

Secondary Publication



Wildner, Raimund; Bosch, Volker; Meinfelder, Florian

Calibrating Nonresponse Bias : A Cautionary Tale

Date of secondary publication: 11.06.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-115542x

Primary publication

Wildner, Raimund; Bosch, Volker; Meinfelder, Florian (2026): Calibrating Nonresponse Bias : A Cautionary Tale, in: Journal of official statistics : JOS, Stockholm: SAGE, Vol. 42, No. 2, pp. 292–311, doi: 10.1177/0282423x261425800.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

Calibrating Nonresponse Bias: A Cautionary Tale

Journal of Official Statistics

2026, Vol. 42(2) 292–311

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0282423X261425800

journals.sagepub.com/home/jof



Raimund Wildner¹, Volker Bosch²
and Florian Meinfelder³ 

Abstract

Low response rates due to unit nonresponse have always been a ubiquitous problem in survey-based empirical research, and calibration is a popular method to adjust for bias caused by unit nonresponse. Typically, some external information on the true population quantities of margins for some calibration variables is available, and sometimes also of higher-order interactions. Weighting algorithms try to adjust the sample to these external benchmarks. It is generally assumed that even if the underlying missingness mechanism of the unit nonresponse is non-ignorable, weighting will at least alleviate the severity of the bias. We discuss data situations where weighting under a missing at random (MAR) assumption adjusts the sample correctly but still increases the bias for the analysis model, and we describe strategies for identifying auxiliary variables that are less susceptible to these unwanted effects.

Keywords

unit nonresponse, calibration, weighting, selection at random, selection not at random

1. Introduction

Weighting is fundamental to survey research. It allows researchers to address heterogeneous inclusion probabilities resulting from both the survey design and unit nonresponse. If all heterogeneity can be attributed to the survey design, the weights are fixed, known quantities. However, calibration weights for unit nonresponse are unknown and must be treated as estimates (see e.g., Little and Vartivarian 2003).

¹University of Erlangen-Nuremberg, Erlangen, Germany

²YouGov Germany, Nuremberg, Germany

³University of Bamberg, Bamberg, Germany

Corresponding author:

Florian Meinfelder, University of Bamberg, Feldkirchenstraße 21, Bamberg 96052, Germany.

Email: florian.meinfelder@uni-bamberg.de



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work

without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

The problem of unit nonresponse has been recognized for decades (see e.g., Hansen and Dedrick 1938; Hansen and Hurwitz 1946), and comprehensive overviews are provided by, for example, Groves et al. (2002), Särndal and Lundström (2005), Groves (2006), Särndal (2007), Groves and Peytcheva (2008), Bethlehem (2011). Weighting methods to mitigate unit-nonresponse bias date back at least to iterative proportional fitting (Deming and Stephan 1940) and have been continuously refined over the years (see e.g., Brick 2013; Lundström and Särndal 1999). Nevertheless, the underlying principle has remained the same: calibrating the sample so that the means of adjustment cells match external information from the full sample or population. While Kott and Chang (2010) compare different strategies for calibrating unit-nonresponse bias, practical recommendations remain scarce in the literature, even though the problem has intensified due to ever-decreasing response rates (see e.g., Curtin et al. 2005).

Adjusting for unit nonresponse relies on a corresponding nonresponse model, which itself is based on rather strong assumptions. One assumption frequently made (either explicitly or implicitly) to handle missing data is ignorability (see e.g., Little and Rubin 2019). If, for instance, we observe a skewed age distribution in our sample, we typically calibrate the sample to the age distribution of the target population. In doing so, we implicitly assume that “Age” is the cause of the unit nonresponse and that conditioning on age removes any associated bias. In contrast, non-ignorability describes situations where it is impossible to fully eliminate nonresponse bias using the available information. In such cases, we might still observe a skewed age distribution, even though the (unknown) cause of nonresponse is merely correlated with it. These assumptions and their underlying mechanisms were first described by Rubin (1976) in the context of item nonresponse, but it took several decades for the concept to be generally adopted for both item and unit nonresponse problems.

Since ignorability cannot be disproved by the observed data, it often seems plausible to rely on this assumption even if it is unlikely to hold. This approach appears intuitively correct: if the factors causing nonresponse are correlated with the variables used in the nonresponse model, then controlling for these variables should, in theory, reduce estimator bias. Several publications support this strategy (see e.g., Deville 2000; Earp et al. 2012; Kott and Liao 2017), although Oh and Scheuren (1983) and Bethlehem (1988) had already noted that weighting can occasionally increase bias. Furthermore, Kreuter and Olson (2011) demonstrated in a simulation study that the choice of predictors significantly impacts the degree of bias reduction, with the potential for bias to actually increase.

In this paper, we investigate the validity of the ignorability assumption in the context of calibration. We introduce data scenarios where, contrary to intuition, bias becomes more severe. Finally, we propose strategies for identifying calibration variables that increase rather than reduce nonresponse bias.

The remainder of this paper is structured as follows: Section 2 describes the underlying assumptions for missing-data mechanisms and their application to our framework, which is detailed in Section 3. Section 4 provides simulation-based demonstrations and develops practical recommendations. The paper concludes

with a brief summary of our findings in Section 5. Corresponding proofs and derivations can be found in the Appendix.

2. Missing Data Mechanisms for Unit Nonresponse

2.1. Historical Background

Rubin (1987) originally described missing data mechanisms from the perspective of item nonresponse, primarily because multiple imputation was developed to alleviate problems associated with item nonresponse. In the original notation, let $Y = [Y_1, Y_2, \dots, Y_k]$ denote a multivariate (and only partially observed) variable (i.e., the analysis dataset), and θ its underlying data-generating process. All observed parts of the data are denoted by Y_{obs} , and missing parts by Y_{mis} . We let R denote the positions of missing values in Y ; specifically, it takes a value of 0 for Y_{mis} and 1 for the observed part Y_{obs} (see e.g., Little and Rubin 2019). The indicator variable R is governed by a set of unknown parameters ξ . For any analysis involving missing data, researchers must consider the joint distribution $p(Y, R|\theta, \xi) = p(R|Y_{obs}, Y_{mis}, \theta, \xi) \cdot p(Y_{obs}, Y_{mis}|\theta, \xi)$. Missing at random (MAR) simplifies this equation to $p(R|Y, \xi) = p(R|Y_{obs}, Y_{mis}, \xi) = p(R|Y_{obs}, \xi) \forall Y_{mis}$, where the missingness depends only on the observed part of the data Y_{obs} . An additional requirement is the distinctness of θ and ξ , meaning we assume that the parameters governing the data model and the missingness model are a priori independent (in the Bayesian sense). The more general assumption of ignorability—the synthesis of MAR and distinctness—implies that valid inferences can be obtained from incomplete data without explicitly specifying a model for the missing data mechanism.

In contrast, Missing not at random (MNAR) means that $p(R|Y_{obs}, Y_{mis}, \xi)$ cannot be further simplified because missingness depends on information contained in the missing variables themselves. In such situations, the missingness mechanism is described as nonignorable.

2.2. (Non-)Ignorability in the Context of Sampling

Andridge et al. (2019) introduce a slight modification of MAR for unit nonresponse in survey sampling, where “missing” is replaced by “selection.” For instance, any calibration model implicitly assumes selection at random (SAR) to produce unbiased point estimators. To illustrate this point, we slightly modify Rubin’s original notation: while Rubin described data using a multivariate variable Y , we introduce a completely observed auxiliary variable X and hereafter refer to Y as the (univariate or multivariate) analysis variable.

A peculiarity of unit nonresponse problems is that they are not immediately recognizable as missing-data problems. A typical empirical survey-based dataset does not contain statistical units whose entire set of variables is missing (e.g., due to failed contact or refusal) such that their values are explicitly set to “NA” or another missing data indicator. Instead, these units are omitted entirely from the released data. The remaining sample of respondents is then often calibrated using external aggregated information, such as population margins from a census. Over

time, this common practice has implicitly embedded the ignorability assumption into the calibration literature (see e.g., Brick 2013).

2.3. Missing Data Mechanisms Based on Unobserved Variables

As stated in Section 2.2, the original definition of missing-data mechanisms focused on information that was (at least theoretically) available within the data itself. In our research, we take this notion one step further by introducing a scenario in which missingness is governed by unobserved variables. While such a scenario is particularly plausible in unit nonresponse, it is also theoretically possible in item nonresponse settings. For instance, a common textbook example of MNAR is missing income values, where high (or low) income increases the probability of nonresponse. Although the underlying causes for missing values are rarely discussed in detail, one could imagine a scenario where the actual reason for missing income data is a latent trait such as “cautiousness.” Cautiousness might not be captured by the survey but is nonetheless related to income.

We introduce such a situation into our unit nonresponse framework, where the cause for (non-)participation lies outside the survey; we denote this unobserved variable as Z . We further denote the analysis variable as Y and the auxiliary (calibration) variable as X . The nonresponse-governing variable Z is not part of the survey; therefore, only X and Y are included in the indicator matrix R . The resulting dataset contains only $[X', Y']$, and unit nonresponse is typically never explicitly indicated, as this schematic overview suggests.

Calibration can only be based on X , since Z is not measured. Any calibration method assumes ignorability—that is, we assume $Pr(R = 1|X, Y, Z)$ can be reduced to $Pr(R = 1|X)$. However, if the analysis variable Y is linked to Z and $p(Z, Y|X) \neq p(Z|X) \cdot p(Y|X)$, the mechanism is nonignorable, a condition referred to as selection not at random (SNAR). In this case, $Pr(R = 1|X, Y, Z)$ cannot be reduced to $Pr(R = 1|X)$.

Moreover, calibration has the potential to introduce bias that would not have existed otherwise, as initially noted by Oh and Scheuren (1983). We aim to investigate the underlying statistical mechanisms to mitigate the risk of increasing bias in practical applications. To illustrate the problem, consider the following example: the unobserved variable Z is correlated with the calibration variable X , but not with the analysis variable Y . Specifically, imagine that “Headache” (Z) governs the propensity to participate in a survey. It is positively correlated with “Age” (X) but marginally uncorrelated with “Income” (Y). In this case, an analysis relying solely on the marginal distribution of Y would yield an unbiased estimator. However, since the propensity to have a headache is positively correlated with age, the marginal distribution of age in the sample will be skewed (as the probability of missingness increases with age). While it is tempting to calibrate by age groups, if “Age” and “Income” are not independent, this calibration will introduce bias that was not previously present. Although this may seem counterintuitive, it is mathematically entirely possible.

We can take this example further by demonstrating that calibration can also exacerbate existing bias. Assume that “Headache” (Z) is negatively correlated with “Income” (Y), while “Age” (X) and “Headache” remain positively correlated, as do “Age” and “Income.” In the uncalibrated sample, “Income” is overestimated due to its negative correlation with “Headache” (wealthier people with fewer headaches are overrepresented). Conversely, “Age” is underestimated because older people (who have more headaches) are underrepresented. Calibrating for “Age” will therefore further increase the overestimation of “Income,” as the wealthier older participants receive even higher weights. We emphasize that although the weighting is performed correctly according to standard procedures, the incorrect ignorability assumption leads to an increase in bias.

Using these “Headache” examples as a starting point, this paper derives a formal description of nonresponse and weighting to understand the conditions under which calibration fails. This requires an appropriate description of data transformations through both nonresponse and weighting. As we shall see, the associations between the three variables X , Y , and Z determine the resulting bias. The respective correlation matrix for $[X, Y, Z]$ is constrained by the Fréchet-Hoeffding bounds to remain positive semidefinite (see e.g., Kiesl and Rässler 2006), a well-known constraint frequently discussed in the context of statistical matching (see e.g., Rässler 2002).

3. A Model of Unit Nonresponse and Weighting

3.1. Latent Population Level, Selection through Nonresponse, and Weighting

In this section, we introduce additional assumptions and constraints regarding the joint distribution $p(X, Y, Z, R)$. Our approach belongs to the class of selection models for missing data (Heckman 1979), where the joint distribution of the data and the indicator variable is decomposed as $p(X, Y, Z, R) = p(R|Z) \cdot p(X, Y, Z)$. We assume throughout this framework that the response propensity depends solely on Z .

The expected impact of nonresponse on survey data and its subsequent calibration via weighting is captured using a linear model. In a linear framework, the effect of a change in variable A on variable B depends on the absolute change ΔA , rather than the initial value of A ; this makes correlation a sufficient measure to describe the association between two variables. A linear model is chosen for several reasons. First, it is parsimonious enough to allow for a comprehensive mathematical formulation with unambiguous results. Second, it is relatively robust to violations of linearity. Third, if unexpected outcomes emerge within a linear model, it is highly probable that more extreme outcomes will occur under non-linear data-generating processes.

In its basic form, the model employs three variables $[X, Y, Z]$. It describes the properties of the total bias $b_{y^w} := E(Y^w) - \mu_y$ for the variable of interest Y that persists after both nonresponse (governed by the unobserved variable Z) and weighting for the auxiliary variable X have taken effect (see Figure 1). Here, $E(\cdot)$ denotes

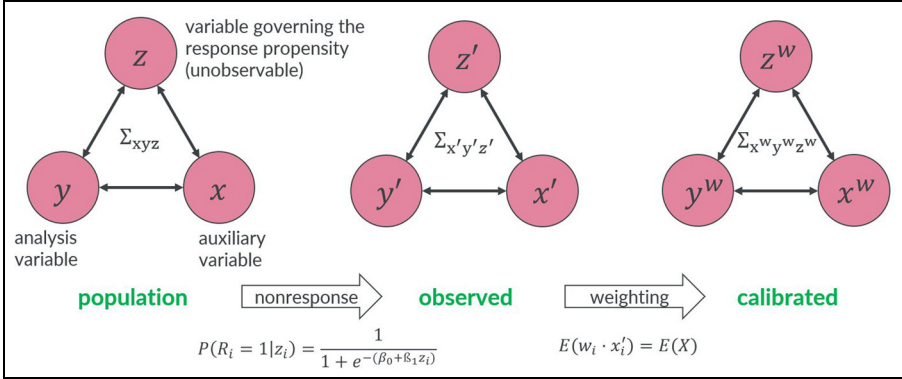


Figure 1. A model describing the mechanisms of nonresponse and weighting.

the expectation value of the underlying random variable, as we focus on unconditional means as the primary analysis model.

In general, Z is unobservable and determines R , the response indicator. Y represents any variable of interest, and X is an auxiliary variable for which the true population mean is known. Each of the variables $[X, Y, Z]$ is assumed to be interval-scaled.

Selection through nonresponse can be formalized as conditioning population-level variables $[X, Y, Z]$ on a positive response, resulting in observed survey-level variables $[X', Y', Z']$, that is, $p(X', Y', Z') = p(X, Y, Z | R = 1)$.

A feasible approach to modeling nonresponse is a logistic regression where a higher value of z_i for unit i results in a higher probability of a positive response $Pr(R_i = 1)$. Each unit i of the contacted population responds according to this unit-specific probability; consequently, only n_r out of n contacted units are measured. The fraction $\frac{n_r}{n}$ defines the response rate ρ (see e.g., Lohr 2021).

In addition to reducing the sample size, nonresponse considerably affects the observed data structure. The multivariate distribution of the survey data $p(X', Y', Z')$ is distorted relative to the population distribution $p(X, Y, Z)$. That is, spurious correlations are introduced such that $\Sigma_{xyz} \neq \Sigma_{x'y'z'}$, and the observed distributions may be significantly biased compared to their respective population distributions. Most notably, the expected values of $[X', Y', Z']$ may shift relative to the population means, resulting in significant biases $b_{x'}, b_{y'}, b_{z'}$ (see Appendix A “Nonresponse Bias” for a detailed proof). This is critical for surveys aiming to estimate population means and proportions—a category that includes nearly all surveys in official statistics, market research, public health, and social research.

Consequently, survey-based research must account for the bias $b_{y'} := E(Y') - \mu_y$. Weighting is employed because it can completely eliminate bias $b_{x'}$ in the auxiliary variables by utilizing reliable external information. Specifically,

unit-specific weights are chosen so that the weighted mean X^w equals the population mean μ_x . As a result, weighting creates a new set of transformed variables $[X^w, Y^w, Z^w]$ with distinct distributions and covariances.

For the purpose of describing biases in mean estimators rather than their standard errors, the specific weighting algorithm used is largely irrelevant, provided the weighted mean of the auxiliary variable matches the population target. However, since interval-scaled auxiliary variables provide a more straightforward initial approach, simple algorithms like iterative proportional fitting are insufficient. Therefore, we assume the application of a GREG weighting algorithm (e.g., Bethlehem and Keller 1987; Deville et al. 1993; Rao and Singh 1997) capable of weighting for mean values (see Appendix B).

3.2. Model Outline

Assuming a simple random sample, the model describes how nonresponse bias propagates from Z to X and Y , and how weighting-induced counter-bias further propagates from X to Y . While nonresponse is the root cause of the bias, the model applies to any unintended deviation from equal-probability sampling. Note that the model does not describe the impact of weighting on the precision of estimates; weighting can, however, inflate the variance of Y , as noted by Kish (1987) and Little and Rubin (2019).

The framework demonstrates that weighting can successfully remove all nonresponse bias in Y when the underlying response process is SAR. It also illustrates why the response rate alone cannot predict the extent of bias before weighting. Instead, the combination of selection bias and variance reduction in the unobserved propensity variable Z constitutes the primary cause. These two features of nonresponse are propagated to observable variables via the joint covariance matrix, which fully determines the nonresponse bias in any observed variable.

3.3. Bias in the Variable of Interest Y

In the following, let σ_x denote the standard deviation of X and r_{xy} the correlation between X and Y (likewise for other variables). For explanatory purposes (WLOG), we assume that the participation propensity is positively associated with Z . The variance of the response propensity variable Z is reduced by nonresponse because units with higher values of Z are more likely to be selected. Moreover, nonresponse disproportionately affects the lower tail of the Z -distribution, shifting the average upward. Hence, $\sigma_z^2 < \sigma_z'^2$ and $b_z := E(Z') - \mu_z > 0$. We define $\delta := 1 - \frac{\sigma_z'^2}{\sigma_z^2}$ as the relative loss of variance in Z .

The parameters δ and b_z are determined by the selection process, which is largely unobservable. For simulation purposes, however, these parameters can be manipulated via the β_0 and β_1 parameters of a logistic regression model. Importantly, δ and b_z do not fully determine each other; both must be considered when calculating nonresponse effects. Knowledge of δ , b_z , and the covariance matrix Σ_{xyz} is sufficient to derive the total bias of Y after weighting, b_{y^w} , as detailed in Appendix A “Weighting Bias”. The final result is given by Equation (1):

$$b_{y^w} = b_{z'} \cdot \frac{\sigma_{y'}}{\sigma_z} \cdot \left[r_{yz} - \frac{r_{xy} \cdot r_{xz} - r_{yz} \cdot r_{xz}^2 \cdot \delta}{1 - r_{xz}^2 \cdot \delta} \right]. \quad (1)$$

This is a non-trivial result with implications that are not immediately obvious. We first analyze this by assuming SAR conditioned on X in Section 3.4, followed by the SNAR case in Section 3.5.

3.4. Weighting under SAR

The definition of SAR states that nonresponse does not affect Y when the data is controlled for X . That is, Z and Y are independent when conditioned on X , which, in a linear framework, implies $r_{yz} = r_{xz} \cdot r_{xy}$. Consequently, under SAR, knowledge of X is sufficient to eliminate the nonresponse bias of Y' through weighting. A proof is provided in Appendix A “Nonresponse with SAR”, Equation (A.11).

To justify SAR, the auxiliary variable(s) X must fully characterize the selection process. For example, in a landline survey, X would need to capture the probability of an individual being home, their willingness to answer the phone, and the internal household selection process. While sociodemographic variables may capture part of this process, they are unlikely to account for it entirely. Thus, SAR is a very strong and often unrealistic assumption, as it implies that all relevant causal influences on response behavior are captured by the available X .

3.5. Weighting under SNAR

Under SNAR, conditional independence no longer holds. Z biases Y without being fully controllable via X , making it impossible to judge the impact of nonresponse and weighting with high precision. To describe the uncertainty under SNAR, we explore the potential space of the total bias b_{y^w} as a function of δ , $b_{z'}$, and Σ_{xyz} .

A tipping point is reached when the bias after weighting b_{y^w} is equal to the initial bias $b_{y'}$ but with the opposite sign:

$$-b_{y'} = b_{y^w}.$$

This point defines the boundary between a decrease and an increase in bias due to weighting. If r_{xy} and r_{xz} have identical signs, weighting increases the bias of Y if and only if (see Appendix A “Determining the Critical Value for r_{yz} ”):

$$r_{yz} < \frac{r_{xy} \cdot r_{xz}}{2 - r_{xz}^2 \cdot \delta}. \quad (2)$$

Conversely, if r_{xy} and r_{xz} have opposite signs, weighting increases the absolute bias of Y if and only if:

$$r_{yz} > \frac{r_{xy} \cdot r_{xz}}{2 - r_{xz}^2 \cdot \delta}. \quad (3)$$

As a first approximation, the term $r_{xz}^2 \cdot \delta$ can be assumed to be negligible.

Weighting always increases the bias of Y when the initial bias $b_{y'}$ is zero ($r_{yz} = 0$), as in the first “Headache” example in Section 2.3. At the Fréchet-Hoeffding bounds ($\delta \rightarrow 1$ and $r_{xz} = r_{xy} = \sqrt{2}/2$), the entire initial bias of Z is transferred to Y with an inverted sign: $\frac{b_{y^{w'}}}{\sigma_y} = -\frac{b_z}{\sigma_z}$. Even when δ is near zero, half of the initial bias b_z is still induced into Y .

Another detrimental mechanism occurs when the counter-bias pushes in the same direction as the initial bias, thereby compounding the total bias. This is akin to the second “Headache” example in Section 2.3. This happens when the product of the correlations is negative ($r_{xy} \cdot r_{xz} \cdot r_{yz} < 0$). In an extreme case ($\delta \rightarrow 1$ and $|r_{xy}| = |r_{xz}| = |r_{yz}| = 0.5$), the initial bias is doubled rather than reduced. Even as $\delta \rightarrow 0$, the total bias remains 50% higher than the initial nonresponse bias.

4. Illustration of Practical Relevance

4.1. Laboratory Examples

Figure 2 illustrates the biases of Y for six scenarios (A–F) with varying parameters β_0, β_1, r_{xy} , and r_{xz} . The bias of the unweighted mean $b_{y'}$ (dashed line) and the bias of the weighted mean $b_{y^{w'}}$ (solid line) of Y are shown as functions of r_{yz} . The parameters ρ, δ , and b_z are also provided for each scenario.

Bias values are only plotted for those r_{yz} for which the resulting correlation matrix remains positive semidefinite. The shaded regions indicate the ranges where the absolute bias of the weighted mean is smaller than that of the unweighted mean. In the unshaded regions, weighting increases rather than reduces the absolute bias.

All biases are linear in r_{yz} , with $b_{y^{w'}}$ exhibiting a steeper positive slope than $b_{y'}$, as derived from Equations (A.4) and (A.10) in Appendix A “Nonresponse Bias” and “Weighting Bias”. Furthermore, the unweighted mean is unbiased if and only if $r_{yz} = 0$, whereas the weighted mean is unbiased if and only if $r_{yz} = r_{xy} \cdot r_{xz}$.

Scenarios A to C in Figure 2 demonstrate that the magnitude of bias $b_{y'}$ is directly influenced by δ and b_z rather than the response rate ρ . Scenarios D and E illustrate that while r_{xz} and r_{xy} do not affect $b_{y'}$, they are primary determinants of $b_{y^{w'}}$. A comparison of scenarios E and F shows that reversing the sign of r_{xz} or r_{xy} results in a point-symmetric transformation relative to the origin.

Finally, the extreme cases discussed in the previous section are reflected in Scenarios D and E at the leftmost limit of the graphs (the Fréchet-Hoeffding bounds, where $\delta = 0.408$). In Scenario D, the weighted bias at $r_{yz} = 0.5$ is approximately two-thirds larger than in the unweighted case. In Scenario E, there is no initial bias at $r_{yz} = 0$, yet weighting induces a substantial bias of $b_{y^{w'}} = 0.63 \cdot \sigma_y$.

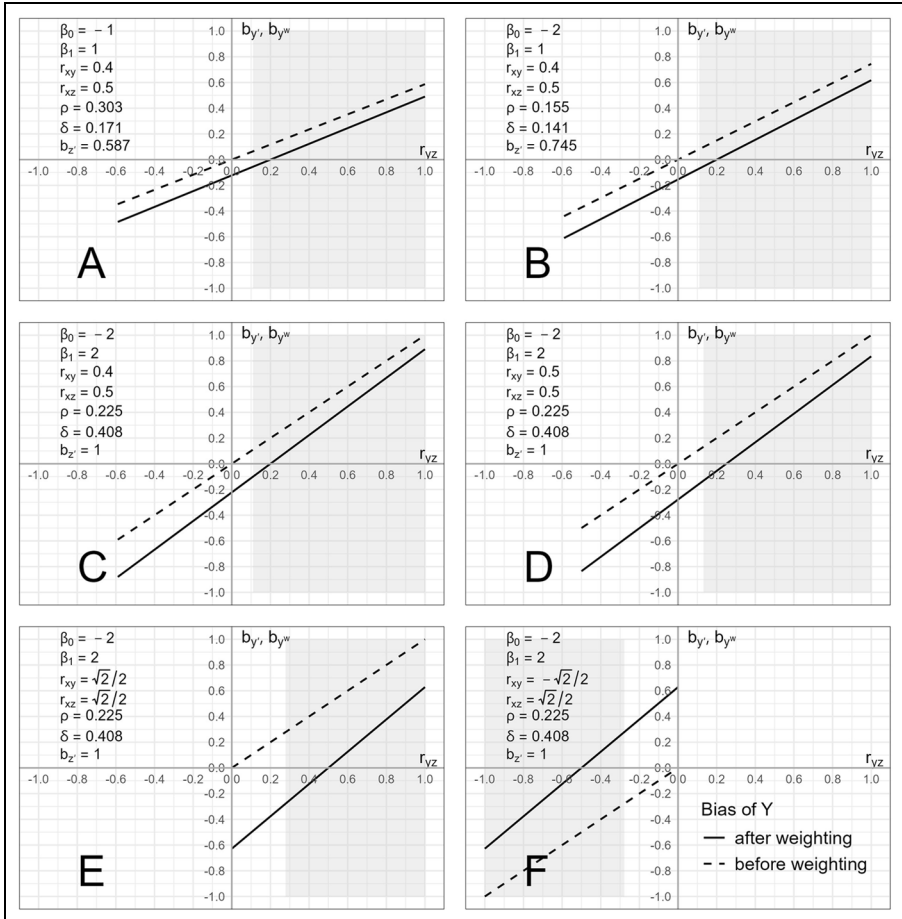


Figure 2. Six scenarios depicting bias before weighting b_y (dashed line) and after weighting b_{y^w} (solid line) as a function of r_{yz} with σ_y conveniently set to unity. The shaded area represents the range of r_{yz} where weighting leads to bias reduction. r_{yz} is constrained by the Fréchet-Hoeffding bounds.

4.2. Practical Examples

The fact that weighting for an auxiliary variable X can increase the bias of a dependent variable Y is counterintuitive. Previously, the prevailing assumption was that eliminating bias in an auxiliary variable would invariably propagate to any dependent variable, reducing bias there to some degree. To better understand the practical implications of counterproductive weighting, consider two illustrative examples.

First, imagine a landline survey where the response driver Z is characterized by “Openness” and “Accessibility”. The variable of interest, Y (“Gambling”), is largely determined by “Openness” and “Risk-Taking”. The auxiliary variable, X

(“Age”), can be viewed as a construct of “Accessibility” minus “Risk-Taking”, as the likelihood of being accessible via landline increases with age, while the propensity for risk-taking generally decreases. Consequently, Y and X are both positively correlated with Z , while Y and X are negatively correlated. Y' is positively biased because open-minded individuals are more likely to respond. Similarly, X' is positively biased due to the higher accessibility of older respondents. However, weighting for X exacerbates the bias in Y . Since the older age group is weighted downward to match population margins, the negative correlation between X and Y further inflates the weighted mean of Y' , leading to $E(Y^w) > E(Y') > \mu_y$.

Second, suppose one aims to estimate the average “Statistical Knowledge” (Y) of scientists. Authors are randomly contacted for an online test, and the researcher controls for X (“Number of Publications per Author”) to ensure “representativeness.” Let the primary driver for participation be Z (“Ambition”), as authors may use the test results to signal their proficiency. Assume that publication frequency is determined by a combination of knowledge and ambition, which are themselves statistically independent. Even if publication frequency is controlled for in the initial sampling, the observed mean X' will be too high because ambitious authors both publish more and are more likely to participate. Consequently, X' must be weighted downward. Since X and Y are positively correlated (knowledge being a component of publication frequency), this downward adjustment propagates to Y , resulting in $\overline{Y^w} < \overline{Y'}$. However, Y' was unbiased initially, as knowledge and ambition are independent. Thus, weighting for X introduces a bias that was not originally present.

4.3. Continuous versus Dichotomous Variables

Categorical variables are frequently used as auxiliary variables; in many surveys, dichotomous or polytomous variables (e.g., sex, age class, region) dominate. Similarly, the analysis variable Y may be categorical. These cases can be addressed by applying transformations, such as McCall’s (1922) area transformation, to derive a standard normal variable X^n from a dichotomous variable X^d . Instead of $[0, 1]$ dummy coding, one can choose values $[x_0, x_1]$ that reflect the means of the respective halves of a standard normal distribution based on the proportions $(1 - p)$ and p . By utilizing the linearity assumption, it remains possible to calculate the resulting biases in these settings.

4.4. Multiple Auxiliary Variables

In household or person surveys, a large number of auxiliary variables are often employed, with each category of a polytomous variable serving as a distinct dummy variable. Each auxiliary variable can contribute to the total bias b_{y^w} . However, this accumulation of bias is constrained by the requirement that the covariance matrix remains positive semidefinite. Each individual counter-bias of an X_k (for $1 < k < K$) shares r_{yz} with other counter-biases, restricting the possible ranges of $r_{x_k y}$ and $r_{x_k z}$ as

K increases. Consequently, while the maximum joint bias imposed by X_1 and X_2 on Y is higher than that of a single variable, it is substantially less than the sum of their individual maximum biases.

The question of an absolute upper limit for total bias may be secondary to the fact that weighting efficiency typically deteriorates long before reaching theoretical extremes. In conclusion, while bias accumulation is possible, it is limited. It is more likely that a few dominant auxiliary variables—specifically those furthest from their population means—drive the weighting procedure and should be the primary focus of the researcher.

4.5. Evaluating Sampling Quality

Sampling quality is typically defined by potential biases resulting from nonresponse or other deviations from probability sampling. Although the response rate ρ is a widely used and accessible metric, it is not a reliable indicator of bias, as it fails to capture the underlying biasing processes (see Section 3.2). Instead, bias is fully determined by $b_{z'}$, δ , and the covariance matrix Σ .

Theoretically, it is possible to estimate the bias parameters $b_{z'}$ and δ , as well as r_{xz} . By deriving empirical estimates for ρ , $\sigma_{x'}$, and $b_{x'}$, one can solve three independent equations for the unknown parameters r_{xz} , β_0 , and β_1 (see Appendix C). This allows for a more accurate determination of sample quality than ρ provides. However, determining $b_{y'}$ or $b_{y^{*w}}$ remains impossible from the sample alone, as r_{yz} cannot be empirically identified.

4.6. Practical Advice

Although Z is unobserved and r_{yz} is not identifiable, one can often deduce whether a specific auxiliary variable X might inflate bias. Researchers can often make reasonable assumptions about the nonresponse process based on expert knowledge of the data collection context. Hence, one may be able to assume the sign or even the expected range of the correlation r_{yz} . Since r_{xy} can be approximated by $r_{x'y'}$ and r_{xz} can be estimated (as described in Section 4.5), researchers can therefore gauge the risk of bias inflation by analyzing the assumed structure of the correlation matrix Σ_{xyz} .

Two scenarios warrant particular caution: when r_{yz} is assumed to be small but weighting imposes a strong counter-bias, or when the product of the three correlations ($r_{xy} \cdot r_{xz} \cdot r_{yz}$) is negative. In these cases, weighting for X is likely counterproductive. Furthermore, to minimize the Root Mean Square Error (RMSE), it is often beneficial to exclude auxiliary variables with limited potential for bias reduction to prevent variance inflation from outweighing any small bias gains.

5. Summary and Outlook

In this paper, we investigated the implications of incorrectly assuming ignorability in sample calibration. We demonstrated that the common practice of calibrating

auxiliary variables to external benchmarks can be counterproductive in specific data scenarios. Using a parsimonious linear model, we outlined the mechanisms under which weighting becomes problematic and provided guidance on identifying detrimental auxiliary variables.

Despite these risks, we remain proponents of sample calibration, as weighting is an efficient method for incorporating population information. However, researchers must remain cognizant of the underlying missing-data assumptions. As Lohr (2023) emphasizes, over-reliance on nonresponse model assumptions can be hazardous.

Looking forward, employing multiple weighting schemes tailored to different analysis models could provide more flexibility, though this may be impractical for general-purpose datasets. A compromise could involve focusing calibration on variables central to the most common analytical needs. Counterintuitive results in science often serve as a catalyst for deeper understanding. Our findings can be extended to related fields, such as non-probability sampling, where similar assumptions regarding selection mechanisms are required.

Appendix A

Modeling Effects of Nonresponse and Weighting

Model Setup. Three random variables $[X, Y, Z]$ and their respective covariance matrix Σ_{xyz} describe the situation in the population before selection through nonresponse:

X := auxiliary variable
 Y := analysis variable
 Z := participation driver

All variables $[X, Y, Z]$ are interval-scaled, and the model assumes linearity throughout. The value of Z drives the likelihood of participation in the survey. This likelihood can be determined by a logistic regression or any other suitable model with appropriate parameters. The superscript “'” indicates a random variable after nonresponse. When such a variable has been weighted with respect to the mean of X , the superscript changes to “w.”

With linearity, nonresponse affects the variables involved such that $\sigma_z^2 < \sigma_z'^2$ and bias $b_{z'} := E(Z') - \mu_z \geq 0$. In short, we define $\delta := 1 - \frac{\sigma_z'^2}{\sigma_z^2}$ as the relative reduction of σ_z^2 due to nonresponse. As Z is conveniently set to standard normal, σ_z^2 is one. However, for the sake of clarity, σ_z will be explicitly stated in the following derivations.

Nonresponse Bias. Knowledge of Σ_{xyz} , $b_{z'}$, and δ is sufficient to derive $\Sigma_{x'y'z'}$, $b_{y'}$, $b_{x'}$, and, most importantly, b_{y^w} := the bias of the variable of interest after weighting.

The fraction of the variance σ_y^2 that is explained by Z is given by r_{yz}^2 . Through nonresponse, this fraction is reduced by a factor δ . As a consequence, $\sigma_{y'}^2$ will be smaller than σ_y^2 . Likewise, $\sigma_{x'}^2$ will be smaller than σ_x^2 .

$$\begin{aligned}\sigma_{y'}^2 &= \sigma_y^2 \cdot (1 - \delta \cdot r_{yz}^2) \\ \sigma_{x'}^2 &= \sigma_x^2 \cdot (1 - \delta \cdot r_{xz}^2)\end{aligned}\tag{A.1}$$

Similarly, the respective covariance terms of Y and X with Z lose a share of δ through nonresponse, that is, a share of $1 - \delta$ is retained. By renorming the terms with the respective standard deviations, the correlations $r_{y'z'}$ and $r_{x'z'}$ can be derived.

$$\begin{aligned}r_{y'z'} &= r_{yz} \cdot \frac{\sigma_z \cdot \sigma_y}{\sigma_{z'} \cdot \sigma_{y'}} \cdot (1 - \delta) = r_{yz} \cdot \frac{\sigma_{z'} \cdot \sigma_y}{\sigma_z \cdot \sigma_{y'}} \\ r_{x'z'} &= r_{xz} \cdot \frac{\sigma_z \cdot \sigma_x}{\sigma_{z'} \cdot \sigma_{x'}} \cdot (1 - \delta) = r_{xz} \cdot \frac{\sigma_{z'} \cdot \sigma_x}{\sigma_z \cdot \sigma_{x'}}\end{aligned}\tag{A.2}$$

r_{xy} is distorted through nonresponse in a different, more subtle way. Only the fraction of $\text{cov}(X, Y)$ that is also shared with Z through r_{yz} and r_{xz} is reduced by δ . Thus, renorming with the respective standard deviations yields

$$r_{x'y'} = (r_{xy} - r_{yz} \cdot r_{xz} \cdot \delta) \cdot \frac{\sigma_y \cdot \sigma_x}{\sigma_{y'} \cdot \sigma_{x'}}\tag{A.3}$$

The bias of Z' propagates to Y and X through the respective correlations.

$$\begin{aligned}b_{y'} &:= E(Y') - \mu_y \\ &= b_{z'} \cdot r_{yz} \cdot \frac{\sigma_y}{\sigma_z} \\ b_{x'} &:= E(X') - \mu_x \\ &= b_{z'} \cdot r_{xz} \cdot \frac{\sigma_x}{\sigma_z}\end{aligned}\tag{A.4}$$

Alternatively, these results may be derived by integration. WLOG, assume that $E(Y) = E(Z) = 0$. Then

$$b_{y'} = E(Y|Z) = \iint_{-\infty}^{\infty} p(Y) \cdot Y \cdot \frac{\rho(Z)}{\rho} dYdZ,\tag{A.5}$$

with $\rho(Z)$ being the probability of response as a function of Z and ρ being the overall response rate. By assuming linearity, Y can be replaced by $\beta_z \cdot Z + \epsilon_z$ with $\beta_z = r_{yz} \cdot \frac{\sigma_y}{\sigma_z}$. Hence,

$$\begin{aligned}
 b_{y'} &= \iint p(\beta_z \cdot Z + \epsilon_z) \cdot (\beta_z \cdot Z + \epsilon_z) \cdot \frac{\rho(Z)}{\rho} dYdZ \\
 &= \int p(Z) \cdot (\beta_z \cdot Z) \cdot \frac{\rho(Z)}{\rho} dZ \\
 &= r_{yz} \cdot \frac{\sigma_y}{\sigma_z} \int p(Z) \cdot Z \cdot \frac{\rho(Z)}{\rho} dZ \\
 &= r_{yz} \cdot \frac{\sigma_y}{\sigma_z} \cdot b_{z'}.
 \end{aligned}
 \tag{A.6}$$

Likewise, by substituting Z instead of Y , this approach can be used to arrive at Bethlehem’s (1988) formula 3.5, stating that $b_{y'} = \frac{cov(Y, \rho(Y))}{\rho}$ (our notation). Specifically,

$$\begin{aligned}
 b_{y'} &= \iint p(Y) \cdot Y \cdot \frac{\rho(\beta_y \cdot Y + \epsilon_y)}{\rho} dYdZ \\
 &= \frac{1}{\rho} \int p(Y) \cdot Y \cdot \rho(Y) dY \\
 &= \frac{1}{\rho} \cdot cov(Y, \rho(Y)).
 \end{aligned}
 \tag{A.7}$$

This holds because with $E(Y) = 0$,

$$\begin{aligned}
 cov(Y, \rho(Y)) &= \int p(Y) \cdot (Y - 0) \cdot (\rho(Y) - \rho) dY \\
 &= \int p(Y) \cdot Y \cdot \rho(Y) dY - \int p(Y) \cdot Y \cdot \rho dY \\
 &= \int p(Y) \cdot Y \cdot \rho(Y) dY - \rho \cdot 0.
 \end{aligned}
 \tag{A.8}$$

Weighting Bias. Weighting induces a “counterbias” $b_{x'}^w$ in X that aims to eliminate $b_{x'}$, that is, $b_{x'}^w = -b_{x'}$. This counterbias propagates to Y by means of $r_{x'y'}$, determining $b_{y'}^w$:

$$\begin{aligned}
 b_{y'}^w &:= E(Y^w) - E(Y') \\
 &= -b_{x'} \cdot r_{x'y'} \cdot \frac{\sigma_{y'}}{\sigma_{x'}} \\
 &= -b_{z'} \cdot r_{xz} \cdot \frac{\sigma_x}{\sigma_z} \cdot r_{x'y'} \cdot \frac{\sigma_{y'}}{\sigma_{x'}}
 \end{aligned}
 \tag{A.9}$$

Finally, $b_{y^w} :=$ total bias of Y^w , is the sum of the nonresponse bias $b_{y'}$ and the counterbias $b_{y'}^w$.

$$\begin{aligned}
b_{y^w} &:= E(Y^w) - \mu_x = b_{y'} + b_{y'}^w \\
&= b_{z'} \cdot r_{yz} \cdot \frac{\sigma_y}{\sigma_z} - b_{z'} \cdot r_{xz} \cdot \frac{\sigma_x}{\sigma_z} \cdot r_{x'y'} \cdot \frac{\sigma_{y'}}{\sigma_{x'}} \\
&= b_{z'} \cdot r_{yz} \cdot \frac{\sigma_y}{\sigma_z} - b_{z'} \cdot r_{xz} \cdot \frac{\sigma_x}{\sigma_z} \cdot (r_{xy} - r_{yz} \cdot r_{xz} \cdot \delta) \cdot \frac{\sigma_y \cdot \sigma_x}{\sigma_{y'} \cdot \sigma_{x'}} \cdot \frac{\sigma_{y'}}{\sigma_{x'}} \\
&= b_{z'} \cdot \frac{\sigma_y}{\sigma_z} \cdot \left[r_{yz} - r_{xz} \cdot \frac{\sigma_x^2}{\sigma_{x'}^2} \cdot (r_{xy} - r_{yz} \cdot r_{xz} \cdot \delta) \right] \\
&\stackrel{eqn.A.1}{=} b_{z'} \cdot \frac{\sigma_y}{\sigma_z} \cdot \left[r_{yz} - \frac{r_{xy} \cdot r_{xz} - r_{yz} \cdot r_{xz}^2 \cdot \delta}{1 - r_{xz}^2 \cdot \delta} \right]
\end{aligned} \tag{A.10}$$

Nonresponse with SAR. With SAR, Y and Z are independent when conditioned on X . Hence, linear independence is also given by $r_{yz} = r_{xz} \cdot r_{xy}$. Consequently, the total bias b_{y^w} as given in Equation (A.10) above simplifies to the following.

$$\begin{aligned}
b_{y^w} &= b_{z'} \cdot \frac{\sigma_y}{\sigma_z} \cdot \left[r_{yz} - r_{xz} \cdot \frac{(r_{xy} - r_{xz} \cdot r_{xy} \cdot r_{xz} \cdot \delta)}{1 - r_{xz}^2 \cdot \delta} \right] \\
&= b_{z'} \cdot \frac{\sigma_y}{\sigma_z} \cdot \left[r_{yz} - r_{xz} \cdot r_{xy} \cdot \frac{1 - r_{xz}^2 \cdot \delta}{1 - r_{xz}^2 \cdot \delta} \right] \\
&= b_{z'} \cdot \frac{\sigma_y}{\sigma_z} \cdot [r_{yz} - r_{yz}] = 0.
\end{aligned} \tag{A.11}$$

Thus, if SAR holds in population data, weighting completely removes the bias inflicted on Y . More generally, SAR also holds for response-selected data, if SAR holds for the population. This means $r_{y'z'} = r_{x'z'} \cdot r_{x'y'}$ holds if $r_{yz} = r_{xz} \cdot r_{xy}$:

$$\begin{aligned}
&r_{y'z'} \stackrel{?}{=} r_{x'z'} \cdot r_{x'y'} \mid \text{SAR} \\
&\Rightarrow r_{yz} \cdot \frac{\sigma_z \cdot \sigma_y}{\sigma_{z'} \cdot \sigma_{y'}} \cdot (1 - \delta) = r_{xz} \cdot \frac{\sigma_z \cdot \sigma_x}{\sigma_{z'} \cdot \sigma_{x'}} \cdot (1 - \delta) \cdot (r_{xy} - r_{yz} \cdot r_{xz} \cdot \delta) \cdot \frac{\sigma_y \cdot \sigma_x}{\sigma_{y'} \cdot \sigma_{x'}} \\
&\stackrel{\text{SAR}}{\Rightarrow} \\
&r_{yz} = r_{xz} \cdot r_{xy} \cdot (1 - r_{xz}^2 \cdot \delta) \cdot \frac{\sigma_x^2}{\sigma_{x'}^2} \\
&\stackrel{eqn.A.1}{\Rightarrow} \\
&r_{yz} = r_{yz} \cdot \frac{(1 - r_{xz}^2 \cdot \delta)}{(1 - r_{xz}^2 \cdot \delta)} \text{ qed.}
\end{aligned} \tag{A.12}$$

Determining the Critical Value for r_{yz} . When $-b_{y'} = b_{y^w}$, the decrease in bias turns into bias increase. With Equation (A.4) this expands to

$$-r_{yz} = r_{yz} - \frac{r_{xy} \cdot r_{xz} - r_{yz} \cdot r_{xz}^2 \cdot \delta}{1 - r_{xz}^2 \cdot \delta}.$$

The solution for r_{yz} then yields the critical value of r_{yz} :

$$r_{yz}^{crit} = \frac{r_{xy} \cdot r_{xz}}{2 - r_{xz}^2 \cdot \delta}.$$

Any r_{yz} going beyond r_{yz}^{crit} results in a bias increase. Specifically, if r_{xy} and r_{xz} have identical signs, the bias of Y is increased if and only if

$$r_{yz} < \frac{r_{xy} \cdot r_{xz}}{2 - r_{xz}^2 \cdot \delta}.$$

Similarly, if r_{xy} and r_{xz} are of opposite signs, the bias of Y is increased if and only if

$$r_{yz} > \frac{r_{xy} \cdot r_{xz}}{2 - r_{xz}^2 \cdot \delta}.$$

Appendix B

Weighting for a Single Metric Auxiliary Variable

When weighting is performed to adjust for the mean of a single metric variable X , each sample unit i can be assigned a weight w_i based on the observed x'_i according to the linear weighting formula

$$w(x'_i) = 1 + (x'_i - \bar{x}') \cdot \frac{\mu_x - \bar{x}'}{\sigma_{x'}^2}.$$

Note that $\sigma_{x'}^2$ denotes the sample variance of X with the empirical correction term omitted, that is, $\sigma_{x'}^2 = \frac{1}{n} \sum_{i=1}^n (x'_i - \bar{x}')^2$. μ_x is the target value for the mean of X provided by an external source. As a result, $\frac{1}{n} \sum_{i=1}^n (w_i \cdot x'_i) = \mu_x$ will hold. For general weighting solutions in the context of GREG models, see, for example, Bethlehem (1988) and Rao and Singh (1997).

Appendix C

Sampling Quality Parameter Estimation through Integration

The general solution to calculate the weighted bias b_{y^w} by integration can be given as

$$b_{y^w} = \frac{1}{\rho} \iiint_{-\infty}^{\infty} \rho(Z) \cdot w(X) \cdot p(X, Y, Z) \cdot Y \, dX \, dY \, dZ - \int p(Y) \cdot Y \, dY,$$

where ρ is the overall response rate, $\rho(Z)$ denotes the response probability for a given Z , $w(X)$ is a function determining the weights, and $p(X, Y, Z)$ is the trivariate probability density function in the population. Note that this general solution does not require assumptions about distributions or linearity. However, in order to derive estimates for sampling quality parameters, such assumptions must be made. Hence, we assume that X, Y, Z are multivariate standard normal. Therefore,

linearity holds. The response probability $\rho(Z)$ is assumed to be the result of a logistic regression with parameters β_0 and β_1 .

μ_x and σ_x are known from external sources. In addition, ρ , $b_{x'}$, and $\sigma_{x'}$ can be estimated from the survey data and metadata. By integration over the probability of response, ρ is calculated:

$$\rho = \int \frac{p(Z)}{1 + e^{-(\beta_0 + \beta_1 \cdot Z)}} dZ \quad (\text{C.1})$$

With ρ and Equation (A.4) we get:

$$b_{z'} = \int p(Z') dZ' = \frac{1}{\rho} \cdot \int \frac{Z \cdot p(Z)}{1 + e^{-(\beta_0 + \beta_1 \cdot Z)}} dZ$$

and

$$b_{x'} = r_{xz} \cdot \sigma_x \cdot \frac{1}{\rho} \cdot \int \frac{Z \cdot p(Z)}{1 + e^{-(\beta_0 + \beta_1 \cdot Z)}} dZ \quad (\text{C.2})$$

From Equation (A.1) follows that

$$\sigma_{x'}^2 = \sigma_x^2 \cdot (1 - \delta \cdot r_{xz}^2) = \sigma_x^2 \cdot (1 - r_{xz}^2 \cdot (1 - \sigma_z^2)).$$

Therefore, with

$$\begin{aligned} \sigma_{z'}^2 &= E(Z'^2) - b_{z'}^2, \\ \sigma_{x'}^2 &= \sigma_x^2 \cdot \left[1 - r_{xz}^2 \cdot \left(1 - \left(\frac{1}{\rho} \cdot \int \frac{Z^2 \cdot p(Z)}{1 + e^{-(\beta_0 + \beta_1 \cdot Z)}} dZ - b_{z'}^2 \right) \right) \right]. \end{aligned} \quad (\text{C.3})$$

The three independent Equations (C.1) to (C.3) form a system with three unknown parameters r_{xy} , β_0 , and β_1 . With the assumptions given above, the system can be solved by utilizing numerical optimization methods. The parameter δ follows from determining Z' and its variance $\sigma_{y'}^2$ by logistically regressing Z with parameters β_0 and β_1 .


Acknowledgements

The authors are grateful for helpful comments from five anonymous reviewers which substantially improved the quality of the paper.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Florian Meinfelder  <https://orcid.org/0000-0001-5959-7612>

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: Series C (Applied Statistics)* 68 (5): 1465–83. DOI: <https://doi.org/10.1111/rssc.12371>.
- Bethlehem, J. G. 1988. "Reduction of Nonresponse Bias Through Regression Estimation." *Journal of Official Statistics* 4 (3): 251–60. <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/reduction-of-nonresponse-bias-through-regression-estimation.pdf>.
- Bethlehem, J. G. 2011. *Handbook of Nonresponse in Household Surveys*. Wiley. <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470891056>.
- Bethlehem, J. G., and W. Keller. 1987. "Linear Weighting of Sample Survey Data." *Journal of Official Statistics* 3: 141–53.
- Brick, J. M. 2013. "Unit Nonresponse and Weighting Adjustments: A Critical Review." *Journal of Official Statistics* 29 (3): 329–53. DOI: <https://doi.org/10.2478/jos-2013-0026>.
- Curtin, R., S. Presser, and E. Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69 (1): 87–98. DOI: <https://doi.org/10.1093/poq/nfi002>.
- Deming, W. E., and F. F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *Annals of Mathematical Statistics* 11 (4): 427–44. DOI: <https://doi.org/10.1214/aoms/1177731829>.
- Deville, J.-C. 2000. "Generalized Calibration and Application to Weighting for Non-Response." In *COMPSTAT: Proceedings in Computational Statistics, 14th Symposium Held in Utrecht, The Netherlands*, edited by J. G. Bethlehem and P. G. M. Van Der Heijden. Physica Verlag, Springer.
- Deville, J.-C., C.-E. Särndal, and O. Sautory. 1993. "Generalized Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88 (423): 1013–20. DOI: <https://doi.org/10.1080/01621459.1993.10476369>.
- Earp, M., P. S. Kott, F. Kreuter, and S. Porter. 2012. "Nonresponse Bias Adjustment in Establishment Surveys: A Comparison of Weighting Methods Using the Agricultural Resource Management Survey (ARMS)." Proceedings of the Survey Research Methods Section. *American Statistical Association*. <https://www.bls.gov/osmr/research-papers/2012/st120240.htm>.
- Groves, R. M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *The Public Opinion Quarterly* 70 (5): 646–75. <https://www.jstor.org/stable/4124220>.
- Groves, R. M., D. A. Dillman, J. L. Eltinge, and R. J. A. Little, eds. 2002. *Survey Nonresponse*. Wiley.
- Groves, R. M., and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72: 167–89. DOI: <https://doi.org/10.1093/poq/nfn011>.
- Hansen, M. H., and C. L. Dedrick. 1938. *Final Report on Total and Partial Unemployment, [1937]*. U.S. Govt. print. off. <https://catalog.hathitrust.org/Record/101671541>.
- Hansen, M. H., and W. N. Hurwitz. 1946. "The Problem of Non-Response in Sample Surveys." *Journal of the American Statistical Association* 41 (236): 517–29. <https://www.jstor.org/stable/2280572>.

- Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61. <https://www.jstor.org/stable/1912352>.
- Kiesl, H., and S. Rässler. 2006. "Quality in Data Fusion." *European Conference on Quality in Survey Statistics (Q2006)*, Cardiff. <https://iab.de/publikationen/publikation/?id=190098>.
- Kish, L. 1987. "Questions and Answers." *The Survey Statistician* 17 (9): 13–7.
- Kott, P. S., and T. Chang. 2010. "Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse." *Journal of the American Statistical Association* 105 (491): 1265–75. <https://www.jstor.org/stable/27920149>.
- Kott, P. S., and D. Liao. 2017. "Calibration Weighting for Nonresponse That Is Not Missing at Random: Allowing More Calibration Than Response-Model Variables." *Journal of Survey Statistics and Methodology* 5 (2): 159–74. DOI: <https://doi.org/10.1093/jssam/smx003>.
- Kreuter, F., and K. Olson. 2011. "Multiple Auxiliary Variables in Nonresponse Adjustment." *Sociological Methods & Research* 40 (2): 311–32. DOI: <https://doi.org/10.1177/0049124111400042>.
- Little, R. J., and S. Vartivarian. 2003. "On Weighting the Rates in Non-Response Weights." *Statistics in Medicine* 22 (9): 1589–99. DOI: <https://doi.org/10.1002/sim.1513>.
- Little, R. J., and D. B. Rubin. 2019. *Statistical Analysis with Missing Data*. 3rd ed. Wiley.
- Lohr, S. L. 2021. *Sampling: Design and Analysis*. 3rd ed. Chapman & Hall/CRC.
- Lohr, S. L. 2023. "Assuming a Nonresponse Model Does Not Make It True." *Harvard Data Science Review* 5 (3). <https://hdsr.mitpress.mit.edu/pub/sqinvh4c/release/1>.
- Lundström, S., and C.-E. Särndal. 1999. "Calibration as a Standard Method for Treatment of Nonresponse." *Journal of Official Statistics* 15 (2): 305–27. <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/calibration-as-a-standard-method-for-treatment-of-nonresponse.pdf>.
- McCall, W. 1922. *How to Measure in Education*. McMillan.
- Oh, L., and F. Scheuren. 1983. "Weighting Adjustment for Unit Nonresponse." In *Incomplete Data in Sample Surveys: Theory and Bibliographies*, edited by W. G. Madow, I. Olkin, and D. B. Rubin. Academic Press.
- Rao, J., and A. Singh. 1997. "A Ridge-Shrinkage Method for Range-Restricted Weight Calibration in Survey Sampling." Proceedings of the Section on Survey Research Methods. American Statistical Association.
- Rässler, S. 2002. *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*, Vol. 168 of Lecture Notes in Statistics. 1st ed. Springer.
- Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika* 63 (3): 581–92. DOI: <https://doi.org/10.1093/biomet/63.3.581>.
- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. Wiley. <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316696>.
- Särndal, C.-E. 2007. "The Calibration Approach in Survey Theory and Practice." *Survey Methodology* 33 (2): 99–119. <https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2007002/article/10488-eng.pdf>.
- Särndal, C.-E., and S. Lundström. 2005. *Estimation in Surveys with Nonresponse*. John Wiley & Sons.

Received: February 6, 2025

Accepted: January 19, 2026