

A Discussion of Weighting Procedures for Unit Nonresponse

David Haziza¹ and Éric Lesage²

Weighting procedures are commonly applied in surveys to compensate for nonsampling errors such as nonresponse errors and coverage errors. Two types of weight-adjustment procedures are commonly used in the context of unit nonresponse: (i) nonresponse propensity weighting followed by calibration, also known as the two-step approach and (ii) nonresponse calibration weighting, also known as the one-step approach. In this article, we discuss both approaches and warn against the potential pitfalls of the one-step procedure. Results from a simulation study, evaluating the properties of several point estimators, are presented.

Key words: Calibration; nonresponse bias; one-step approach; propensity-score adjusted estimator; two-step approach; unit nonresponse.

1. Introduction

Weighting procedures are commonly applied in surveys to compensate for nonsampling errors such as nonresponse errors and coverage errors. [Brick \(2013\)](#) provides an excellent overview of weighting in the presence of unit nonresponse; see also [Kalton and Flores-Cervantes \(2003\)](#). Two types of weight-adjustment procedures are commonly used in the context of unit nonresponse: (i) nonresponse propensity weighting followed by calibration, also known as the two-step approach and (ii) nonresponse calibration weighting, also known as the one-step approach. In this article, our focus is to warn against the potential pitfalls of the one-step procedure.

The two-step approach consists of adjusting the weights in two distinct steps: the basic (design) weights of respondents are first multiplied by a nonresponse adjustment factor, which is defined as the inverse of the estimated response probability. The adjusted weights are further modified so that survey-weighted estimates agree with known population totals. In the first step, survey statisticians aim at reducing the nonresponse bias, which may be appreciable when respondents and nonrespondents are different with respect to the survey variables. Whether or not one will succeed in achieving an efficient bias reduction depends on the availability of powerful auxiliary information ([Särndal and Lundström 2005](#)), which is a set of variables available for both respondents and nonrespondents. In the second step, some form of calibration (e.g., poststratification) is performed in order to ensure consistency between survey-weighted estimates and known population totals.

¹ Université de Montréal, C.P. 6128, Centre-ville Montreal Quebec H3C 3J7 Canada. Email: david.haziza@umontreal.ca

² INSEE, 18 bd Adolphe Pinard 75 014 Paris, France. Email: eric.lesage@insee.fr

Acknowledgments: The authors are grateful to an associate editor and three reviewers for their comments and suggestions, which substantially improved the quality of this article. David Haziza's research was funded by a grant from the Natural Sciences and Engineering Research Council of Canada.

Calibration procedures require that the auxiliary variables (called calibration variables) are available for the respondents and that their population totals are known. In practice, the calibration variables are often specified by survey managers, who wish to ensure consistency with respect to some important variables (e.g., age and sex). Moreover, if the calibration variables are related to the characteristics of interest, the resulting calibration estimators tend to be more efficient than the noncalibrated ones.

The one-step approach pursues the same three goals as the two-step approach: reduce the nonresponse bias, ensure consistency between survey estimates and known population totals and, possibly, reduce the variance of point estimators. However, the weighting process is performed in a single step and does not require explicit estimation of the response probabilities.

In the absence of nonsampling errors, calibration consists of determining a set of calibrated (or final) weights as close as possible to the basic weights, while satisfying calibration constraints. A calibrated weight is expressed as the basic weight multiplied by a calibration adjustment factor, which depends on a calibration function. Commonly used calibration functions include the linear function, the exponential function, the truncated linear function and the logit function; see Section 2. [Deville and Särndal \(1992\)](#) showed that calibration estimators are asymptotically design consistent and that all the distance functions are asymptotically equivalent in the sense that they all lead to calibration estimators that are asymptotically equivalent to the calibration estimator based on the linear calibration function. The calibration function is usually chosen so that the distribution of the calibrated weights is “cosmetically attractive”. For example, a problem that can be encountered with the linear function is the occurrence of negative weights, which can be prevented by using the exponential function that ensures positive weights. However, the latter may lead to extreme weights, which in turn may contribute to increase the instability of point estimators for characteristics of interest weakly correlated with the calibration variables. In this case, functions such as the truncated linear function or the logit function can be used in order to ensure that the calibration adjustment factors lie between prespecified lower and upper bounds.

How to choose the calibration function in the presence of unit nonresponse? In the case of the two-step approach, calibration is performed after the weights have been adjusted for nonresponse. As a result, the choice of the calibration function can be essentially made using the same criteria as in the complete response case. This is discussed further in Section 3. The situation is more intricate with the one-step approach, as different calibration functions may lead to calibration estimators with substantially different properties in terms of bias and mean square error. As a result, the choice of the calibration function is generally important when calibration is used for treating nonresponse. While the choice of calibration variables has been widely discussed in the literature (e.g., [Särndal and Lundström 2005](#) and [Särndal 2011](#)), the issues of how to select an appropriate calibration function in the context of the one-step approach and the effect of function misspecification on the properties of the resulting estimators have not received a lot of attention. Two notable exceptions are [Kott \(2006\)](#) and [Kott and Liao \(2012\)](#). In this article, we argue that, even though the one-step approach does not use estimated response probabilities in the construction of point estimators explicitly, a wrong choice of the calibration function can have inadvertent and detrimental effects, even in the presence of

high association between the auxiliary variables and the study variable. The matter deserves more careful attention than what it seems has hitherto been noticed in the literature; see Section 4. In Section 5, we show empirically that an inappropriate calibration function may lead to biased calibration estimators (sometimes exhibiting a bias larger than that of unadjusted estimators). This is especially true in the presence of quantitative auxiliary variables. The paper ends with a discussion in Section 6.

2. Calibration Weighting in the Complete Data Case

Let $U = \{1, 2, \dots, N\}$ be a finite population consisting of N elements. Most surveys conducted by statistical agencies are multipurpose surveys, which are designed to provide statistics for a possibly large number of variables. For simplicity, we use the generic notation y to denote a characteristic of interest. In this paper, we are interested in estimating a population total $t_y = \sum_{k \in U} y_k$, where y_k denotes the k -th value of the characteristic of interest y , $k = 1, \dots, N$. A sample s , of size n , is selected from U according to a given sampling design $p(s)$. Let π_k denote the first-order inclusion probability of unit k in the sample and $d_k = 1/\pi_k$ denote its design weight. Applying the basic weighting system, $\{d_k; k \in s\}$, to a y -variable leads to the well-known Horvitz-Thompson estimator

$$\hat{t}_{y\pi} = \sum_{k \in s} d_k y_k. \quad (1)$$

The estimator (1) is design unbiased for t_y regardless of the characteristic of interest y being estimated. That is, $E_p(\hat{t}_{y\pi}) = t_y$, where $E_p(\cdot)$ denotes the expectation with respect to the sampling design.

In practice, auxiliary information is often available at the estimation stage. Let $\mathbf{x}_i = (x_{1i}, \dots, x_{ji})^\top$ be a J -vector of auxiliary variables attached to unit i . We assume that the vector of population totals, $\mathbf{t}_x = (t_{x_1}, \dots, t_{x_j})^\top$, is known without error, where $t_{x_j} = \sum_{i \in U} x_{ji}$. While the basic weighting system ensures unbiasedness, that is, $E_p(\hat{\mathbf{t}}_{x\pi}) = \mathbf{t}_x$, it does not generally produce an exact estimate for each of the J auxiliary variable; that is, $\hat{\mathbf{t}}_{x\pi} \neq \mathbf{t}_x$, in general. To overcome the problem, we seek a calibrated weighting system $\{w_k; k \in s\}$ such that the weights w_k are “as close as possible” to the design weights d_k while satisfying the calibration constraints

$$\sum_{k \in s} w_k \mathbf{x}_k = \mathbf{t}_x.$$

The resulting calibrated weight w_k is given by

$$w_k = d_k F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k), \quad (2)$$

where $F(\cdot)$ is a monotonic and twice-differentiable function such that $F(0) = 1$ and $F'(0) = 1$ and $\hat{\boldsymbol{\lambda}}$ is a J -vector of estimated coefficients (Deville and Särndal 1992). The weight w_k in (2) is the product of the design weight d_k and the calibration adjustment factor $F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k)$. The calibration factor $F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k)$ depends on (i) the calibration function $F(\cdot)$, (ii) the characteristics of unit k through \mathbf{x}_k and (iii) the vector of estimated coefficients $\hat{\boldsymbol{\lambda}}$, which can be viewed as a measure of sample imbalance. Under mild

regularity conditions, [Deville and Särndal \(1992\)](#) showed that $\hat{\lambda} \rightarrow 0$ in probability as $n \rightarrow \infty$ and $N \rightarrow \infty$.

The resulting calibration estimator is

$$\hat{t}_C = \sum_{k \in s} w_k y_k. \quad (3)$$

Several calibration functions $F(\cdot)$ are used in practice, each corresponding to a particular calibration method. The most popular calibration methods are: (i) the linear method

$$F(u) = 1 + u; \quad (4)$$

(ii) the exponential method

$$F(u) = \exp(u); \quad (5)$$

(iii) the truncated linear method

$$F(u) = \begin{cases} 1 + u & L - 1 \leq u \leq M - 1 \\ M & u > M - 1 \\ L & u < L - 1, \end{cases} \quad (6)$$

where L and M are the prespecified lower and upper bounds, respectively; and (iv) the logit method

$$F(u) = \frac{L(M - 1) + M(1 - L) \exp(Au)}{M - 1 + (1 - L) \exp(Au)}, \quad (7)$$

where

$$A = \frac{M - L}{(1 - L)(M - 1)}.$$

Assuming that the inverse of $\sum_{k \in s} d_k \mathbf{x}_k \mathbf{x}_k^\top$ exists, the linear method leads to a closed-form solution. In contrast, Methods (5)–(7) require some numerical methods that may fail to converge in some situations. However, the linear method may produce negative calibration adjustment factors, $F(\hat{\lambda}^\top \mathbf{x}_k)$, resulting in negative calibrated weights. On the other hand, the exponential method ensures that the calibration adjustment factors are positive, although some could be extreme. To avoid unduly large calibration adjustment factors, one can specify lower and upper bounds through the use of the truncated linear and logit methods. [Deville and Särndal \(1992\)](#) showed that the calibration estimator (3) is design consistent and approximately design unbiased for t_y regardless of the characteristic y being estimated and that all the calibration methods are asymptotically equivalent in the sense that they all lead to the calibration estimator based on the linear method.

We now discuss two important situations that are frequently encountered in practice. Let x_1 and x_2 be two categorical variables with J_1 and J_2 categories, respectively. The population U is then divided into $J_1 \times J_2$ cells. Let $N_{j_1 j_2}$ be the population count corresponding to the (j_1, j_2) cell, $j_1 = 1, \dots, J_1$ and $j_2 = 1, \dots, J_2$. Two cases may occur in practice: (i) the population counts $N_{j_1 j_2}$ are known. This case corresponds to a standard

poststratification based on a vector of auxiliary information of size $J = J_1 \times J_2$. It is worth noting that, in this case, the choice of the calibration function $F(\cdot)$ is unimportant as all the calibration functions lead to the same calibrated weighting system $\{w_k; k \in s\}$. (ii) The individual cell counts $N_{j_1 j_2}$ are not known but the population margins $N_{j_1 \bullet} = \sum_{j_2=1}^{J_2} N_{j_1 j_2}$ and $N_{\bullet j_2} = \sum_{j_1=1}^{J_1} N_{j_1 j_2}$ are known, leading to a vector of auxiliary information of size $J = J_1 + J_2$. In this context, [Deville et al. \(1993\)](#) showed that the use of the exponential method (5) leads to the raking ratio estimator. Unlike case (i), different calibration functions generally lead to different calibrated weighting systems in case (ii). This discussion can be extended to more than two categorical variables. In this instance, case (ii) is often referred to as generalized raking procedures. We revisit both situations in Section 6 in the context of nonresponse adjustment.

3. The Two-Step Approach: Nonresponse Propensity Weighting Followed by Calibration

In the presence of unit nonresponse, the characteristics of interest are observed for a subset, s_r , of the original sample s . Let ϕ_k be the unknown response propensity attached to unit k . We assume that $\phi_k > 0$ for all k and that units respond independently of one another. We postulate the following nonresponse model

$$\phi_k = m(\mathbf{z}_k, \gamma), \quad (8)$$

where $m(\cdot)$ is a given function, \mathbf{z}_k is a vector of auxiliary variables available for both respondents and nonrespondents and γ is a vector of unknown parameters. In this article, we assume that the \mathbf{z} -vector is correctly specified but not necessarily the functional form of (8). The choice of the \mathbf{z} -vector is discussed in [Little and Vartivarian \(2005\)](#).

In the first step, an estimate of ϕ_k is $\hat{\phi}_k = m(\mathbf{z}_k, \hat{\gamma})$, where $\hat{\gamma}$ is a suitable estimator of γ . The adjusted weight for nonresponse attached to unit k is defined as $\tilde{w}_k = d_k / \hat{\phi}_k$ for $k \in s_r$, leading to a weighting system adjusted for nonresponse, $\{\tilde{w}_k; k \in s_r\}$. The factor $\hat{\phi}_k^{-1}$ is often called the nonresponse adjustment factor for unit k . Applying the weighting system $\{\tilde{w}_k; k \in s_r\}$ to a characteristic of interest y leads to the Propensity-Score Adjusted (PSA) estimator of t_y (e.g., [Lee 2006](#)):

$$\hat{t}_{PSA} = \sum_{k \in s_r} d_k \hat{\phi}_k^{-1} y_k = \sum_{k \in s_r} \tilde{w}_k y_k. \quad (9)$$

The rationale behind this type of weighting procedure is similar in spirit to weighting for two-phase sampling.

Estimates of the ϕ_k 's may be obtained through the use of a parametric model; for example, a logistic regression model as found in [Ekholm and Laaksonen \(1991\)](#). In the context of parametric nonresponse models, [Kim and Kim \(2007\)](#) showed that the PSA estimator (9) is asymptotically unbiased and consistent for t_y regardless of the characteristic y being estimated if (8) is correctly specified. However, parametric methods are rarely used in practice because some estimates $\hat{\phi}_k$ may be very small, leading to extreme nonresponse adjustment factors, ultimately resulting in highly dispersed weights \tilde{w}_k . Moreover, parametric methods are vulnerable to the misspecification of $m(\cdot)$.

In practice, nonparametric methods are preferred. A popular method, called the score method (Haziza and Beaumont 2007), consists of first obtaining preliminary estimated response probabilities $\tilde{\phi}_k$ using a parametric model (e.g., the logistic regression model) and partitioning the sample into homogeneous weighting classes formed on the basis of the $\tilde{\phi}_k$'s. The basic weight of a respondent in a given class is then adjusted using the observed response rate within the same class (e.g., Little 1986; Eltinge and Yanaseh 1997). Other nonparametric methods include smoothing methods such as kernel and local polynomial methods (e.g., Giommi 1987; Da Silva and Opsomer 2006, 2009) and regression trees (e.g., Phipps and Toth 2012). Nonparametric methods are expected to provide some robustness if the form of $m(\cdot)$ is misspecified and protect (to some extent) against the noninclusion of predictors accounting for curvature or interactions in the \mathbf{z} -vector.

In the second step, the adjusted weights \tilde{w}_k are further modified so that survey-weighted estimates agree with known population totals. More specifically, we assume that a vector of calibration variables \mathbf{x}^* is available for $k \in s_r$, and that the vector of population totals $\mathbf{t}_{\mathbf{x}^*} = \sum_{k \in U} \mathbf{x}_k^*$ is known. The \mathbf{x}^* -vector may contain one or more z -variables that were used in (8). The final weighting system is given by $\{w_k; k \in s_r\}$, where

$$w_k = \tilde{w}_k F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k^*) \quad (10)$$

and $\hat{\boldsymbol{\lambda}}^\top$ is a vector of estimated coefficients. The final weights w_k satisfy the calibration constraints

$$\sum_{k \in s_r} w_k \mathbf{x}_k^* = \mathbf{t}_{\mathbf{x}^*}. \quad (11)$$

The weight w_k in (10) is the product of the adjusted weight \tilde{w}_k and the calibration adjustment factor $F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k^*)$.

For example, the linear method (4) leads to

$$w_k = \tilde{w}_k (1 + \hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k^*),$$

whereas the exponential method leads to

$$w_k = \tilde{w}_k \exp(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k^*).$$

Alternative weighting methods are discussed in Kott and Liao (2012). Applying the final weighting system, $\{w_k; k \in s_r\}$, to a characteristic of interest y leads to the two-step calibration estimator

$$\hat{t}_{C,2} = \sum_{k \in s_r} w_k y_k = \sum_{k \in s_r} d_k \hat{\phi}_k^{-1} F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k^*) y_k. \quad (12)$$

We make the following remarks: (i) if the nonresponse model (8) is correctly specified (and so the estimator (9) is asymptotically unbiased for t_y for every characteristic of interest), the two-step calibration estimator $\hat{t}_{C,2}$ is asymptotically unbiased for t_y regardless of the characteristic y being estimated. (ii) If the \mathbf{x}^* -vector is linearly related to y , then $\hat{t}_{C,2}$ is expected to be more efficient than \hat{t}_{PSA} . (iii) As for the complete data case, $\hat{\boldsymbol{\lambda}} \rightarrow 0$ in probability as $n \rightarrow \infty$ and $N \rightarrow \infty$ if the nonresponse model (8) is correctly specified. (iv) In the two-step approach, the calibration function is chosen using the same criteria as those

encountered in the complete-data case. Most often, the distribution of the calibration adjustment factors $F(\hat{\boldsymbol{\lambda}}^\top \mathbf{x}_k^*)$ drives the choice of the function $F(\cdot)$.

4. The One-Step Approach: Nonresponse Calibration Weighting

Following Särndal and Lundström (2005), we distinguish between two levels of auxiliary information:

- (1) U -level: a vector of auxiliary variables \mathbf{x}_k^* is minimally available for $k \in s_r$ and the vector of population totals $\mathbf{t}_{\mathbf{x}^*} = \sum_{k \in U} \mathbf{x}_k^*$ is known.
- (2) s -level: a vector of auxiliary variables \mathbf{x}_k^o is available for $k \in s$ but the vector of population totals, $\sum_{k \in U} \mathbf{x}_k^o$, is unknown. Instead, the vector of complete-data estimators, $\hat{\mathbf{t}}_{\mathbf{x}^o} = \sum_{k \in s} d_k \mathbf{x}_k^o$, is available.

We define the stacked vector of auxiliary variables for unit k as $\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_k^* \\ \mathbf{x}_k^o \end{pmatrix}$ and the corresponding vector of totals $\mathbf{t}_{\mathbf{x}} = \begin{pmatrix} \mathbf{t}_{\mathbf{x}^*} \\ \hat{\mathbf{t}}_{\mathbf{x}^o} \end{pmatrix}$. The x^o -variables are believed to be associated with nonresponse and, possibly, with some characteristics of interest. Their role is similar to that of the z -variables in the two-step approach: contribute to reducing the nonresponse bias.

The final weighting system is $\{w_k; k \in s_r\}$, where

$$w_k = d_k F(\hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k), \quad (13)$$

and $\hat{\boldsymbol{\lambda}}_r$ is determined so that the calibration constraints

$$\sum_{k \in s_r} w_k \mathbf{x}_k = \mathbf{t}_{\mathbf{x}}$$

are satisfied. The final weight w_k in (13) is the product of the design weight d_k and the nonresponse/calibration adjustment factor $F(\hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k)$. Applying the final weighting system, $\{w_k; k \in s_r\}$, to a characteristic of interest y leads to the one-step calibration estimator

$$\hat{t}_{C,1} = \sum_{k \in s_r} w_k y_k = \sum_{k \in s_r} d_k F(\hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k) y_k. \quad (14)$$

Note that, unlike the two-step approach, the vector of estimated coefficient $\hat{\boldsymbol{\lambda}}_r$ does not converge towards 0 as $n \rightarrow \infty$ and $N \rightarrow \infty$. This is due to the fact that $F(\hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k)$ is essentially an estimate of ϕ_k^{-1} .

We now compare the one-step and the two-step approaches. To that end, note that it is sufficient to compare the PSA estimator (which is the estimator resulting from the first step in the two-step approach) and a calibration estimator based on the x^o -variables only. The second step in the two-step approach or the use of the \mathbf{x}^* -variables in the one-step approach strive to make survey estimates and known population totals agree, which is not the focus here. Below, we argue that the one-step based on the x^o -variables imposes a parametric model for the relationship between the response propensity and the vector of auxiliary variables, which makes the resulting estimator vulnerable to a misspecification of the calibration function.

Recall that \hat{t}_{PSA} is asymptotically unbiased for t_y regardless of the characteristic y being estimated, provided that the nonresponse model (8) is correctly specified. Therefore, for $\hat{t}_{C,1}$ in (14) to be asymptotically unbiased for t_y regardless of the characteristic y being estimated, we require

$$F(\hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k) = \hat{\phi}_k^{-1}.$$

The previous expression suggests that the adjustment factor $F(\hat{\boldsymbol{\lambda}}_r^\top \mathbf{x}_k)$ can be viewed as an implicit estimate of ϕ_k^{-1} .

Next, we examine the bias of $\hat{t}_{C,1}$, where the bias is defined as $Bias(\hat{t}_{C,1}) = E_p E_q(\hat{t}_{C,1}|s) - t_y$, and the subscripts p and q refer to the sampling design and the nonresponse mechanism respectively. Using a first-order Taylor expansion and ignoring the higher-order terms, the bias of $\hat{t}_{C,1}$ can be approximated by

$$Bias(\hat{t}_{C,1}) \approx - \sum_{k \in U} (1 - \phi_k F_k)(y_k - \mathbf{x}_k^\top \mathbf{B}_{\phi f}), \quad (15)$$

where

$$\mathbf{B}_{\phi f} = \left(\sum_{k \in U} \phi_k f_k \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \sum_{k \in U} \phi_k f_k \mathbf{x}_k y_k$$

with $F_k \equiv F(\boldsymbol{\lambda}_N^\top \mathbf{x}_k)$, $f_k \equiv F'(\boldsymbol{\lambda}_N^\top \mathbf{x}_k)$ and $\boldsymbol{\lambda}_N$ denotes the probability limit of $\hat{\boldsymbol{\lambda}}_r$.

In the case of linear weighting (4), Expression (15) reduces to

$$Bias(\hat{t}_{C,1}) \approx - \sum_{k \in U} (1 - \phi_k)(y_k - \mathbf{x}_k^\top \mathbf{B}_\phi), \quad (16)$$

where

$$\mathbf{B}_\phi = \left(\sum_{k \in U} \phi_k \mathbf{x}_k \mathbf{x}_k^\top \right)^{-1} \sum_{k \in U} \phi_k \mathbf{x}_k y_k.$$

Expression (16) is identical to Expression (9.14) in [Särndal and Lundström \(2005\)](#). Note that the more general expression (15) does not appear in [Särndal and Lundström \(2005\)](#), where the focus is placed on linear weighting.

Expression (15) is interesting because it sheds some light on the conditions required for asymptotic unbiasedness:

- (1) On the one hand, the asymptotic bias (15) vanishes if the finite population covariance between the residuals $e_k = (y_k - \mathbf{x}_k^\top \mathbf{B}_{\phi f})$ and $\delta_k = \phi_k F_k - 1$ is equal to zero.

This condition is satisfied if

$$y_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k \quad (17)$$

with

$$E(\epsilon_k | \mathbf{x}_k) = 0 \quad (18)$$

and if the response probability ϕ_k is not related to y_k after conditioning on \mathbf{x}_k . The latter condition is essentially the customary MAR assumption (Rubin 1976).

In multipurpose surveys, it is unrealistic to presume that Model (17) holds for every characteristic of interest y , in which case some estimates may suffer from bias. In fact, in household and social surveys, most characteristics of interest are categorical, in which case (17) is generally not appropriate.

(2) On the other hand, the asymptotic bias of $\hat{t}_{C,1}$ is equal to zero if

$$F_k = \phi_k^{-1}. \quad (19)$$

Hence, selecting a calibration function $F(\cdot)$ such that (19) is satisfied ensures that the one-step calibration estimator is asymptotically unbiased regardless of the characteristic of interest y being estimated, even if (17) and (18) do not hold. For linear weighting, it follows from (19) that $\hat{t}_{C,1}$ is asymptotically unbiased for t_y for every y if

$$\phi_k^{-1} = 1 + \boldsymbol{\lambda}^\top \mathbf{x}_k \quad \text{for all } k \in U, \quad (20)$$

for a vector of unknown constants $\boldsymbol{\lambda}$ (see Särndal and Lundström 2005, ch. 9). For exponential weighting, we require

$$\phi_k^{-1} = \exp(\boldsymbol{\lambda}^\top \mathbf{x}_k) \quad \text{for all } k \in U; \quad (21)$$

see also Kott and Liao (2012) for a discussion of alternative weighting methods. In other words, both the linear and exponential methods correspond to specific parametric nonresponse models, which suggests that selecting either one is somehow equivalent to (implicitly) selecting a nonresponse model. This begs the following question: how is $\hat{t}_{C,1}$ affected if (20) (respectively (21)) is not an appropriate description of the relationship linking the \mathbf{x} -vector and the ϕ_k 's, that is, if the calibration function is misspecified? This aspect is investigated in Section 5.

A key aspect here is to realize that each calibration function corresponds to a specific parametric nonresponse model. By choosing a given calibration function, one is effectively making a strong statement about the underlying nonresponse mechanism. Therefore, in order to avoid an incorrect functional form, a complete modeling exercise is needed to validate the form of the function linking the response propensity ϕ_k to the vector of auxiliary variable \mathbf{x}_k . Failing to do so may result in biased estimators. Furthermore, there may be no calibration that corresponds to the inverse of the estimated response probabilities. For instance, suppose that the relationship between the response probability and a single auxiliary variable x is described by a nonmonotonic function. In this case, it may be difficult to find a calibration function that provides an adequate description of the relationship between the inverse of the response propensity and the \mathbf{x} -vector.

5. Simulation Study

We conducted a simulation study to illustrate the importance of carefully selecting a calibration function $F(\cdot)$ in the context of a one-step approach. We generated a population of size $N = 1,000$, which consisted of an auxiliary variable x and four variables of interest y_1, y_2, y_3 and y_4 . The x -values were first generated from a uniform distribution $(0, 80)$. The y_1 -values were generated according to the linear model

$$y_{k1} = 1,000 + 10x_k + \varepsilon_{k1},$$

where the errors ε_{k1} were generated from a normal distribution with mean 0 and variance 300. The y_2 -values were generated according to the exponential model

$$y_{k2} = \exp(-0.1 + 0.1x_k) + \varepsilon_{k2},$$

where the errors ε_{k2} were generated from a normal distribution with mean 0 and variance 300. The y_3 -values were generated according to the logistic model

$$y_{k3} \sim B(1, p_k),$$

where $p_k = [\exp\{-0.5 (x_k - 55)\} + 1]^{-1}$. The y_4 -values were generated according to the quadratic model

$$y_{k4} = 1,300 - (x_k - 40)^2 + \varepsilon_{k4},$$

where the errors ε_{k4} were generated from a normal distribution with mean 0 and standard deviation 300. The relationships between y_j and x are displayed in Figure 1, $j = 1, \dots, 4$.

In order to focus on the nonresponse error, we considered the census case; that is, $n = N = 1,000$ and $d_k = 1$ for all k . In each population, units were assigned a response probability ϕ_k according to a given nonresponse mechanism. We simulated nonresponse according to four nonresponse mechanisms, all presented in Table 1; see also Figure 2. For each mechanism, the parameters were set so that the overall response rate was

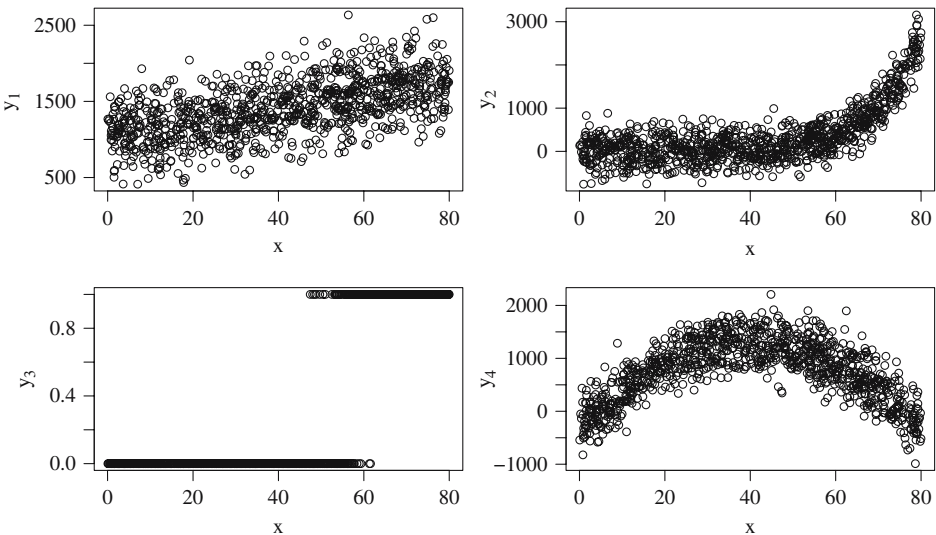


Fig. 1. Relationships between the characteristics of interest and x

Table 1. Nonresponse mechanisms used for generating nonresponse

Nonresponse mechanism	Name	ϕ_k
1	Inverse linear	$(1.2 + 0.024 \ x_k)^{-1}$
2	Exponential	$\exp(-0.2 - 0.014x_k)$
3	Logistic type	$0.2 + 0.6\{1 + \exp(-5 + x_k/8)\}^{-1}$
4	Quadratic	$0.7 + 0.45 \ (x_k/40 - 1)^2 + 0.0025 \ x_k$

approximately equal to 50%. The response indicators R_k for $k \in U$ were generated independently from a Bernoulli distribution with parameter ϕ_k , resulting in a population of respondents U_r of size N_r . The nonresponse process was repeated $M = 5,000$ times, leading to $M = 5,000$ sets of respondents for each nonresponse mechanism. From Figure 1 and Figure 2, we note that both the response propensity and the characteristics of interest $y_1 - y_4$ are highly related to x in all the scenarios.

We were interested in estimating the populations totals $t_{y_j}, j = 1, 2, 3, 4$. For each total, we computed three estimators: (i) The unadjusted estimator $\hat{t}_{un} = N\bar{y}_r$ where $\bar{y}_r = \sum_{k \in U_r} y_k / N_r$; (ii) The one-step calibration estimator $\hat{t}_{C,1}$ given by (14) based on different calibration functions: linear, exponential and logit, given by (4), (5) and (7), respectively, using $\mathbf{x}_k = (1, x_k)$ as the auxiliary vector. In other words, the estimator $\hat{t}_{C,1}$ was calibrated on the population size N as well as the population total of x -values, t_x ; (iii) the Propensity-Score Adjusted estimator \hat{t}_{PSA} , where the response propensities were estimated using the score method described in Section 3. To that end, preliminary response probabilities ϕ_k were first obtained using a logistic regression model with $(1, x_k)^\top$ as the vector of predictors. Then, the sample was partitioned into 20 weighting classes according the $\hat{\phi}_k$'s and the response propensity of a unit in a given class was estimated using the response rate observed within the same class. Although five imputation classes are often sufficient for an effective bias reduction (Eltinge and Yansaneh 1997; Rosenbaum and

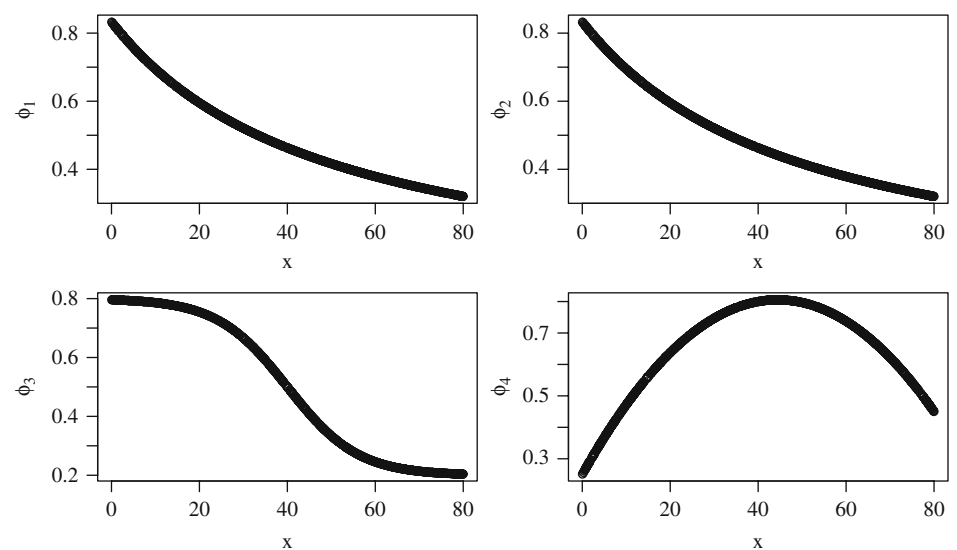


Fig. 2. Relationships between the response probability and x

Rubin 1983), it may not be appropriate when the relationship between the characteristic of interest and the auxiliary variable is highly nonlinear or contains a quadratic terms as it is the case for y_1 and y_4 , respectively (see Haziza and Beaumont 2007). This is why we used 20 imputation classes.

As we argued in Section 4, in order to show that one-step calibration is vulnerable to the misspecification of the calibration function, it is sufficient to compare \hat{t}_{PSA} and $\hat{t}_{C,1}$ based on the x^o -variables only. In other words, there is no need to perform the second step in the two-step approach or to use x^* -type variables in the one-step approach.

As a measure of bias of an estimator $\hat{\theta}$ of a parameter θ , we used the Monte Carlo percent relative bias (RB)

$$RB_{MC}(\hat{\theta}) = \frac{100}{M} \sum_{m=1}^M \frac{(\hat{\theta}_{(m)} - \theta)}{\theta},$$

where $\hat{\theta}_{(m)}$ denotes the estimator $\hat{\theta}$ in the m -th repetition, $m = 1, \dots, M$. We also computed the percent relative root mean square error (RRMSE) of $\hat{\theta}$:

$$RRMSE_{MC}(\hat{\theta}) = 100 \times \frac{\left\{ M^{-1} \sum_{m=1}^M (\hat{\theta}_{(m)} - \theta)^2 \right\}^{1/2}}{\theta}.$$

The results are shown in Tables 2–5. As expected, the unadjusted estimator was biased in all the scenarios. This can be explained by the fact that the response probability was related to the characteristics of interest via the auxiliary variable x and that the unadjusted estimator did not account for x .

We now turn to the variable y_1 , which was linearly related to the variable x . We see from Tables 2–5 that the resulting one-step calibration estimator $\hat{t}_{C,1}$ showed negligible bias regardless of the calibration method $F(\cdot)$ used. This is consistent with the Expressions (15)–(18). Furthermore, the choice of calibration method did not affect the efficiency of the estimator for a given nonresponse mechanism. For example, in Table 2

Table 2. Monte Carlo percent relative bias and percent relative root mean square error (in parenthesis) of several estimators under the inverse linear nonresponse mechanism: $\Phi_k = (1.2 + 0.024 x_k)^{-1}$

	\hat{t}_{un}	$\hat{t}_{yC,1}$ $F(u) =$			\hat{t}_{PSA}
		$1 + u$	$\exp(u)$	$\frac{L(M - 1) + M(1 - L) \exp(Au)}{M - 1 + (1 - L) \exp(Au)}$	
y_1 (linear)	−4.1 (4.2)	0.0 (0.7)	0.0 (0.7)	0.0 (0.7)	−0.0 (0.8)
y_2 (exponential)	−28.1 (28.7)	−0.1 (5.5)	2.8 (6.1)	3.3 (6.4)	−0.1 (3.0)
y_3 (logistic)	−27.5 (27.9)	−0.1 (3.4)	1.7 (3.6)	2.1 (3.8)	−0.1 (2.3)
y_4 (quadratic)	−4.8 (5.3)	0.1 (2.8)	−2.0 (3.3)	−2.4 (3.5)	−0.1 (1.4)

Table 3. Monte Carlo percent relative bias and percent relative root mean square error (in parenthesis) of several estimators under the exponential nonresponse mechanism: $\Phi_k = \exp(-0.2 - 0.014x_k)$

	\hat{t}_{un}	$\hat{t}_{yC,1}$ $F(u) =$			\hat{t}_{PSA}
		$1 + u$	$\exp(u)$	$\frac{L(M - 1) + M(1 - L) \exp(Au)}{M - 1 + (1 - L) \exp(Au)}$	
y_1 (linear)	− 4.9 (4.9)	− 0.0 (0.8)	0.0 (0.8)	0.0 (0.8)	− 0.0 (0.8)
y_2 (exponential)	− 35.1 (35.5)	− 4.0 (7.1)	− 0.0 (5.8)	0.7 (5.9)	− 0.1 (3.2)
y_3 (logistic)	− 33.8 (34.1)	− 2.5 (4.3)	0.0 (3.3)	0.6 (3.3)	− 0.1 (2.3)
y_4 (quadratic)	− 3.6 (4.3)	2.9 (4.2)	0.0 (2.7)	− 0.6 (2.8)	− 0.0 (1.6)

(which corresponds to the inverse linear nonresponse mechanism), the RRMSE of $\hat{t}_{C,1}$ was equal to 0.7 for all the calibration methods. Finally, the PSA estimator showed virtually no bias in all the scenarios corresponding to the y_1 -variable and showed the same efficiency as that of $\hat{t}_{C,1}$, except in Table 2, where we note a slight loss of efficiency.

For the variables y_2 - y_4 that were not linearly related to the x -variable, we note that the resulting one-step calibration estimator was generally biased, except when the calibration method $F(\cdot)$ was appropriate; see Expression (19). For example, in Table 2 (which corresponds to the inverse linear nonresponse mechanism), the one-step calibration estimator $\hat{t}_{C,1}$ showed no bias for the three variables under the linear calibration method $F(u) = 1 + u$. These results are consistent with (20). On the other hand, the other calibration methods (exponential and logit) led to some bias with an absolute RB ranging

Table 4. Monte Carlo percent relative bias and percent relative root mean square error (in parenthesis) of several estimators under the logistic nonresponse mechanism: $\Phi_k = 0.2 + 0.6\{1 + \exp(-5 + x_k/8)\}^{-1}$

	\hat{t}_{un}	$\hat{t}_{yC,1}$ $F(u) =$			\hat{t}_{PSA}
		$1 + u$	$\exp(u)$	$\frac{L(M - 1) + M(1 - L) \exp(Au)}{M - 1 + (1 - L) \exp(Au)}$	
y_1 (linear)	− 7.3 (7.3)	− 0.3 (0.9)	− 0.2 (0.9)	− 0.2 (0.9)	− 0.1 (0.9)
y_2 (exponential)	− 51.5 (51.7)	− 10.0 (12.3)	− 0.4 (7.0)	0.9 (7.1)	− 0.2 (3.7)
y_3 (logistic)	− 53.4 (53.5)	− 12.1 (12.9)	− 5.6 (6.7)	− 4.5 (5.8)	− 0.3 (3.0)
y_4 (quadratic)	− 1.0 (2.3)	11.7 (12.3)	4.3 (5.3)	3.1 (4.3)	− 0.0 (1.8)

Table 5. Monte Carlo percent relative bias and percent relative root mean square error (in parenthesis) of several estimators under the quadratic nonresponse mechanism: $\Phi_k = 0.7 + 0.45 (x_k/40 - 1)^2 + 0.0025 x_k$

	\hat{t}_{un}	$\hat{t}_{yC,1}$ $F(u) =$			\hat{t}_{PSA}
		$1 + u$	$\exp(u)$	$\frac{L(M - 1) + M(1 - L) \exp(Au)}{M - 1 + (1 - L) \exp(Au)}$	
y_1 (linear)	1.3 (1.4)	-0.2 (0.5)	-0.2 (0.5)	-0.2 (0.5)	0.0 (0.5)
y_2 (exponential)	-8.3 (9.4)	-19.7 (19.9)	-19.2 (19.5)	-19.0 (19.3)	-0.4 (2.3)
y_3 (logistic)	-0.5 (3.2)	-11.8 (12.0)	-11.5 (11.7)	-11.4 (11.6)	-0.1 (1.0)
y_4 (quadratic)	13.1 (13.2)	13.7 (13.8)	13.4 (13.5)	13.2 (13.4)	0.2 (1.1)

from 1.7% to 3.3%. Similarly, in Table 3 (which corresponds to the exponential nonresponse mechanism), the one-step calibration estimator $\hat{t}_{C,1}$ showed no bias for the three variables under the exponential calibration method $F(u) = \exp(u)$. These results are consistent with (21). On the other hand, the other calibration methods (linear and logit) led to some bias with an absolute RB ranging from 0.6% to 4.0%.

In Tables 4 and 5, we note that the one-step calibration estimator showed some bias in all the scenarios, which can be explained by the fact that none of the calibration methods (linear, exponential or logit) provided an adequate description of the relationship between the inverse of the response probability and the x -variable. For example, in Table 5, all the calibration methods led to substantial bias with an absolute RB ranging from 11.4% to 19.7%. It is worth noting that the one-step calibration estimator was significantly more biased than the unadjusted estimator for the variables y_2 and y_3 , which illustrates that a poor choice of $F(\cdot)$ may result in significant biases, which can be larger than that of the unadjusted estimator. Finally, the PSA estimator showed negligible biases in all the scenarios corresponding to y_2 - y_4 . Moreover, its RRMSE was considerably smaller than that of the one-step calibration estimator for these variables. These results suggest that the score method, which is nonparametric in nature, is robust to the misspecification of the form of the function $m(\cdot)$ in (8).

The results presented here suggest that a high association between the characteristic variable and the auxiliary variables is not necessarily enough for the one-step calibration method to yield good results, as in the cases of the variables y_2 and y_3 . Also, as shown in Table 5, the fact that various calibration functions yield about the same estimate is not necessarily a sign that any of the choices will work well.

6. Discussion

In this article, we have discussed two weighting approaches in the presence of unit nonresponse: the one-step approach and the two-step approach, the latter being the

customary approach to weighting in statistical agencies. Although it is more complex to implement than the one-step approach as two distinct weighting procedures must be applied, the two-step approach offers several advantages: first, it makes it possible to assess the impact of nonresponse adjustment and calibration adjustment on the distribution of the weights separately. Furthermore, when multiple characteristics are collected, survey statisticians prefer modeling the response probability to the survey as it does not require a different model for each characteristic of interest. In this case, complete reliance is placed on the nonresponse model in order to achieve an efficient bias reduction for every y . In statistical agencies, the response propensities are typically estimated through nonparametric methods such as weighting classes based on estimated response probabilities or regression trees, as both types of methods provide protection against misspecification of the functional and account for curvature and interactions. This is especially important when the auxiliary variables are continuous and their association with the response probability is not monotonic.

In contrast, the single-step calibration approach is simple to implement as the whole weighting process is performed in a single step. Furthermore, it does not make explicit use of estimated response probabilities, unlike the two-step approach. However, as we have illustrated empirically, the choice of the calibration function $F(\cdot)$ is generally important. In the simulation study conducted in Section 5, where we considered the case of a quantitative variable x , the results suggested that the one-step calibration estimator suffered from significant bias if the calibration function is inappropriate. Would the results be similar if the calibration results were categorical? We revisit the case of two categorical variables x_1 and x_2 described in Section 2. Let N_{j_1, j_2} be the individual cell counts available at the sample level (Info- s). Matching the individual cell counts results in a poststratified-type estimator, in which case the choice of the calibration function is unimportant as different $F(\cdot)$ would result in the same estimator. In other words, as long as the variables x_1 and x_2 are related to nonresponse, the one-step calibration estimator should exhibit no bias. In fact, in this case, the latter is identical to the PSA estimator based on weighting classes obtained by cross classifying x_1 and x_2 . On the other hand, if calibration is performed to match the margins $N_{j_1 \bullet}$ and $N_{\bullet j_2}$ available at the sample level, choosing the appropriate calibration function becomes an issue once again, as different $F(\cdot)$ would lead to different one-step calibration estimators. In their Remark 10.1, [Särndal and Lundström \(2005\)](#) suggest that categorizing the x -variables when the latter are quantitative may bring some robustness. For a poststratification-type situation, we agree with this recommendation. However, when calibration is performed on margins only, the extent to which the one-step calibration estimators would be robust to the misspecification of the calibration function $F(\cdot)$ is not so clear-cut.

Although the PSA estimator based on the score method performed well in all the scenarios presented in Section 5, we are not suggesting that it would perform well in any type of situations. If a causal relationship exists between one or more characteristics of interest and the response propensity, some residual nonresponse bias will remain. Furthermore, we have considered the case of a single quantitative variable x . Additional studies are needed to investigate how the score method would perform in the presence of multiple quantitative variables with, possibly, quadratic or cubic terms. The results simply suggest that nonparametric methods are attractive from a practical point of view as they

bring some robustness if the nonresponse model is not correctly specified. This is not true, in general, for the one-step approach that imposes an implicit parametric model.

7. References

- Brick, M.J. 2013. "Unit Nonresponse and Weighting Adjustments: A Critical Review." *Journal of Official Statistics* 29: 329–353. Doi: <http://dx.doi.org/10.2478/jos-2013-0026>.
- Da Silva, D.N. and J.D. Opsomer. 2006. "A Kernel Smoothing Method of Adjusting for Unit Non-Response in Sample Surveys." *The Canadian Journal of Statistics* 34: 563–579.
- Da Silva, D.N. and J.D. Opsomer. 2009. "Nonparametric Propensity Weighting for Survey Nonresponse through Local Polynomial Regression." *Survey Methodology* 35: 165–176.
- Deville, J.-C. and C.-E. Särndal. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association* 87: 376–382. Doi: <http://dx.doi.org/10.1080/01621459.1992.10475217>.
- Deville, J.-C., C.-E. Särndal, and O. Sautory. 1993. "Generalized Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88: 1013–1020. Doi: <http://dx.doi.org/10.1080/01621459.1993.10476369>.
- Ekhholm, A. and S. Laaksonen. 1991. "Weighting via Response Modeling in the Finnish Household Budget Survey." *Journal of Official Statistics* 7: 325–337.
- Eltinge, J.L. and I.S. Yansaneh. 1997. "Diagnostics for Formation of Nonresponse Adjustment Cells, with an Application to Income Nonresponse in the U.S. Consumer Expenditure Survey." *Survey Methodology* 23: 33–40.
- Giommi, A. 1987. "Nonparametric Methods for Estimating Individual Response Probabilities." *Survey Methodology* 13: 127–134.
- Haziza, D. and J.-F. Beaumont. 2007. "On the Construction of Imputation Classes in Surveys." *International Statistical Review* 75: 25–43. Doi: <http://dx.doi.org/10.1111/j.1751-5823.2006.00002.x>.
- Kalton, G. and I. Flores-Cervantes. 2003. "Weighting Methods." *Journal of Official Statistics* 19: 81–97.
- Kim, J.K. and J.J. Kim. 2007. "Nonresponse Weighting Adjustment Using Estimated Response Probability." *The Canadian Journal of Statistics* 35: 501–514. Doi: <http://dx.doi.org/10.1002/cjs.5550350403>.
- Kott, P. 2006. "Using Calibration Weighting to Adjust for Nonresponse and Undercoverage." *Survey Methodology* 32: 133–142.
- Kott, P.S. and D. Liao. 2012. "Providing Double Protection for Unit Nonresponse With a Nonlinear Calibration-Weighting Routine." *Survey Research Methods* 6: 105–111.
- Lee, S. 2006. "Propensity Score Adjustments as a Weighting Scheme for Volunteer Panel Web Surveys." *Journal of Official Statistics* 22: 329–349.
- Little, R.J.A. 1986. "Survey Nonresponse Adjustments for Estimates of Means." *International Statistical Review* 54: 139–157.
- Little, R.J.A. and S. Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31: 161–168.

- Phipps, P. and D. Toth. 2012. “Analyzing Establishment Nonresponse Using an Interpretable Regression Tree Model With Linked Administrative Data.” *Annals of Applied Statistics* 6: 772–794. Doi: <http://dx.doi.org/10.1214/11-AOAS521>.
- Rosenbaum, P.R. and D.B. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70: 41–55. Doi: <http://dx.doi.org/10.1093/biomet/70.1.41>.
- Rubin, D.B. 1976. “Inference and Missing Data.” *Biometrika* 63: 581–590. Doi: <http://dx.doi.org/10.1093/biomet/63.3.581>.
- Särndal, C.-E. 2011. “Three Factors to Signal Non-Response Bias With Applications to Categorical Auxiliary Variables.” *International Statistical Review* 79: 233–254. Doi: <http://dx.doi.org/10.1111/j.1751-5823.2011.00142.x>.
- Särndal, C.-E. and S. Lundström. 2005. *Estimation in Surveys with Nonresponse*. New York: John Wiley and Sons.