tag{11}$$

As noted in Appendix (Subsection 8.2), the estimator of total can be approximated by

$$\hat{Y}_{M2} \doteq \bar{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)\tilde{\mathbf{B}}_U \tag{12}$$

and the approximate $R\pi$ variance is

$$V_{R\pi}\left(\hat{Y}_{M2}\right) \doteq V_R\left(\hat{Y}_{np}(\pi)\right) + \tilde{\mathbf{B}}_U^T V_\pi\left(\hat{\mathbf{X}}_p\right)\tilde{\mathbf{B}}_U, \tag{13}$$

which can be estimated as

$$\upsilon_{R\pi}\left(\hat{Y}_{M2}\right) = \upsilon_R\left(\hat{Y}_{np}(\pi)\right) + \hat{\tilde{\mathbf{B}}}_{np}(\pi)^T \upsilon_\pi\left(\hat{\mathbf{X}}_p\right) \hat{\tilde{\mathbf{B}}}_{np}(\pi). \tag{14}$$

Note that, in the situation studied here, both terms of the variance in Equation (13) have the same order of magnitude, $O(N^2/n)$, since they are based on samples of the same size. Thus,

the $R\pi$-variance is the variance in the nonprobability sample of the estimator with inverse pseudo-inclusion probability weights plus a term reflecting the variance of the estimator of the **x**-totals in the probability sample.

Since $\hat{Y}_{np}(\pi) = \hat{Y}_{M1}$, Equation (13) also implies that the $R\pi$-variance of $\hat{Y}_{M2}$ with calibrated S_p weights is larger than that of the uncalibrated $\hat{Y}_{M1}$. This disagrees with the usual expectation that calibration on an effective predictor of Y reduces variances. To better understand this, note that if the matched x's in S_p and S_{np} were identical, then $\hat{\mathbf{X}}_p = \hat{\mathbf{X}}_{np}$ and the variable part in Equation (12) could be written as a weighted sum over S_{np} of residuals, which can then be used to show that $\hat{Y}_{M2}$ can have a smaller variance than $\hat{Y}_{M1}$. However, with imperfect matching the relationship in Equation (12) becomes more realistic.

As shown in Appendix (Subsection 8.2) the approximate $R\pi\xi$-variance when $R_j = \pi_j$ is

$$V_{R\pi\xi}\left(\hat{Y}_{M2}\right) \doteq \sum_{U} \frac{\sigma_j^2}{\pi_j} + \beta^T V_R\left(\hat{\mathbf{X}}_{np}\right)\beta + \beta^T V_\pi\left(\hat{\mathbf{X}}_p\right)\beta. \tag{15}$$

A natural estimator of Equation (15) is then

$$\upsilon_{R\pi\xi}\left(\hat{Y}_{M2}\right) = \sum_{S_{np}} \left(e_j/\pi_j\right)^2 + \hat{\mathbf{B}}_{np}(\pi)^T \upsilon_R\left(\hat{\mathbf{X}}_{np}\right)\hat{\mathbf{B}}_{np}(\pi) + \hat{\mathbf{B}}_{np}(\pi)^{\mathrm{T}} \upsilon_\pi\left(\hat{\mathbf{X}}_p\right)\hat{\mathbf{B}}_{np}(\pi). \tag{16}$$

Consequently, there are several options for variance estimation for $\hat{Y}_{M2}$ for cases (1) and (2). They can be summarized as:

- Case (1), $\tilde{w}_j = \pi_j^{-1}$
  - Estimate the ξ-variance in Equation (7)
  - Estimate the quasi-randomization ($R\pi$) variance with the with-replacement estimator in Equation (6);
  - Estimate the $R\pi\xi$-model variance with $v_{R\pi\xi}$ in Equation (10);
- Case (2), $\tilde{w}_j =$ GREG weight from S_p
  - Estimate the ξ-variance with Equation (11)
  - Estimate the $R\pi$-variance with Equation (14);
  - Estimate the $R\pi\xi$-variance with Equation (16);

4. Calibration Adjustment After Matching

The $R-$, $R\pi-$, or $R\pi\xi$-bias of the matched estimators, $\hat{Y}_{M1}$ and $\hat{Y}_{M2}$, in Section 3 depend critically on whether $Pr\,(j \in S_{np}|\mathbf{x}_j) = Pr\,(i \in S_p)$ for matched units i and j. Matching on covariates attempts to ensure this; however, there is no guarantee that the condition is satisfied regardless of how extensive the set of covariates is.

Consequently, one might hope that calibrating the weights for the nonprobability sample will provide some bias protection. Suppose that the $\{\tilde{w}_j\}_{j\in S_{np}}$ weights are calibrated to the $\mathbf{X}_U$ population totals using the chi-square distance function associated with a GREG. Using the standard formula from Särndal et al. (1992, eq. 6.5.10), the resulting weight for unit j is

$$w_j^* = \tilde{w}_j\left[1 + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w})\right)^T\left[\mathbf{A}_{np}^*(\tilde{w})\right]^{-1}\mathbf{x}_j/\sigma_j^{*2}\right]$$
$$\equiv \tilde{w}_j g_j^*, \tag{17}$$

where $\hat{\mathbf{X}}_{np}(\tilde{w}) = \sum_{S_{np}} \tilde{w}_j\mathbf{x}_j$ and $\mathbf{A}_{np}^*(\tilde{w}) = \sum_{S_{np}} \tilde{w}_j\mathbf{x}_j\mathbf{x}_j^T/\sigma_j^{*2}$. (Note that σ_j^{*2} does not have to be the same as $\tilde{\sigma}_j^2$ used in constructing the GREG weight in S_p.) As in Section 3, σ_j^{*2} is often set to a constant in which case it drops out of the formula for w_j^*. The matched, calibrated estimator is then

$$\hat{Y}_{MC} = \sum_{S_{np}} w_j^* y_j$$
$$= \sum_{S_{np}} \tilde{w}_j y_j + \left(\mathbf{X}_U - \hat{\mathbf{X}}_M\right)^T\left[\mathbf{A}_{np}^*(\tilde{w})\right]^{-1}\sum_{S_{np}} \tilde{w}_j\mathbf{x}_j y_j/\sigma_j^{*2}$$
$$= \hat{Y}_M + \left(\mathbf{X}_U - \hat{\mathbf{X}}_M\right)^T\hat{\mathbf{B}}_{np}^*(\tilde{w}), \tag{18}$$

where $\hat{\mathbf{B}}_{np}^*(\tilde{w}) = \left[\mathbf{A}_{np}^*(\tilde{w})\right]^{-1}\sum_{S_{np}} \tilde{w}_j\mathbf{x}_j y_j/\sigma_j^{*2}$. As in Section 3, calculations depend on cases (1) and (2) of the $\tilde{w}_j$ weights. When case (1) weights are used from S_p, the calibrated estimator will be denoted by $\hat{Y}_{MC1}$; when case (2) weights are used, $\hat{Y}_{MC2}$ denotes the calibrated estimator in subsequent sections.

4.1. Biases in Case (1)

When $\tilde{w}_j = \pi_j^{-1}$, $\hat{\mathbf{X}}_M = \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)$ and, after calibration, the estimator of the total can be written as

$$\hat{Y}_{MC1} = \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)^T\hat{\mathbf{B}}_{np}^*(\pi),$$

where $\hat{\mathbf{B}}_{np}^*(\pi)$ is the special case of $\hat{\mathbf{B}}_{np}^*(\tilde{w})$ with $\tilde{w}_j = \pi_j^{-1}$. Since $E_\xi\left(\hat{Y}_{np}(\pi)\right) = \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\beta$ under model (1) and $E_\xi\left(\hat{\mathbf{B}}_{np}^*(\pi)\right) = \beta$, $E_\xi\left(\hat{Y}_{MC1} - Y_U\right) = 0$, that is, $\hat{Y}_{MC1}$ is ξ-unbiased. Thus, calibrating on the x's in the ξ-model yields an ξ-unbiased estimator even if $R_j \neq \pi_j$.

To calculate the $R\pi$-expectation, define $\mathbf{B}_U^* = \mathbf{A}_U^{*-1}\left(\sum_U \frac{R_j}{\pi_j}\mathbf{x}_j y_j/\sigma_j^{*2}\right)$ with $\mathbf{A}_U^* = \sum_U \frac{R_j}{\pi_j}\mathbf{x}_j\mathbf{x}_j^T/\sigma_j^{*2}$. By the same type of Taylor series argument as in Särndal et al. (1992, sec. 6.5),

$$\hat{Y}_{MC1} \doteq \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)^T\mathbf{B}_U^*. \tag{19}$$

It follows that $E_R E_\pi\left(\hat{Y}_{MC1}\right) = E_R\left(\hat{Y}_{MC1}\right) \doteq \sum_U R_j y_j/\pi_j + \left(\mathbf{X}_U - \sum_U R_j\mathbf{x}_j/\pi_j\right)^T\mathbf{B}_U^*$. If $R_j = \pi_j$, then $\hat{Y}_{MC1}$ is approximately $R\pi$-unbiased. Another consequence is that, when S_{np} is calibrated with the x's in model if (1) and S_p has case (1) weights, $\hat{Y}_{MC1}$ is $R\pi\xi$-unbiased if $R_j = \pi_j$.

4.2. Biases in Case (2)

In case (2) with $\tilde{w}_j = g_j/\pi_j$ and g_j defined in Equation (3), the matched estimator after calibration equals

$$\hat{Y}_{MC2} = \sum_{S_{np}} g_j^* g_j y_j / \pi_j,$$

where

$$g_j^* = 1 + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\tilde{w})\right)^T \left[\tilde{\mathbf{A}}_{np}^*(\tilde{w})\right]^{-1} \mathbf{x}_j / \sigma_j^{*2}.$$

As show in Appendix (Subsection 8.3), the calibrated estimator of the total is approximately

$$\hat{Y}_{MC2} \doteq \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{B}}_U + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\tilde{w})\right)^T \mathrm{B}_U^*. \tag{20}$$

Using Equation (19), the ξ-expectation is

$$E_\xi\left(\hat{Y}_{MC2}\right) \doteq \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\beta + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \beta + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w})\right)^T \beta,$$

which is not $E_\xi(Y_U) = \mathbf{X}_U^T \beta$. That is, $\hat{Y}_{MC2}$ is ξ-biased. This bias occurs even though the nonprobability sample is calibrated on the x's in the model for Y.

If $R_j = \pi_j$, then $E_R E_\pi\left(\hat{Y}_{MC2}\right) \doteq Y_U$ and $E_R E_\pi E_\xi\left(\hat{Y}_{MC2} - Y_U\right)$ is approximately zero.

The bias results for the matched, calibrated estimators $\hat{Y}_{MC1}$ and $\hat{Y}_{MC2}$ can be summarized as follows:

- Case (1) $\tilde{w}_j = \pi_j^{-1}$ and the $\tilde{w}_j$ are then calibrated to population x-totals
  - When the linear model (1) holds, $\hat{Y}_{MC1}$ is ξ-unbiased regardless of whether $R_j = \pi_j$;
  - $\hat{Y}_{MC1}$ is approximately R-, $R\pi$-, and $R\pi\xi$-unbiased in large samples if $R_j = \pi_j$;
- Case (2), $\tilde{w}_j$ is the GREG weight for its matched unit in S_p and the $\tilde{w}_j$ are then calibrated to population x-totals
  - $\hat{Y}_{MC2}$ is ξ-biased even if Equation (1) holds and the nonprobability sample S_{np} is calibrated on the x's in the model;
  - $\hat{Y}_{MC2}$ is approximately $R\pi$-unbiased in large samples if $R_j = \pi_j$;
  - $\hat{Y}_{MC2}$ is approximately $R\pi\xi$-unbiased in large samples if $R_j = \pi_j$;

If case (1) holds where the weights assigned to matched units are inverses of selection probabilities from S_p, the situation is more straightforward than case (2). R-unbiasedness in case (1) requires that the pseudo-inclusion probabilities can be taken from the probability sample, that is, $R_j = \pi_j$. Nonetheless, in case (1) calibrating the nonprobability sample does produce an ξ-unbiased estimator even if $R_j \neq \pi_j$, as one would hope. However, in case (2) when the weights from the probability sample are calibrated and the nonprobability sample is further calibrated on the same x's, the resulting estimator is not ξ-unbiased.

4.3. Variance of the Matched, Calibrated Estimator in Case (1)

To compute the ξ-model variance, note that the estimator of total can also be written as $\hat{Y}_{MC1} = \sum_{S_{np}} g_j^* y_j / \pi_j$ with g_j^* defined in Equation (17) with $\tilde{w}_j = 1/\pi_j$. The ξ-variance is then

$$V_\xi\left(\hat{Y}_{MC1}\right) = \sum_{S_{np}} \left(\frac{g_j^*}{\pi_j}\right)^2 \sigma_j^2.$$

It follows that the $R\pi\xi$-variance is $V_{R\pi\xi}(\hat{Y}_{MC1}) = \sum_U \left(g_j^{*2}/\pi_j\right)\sigma_j^2$. The ξ-variance can be estimated with

$$v_\xi(\hat{Y}_{MC1}) = \sum_{S_{np}} \left(\frac{g_j^*}{\pi_j}\right)^2 \hat{e}_j^{*2}, \tag{21}$$

where $\hat{e}_j^{*2} = y_j - \mathbf{x}_j^T \hat{\mathbf{B}}_{np}^*(\pi)$ with $\hat{\mathbf{B}}_{np}^*(\pi) = \left(\sum_{S_{np}} \mathbf{x}_j \mathbf{x}_j^T / \left(\pi_j \sigma_j^{*2}\right)\right)^{-1} \sum_{S_{np}} \mathbf{x}_j y_j / \left(\pi_j \sigma_j^{*2}\right)$.

To compute the $R-$ and $R\pi$-variance, we use the approximation in Equation (19). Assume that $R_j = \pi_j$ so that $\hat{Y}_{MC1}$ is R-unbiased. Based on results in Subsection 4.1, the estimator can be approximated as

$$\begin{aligned}
\hat{Y}_{MC1} &= \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)^T \hat{\mathbf{B}}_{np}^*(\pi) \\
&\doteq \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)^T \mathbf{B}_U^* \\
&= \sum_{S_{np}} \pi_j^{-1} e_j^* + \mathbf{X}_U^T \mathbf{B}_U^*,
\end{aligned} \tag{22}$$

where $e_j^* = y_j - \mathbf{x}_j^T \mathbf{B}_U^*$. The R- (and $R\pi$-) variance is, thus, equal to the variance of the first term in the last line of Equation (22). If the sample S_{np} is treated as being selected with replacement, then a variance estimator is

$$v_{R\pi}(\hat{Y}_{MC1}) = \frac{n}{n-1} \sum_{j \in S_{np}} \left(\tilde{w}_j \hat{e}_j^* - \frac{1}{n} \sum_{j' \in S_{np}} \tilde{w}_{j'} \hat{e}_{j'}* \right)^2. \tag{23}$$

4.4. Variance of the Matched, Calibrated Estimator in Case (2)

As shown in Appendix (Subsection 8.4) approximation (20) can be rewritten as

$$\hat{Y}_{MC2} \doteq \sum_{S_{np}} y_j \left(\frac{1}{\pi_j} + F_j\right) + \sum_{U - S_{np}} y_j F_j,$$

where F_j is a term that is $O_p\left(n^{-1/2}\right)$. As a result, $V_\xi(\hat{Y}_{MC2}) \doteq \sum_{S_{np}} \left(\sigma_j/\pi_j\right)^2$, which can be estimated with

$$v_\xi(\hat{Y}_{MC2}) = \sum_{S_{np}} \left(\frac{\hat{e}_j^*}{\pi_j}\right)^2. \tag{24}$$

Rewriting Equation (20), the calibrated estimator of the total is also

$$\hat{Y}_{MC2} \doteq \sum_{S_{np}} \frac{e_j^*}{\pi_j} + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{B}}_U + \mathbf{X}_U^T \mathbf{B}_U^*, \tag{25}$$

where e_j^* was defined above. Using the total variance formula, the $R\pi$-variance can be derived as

$$V_{R\pi}(\hat{Y}_{MC2}) = V_R E_\pi(\hat{Y}_{MC2} | S_{np}) + E_R V_\pi(\hat{Y}_{MC2} | S_{np})$$

$$= V_R\left(\sum_{S_{np}} \frac{e_j^*}{\pi_j}\right) + E_R V_\pi\left[\left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{B}}_U\right]$$

$$= V_R\left(\sum_{S_{np}} \frac{e_j^*}{\pi_j}\right) + \tilde{\mathbf{B}}_U^T V_\pi\left(\hat{\mathbf{X}}_p\right)\tilde{\mathbf{B}}_U.$$

An estimator of this variance is

$$\upsilon_{R\pi}\left(\hat{Y}_{MC2}\right) = \upsilon_R\left(\sum_{S_{np}} \frac{e_j^*}{\pi_j}\right) + \hat{\tilde{\mathbf{B}}}_{np}(\pi)^T \upsilon_\pi\left(\hat{\mathbf{X}}_p\right)\hat{\tilde{\mathbf{B}}}_{np}(\pi) \tag{26}$$

with $\upsilon_R\left(\sum_{S_{np}} \frac{e_j^*}{\pi_j}\right)$ being a variance estimator of an estimated total appropriate to how the nonprobability sample is handled. We use $\hat{\tilde{\mathbf{B}}}_{np}(\pi)$ in Equation (26) rather than an estimator with $\tilde{w}$ weights since the former is expected to be somewhat more stable. If S_{np} is treated as being with-replacement, the first component in Equation (26) can be computed with Equation (23).

Details of calculating $V_{R\pi\xi}\left(\hat{Y}_{MC2}\right)$ are in Appendix (Subsection 8.4). This variance can be estimated with

$$\upsilon_{R\pi\xi}\left(\hat{Y}_{MC2}\right) = \sum_{S_{np}}\left(\frac{\hat{e}_j^*}{\pi_j}\right)^2 + \hat{\tilde{\mathbf{B}}}_{np}(\pi)^T \upsilon_\pi\left(\hat{\mathbf{X}}_p\right)\hat{\tilde{\mathbf{B}}}_{np}(\pi). \tag{27}$$

For each of the variance estimators above for the matched, calibrated estimator in cases (1) and (2), it is important to remember that unless $R_j = \pi_j$ the estimator of total itself will be biased. If so, the mean square error will have a bias-squared component that none of the variance estimators will reflect.

In the combination above, both the weights in S_p and those in S_{np} are calibrated to a given set of x's. This is similar to the situation studied by Rao et al. (2002, 368), who noted that in a regression with calibration weights, GREG residuals are based on the regression of model residuals on $\mathbf{X}$. If the model fits well, there will be very little association between those residuals and $\mathbf{X}$ leading to no gain compared to an estimator not using calibration weights. In our situation, when the estimators of totals are unbiased, we can expect $\hat{Y}_{M2}$ with calibration in S_p, $\hat{Y}_{MC1}$ with no calibration in S_p and calibration in S_{np}, and $\hat{Y}_{MC2}$ with calibration in both S_p and S_{np} to be about equally precise—a point borne out by the simulation in Section 5.

5. Simulation Studies

To study the performance of the proposed estimators described above, we performed two simulation studies with an artificial population. In the first, conditions are created where close matches can be found between units in the probability sample and the nonprobability sample. In the second simulation, close matches are much less likely.

5.1. Simulation Study I

In the simulation, a finite population of size $N = 100{,}000$ was based on the following model:

$$E_\xi(Y) = \alpha + \beta X, V_\xi(Y) = \sigma^2 X^{3/2},$$

where $\alpha = 0.4$, $\beta = 0.25$, $\sigma^2 = 0.0625$, and X follows a gamma distribution with density function $f(x) = 0.04x \exp(-x/5)$. This is the same model as used by Hansen et al. (1983); the function `HMT` in the `R PracTools` package (Valliant et al. 2020) was used to generate the population. Conditional on X, Y follows a gamma distribution with density function $g(y;\ x) = (1/b^c\Gamma(c))y^{c-1} \exp(-y/b)$, where $b = 1.25x^{3/2}(8 + 5x)^{-1}$, $c = 0.04x^{-3/2}(8 + 5x)^2$ and $\Gamma(\cdot)$ is the gamma function. The finite population is stratified into five strata by ranges of the variable X, such that each stratum has approximately the same total of X. A stratified, probability sample S_p of size $n = 250$ is then selected from the population using stratified, simple random sampling (*stsrs*) without replacement, in which the sample stratum sizes are given by (50, 50, 50, 50, 50). Further, a stratified, volunteer panel of size $M = 1{,}250$ is selected from the population with stratum sample sizes (250, 250, 250, 250, 250) using stratified, simple random sampling. Although the volunteer panel is a probability sample, their weights are treated as unknown for the simulation. Note that the sampling fractions of both S_p and S_{np} are small and, thus, will not affect the empirical variances of estimates.

For each unit of the probability sample of $n = 250$, we find the closest matching unit of the volunteer panel to obtain the matched, non-probability sample S_{np} of size $n = 250$, using single nearest neighbor matching without replacement based on the single auxiliary variable X. That is, it matches each unit i in the probability sample with a unit j in the volunteer panel with the closest distance $d(i, j) = \|X_i - X_j\|$ and the unit j in the volunteer panel can be used only once, where $\|\cdot\|$ is a norm, such as 1-norm, 2-norm and ∞-norm. The units in the volunteer panel are then assigned the weight of their nearest neighbor match from the probability sample using the `R` package `Matching` (Sekhon 2011). In this example, finding close matches is fairly easy, and we should have $R_j = \pi_j$, $j \in S_{np}$, in almost all cases because both S_p and S_{np} are *stsrs*. The parameter of interest is the population total of Y. Finally, the matched estimator and the matched, calibrated estimator under cases (1) and (2) in Section 3 are computed, denoted by

- $\hat{Y}_{M1}$, estimator (2) with $1/\pi$ weights from the matched units in S_p,
- $\hat{Y}_{M2}$, estimator (4) with GREG weights from the matched units in S_p,
- $\hat{Y}_{MC1}$, estimator (18) with $1/\pi$ weights from the matched units in S_p followed by calibration in S_{np}, and
- $\hat{Y}_{MC2}$, estimator (18) with GREG weights from the matched units in S_p followed by calibration in S_{np}.

The above process is repeated 5,000 times. The percentage relative biases (relbiases), the variances and the mean squared errors of the matched estimator and the matched, calibrated estimator under cases (1) and (2), are presented in Table 1. The empirical percent relative bias is defined as $100 \times bias(\hat{Y})/Y$.

Table 1. Simulation Study I: Monte Carlo percent relative biases, variances and mean squared errors of the point estimators

Estimators	Relative bias (%)	Variance ($\div 10^7$)	MSE ($\div 10^7$)	Ratio to min MSE
$\hat{Y}_{M1}$	-0.0318	9.02	9.02	1.17
$\hat{Y}_{M2}$	-0.0078	7.68	7.68	1.00
$\hat{Y}_{MC1}$	-0.0274	7.68	7.68	1.00
$\hat{Y}_{MC2}$	-0.0274	7.68	7.68	1.00
$\hat{Y}_{DR}$	-0.1062	9.23	9.24	1.20

For comparison we included a doubly robust estimator, denoted by $\hat{Y}_{DR}$, that was computed without matching. This estimator was computed in two steps as described in Elliott and Valliant (2017). First, an equal probability subsample of $n = 250$ was selected from the volunteer panel of $m = 1{,}250$. Then, S_p and S_{np} are combined. Units in S_{np} are given a weight of 1 while units in S_p were assigned their sampling weight of $1/\pi_i$. A logistic regression with X as the covariate was run to predict the probability of being in S_{np}. The weight for unit j in S_{np} was then calculated as $w_j = \left(1 - \hat{R}_j\right)/\hat{R}_j$ where $\hat{R}_j$ is the predicted probability of being in S_{np} (see Wang et al. 2021). Without the odds transformation, the estimator would be somewhat biased (Chen et al. 2020), but in this case the bias was negligible since S_{np} is a small fraction of the population (Wang et al. 2021). Finally, the estimator was calibrated with a model having an intercept and X. The general formulas for $\hat{Y}_{DR}$ and its variance estimator are

$$\hat{Y}_{DR} = \sum_{S_{np}} g_{DRj} w_j y_j; \quad v_{DR} = \sum_{S_{np}} (g_{DRj} w_j)^2 e_{DRj}^2 \tag{28}$$

where w_j is the estimated weight defined above, $e_{DRj} = y_j - \mathbf{x}_j^T \hat{\mathbf{B}}_{DR}$ with $\hat{\mathbf{B}}_{DR}$ being the weighted least squares estimator of slope with weights equal to w_j, and g_{DRj} is the g-weight defined using the w_j weights. The estimator $\hat{Y}_{DR}$ will be approximately unbiased if either the model for the propensities of being in S_{np} or the model for Y is correct. Two other things to note are that (a) $\hat{Y}_{DR}$ is not the same as the DR estimator considered by Chen et al. (2020) and (b) v_{DR} is theoretically incorrect because it treats the weights w_j as constants rather than estimates. Although Chen et al. (2020) present a variance estimator for their DR estimator, it is not appropriate here since our $\hat{Y}_{DR}$ differs from theirs.

Simulation results in Table 1 show that the absolute relative biases of the matched estimators under the two cases of weights from S_p are small and close to those of the corresponding matched, calibrated estimators under the two cases. The t-statistics, computed across the 5,000 samples, for testing whether the biases of the estimated totals are zero are 0.68 in absolute value or less with the exception of $\hat{Y}_{DR}$ which is -2.26. Since the relbias of the DR estimator is only -0.1062%, a t-statistic of this size is of no practical importance. Thus, both the matched estimators and the matched, calibrated estimators are unbiased when $R_j \cong \pi_j$, $j \in S_{np}$ in both cases (1) and (2) as predicted by the theory in Subsections 3.1 and 3.2. The variances and MSEs of $\hat{Y}_{M2}$, $\hat{Y}_{MC1}$, and $\hat{Y}_{MC2}$ are all equal as anticipated in the comment at the end of Subsection 4.4 and are about 17% smaller than those of $\hat{Y}_{M1}$. Consequently, while all estimates are approximately unbiased, the calibration adjustment after matching produces more efficient estimators compared to only

matching under case (1). Also noteworthy is the fact that the doubly robust estimator, $\hat{Y}_{DR}$, has a 20% larger MSE than the best of the matching estimates. This is a consequence of the logistic model used to estimate $Pr\ (j \in S_{np})$ being a misspecification.

In addition to the point estimators, the variance estimators of the matched estimator, the matched, calibrated estimator under cases (1) and (2), and the doubly robust estimator are also computed according to Equations (6), (7), (10), (11), (14), (16), (21), (23), (24), (26), (27), and (28). In all cases S_{np} is treated as an unstratified, with replacement sample. Percent relative biases (RB) are computed for the variance estimators with respect to the empirical variances (Empvar) and MSEs of the point estimators across the 5,000 simulations:

$$RB.Empvar = \frac{100 \times \left(\sum_{b=1}^{B} v^{(b)}(\hat{Y})/B - V(\hat{Y})\right)}{V(\hat{Y})},$$

$$RB.MSE = \frac{100 \times \left(\sum_{b=1}^{B} v^{(b)}(\hat{Y})/B - MSE(\hat{Y})\right)}{MSE(\hat{Y})},$$

where $V(\hat{Y})$ is the empirical or monte carlo variance of a point estimator $\hat{Y}$, $MSE(\hat{Y})$ is MSE of the point estimator $\hat{Y}$, $v^{(b)}(\hat{Y})$ is a variance estimator of $\hat{Y}$ computed from the b^{th} simulated sample, and $B = 5{,}000$ is the total number of simulation runs. The percent relative biases (RB) and 95% confidence interval (CI) coverages using the normal approximation and the different variance estimates, are presented in Table 2.

With three exceptions, the relbiases in Table 2 are small, ranging from -1.1% to 3.1%. An exception is $v_{\xi}(\hat{Y}_{M1})$ which is a 15.7% underestimate due to the fact that it does not account for the variability of $\hat{X}_{np}$ as shown in Subsection 3.3. The $R\pi$ and $R\pi\xi$ estimators for $\hat{Y}_{M2}$ are about 24% overestimates. As explained in Appendix (Subsection 8.2), these estimators will not fully account for precision gains due to calibration of weights in S_p when the x-matches are extremely close. Confidence interval coverage ranges from 94.8% to 97.0% except for $v_{\xi}(\hat{Y}_{M1})$ which covers in 92.7% of samples due to its underestimation.

Table 2. Simulation Study I: Percent relative biases and 95% confidence interval coverages of the variance estimators

Estimators	RB.Empvar (%)	RB.MSE (%)	CI coverage (%)
$v_{\xi}(\hat{Y}_{M1})$	-15.7	-15.7	92.7
$v_{R}(\hat{Y}_{M1})$	3.1	3.1	95.0
$v_{R\pi\xi}(\hat{Y}_{M1})$	2.8	2.8	96.8
$v_{\xi}(\hat{Y}_{M2})$	-0.9	-0.9	94.8
$v_{R\pi}(\hat{Y}_{M2})$	23.9	23.9	96.8
$v_{R\pi\xi}(\hat{Y}_{M2})$	23.6	23.6	97.0
$v_{\xi}(\hat{Y}_{MC1})$	-1.0	-1.0	94.9
$v_{R}(\hat{Y}_{MC1})$	-0.7	-0.7	94.8
$v_{\xi}(\hat{Y}_{MC2})$	-1.1	-1.1	94.8
$v_{R\pi}(\hat{Y}_{MC2})$	2.2	2.2	95.2
$v_{R\pi\xi}(\hat{Y}_{MC2})$	1.8	1.9	95.2
$v(\hat{Y}_{DR})$	0.4	0.4	94.9

5.2. Simulation Study II

In this simulation, we consider a case in which $R_j \neq \pi_j, j \in S_{np}$. The same finite population of size $N = 100{,}000$ is used as in simulation study I along with a stratified, probability sample S_p of size $n = 250$. A volunteer panel of expected size $m = 1{,}250$ is selected from the population using Poisson sampling with selection probabilities $\pi_i{}'$ defined as follows:

$$\pi_i = 0.085 \exp(-0.085X),$$
$$\pi_i{}' = \frac{m\pi_i}{\sum_{i=1}^{N} \pi_i}.$$

With this definition of π_i, the probability of being in S_{np} decreases with increasing X. This kind of selection for the volunteer panel will generally result in $R_j \neq \pi_i$, for a unit $j \in S_{np}$ matched to a unit $i \in S_p$.

As in simulation I, single nearest neighbor matching without replacement based on the variable X is adopted to conduct matching for the probability sample. The matched estimator, the matched, calibrated estimator, the doubly robust estimator and their variance estimators under cases (1) and (2) are computed. The above procedure is repeated 5,000 times. The relative biases, the variances and the mean squared errors are listed in Table 3. Also, the same relative biases and 95% CI coverages of variance estimators as those in simulation study I are displayed in Table 4.

In Table 3 the matched estimators, $\hat{Y}_{M1}$ and $\hat{Y}_{M2}$, have biases of about -5%. These biases are largely corrected by calibrating with $\hat{Y}_{MC1}$ and $\hat{Y}_{MC2}$. The t-statistics for testing whether the biases are zero are over 100 in absolute value for $\hat{Y}_{M1}$ and $\hat{Y}_{M2}$ but are only about -3 for the calibrated estimators. Despite the somewhat large t-statistics for the calibrated estimators, calibration is very effective in reducing bias due to using incorrect S_{np} inclusion probabilities. The calibrated estimates, consequently, have substantially smaller MSEs than $\hat{Y}_{M1}$ and $\hat{Y}_{M2}$ because of their reduced bias. The doubly robust estimator, $\hat{Y}_{DR}$, is also approximately unbiased; however, its variance and MSE are 50% higher than those of $\hat{Y}_{MC1}$ and $\hat{Y}_{MC2}$.

In Table 4 the variance estimates for $\hat{Y}_{M1}$ and $\hat{Y}_{M2}$ are biased estimates of the empirical variance and severe underestimates of the MSEs. This leads to CIs that cover only about 56% to 66% of the time for the first four variance estimates in Table 4. Since $v_{R\pi}(\hat{Y}_{M2})$ and $v_{R\pi\xi}(\hat{Y}_{M2})$ overestimate the empirical variances by about 22%, their CIs do cover the population totals in 96.8% and 96.9% of samples. The fact that calibrating removes the bias of the matching estimators plus the low biases of the variance estimators for $\hat{Y}_{MC1}$ and

Table 3. Simulation Study II: Percent relative biases, variances and mean squared errors of the point estimators

Estimators	Relative bias (%)	Variance ($\div 10^7$)	MSE ($\div 10^7$)	Ratio to min MSE
$\hat{Y}_{M1}$	-5.2	8.83	31.04	3.9
$\hat{Y}_{M2}$	-5.1	7.99	29.86	3.8
$\hat{Y}_{MC1}$	-1.3	7.66	7.67	1.0
$\hat{Y}_{MC2}$	-1.3	7.66	7.67	1.0
$\hat{Y}_{DR}$	-0.2	11.31	11.35	1.5

Table 4. Simulation Study II: Percent relative biases and 95% confidence interval coverages of the variance estimators

Estimators	RB.Empvar (%)	RB.MSE (%)	CI coverage (%)
$v_{\xi}(\hat{Y}_{M1})$	-18.2	-76.7	56.2
$v_{R}(\hat{Y}_{M1})$	8.7	-69.1	64.9
$v_{R\pi\xi}(\hat{Y}_{M1})$	8.4	-69.2	66.1
$v_{\xi}(\hat{Y}_{M2})$	-9.5	-75.8	57.0
$v_{R\pi}(\hat{Y}_{M2})$	22.9	-67.1	96.8
$v_{R\pi\xi}(\hat{Y}_{M2})$	22.5	-67.2	96.9
$v_{\xi}(\hat{Y}_{MC1})$	1.6	1.4	95.1
$v_{R}(\hat{Y}_{MC1})$	-5.3	-5.4	94.2
$v_{\xi}(\hat{Y}_{MC2})$	-5.6	-5.8	94.5
$v_{R\pi}(\hat{Y}_{MC2})$	-2.4	-2.5	94.5
$v_{R\pi\xi}(\hat{Y}_{MC2})$	-2.7	-2.9	94.5
$v(\hat{Y}_{DR})$	-3.1	-3.4	94.3

$\hat{Y}_{MC2}$ leads to CI coverage of 94.2% to 95.1%. The variance estimator for $\hat{Y}_{DR}$ has CI coverage of 94.3% and performs well. These coverage probabilities are estimated reasonably precisely in the simulation of 5,000 samples. The half-widths of normal approximation 95% CIs on the empirical coverage rates themselves range from 0.62 percentage points when coverage is 95% to 1.39 percentage points when coverage is 60%. (These half-width lengths also apply to the results for Simulation I and for the real data simulation presented in Section 6.)

6. Illustration with Real Population

To further assess the performance of the matching estimators, they are applied to data obtained from the 2015 US Behavioral Risk Factor Surveillance Survey (BRFSS) (Center for Disease Control and Prevention 2023), which is a sample from the US population 18 years and older. The file contains information about whether persons used the internet in the past 30 days (`INTERNET`). The BRFSS is part of a national state-by-state system of surveys used to monitor health conditions in the United States. Data are collected through telephone household interviews. The analytic variables Y in this study are whether respondents were ever diagnosed with a heart attack (`CVDINFR4`), were ever told by a medical professional that they have diabetes (`DIABETE3`), and were ever told they had a stroke (CVDSTRK3). Although each of these analysis variables is binary, use of linear estimators, as studied in previous sections, is standard survey practice, largely because of their convenience for data analysts.

Covariates associated with Y are sex, age, race, marital status, physical weight, employment status, education level, income level, whether respondents smoked at least 100 cigarettes in their entire life, and whether respondents participated in any physical activities or exercises in the past 30 days in 2015. All of the variables are shown in Table 5.

After deleting cases with either a missing, a don't know or a refused response to any of these variables, 315,669 persons are available for this study. Two weights are provided with the data set: `X_WT2RAKE`, which is a design weight and `X_LLCPWT`, which is a raked, final weight. According to the documentation Center for Disease Control and

Table 5. Covariates used in the BRFSS simulation study

Variables	Type	Description
SEX	2 categories	Respondents sex: 1=Male; 2=Female
X_AGE	6 categories	Imputed age in six groups: 1=Age 18 to 24; 2=Age 25 to 34; 3=Age 35 to 44; 4=Age 45 to 54; 5=Age 55 to 64; 6=Age 65 or older
X_RACE	8 categories	Computed race-ethnicity grouping: 1=White only, non-Hispanic; 2=Black only, non-Hispanic; 3=American Indian or Alaskan Native only, Non-Hispanic; 4=Asian only, non-Hispanic; 5=Native Hawaiian or other Pacific Islander only, Non-Hispanic; 6=Other race only, non-Hispanic; 7=Multiracial, non-Hispanic; 8=Hispanic
MARITAL	6 categories	Marital status: 1=Married; 2=Divorced; 3=Widowed; 4=Separated; 5=Never married; 6=A member of an unmarried couple
WEIGHT2	Continuous	Reported weight in pounds: 50–999
EMPLOY1	8 categories	Employment status: 1=Employed for wages; 2=Self-employed; 3=Out of work for 1 year or more; 4=Out of work for less than 1 year; 5=A homemaker; 6=A student; 7=Retired; 8=Unable to work
EDUCA	6 categories	Education level: 1=Never attended school or only kindergarten; 2=Grades 1 through 8 (Elementary); 3=Grades 9 through 11 (Some high school); 4=Grade 12 or GED (High school graduate); 5=College 1 year to 3 years (Some college or technical school); 6=College 4 years or more (College graduate)
INCOME2	8 categories	Income level: 1=Less than USD 10,000; 2=USD 10,000 to less than USD 15,000; 3=USD 15,000 to less than USD 20,000; 4=USD 20,000 to less than USD 25,000; 5=USD 25,000 to less than USD 35,000; 6=USD 35,000 to less than USD 50,000; 7=USD 50,000 to less than USD 75,000; 8=USD 75,000 or more
SMOKE100	2 categories	Smoked at least 100 cigarettes?: 1=Yes; 2=No
EXERANY2	2 categories	Exercise in past 30 days?: 1=Yes; 2=No
INTERNET	2 categories	Internet use in the past 30 days?: 1=Yes; 2=No
CVDINFR4	2 categories	Ever diagnosed with heart attack?: 1=Yes; 2=No
DIABETE3	2 categories	Ever told you have diabetes?: 1=Yes; 2=No
CVDSTRK3	2 categories	Ever told you had a stroke?: 1=Yes; 2=No

Prevention (2017). BRFSS rakes the design weight to eight margins (gender by age group, race/ethnicity, education, marital status, tenure, gender by race/ethnicity, age group by race/ethnicity, and phone ownership). The raking also serves as a noncoverage/nonresponse adjustment. Because of the asymptotic equivalence of the GREG and raked estimators shown by Deville and Särndal (1992), the earlier theory in Sections 3 and 4 should apply to estimators based on `X_LLCPWT`.

In this data set of 315,669 persons, 256,949 people who had used the internet in the past 30 days are considered as the web (nonprobability) subset. Using the `X_LLCPWT` weights, the web population is only 84% (81% unweighted) of the target population, indicating that the effect of coverage error could be substantial. Moreover, the weighted distributions of the

categorical covariates among all respondents in the web, non-web, and full populations are given in Table 6. Categories of some variables are combined in Table 6 and in the simulation compared to the categories in Table 5 because they are small. Table 7 gives the proportions that reported a heart attack, diabetes, or a stroke in the web, non-web, and full populations.

As shown in Tables 6 and 7, there are differences between the target population and the web and non-web populations in the estimated distributions of some of the covariates. For example, 0.19 of the full population are age 65 or older, 0.14 of the web population

Table 6. Distributions of the categorical variables and means of the continuous variable, body weight, in the web, non-web, and full populations

	Variables	Web Population	Non-web Population	Target Population
SEX	Male	0.50	0.49	0.50
	Female	0.50	0.51	0.50
X_AGE	Age 18 to 24	0.13	0.02	0.11
	Age 25 to 34	0.20	0.05	0.18
	Age 35 to 44	0.19	0.09	0.17
	Age 45 to 54	0.19	0.16	0.18
	Age 55 to 64	0.16	0.22	0.17
	Age 65 or older	0.14	0.45	0.19
X_RACE	Non-black, non-Hispanic	0.90	0.84	0.89
	Black only, non-Hispanic	0.10	0.16	0.11
MARITAL	Married or member of an unmarried couple	0.59	0.47	0.58
	Divorced	0.11	0.15	0.11
	Widowed, separated, never married	0.30	0.38	0.31
EMPLOY1	Employed for wages, self-employed	0.65	0.30	0.59
	Out of work	0.05	0.06	0.05
	Other (homemaker, student, retired, unable to work)	0.30	0.63	0.35
EDUCA	Grade 11 or less	0.08	0.39	0.13
	Grade 12 or equivalent	0.25	0.37	0.27
	College 1 year to 3 years	0.35	0.18	0.32
	College 4 years or more	0.33	0.06	0.29
INCOME2	Less than USD 25,000	0.21	0.60	0.27
	USD 25,000 to less than USD 50,000	0.24	0.27	0.24
	USD 50,000 to less than USD 75,000	0.17	0.07	0.16
	USD 75,000 or more	0.38	0.06	0.33
SMOKE100	Smoked at least 100 cigarettes	0.59	0.51	0.58
	Not smoked at least 100 cigarettes	0.41	0.49	0.42
EXERANY2	Exercise in past 30 days	0.22	0.40	0.25
	No exercise in past 30 days	0.78	0.60	0.75
WEIGHT2	Body weight in pounds	180.50	176.90	180.00

Table 7. Proportions of the web, non-web, and total populations that have been told by a medical professional that they have three health conditions

Condition	Web	Non-web	Total pop
Heart attack (CVDINFR4)	0.031	0.107	0.043
Diabetes (DIABETE3)	0.093	0.233	0.116
Stroke (CVDSTRK3)	0.020	0.076	0.029

are, and 0.45 of the non-web are 65+. In the target population, 0.59 are employed for wages, 0.65 are in the web population, but only 0.30 of the non-web are. About 8% of the web population have a grade 11 education or less while 13% of the full population does; 33% of the web population attended four or more years of college while 29% of the full population did. For the analysis variables in Table 7, 4.3% of the target population have ever been diagnosed with a heart attack while 3.1% of the web population and 10.7% of the non-web population have. Similar differences occur for diabetes and stroke. Although the percentage point differences are small between the web and full populations, the relative differences are substantial. For example, heart attacks in the web population are 72% (0.031/0.043) of those in the full population; diabetes in the web population is 80% of the full population rate; strokes in the web population are 72% of those in the full population. Consequently, calibrating the matched sample may reduce bias and variance as long as the covariates in Table 6 are predictive of the Y's. However, it is clear that weighting a sample from the web population will have to achieve a considerable amount of bias correction in order to produce good estimates for the full, target population.

Also noteworthy are the substantial differences between the web and non-web subpopulations. The non-web people are older, more likely to be Black and non-Hispanic, more likely to not be in the labor force, less educated, lower income, and more likely to have smoked than the web persons. The non-web people are also much more likely to have had heart attacks, diabetes, and strokes. Our focus is on using a sample from the web population to make estimates for the full population, but any attempt to use a sample from the web population to represent the non-web population seems doomed to failure. In general, a nonprobability sample that has serious coverage problems cannot be expected to produce good estimates for poorly covered domains.

To apply the proposed matching method, simple random samples are selected from the BRFSS web subsample and from the BRFSS full sample. Using equal probability sampling preserves any differences between the web and full samples and, in particular, any coverage defects in the web sample. The size of the S_p probability sample was $n = 500$ while the size of the initial S_{np} web sample was $M = 3{,}000$. The BRFSS raked weights for persons in S_p were adjusted to equal $\tilde{w}_j = (N/n) * \texttt{X_LLCPWT}$ where $N = 315{,}669$. Since the BRFSS design weights did not include a nonresponse adjustment and, consequently, did not sum to an estimate of the size of the target population, we computed a nonresponse-adjusted design weight for each person in S_p as $\tilde{w}_{\pi j} = (N/n) * \texttt{X_WT2RAKE} * f_{NR}$ where f_{NR} is the sum of `X_LLCPWT` over the sum of `X_WT2RAKE`.

The samples, S_p and S_{np}, are combined and the propensity of being in S_p is estimated via logistic regression. The n closest matches in S_{np}, found using the R package Matching with the same matching criterion as described in Subsection 5.1, are retained for estimation. The

matching reduces the size of S_{np} to be the same ($n = 500$) as that of S_p. The weights $\tilde{w}_j$ and $\tilde{w}_{\pi j}$ from the matching person in S_p are assigned to person j in S_{np}. These weights were used to calculate estimated proportions, $\hat{\bar{Y}}_{M1}$, $\hat{\bar{Y}}_{M2}$, $\hat{\bar{Y}}_{MC1}$, $\hat{\bar{Y}}_{MC2}$, and their associated variance estimators. Estimators of the proportions of persons who reported heart attacks, diabetes, or strokes were computed based on the estimators of totals divided by $\hat{N} = \sum_{S_{np}} \tilde{w}_j$. Because of the way full-sample BRFSS weights are constructed, the variation $\hat{N}$ from sample to sample is minimal so that $\hat{N}$ is treated as a constant for variance estimation.

For $\hat{\bar{Y}}_{MC1}$ and $\hat{\bar{Y}}_{MC2}$ the calibration model used main effects for `SEX`, `X_AGE`, `MARITAL`, `EMPLOY1`, `EDUCA`, `INCOME2`, `EXERANY2`, and `SMOKE100` plus the continuous variable `WEIGHT2`. After some testing, the race variable was not included since it did not improve predictions once the other covariates were in the model. Calibration was done with the R survey package (Lumley 2020).

We also computed two versions of the doubly robust estimator for comparison. The two alternatives differed in the covariates used in the propensity and calibration models. The first, $\hat{\bar{Y}}_{DR1}$, used propensity and calibration models with the same covariates as the calibration model for $\hat{\bar{Y}}_{MC1}$ and $\hat{\bar{Y}}_{MC2}$. The second, $\hat{\bar{Y}}_{DR2}$, used propensity and calibration models that have a much smaller set of covariates: an intercept, `SEX`, `X_AGE`, `EMPLOY1`, and `EDUCA`. Although not reported here, we also ran simulations for a DR estimator that had a propensity model with an intercept, the interactions of `INCOME2` with `X_AGE`, `EDUCA` with `X_AGE`, and `INCOME2` with `EDUCA`. These interactions were determined from a regression tree analysis, and the covariates were recoded for the interactions to be binary. `INCOME2` was recoded to less than or greater than or equal to USD 25,000; `X_AGE` to less than 55 years or greater than or equal to 55 years; `EDUCA` to less than high school or high school or more. The logistic propensity model for being in S_{np} based on the merged data set of S_p and S_{np} was estimated using the method described in Wang et al. (2021). For this DR alternative, the same calibration model was used as f*or* $\hat{\bar{Y}}_{MC1}$ and $\hat{\bar{Y}}_{MC2}$. Because the summary results for this DR estimator were similar to those for $\hat{\bar{Y}}_{DR1}$, we have omitted them.

This process was repeated 5,000 times for each of the three analysis variables. The relative biases, the variances and the mean squared errors (MSEs) of the three point estimators across the 5,000 samples are summarized in Table 8. For all three analysis variables the biases of $\hat{\bar{Y}}_{M1}$ and $\hat{\bar{Y}}_{M2}$ are positive, ranging from 4.8% for diabetes with $M1$ to 15.7% for heart attack for $M2$. Recall that $M1$ is a type of π-estimator with the π-weight taken from the matched case in the probability sample. In this example, $M2$ is a raked estimator with the weight being the raked weight from the matched case in S_p. In contrast, the $MC1$, $MC2$, and $DR1$ estimators have serious negative biases, ranging from -21.6% to -17.5%. The DR estimator with fewer covariates, $\hat{\bar{Y}}_{DR2}$, has the smallest absolute relbias for two of the three variables: -6.5% for heart attack and -9.1% for stroke. Its -9.8% relbias for diabetes is an improvement over that of $\hat{\bar{Y}}_{DR1}$ but is still undesirably large. Although modeling suggested that the extensive list of covariates used in $MC1$, $MC2$, and $DR1$ was needed to correct for coverage bias, the simulations illustrate the well-known phenomenon that overloading a regression model with predictors leads to instability. $DR2$ has the smallest MSE for all three variables and might be preferred on that basis, but its bias results in poor confidence interval coverage as shown next in Table 9. All of the relbiases in Table 8 are highly significantly different from zero based on t-tests.

Table 8. Simulation study with BRFSS population: Monte Carlo percent relative biases, variances and mean squared errors of the point estimators

Estimator	Relative bias (%)	Variance ($\times 10^4$)	MSE ($\times 10^4$)	Ratio to min MSE
Heart attack				
$\hat{\bar{Y}}_{M1}$	12.4	2.57	2.85	1.51
$\hat{\bar{Y}}_{M2}$	15.7	3.76	4.22	2.23
$\hat{\bar{Y}}_{MC1}$	-20.5	1.73	2.51	1.32
$\hat{\bar{Y}}_{MC2}$	-20.3	2.03	2.78	1.47
$\hat{\bar{Y}}_{DR1}$	-21.6	1.61	2.47	1.30
$\hat{\bar{Y}}_{DR2}$	-6.5	1.81	1.89	1.00
Diabetes				
$\hat{\bar{Y}}_{M1}$	4.8	5.71	6.02	1.09
$\hat{\bar{Y}}_{M2}$	6.4	8.07	8.62	1.56
$\hat{\bar{Y}}_{MC1}$	-20.1	4.47	9.88	1.79
$\hat{\bar{Y}}_{MC2}$	-19.8	5.26	10.53	1.91
$\hat{\bar{Y}}_{DR1}$	-20.6	3.78	9.48	1.72
$\hat{\bar{Y}}_{DR2}$	-9.8	4.22	5.52	1.00
Stroke				
$\hat{\bar{Y}}_{M1}$	11.2	1.73	1.83	1.38
$\hat{\bar{Y}}_{M2}$	15.5	2.59	2.80	2.10
$\hat{\bar{Y}}_{MC1}$	-18.4	1.33	1.62	1.21
$\hat{\bar{Y}}_{MC2}$	-17.5	1.59	1.85	1.39
$\hat{\bar{Y}}_{DR1}$	-20.3	1.17	1.52	1.14
$\hat{\bar{Y}}_{DR2}$	-9.1	1.26	1.33	1.00

Table 9. Simulation study with BRFSS population: Percent relative biases and 95% confidence interval coverages of the variance estimators

Estimator	RB.Empvar (%)	RB.MSE (%)	CI coverage (%)
Heart attack			
$v_\xi(\hat{\bar{Y}}_{M1})$	-28.4	-35.5	89.6
$v_{R\pi}(\hat{\bar{Y}}_{M1})$	-4.0	-13.4	92.8
$v_{R\pi\xi}(\hat{\bar{Y}}_{M1})$	-10.1	-18.9	92.6
$v_\xi(\hat{\bar{Y}}_{M2})$	-26	-33.9	89.6
$v_{R\pi}(\hat{\bar{Y}}_{M2})$	-22.7	-31.1	89.5
$v_{R\pi\xi}(\hat{\bar{Y}}_{M2})$	-26.9	-34.8	89.4
$v_\xi(\hat{\bar{Y}}_{MC1})$	7.5	-25.8	81.6
$v_R(\hat{\bar{Y}}_{MC1})$	-14.4	-40.9	73.1
$v_\xi(\hat{\bar{Y}}_{MC2})$	-8.1	-33.1	80.4
$v_{R\pi}(\hat{\bar{Y}}_{MC2})$	-5.0	-30.9	77.2

Table 9. Continued.

Estimator	RB.Empvar (%)	RB.MSE (%)	CI coverage (%)
$v_{R\pi\xi}(\hat{\bar{Y}}_{MC2})$	13.8	-17.2	84.4
$v(\hat{\bar{Y}}_{DR1})$	-22.4	-49.3	68.4
$v(\hat{\bar{Y}}_{DR2})$	-19.6	-22.9	82.2
Diabetes			
$v_{\xi}(\hat{\bar{Y}}_{M1})$	-26.3	-30.1	89.8
$v_{R\pi}(\hat{\bar{Y}}_{M1})$	3.1	-2.2	94.8
$v_{R\pi\xi}(\hat{\bar{Y}}_{M1})$	2.1	-3.1	94.9
$v_{\xi}(\hat{\bar{Y}}_{M2})$	-25.0	-29.8	89.9
$v_{R\pi}(\hat{\bar{Y}}_{M2})$	-8.9	-14.8	92.9
$v_{R\pi\xi}(\hat{\bar{Y}}_{M2})$	-9.7	-15.5	92.9
$v_{\xi}(\hat{\bar{Y}}_{MC1})$	-5.3	-57.2	72.9
$v_{R}(\hat{\bar{Y}}_{MC1})$	-13.4	-60.9	68.4
$v_{\xi}(\hat{\bar{Y}}_{MC2})$	-19.6	-59.8	71.7
$v_{R\pi}(\hat{\bar{Y}}_{MC2})$	1.5	-49.3	75.6
$v_{R\pi\xi}(\hat{\bar{Y}}_{MC2})$	8.3	-45.9	79.4
$v(\hat{\bar{Y}}_{DR1})$	-19.2	-67.8	61.6
$v(\hat{\bar{Y}}_{DR2})$	-18.8	-37.9	80.0
Stroke			
$v_{\xi}(\hat{\bar{Y}}_{M1})$	-25.6	-29.9	87.9
$v_{R\pi}(\hat{\bar{Y}}_{M1})$	-1.8	-7.5	90.1
$v_{R\pi\xi}(\hat{\bar{Y}}_{M1})$	-9.7	-14.9	90.5
$v_{\xi}(\hat{\bar{Y}}_{M2})$	-24.7	-30.2	87.8
$v_{R\pi}(\hat{\bar{Y}}_{M2})$	-24.6	-30.1	87.4
$v_{R\pi\xi}(\hat{\bar{Y}}_{M2})$	-29.8	-35.0	87.4
$v_{\xi}(\hat{\bar{Y}}_{MC1})$	-2.3	-19.8	79.7
$v_{R}(\hat{\bar{Y}}_{MC1})$	-16.1	-31.1	73.2
$v_{\xi}(\hat{\bar{Y}}_{MC2})$	-18.3	-29.8	78.8
$v_{R\pi}(\hat{\bar{Y}}_{MC2})$	-13.3	-25.6	76.9
$v_{R\pi\xi}(\hat{\bar{Y}}_{MC2})$	-1.7	-15.6	82.2
$v(\hat{\bar{Y}}_{DR1})$	-19.7	-38.3	70.1
$v(\hat{\bar{Y}}_{DR2})$	-17.7	-22.1	80.1

Table 9 shows the percent relative biases of the variance estimators with respect to the empirical variance of each estimator of the proportion and with respect to the empirical MSE. These are labeled RB.Empvar (%) and RB.MSE (%). For the most part, the relative biases are negative. With respect to the MSE, all are negative owing to the biases of the point estimators of the proportions which inflate the MSEs. The coverage rates for 95% normal approximation confidence intervals is generally poor because the intervals are centered at the wrong place due to the biases of the estimators of proportions. Only the combinations of $\hat{\bar{Y}}_{M1}$ with $v_{R\pi}$ and $v_{R\pi\xi}$ have coverage rates above 90%.

Finally, as an experiment we also increased the sample sizes to $n = 1{,}000$ for the nonprobability sample and $M = 5{,}000$ for the initial probability sample. The increased sample sizes had no effect on the biases of the point estimates of means. (Results are omitted here.)

In summary, substantial differences between the web and non-web subpopulations result in serious coverage problems for the nonprobability sample S_{np}, causing all estimators to be biased. Furthermore, calibration adjustment after matching does not reduce the biases of the matched estimators. The better of the doubly robust estimators does produce the smallest MSEs in the simulations, but, by no means does it reduce the biases of estimated means to a negligible level.

7. Conclusion

In this article we present several alternative estimators when a nonprobability sample, S_{np}, is matched to a probability sample, S_p. The general setting is that the nonprobability sample is weighted by assigning the weight from an S_p unit to its matched unit in the nonprobability sample. Particular cases are (A) the weight from S_p is its π-weight, (B) the weight from S_p is a GREG weight, (C) case (A) with the nonprobability sample being calibrated with a linear model, and (D) case (B) with S_{np} calibrated with a linear model. Under some restrictive conditions that may be difficult to satisfy in practice, these estimators can be approximately unbiased. The key requirement is that the actual propensity of a unit's being observed in the nonprobability sample should be equal or close to the inclusion probability of the unit to which it is matched in the probability sample.

Three simulation studies illustrated several points about the matched estimator and the doubly robust estimator, which is included for comparison. Study I used artificial data where the variable to be analyzed follows a linear model with a single covariate X, which was also used to create strata. The sample designs for both S_p and S_{np} were stratified simple random sampling with the design for S_{np} treated as unknown. In this case, matching on X was reliable and all estimators were unbiased. In fact, three of four of the matching estimators had a smaller MSE than the doubly robust estimator.

The second simulation used the same artificial population and S_p sample design as Study I, but S_{np} was selected with probabilities (treated as unknown) that decreased with X. In this example, the inclusion probabilities for the nonprobability sample are far from those in the probability sample used for matching. Consequently, the matched estimators without calibration are biased. However, calibration corrects the biases and the matched, calibrated estimator has a smaller MSE than the doubly robust estimator.

The third simulation used a real population (BRFSS) in which persons who had accessed the internet in the previous 30 days were treated as a nonprobability sample from the full US adult population. Since there was no control over how the nonprobability units were selected, this mirrored a situation that would be faced in practice. The prevalence of three health conditions was estimated. The prevalences differed considerably between the part of the population that was covered by S_{np} and the part that was not. The persons who did not use the internet were older, less educated, lower income, and less healthy than the internet users. These differences led to all estimators in the study being biased. Calibrating the matching estimators on a list of covariates did not correct their biases. In addition,

doubly robust (DR) estimation, which has been touted as one of the better options, needs to be carefully implemented to be effective. Including too many covariates in a DR estimator can destabilize the estimator (as in any regression) even though an extensive list of predictors may, in principle, be related to undercoverage of the population by the nonprobability sample.

The failure in the real data study has several, potential contributing factors, including poor matches between the nonprobability and probability units, inadequate models for the propensity of being observed in the nonprobability sample, and poor calibration models for predicting the health characteristics analyzed. However, the facts that the nonprobability sample does not cover the target population, and the noncovered units differ both on the distributions of the analytic variables and covariates is the critical problem. Some diagnostics have been devised for detecting nonignorability of selection of a nonprobability sample (e.g., see Andridge et al. 2019; Little et al. 2019). These diagnostics will signal non-ignorability if the means of covariates in S_{np} and the target population are sufficiently different. Thus, they might be a way forward in the BRFSS application.

However, if the variables to be analyzed differ between S_{np} and the target population but covariate distributions do not, the diagnostics will not alert an analyst to trouble, and poor inferences will still be made from the nonprobability sample. The type of coverage error in the BRFSS study is an example of what can happen in nonprobability samples, generally, and may be a problem that no amount of sophisticated mathematics is likely to correct.

8. Appendix

This appendix shows the details of variance calculations given in earlier sections. Several assumptions are used in the results below. These apply as N and $n \to \infty$.

(1) $\pi_j = O(N/n)$, $R_j = O(N/n)$ and $n/N \to 0$.
(2) $\tilde{\mathbf{A}}_U$ and $\tilde{\mathbf{A}}_U^*$ are $O(N)$.
(3) $V_\pi(\mathbf{X}_p) = O(N^2/n)$.
(4) $V_R(\mathbf{X}_{np}) = O(N^2/n)$.
(5) When $R_j = \pi_j$, $N^{-1}\tilde{\mathbf{A}}_p$ and $N^{-1}\tilde{\mathbf{A}}_{np}(\pi)$ both converge in probability to $N^{-1}\tilde{\mathbf{A}}_u = N^{-1}\sum_U \mathbf{x}_j\mathbf{x}_j^T/\tilde{\sigma}_j^2$.
(6) $N^{-1}\mathbf{A}_{np}^*(\tilde{w})$ converges in probability to $N^{-1}\tilde{\mathbf{A}}_u^* = N^{-1}\sum_U \mathbf{x}_j\mathbf{x}_j^T/\tilde{\sigma}_j^*$.
(7) When $R_j = \pi_j$, $\tilde{\mathbf{A}}_p^{-1}\sum_{S_{np}} \frac{\mathbf{x}_j y_j}{\pi_j \tilde{\sigma}_j^2} \xrightarrow{p} \tilde{\mathbf{B}}_U$ and $\left[\tilde{\mathbf{A}}_{np}^*(\tilde{w})\right]^{-1}\sum_{S_{np}} \frac{\mathbf{x}_j y_j}{\pi_j \sigma_j^{*2}} \xrightarrow{p} \mathbf{B}_U^*$.
(8) When $R_j = \pi_j$, $\sqrt{n}\,(\hat{\mathbf{X}}_p - \mathbf{X}_U)/N$, $\sqrt{n}\,(\hat{\mathbf{X}}_{np}(\pi) - \mathbf{X}_U)/N$, and $\sqrt{n}\,(\hat{\mathbf{X}}_{np}(\tilde{w}) - \mathbf{X}_U)/N$ are asymptotically multivariate normal with mean $\mathbf{0}$.

8.1. ξ-expectation of the With-Replacement Variance Estimator under Case (1)

To compute the ξ-expectation of $v_R\left(\hat{Y}_{M1}\right)$ in Subsection 3.3 under case (1), define $r_j = \tilde{w}_j y_j - \frac{1}{n}\sum_{j' \in S_{np}} \tilde{w}_{j'} y_{j'}$. Since $\tilde{w}_j = \pi_j^{-1}$, this can be rewritten as

$$r_j = \frac{n-1}{n}\frac{y_j}{\pi_j} - \frac{1}{n}\sum_{j' \neq j \in S_{np}} \frac{y_{j'}}{\pi_{j'}}.$$

The ξ-expectation of r_j^2 is then

$$E_\xi\left(r_j^2\right) = V_\xi(r_j) + \left[E_\xi(r_j)\right]^2$$

$$= \left(\frac{n-1}{n}\right)^2 \frac{\sigma_j^2}{\pi_j^2} + \frac{1}{n^2}\left(\sum_{j' \neq j \in S_{np}} \frac{\sigma_{j'}^2}{\pi_{j'}^2}\right) + \left\{\left(\frac{\mathbf{x}_j}{\pi_j} - \frac{1}{n}\sum_{j' \in S_{np}} \frac{\mathbf{x}_{j'}}{\pi_{j'}}\right)^T \beta\right\}^2.$$

Adding and subtracting σ_j^2/π_j^2 in the second term, summing over S_{np}, and doing some algebra leads to

$$E_\xi(v_{R\pi}) = \sum_{S_{np}} \frac{\sigma_j^2}{\pi_j^2} + \frac{n}{n-1}\sum_{j \in S_{np}}\left\{\left(\frac{\mathbf{x}_j}{\pi_j} - \frac{1}{n}\sum_{j' \in S_{np}} \frac{\mathbf{x}_{j'}}{\pi_{j'}}\right)^T \beta\right\}^2$$

as noted in Subsection 3.3. That is, $v_{R\pi}$ is an overestimate of the model variance under Equation (1). However, because $\hat{Y}_{M1}$ is model-biased, $v_{R\pi}$ will not appropriately estimate the ξ mean square error despite its overestimating the ξ-variance.

To derive the $R\pi\xi$-variance, note that

$$V_{R\pi\xi}\left(\hat{Y}_{M1}|S_p, S_{np}\right) = V_{R\xi}\left(\hat{Y}_{M1}|S_p, S_{np}\right)$$

$$= E_R\left\{V_\xi\left(\hat{Y}_{M1}|S_p, S_{np}\right)\right\} + V_R\left\{E_\xi\left(\hat{Y}_{M1}|S_p, S_{np}\right)\right\}.$$

Using the independence of the Y's under Equation (1), the first term is $\sum_U \sigma_j^2/\pi_j$. The second term is $V_R\left\{E_\xi\left(\hat{Y}_{M1}|S_p, S_{np}\right)\right\} = V_R\left(\hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)^T\beta\right) = \beta^T V_R\left(\hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)\beta$. Combining gives the expression shown in Equation (9).

8.2. *Variance of Matched Estimator $\hat{Y}_{M2}$ Under Case (2)*

Following similar steps to those in Särndal et al. (1992, sec. 6.6) and using condition (7), $\hat{Y}_{M2}$ can be approximated as

$$\hat{Y}_{M2} \doteq \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)\tilde{\mathbf{B}}_U, \tag{29}$$

where $\tilde{\mathbf{B}}_U = \left(\sum_U \frac{\mathbf{x}_j\mathbf{x}_j^T}{\pi_j\sigma_j^2}\right)^{-1}\left(\sum_U \frac{\mathbf{x}_j y_j}{\pi_j\sigma_j^2}\right)$.

Using the formula for total variance across the R and π distributions (denoted by $V_{R\pi}$) gives

$$V_{R\pi}(\hat{Y}_{M2}) = E_R V_\pi(\hat{Y}_{M2}|S_{np}) + V_R E_\pi(\hat{Y}_{M2}|S_{np}). \tag{30}$$

Working term by term in Equation (30) and using the approximation to $\hat{Y}_{M2}$ in Equation (29), we have

$$E_R V_\pi(\hat{Y}_{M2}|S_{np}) \doteq E_R V_\pi\left(\hat{Y}_{np}(\pi) + (\mathbf{X}_U - \hat{\mathbf{X}}_p)^T\tilde{\mathbf{B}}_U|S_{np}\right) = \tilde{\mathbf{B}}_U^T V_\pi(\hat{\mathbf{X}}_p)\tilde{\mathbf{B}}_U,$$

because $\hat{Y}_{np}(\pi)$ has zero R-variance given that S_{np} is fixed. To get the second term in Equation (30), note that $V_R E_\pi\left(\hat{Y}_{M2}|S_{np}\right) \doteq V_R\left(\hat{Y}_{np}(\pi)\right)$ assuming that $\hat{\mathbf{X}}_p$ is π-unbiased.

Combining these results, the variance across the R- and π-distributions is

$$V_{R\pi}\left(\hat{Y}_{M2}\right) \doteq V_R\left(\hat{Y}_{np}(\pi)\right) + \tilde{\mathbf{B}}_U^T V_\pi\left(\hat{\mathbf{X}}_P\right)\tilde{\mathbf{B}}_U.$$

as shown in Equation (13).

Turning to the $R\pi\xi$-variance, the total variance formula is given by Equation (8). The $E_\pi V_\xi$ term is $\sum_{S_{np}} \tilde{\sigma}_j^2 E_\pi\left(g_j^2\right)/\pi_j^2$. Using a Taylor series approximation as in Särndal et al. (1992, sec. 6.6), we have

$$g_j \doteq \pi_j^{-1}\left[1 + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_U^{-1}\mathbf{x}_j/\tilde{\sigma}_j^2\right]. \tag{31}$$

It follows that

$$\begin{aligned} E_\pi V_\xi\left(\hat{Y}_{M2}\right) &\doteq \sum_{S_{np}} \frac{\sigma_j^2}{\pi_j^2}\left\{1 + \frac{\mathbf{x}_j^T}{\tilde{\sigma}_j^2}\tilde{\mathbf{A}}_U^{-1} E_\pi\left[\left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)\left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T\right]\tilde{\mathbf{A}}_U^{-1}\frac{\mathbf{x}_j}{\tilde{\sigma}_j^2}\right\} \\ &= \sum_{S_{np}} \frac{\sigma_j^2}{\pi_j^2}\left\{1 + \frac{\mathbf{x}_j^T}{\tilde{\sigma}_j^2}\tilde{\mathbf{A}}_U^{-1} V_\pi\left(\hat{\mathbf{X}}_p\right)\tilde{\mathbf{A}}_U^{-1}\frac{\mathbf{x}_j}{\tilde{\sigma}_j^2}\right\}. \end{aligned} \tag{32}$$

Thus,

$$E_R E_\pi V_\xi\left(\hat{Y}_{M2}\right) = \sum_U R_j \frac{\sigma_j^2}{\pi_j^2} + \sum_U R_j \frac{\sigma_j^2}{\pi_j^2}\frac{\mathbf{x}_j^T}{\tilde{\sigma}_j^2}\tilde{\mathbf{A}}_U^{-1} V_\pi\left(\hat{\mathbf{X}}_p\right)\tilde{\mathbf{A}}_U^{-1}\frac{\mathbf{x}_j}{\tilde{\sigma}_j^2}.$$

Under the order assumptions at the beginning of this appendix, the first term above is $O(N^2/n)$ while the second is $O(N^2/n^2)$. Thus, we use the approximation $E_R E_\pi V_\xi\left(\hat{Y}_{M2}\right) \doteq \sum_U R_j\sigma_j^2/\pi_j^2$.

The second term in Equation (8) is $E_R V_\pi E_\xi\left(\hat{Y}_{M2}\right)$. Expanding and collecting terms gives

$$\begin{aligned} V_\pi E_\xi\left(\hat{Y}_{M2}\right) &= V_\pi\left(\sum_{S_{np}} \frac{g_j}{\pi_j}\mathbf{x}_j\beta\right) \\ &= V_\pi\left(\sum_{S_{np}} \frac{\mathbf{x}_j}{\pi_j}\beta + \left(\mathbf{X}_U - \mathbf{X}_p\right)^T \tilde{\mathbf{A}}_p^{-1}\tilde{\mathbf{A}}_{np}\beta\right). \end{aligned}$$

Under condition (v) above, $\tilde{\mathbf{A}}_p^{-1}\tilde{\mathbf{A}}_{np}$ converges to the $C \times C$ identity matrix and $E_R V_\pi E_\xi\left(\hat{Y}_{M2}\right) = \beta^T V_\pi\left(\hat{\mathbf{X}}_p\right)\beta$.

The third term in Equation (8) gives $V_R E_\pi E_\xi\left(\hat{Y}_{M2}\right)$. Frist, compute $E_\pi E_\xi\left(\hat{Y}_{M2}\right) = E_\pi\left(\sum_{S_{np}} \frac{g_j}{\pi_j}\mathbf{x}_j\beta\right)$. Using the approximation to g_j in Equation (31), $E_\pi\left(g_j\right) \doteq 1$ and $E_\pi E_\xi\left(\hat{Y}_{M2}\right) \doteq \hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\beta$. Consequently, the third term is $V_R E_\pi E_\xi\left(\hat{Y}_{M2}\right) \doteq \beta^T V_R\left(\hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)\beta$.

Combining results for the three terms in Equation (8) gives

$$V_{R\pi\xi}\left(\hat{Y}_{M2}\right) \doteq \sum_{U} R_j \frac{\sigma_j^2}{\pi_j^2} + \beta^T V_{\pi}\left(\hat{\mathbf{X}}_p\right)\beta + \beta^T V_R\left(\hat{\mathbf{X}}_{\boldsymbol{np}}(\pi)\right)\beta$$

as shown in Equation (15).

8.3. Approximation to $\left(\hat{Y}_{MC2}\right)$ in Case (2)

When S_p has case (2) weights, $\tilde{w}_j = g_j/\pi_j$ with g_j defined in Equation (3). The matched estimator after calibration then equals

$$\hat{Y}_{MC2} = \sum_{S_{np}} g_j^* g_j y_j/\pi_j,$$

where

$$g_j^* = 1 + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\tilde{w})\right)^T \left[\tilde{\mathbf{A}}_{np}^*(\tilde{w})\right]^{-1} \mathbf{x}_j/\sigma_j^{*2}.$$

Multiplying g_j^* by g_j defined in Equation (3) and substituting in the formula for $\hat{Y}_{MC2}$ gives

$$\hat{Y}_{MC2} = \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_p^{-1} \sum_{S_{np}} \frac{\mathbf{x}_j y_j}{\pi_j \tilde{\sigma}_j^2} + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w})\right)^T \left[\tilde{\mathbf{A}}_{np}^*(\tilde{w})\right]^{-1} \sum_{S_{np}} \frac{\mathbf{x}_j y_j}{\pi_j \sigma_j^{*2}}$$

$$+ \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_p^{-1} \sum_{S_{np}} \frac{\mathbf{x}_j \mathbf{x}_j^T y_j}{\pi_j \tilde{\sigma}_j^2 \sigma_j^{*2}} \left[\tilde{\mathbf{A}}_{np}^*(\tilde{w})\right]^{-1} \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w})\right). \tag{33}$$

Using conditions (5), (6), and (7), the orders of the second, third, and fourth terms in Equation (33) are $O_p(N/\sqrt{n})$, $O_p(N/\sqrt{n})$, and $O_p(N\sqrt{n})$. The calibrated estimator can then be approximated by

$$\hat{Y}_{MC2} \doteq \hat{Y}_{np}(\pi) + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \mathrm{B}_U + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{\boldsymbol{np}}(\tilde{w})\right)^T \mathrm{B}_U^*. \tag{34}$$

8.4. Variance of Matched Estimator $\hat{Y}_{MC2}$ in Case (2)

To compute the ξ model variance under case (2), we break $\sum_U \mathbf{x}_j y_j/\tilde{\sigma}_j^2$ and $\sum_U \mathbf{x}_j y_j/\sigma_j^{*2}$ into sums over S_{np} and $U - S_{np}$. Equation (34) can then be expressed as

$$\hat{Y}_{MC2} \doteq \sum_{S_{np}} y_j \left(\frac{1}{\pi_j} + F_j\right) + \sum_{U-S_{np}} y_j F_j,$$

where

$$F_j = \left(\mathbf{X}_U - \hat{\mathbf{X}}_p\right)^T \tilde{\mathbf{A}}_U^{-1} \frac{\mathbf{x}_j}{\tilde{\sigma}_j^2} + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w})\right)^T \tilde{\mathbf{A}}_U^{*-1} \frac{\mathbf{x}_j}{\tilde{\sigma}_j^2}\ .$$

Applying conditions (2) and (8), $F_j = O_p(n^{-1/2})$. Since units in S_{np} and $U - S_{np}$ are independent under model (1), the ξ-variance is

$$V_{\xi}\left(\hat{Y}_{MC2}\right) \doteq \sum_{S_{np}} \sigma_j^2 \left(\frac{1}{\pi_j} + F_j\right)^2 + \sum_{U-S_{np}} \sigma_j^2 F_j^2 \tag{35}$$

$$= \sum_{S_{np}} \left(\frac{\sigma_j^2}{\pi_j} \right)^2 \left[1 + O_p\left(N/n^{3/2} \right) \right]. \tag{36}$$

The $R\pi\xi$-variance can be calculated using the total variance formula in Equation (8). First, when $R_j = \pi_j$, $E_R E_\pi V_\xi \left(\hat{Y}_{MC2} | S_p, S_{np} \right) = \sum_U \left(\sigma_j^2 / \pi_j \right)$. The second term in Equation (8) is

$$E_R V_\pi E_\xi \left(\hat{Y}_{MC2} \right) \doteq E_R V_\pi \left[\hat{\mathbf{X}}_{np}(\pi)^T \beta + \left(\mathbf{X}_U - \hat{\mathbf{X}}_p \right)^T \beta + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w}) \right)^T \beta \right]$$

$$= \beta^T V_\pi \left(\hat{\mathbf{X}}_p \right) \beta.$$

The third term in Equation (8) is

$$V_R \left[E_\pi E_\xi \left(\hat{Y}_{MC2} \right) \right] = V_R E_\pi \left[\hat{\mathbf{X}}_{np}(\pi)^T \beta + \left(\mathbf{X}_U - \hat{\mathbf{X}} p \right)^T \beta + \left(\mathbf{X}_U - \hat{\mathbf{X}}_{np}(\tilde{w}) \right)^T \beta \right]$$

$$= V_R \left\{ E_\pi \left[\hat{\mathbf{X}}_{np}(\pi) - \hat{\mathbf{X}}_{np}(\tilde{w}) \right]^T \beta \right\}.$$

Rewriting the term in brackets above leads to

$$\hat{\mathbf{X}}_{np}(\pi) - \hat{\mathbf{X}}_{np}(\tilde{w}) = \sum_{S_{np}} \frac{(1 - g_j)\mathrm{x}_j}{\pi_j}$$

$$= \left(\hat{\mathbf{X}}_p - \mathbf{X}_U \right)^T \tilde{\mathbf{A}}_p^{-1} \sum_{S_{np}} \frac{\mathbf{x}_j \mathbf{x}_j^T}{\pi j \tilde{\sigma}_j^2}$$

$$= \left(\hat{\mathbf{X}}_p - \mathbf{X}_U \right)^T \tilde{\mathbf{A}}_p^{-1} \tilde{\mathbf{A}}_{np}(\pi).$$

Applying condition (5) implies that $V_R \left[E_\pi E_\xi \left(\hat{Y}_{MC2} \right) \right] \doteq 0$. Combining results for the three terms in Equation (8) yields

$$V_{R\pi\xi} \left(\hat{Y}_{MC2} \right) \doteq \sum_U \frac{\sigma_j^2}{\pi_j} + \beta^T V_\pi \left(\hat{\mathbf{X}}_p \right) \beta.$$

An estimator of this variance is

$$\upsilon_{R\pi\xi} \left(\hat{Y}_{MC2} \right) \doteq \sum_{S_{np}} \left(\frac{\hat{e}_j^*}{\pi_j} \right)^2 + \hat{\hat{B}}_{np}(\pi)^T \upsilon_\pi \left(\hat{\mathbf{X}}_p \right) \hat{\hat{B}}_{np}(\pi)$$

as shown in Equation (27).

9. References

Andridge, R.R., B.T. West, R.J.A. Little, P.S. Boonstra, and F. Alvarado-Leiton. 2019. "Indices of non-ignorable selection bias for proportions estimated from non-probability samples." *Journal of the Royal Statistical Society,* 68(5): 1465–1483. Available at: https://doi.org/10.1111/rssc.12371 (accessed September 2022).

Baker, R., J.M. Brick, N.A. Bates, M.P. Battaglia, M.P. Couper, J.A. Dever, K. Gile, and R. Tourangeau. 2013. *Report of the AAPOR task force on non-probability sampling.* The American Association for Public Opinion Research, Deerfield, IL. Available at: https://www.aapor.org/aapor_main/media/mainsitefiles/nps_tf_report_final_7_revised_fnl_6_22_13.pdf (accessed September 2022).

Brick, J.M., and D. Williams. 2013. "Explaining rising nonresponse rates in cross-sectional surveys." *The Annals of the American Academy of Political and Social Science* 645(1): 36–59. DOI: https://doi.org/10.1177/0002716212456834.

Caliendo, M, and S. Kopeinig. 2008. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 2(1): 31–72. DOI: https://doi.org/10.1111/j.1467-6419.2007.00527.x.

Center for Disease Control and Prevention. 2017. *Weighting the BRFSS Data*. Available at: https://www.cdc.gov/brfss/annual_data/2017/pdf/weighting-2017-508.pdf.

Center for Disease Control and Prevention. 2023. *Behavioral Risk Factor Surveillance Survey*. Available at: http://www.cdc.gov/BRFSS.

Chen, Y., P. Li, and C. Wu. 2020. "Doubly robust inference with non-probability survey samples." *Journal of the American Statistical Association* 115: 2011–2021. DOI: https://doi.org/10.1080/01621459.2019.1677241.

Cochran, W.G. 1953. "Matching in analytical studies." *American Journal of Public Health* 43: 684–691. DOI: https://doi.org/10.2105/AJPH.43.6_Pt_1.684.

Dehejia, R. and S. Wahba. 2002. "Propensity-score matching methods for nonexperimental causal studies." *The Review of Economic and Statistics* 84(1): 151–161. Available at: http://www.mitpressjournals.org/doi/pdf/10.1162/003465302317331982 (accessed September 2022).

Deville, J.C., and C. Särndal. 1992. "Calibration estimators in survey sampling." *Journal of the American Statistical Association* 87(418): 376–382. DOI: https://doi.org/10.2307/2290268.

Elliott, M.R. and R. Valliant. 2017. "Inference for nonprobability samples." *Statistical Science* 32: 249–264. Available at: https://projecteuclid.org/journals/statistical-science/volume-32/issue-2/Inference-for-Nonprobability-Samples/10.1214/16-STS598.full. (accessed September 2022).

Gessendorfer, J., Beste, J., J. Drechsler, and J. Sakshaug. 2018. "Statistical Matching as a Supplement to Record Linkage: A Valuable Method to Tackle Nonconsent Bias?. *Journal of Official Statistics* 34(4): 909–933. DOI: https://doi.org/10.2478/jos-2018-0045.

Hansen, M.H., W.G. Madow, and B.J. Tepping. 1983. "An evaluation of model-dependent and probability sampling inferences in sample surveys." *Journal of the American Statistical Association* 78: 776–793.

Kennedy, C., A. Mercer, S. Keeter, N. Hatley, K. McGeeney, and A. Gimenez. 2016. *Evaluating online nonprobability surveys, vendor choice matters: widespread errors found for estimates based on blacks and hispanics.* Technical report, Pew Research. Available at: https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys/ (accessed September 2022).

Kim, J.-K., S. Park, Y. Chen, and C. Wu. 2021. "Combining non-probability and probability survey samples through mass imputation." *Journal of the Royal Statistical*

Society, Series A: Statistics in Society 184: 941–963. DOI: https://doi.org/10.1111/rssa.12696.

Lee, S. 2006. "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys." *Journal of Official Statistics* 22(2): 329–349. Available at: http://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/propensity-score-adjustment-as-a-weighting-scheme-for-volunteer-panel-web-surveys.pdf (accessed September 2022).

Little, R.J.A., B. West, P.S. Boonstra, and J. Hu. 2019. "Measures of the degree of departure from ignorable sample selection." *Journal of Survey Statistics and Methodology* 8(5): 932–964. DOI: https://doi.org/10.1093/jssam/smz023.

Lumley, T. 2020. "Survey: analysis of complex survey samples." R package version 4.1-1. Available at: https://cran.r-project.org/web/packages/survey/index.html (accessed September 2022).

MacKinnon, J.G., and H. White. 1985. "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties." *Journal of Econometrics* 29(3): 305–325. DOI: https://doi.org/10.1016/0304-4076(85)90158-7.

Rao, J., W. Yung, and M Hidiroglou. 2002. "Estimating equations for the analysis of survey data using poststratification information." *Sankhyā Series A* 64: 364–378.

Rivers, D. 2007. "Sample matching for web surveys: Theory and application." In Proceedings of the Section on Survey Research Methods, July, Salt Lake City, Utah, USA. Available at: http://www.websm.org/uploadi/editor/1368187629Rivers_2007_Sampling_for_web_surveys.pdf (accessed September 2022).

Rivers, D., and D. Bailey. 2009. "Inference from matched samples in the 2008 U.S. national elections." In Proceedings of the American Statistical Association, Section on Survey Research Methods: 627–639. August, Washington D.C., USA. Available at: http://www.asasrms.org/Proceedings/y2009f.html (accessed September 2022).

Rosenbaum, P.R., and D.B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70(1): 41–55. Available at: https://doi.org/10.1093/biomet/70.1.41.

Rothman, K.J., S. Greenland, and T.L. Lash. 2008. *Modern Epidemiology*. Lippincott, Williams and Wilkins, 3rd edition.

Rubin, D. 1973. "Matching to remove bias in observational studies." *Biometrics* 29(1): 159–183. DOI: https://doi.org/10.2307/2529684.

Särndal, C.E., B. Swensson, and J.H. Wretman. 1992. *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer-Verlag, New York.

Schonlau, M., Soest, A.V., A. Kapteyn, and M. Couper. 2009. "Selection bias in web surveys and the use of propensity scores. *Sociological Methods a Research* 37(3): 291–318. DOI: https://doi.org/10.1177/0049124108327128.

Sekhon, J.S. 2011. "Multivariate and propensity score matching software with automated balance optimization: The Matching package for R." *Journal of Statistical Software* 42(7): 1–52. DOI: https://doi.org/10.18637/jss.v042.i07.

Smith, J., and P. Todd. 2005. "Does matching overcome LaLonde's critique of non-experimental estimators?" *Journal of Econometrics* 125(1): 305–353. DOI: https://doi.org/10.1016/j.jeconom.2004.04.011.

Terhanian, G., and J. Bremer. 2012. "A smarter way to select respondents for surveys?" *International Journal of Market Research* 54(6): 751–780. DOI: https://doi.org/10.2501/IJMR-54-6-751-780.

Toluna. 2023. *Toluna delivers real-time consumer insights at the speed of the on-demand economy*. Available at: https://www.greenbook.org/company/Toluna.

U.S. Census Bureau. 2023. *American Community Survey*. Available at: https://www.census.gov/programs-surveys/acs.

Valliant, R. 2020. "Comparing alternatives for estimation from nonprobability samples." *Journal of Survey Statistics and Methodology* 8: 231–263. DOI: https://doi.org/10.1093/jssam/smz003.

Valliant, R. and J.A. Dever. 2011. "Estimating propensity adjustments for volunteer web surveys." *Sociological Methods and Research* 40: 105–137. DOI: https://doi.org/10.1177/0049124110392533.

Valliant, R., J.A. Dever, and F. Kreuter. 2018. *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York, 2nd edition.

Valliant, R., J.A. Dever, and F. Kreuter. 2020. *PracTools: Tools for Designing and Weighting Survey Samples*. R package version 1.2.8. Available at: https://CRAN.R-project.org/package = PracTools (accessedSeptember 2022).

Vavreck, L., and D. Rivers. 2008. "The 2006 cooperative congressional election study." *Journal of Elections, Public Opinion and Parties* 18(4): 35–66. DOI: https://doi.org/10.1080/17457280802305177.

Wang, L. 2020. *Improving External Validity of Epidemiologic Analyses by Incorporating Data from Population-Based Surveys*. PhD thesis, University of Maryland. Available at: http://hdl.handle.net/1903/26125, doi:10.13016/pogq-glbs (accessed September 2022).

Wang, L., Graubard, B.I., H. Katki, and Y. Li. 2020. "Improving external validity of epidemiologic cohort analyses: a kernel weighting approach." *Journal of the Royal Statistical Society, Statistics in Society, Series A* 183(3):1293–1311. Available at: https://rss.onlinelibrary.wiley.com/doi/10.1111/rssa.12564 (accessed September 2022).

Wang, L., R. Valliant, and Y. Li. 2021. "Adjusted logistic propensity weighting methods for population inference using nonprobability volunteer-based epidemiologic cohorts." *Statistics in Medicine*: 1–14. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9122 (accessed September 2022).

Received September 2021
Revised May 2022
Accepted October 2022