Doubly robust inference with missing data in survey sampling

Jae Kwang Kim * *   David Haziza †

Abstract
Statistical inference with missing data requires assumptions about the population or about the response probability. Doubly robust (DR) estimators use both relationships to estimate the parameters of interest, so that they are consistent even when one of the models is misspecified. In this paper, we propose a method of computing propensity scores that leads to DR estimation. In addition, we discuss DR variance estimation so that the resulting inference is doubly robust. Some asymptotic properties are discussed and results from two limited simulation studies are also presented.

Key Words: Calibration; Imputation; Nonresponse; Variance estimation.

1. Introduction

Missing data occurs in surveys because some of the sampled units refuse to respond to the survey or because of the inability to contact them. Dropout or noncompliance in clinical trials may also lead to missing responses for some subjects. It is well known that unadjusted estimators may be heavily biased if the respondents differ from the nonrespondents systematically with respect to the study variables. It is thus desirable to develop estimation procedures exhibiting low biases.

Doubly robust estimation procedures have attracted a lot of attention in mainstream statistics in recent years; e.g., Robins et al. (1994), Scharfstein et al. (1999), Tan (2006), Bang & Robins (2005), Kang & Schafer (2008), Robins et al. (2008), Cao et al. (2009), among others. In the context of finite population sampling, doubly robust estimation has been studied in Kott (1994), Kim & Park (2006), and Haziza & Rao (2006). In doubly robust estimation, two models are introduced: (i) the nonresponse model that requires the specification of a nonresponse model describing the unknown nonresponse mechanism and (ii) the outcome regression model approach that requires the specification of a model describing the distribution of the study variable. An estimator is said to be doubly robust if it remains asymptotically unbiased and consistent if either model (nonresponse or outcome regression) is true. Doubly robust procedures offer some protection against misspecification of one model or the other. In the context of finite population sampling Haziza & Rao (2006) and Kim & Park (2006) proposed doubly robust variance estimators, provided the overall sampling fraction is negligible. In the first paper, the authors considered Taylor linearisation procedures, whereas jackknife variance estimation was considered in the second paper.

In this paper, we consider doubly robust inference in the sense that the inference based on point estimator and variance estimator is justified of either one of the two models, nonresponse model or outcome regression model, holds. The proposed doubly robust variance estimator has a simple form that can be easily implemented using the software for complete sample data.

2. Basic Setup

For simplicity, assume that we have n independent realizations of a random variable Y, denoted by $y_1, \cdots, y_n$, from a distribution and we are interested in estimating $\theta = E(Y)$. In the absence of nonresponse to the study variable y, the parameter θ is consistently estimated by the sample mean

$$\hat{\theta}_n = \sum_{i=1}^{n} w_i y_i, \tag{1}$$

where $w_i = 1/n$. In Section2 and Section3, we set $w_i = 1/n$. In Section4, we use a different set of weights w_i as we treat the problem of doubly robust inference in the finite population sampling context. In addition to the study variable y, assume that a vector of auxiliary variables, denoted by x, is also available in the sample. Let δ_i be a response indicator attached to unit i such that $\delta_i = 1$ if y_i is observed and $\delta_i = 0$, otherwise. Instead of observing (x_i, y_i) for the whole sample, we observe (x_i, y_i) for $\delta_i = 1$ and observe only x_i for $\delta_i = 0$.

In this case, a natural approach for estimating θ consists of first postulating a model for the conditional distribution of y_i given x_i. In particular, if we are only interested in the mean of the y-values, we consider the following

*Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.
†Département de mathématiques et de statistique, Université de Montréal, Montreal, H3C 3J7, Canada

model

$$E\left(y_i \mid x_i, \delta_i = 0\right) = m\left(x_i; \beta_0\right), \tag{2}$$

where $m\left(x_i, \beta\right)$ is a continuous differentiable function of β. The model (2) is called the outcome regression model.

A natural estimator of θ is the so-called imputed estimator given by

$$\hat{\theta}_I = \sum_{i=1}^{n} w_i \left\{ \delta_i y_i + \left(1 - \delta_i\right) m(x_i, \hat{\beta}) \right\}, \tag{3}$$

where $\hat{\beta}$ is a consistent estimator of the true parameter β_0. Since

$$\hat{\theta}_I - \hat{\theta}_n = -\sum_{i=1}^{n} w_i \left(1 - \delta_i\right) \left\{ y_i - m(x_i, \hat{\beta}) \right\}$$

we have

$$E\left\{ \hat{\theta}_I - \hat{\theta}_n \mid \delta_1, \cdots, \delta_n, x_1, \cdots, x_n \right\} = -\sum_{i=1}^{n} w_i \left(1 - \delta_i\right) \left\{ E\left(y_i \mid x_i, \delta_i = 0\right) - m(x_i, \hat{\beta}) \right\},$$

where $E(\cdot \mid \delta_1, \cdots, \delta_n, x_1, \cdots, x_n)$ denotes the conditional expectation with respect to the outcome regression conditionally given δ_i and x_i. Thus, the validity of the imputed estimator (3) follows if the outcome regression model (2) is true and $\hat{\beta}$ is a consistent estimator of β_0. Often, a consistent estimator of β is obtained by solving

$$\sum_{i=1}^{n} w_i \delta_i \left\{ y_i - m\left(x_i; \beta\right) \right\} h(x_i; \beta) \ = 0, \tag{4}$$

for some $h(x_i; \beta)$, which is justified under the missing at random assumption that can be expressed as

$$E\left(y_i \mid x_i, \delta_i = 1\right) = E\left(y_i \mid x_i, \delta_i = 0\right). \tag{5}$$

Note that, under some regularity conditions, the solution $\hat{\beta}$ to (4) is consistent for β_0 if model (2) and the missing-at-random condition (5) hold.

Now, suppose that the probability of response to the study variable y, denoted by $p_i = \Pr\left(\delta_i = 1 \mid i\right)$, follows a logistic regression model

$$p_i = p_i(\phi_0) = \frac{\exp\left(\phi_0' x_i\right)}{1 + \exp\left(\phi_0' x_i\right)} \tag{6}$$

for some ϕ_0. The model (6) is called the nonresponse model. We assume that the intercept term is included in (6). In the classical two-phase sampling setup, where the second-phase sample corresponds to the set of respondents, the second-phase conditional inclusion probability p_i is known and the two-phase regression estimator, given by

$$\begin{aligned} \hat{\theta}_{tp} &= \sum_{i=1}^{n} w_i \left[m(x_i; \hat{\beta}) + \frac{\delta_i}{p_i} \left\{ y_i - m(x_i; \hat{\beta}) \right\} \right] \\ &= \hat{\theta}_n + \sum_{i=1}^{n} w_i \left(\frac{\delta_i}{p_i} - 1 \right) \left\{ y_i - m\left(x_i; \hat{\beta}\right) \right\}, \end{aligned} \tag{7}$$

is approximately unbiased for θ under the nonresponse model $p_i = \Pr\left(\delta_i = 1 \mid i\right)$ (Cochran, 1977) regardless of whether or not the outcome regression model (2) holds. Also, when the nonresponse model is not correct, the estimator is still approximately unbiased if (2) and (5) hold and $\hat{\beta}$ is consistent for β_0. Thus, $\hat{\theta}_{tp}$ is doubly robust in the sense that it remains valid if either one of the two models holds.

When the response probability is estimated, rather than known, we consider a class of estimators of the form

$$\hat{\theta}_{DR}(\hat{\beta}, \hat{\phi}) \quad = \quad \hat{\theta}_n + \sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\hat{\phi})} - 1 \right\} \left\{ y_i - m\left(x_i; \hat{\beta}\right) \right\}, \tag{8}$$

indexed by $(\hat{\beta}, \hat{\phi})$, where $\hat{\beta}$ is consistent for β_0 under the assumed outcome regression model and $\hat{\phi}$ is consistent for ϕ_0 under the assumed nonresponse model. As noted by Scharfstein et al. (1999), the double robustness property also follows if p_i is replaced by $\hat{p}_i = p_i(\hat{\phi})$ using a consistent estimator $\hat{\phi}$ for ϕ_0. Note that the doubly robust estimator, $\hat{\theta}_{tp}(\hat{\beta}, \hat{\phi})$, in (8) is a class of estimators and different choice of $(\hat{\beta}, \hat{\phi})$ leads to different doubly robust estimators.

Scharfstein et al. (1999) and Haziza & Rao (2006) used $\hat{\phi}$ estimated by the maximum likelihood method and $\hat{\beta}$ estimated by ordinary or iteratively reweighted least squares. Recently, Cao et al. (2009) proposed a doubly robust estimator using the optimal score equation based on influence function theory. However, the optimal estimator of Cao et al. (2009) is sub-optimal because they first estimate $\hat{\phi}$ by $\hat{\phi}_{MLE}$ obtained from the maximum likelihood method and then seek for the optimal estimator in the class of estimators $\hat{\theta}^*_{tp}(\hat{\beta}) = \hat{\theta}_{tp}(\hat{\beta}, \hat{\phi}_{MLE})$ as a function of $\hat{\beta}$. As discussed in Kim & Kim (2007), the choice of $\hat{\phi}_{MLE}$ does not necessarily lead to the optimal propensity score estimators. Thus, we expect that the efficiency of the sub-optimal estimator of Cao et al. (2009) can be improved for a suitable choice of $\hat{\phi}$.

We propose a new doubly robust estimator of the form (8) using a different choice of $(\hat{\beta}, \hat{\phi})$. Some asymptotic properties of the resulting doubly robust estimator are discussed in Section3. Also, we propose a new variance estimator that is doubly robust in the sense that it remains consistent even when one of the outcome regression or nonresponse models, is misspecified. Thus, the proposed point and variance estimation procedure leads to doubly robust inference.

3. Main Results

Under the setup described in Section2, we propose a new imputed estimator $\hat{\theta}_I$ of the form (3) using $(\hat{\beta}^*, \hat{\phi}^*)$, where $(\hat{\beta}^*, \hat{\phi}^*)$ is obtained by solving

$$\sum_{i=1}^{n} w_i \delta_i \left\{ \frac{1}{p_i(\phi)} - 1 \right\} \{ y_i - m(x_i; \beta) \} x_i = \mathbf{0} \tag{9}$$

and

$$\sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\phi)} - 1 \right\} \dot{m}(x_i; \beta) = \mathbf{0}. \tag{10}$$

Because an intercept term is included in x, condition (9) implies that

$$\sum_{i=1}^{n} w_i \delta_i \frac{1}{p_i(\hat{\phi}^*)} \left\{ y_i - m(x_i; \hat{\beta}^*) \right\} x_i = \sum_{i=1}^{n} w_i \left\{ y_i - m(x_i; \hat{\beta}^*) \right\} x_i$$

and the imputed estimator (3) can be expressed as a doubly robust estimator of the form (8). Condition (9) has been used in Scharfstein et al (1999) and Haziza & Rao (2006). Condition (10) is a calibration condition in the sense that the propensity score adjusted estimator applied to $\dot{m}(x_i; \beta)$ leads to the complete sample estimator. For example, consider the linear outcome regression model for which $m(x_i; \beta) = x_i' \beta$. Then, condition (10) is equivalent to

$$\sum_{i=1}^{n} w_i \frac{\delta_i}{p_i(\phi)} x_i = \sum_{i=1}^{n} w_i x_i. \tag{11}$$

Condition (11) has been used by Lannacchione et al. (1991) and Chang & Kott (2008) in the context of unit nonresponse in the survey sampling context. From (11), it follows that estimates corresponding to the x-variables do not suffer from nonresponse error. Thus, writing $y_i = x_i' \beta_0 + e_i$, the imputed estimator $\hat{\theta}_I$ can be written as

$$\hat{\theta}_I = \hat{\theta}_n + \sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\hat{\phi}^*)} - 1 \right\} \left\{ x_i' \beta_0 - x_i' \hat{\beta}^* \right\} + \sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\hat{\phi}^*)} - 1 \right\} e_i.$$

Note that the second term on the right hand side of the previous expression is equal to

$$\sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\hat{\phi}^*)} - 1 \right\} x_i' \left\{ \beta_0 - \hat{\beta}^* \right\} = 0 \tag{12}$$

if (10) holds. Thus, under (10),

$$\hat{\theta}_I = \hat{\theta}_n + \sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\hat{\phi}^*)} - 1 \right\} e_i.$$

and the variability associated with $\hat{\beta}$ can be safely ignored. On the other hand, using the fact $\partial p_i^{-1}(\phi)/\partial\phi = -\left\{p_i^{-1}(\phi) - 1\right\} x_i$ under the nonresponse model (6), we can apply the Taylor expansion to get

$$\hat{\theta}_I = \hat{\theta}_n + \sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\phi^*)} - 1 \right\} e_i - \sum_{i=1}^{n} w_i \delta_i \left\{ \frac{1}{p_i(\phi^*)} - 1 \right\} e_i x_i \left(\hat{\phi} - \phi^* \right) + O_p\left(n^{-1}\right), \tag{13}$$

where ϕ^* is the probability limit of $\hat{\phi}^*$. Using condition (9), it can be shown that

$$\sum_{i=1}^{n} w_i \delta_i \left\{ \frac{1}{p_i(\phi^*)} - 1 \right\} e_i x_i = o_p(1)$$

and (13) reduces to

$$\hat{\theta}_I = \hat{\theta}_n + \sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\phi^*)} - 1 \right\} e_i + o_p\left(n^{-1/2}\right) \tag{14}$$

and the variability associated with $\hat{\phi}^*$ can also be safely ignored.

The following theorem extends the above results to the general form of $E(y_i \mid x_i) = m(x_i; \beta_0)$. The proof of Theorem 1 is stated in Appendix A.

Theorem 1 *Under the regularity conditions stated in Appendix A, we have*

$$\sqrt{n}\left\{ \hat{\theta}_I - \tilde{\theta}_I \right\} = o_p(1) \tag{15}$$

where

$$\tilde{\theta}_I = \sum_{i=1}^{n} w_i \left[m(x_i; \beta^*) + \frac{\delta_i}{p_i(\phi^*)} \left\{ y_i - m(x_i; \beta^*) \right\} \right] \tag{16}$$

and β^* *is the probability limit of* $\hat{\beta}$.

Note that the probability statement in (15) is made in the doubly robust sense that the convergence in probability holds if one of the two models is true. If the reference distribution in (15) is with respect to the outcome regression model (2), then $\beta^* = \beta_0$. If the reference distribution in (15) is with respect to the nonresponse model (6), then $\phi^* = \phi_0$. When the two models are true, then $(\beta^*, \phi^*) = (\beta_0, \phi_0)$ and the variance of $\tilde{\theta}$ is equal to

$$V\left(\tilde{\theta}_I\right) = V\left(\hat{\theta}_n\right) + E\left[\sum_{i=1}^{n} w_i^2 \left\{ p_i(\phi_0)^{-1} - 1 \right\} e_i^2 \right] \tag{17}$$

where $e_i = y_i - m(x_i; \beta_0)$. Under simple random sampling, the variance (17) is equal to the semiparametric lower bound of the asymptotic variance and, as a result, is locally efficient (Robins et al. 1994).

If we define

$$\eta_i(\beta, \phi) = m(x_i; \beta) + \frac{\delta_i}{p_i(\phi)} \left\{ y_i - m(x_i; \beta) \right\}, \tag{18}$$

then (15) means that

$$\sum_{i=1}^{n} w_i \eta_i(\hat{\beta}^*, \hat{\phi}^*) = \sum_{i=1}^{n} w_i \eta_i(\beta^*, \phi^*) + o_p\left(\frac{1}{\sqrt{n}} \right).$$

Thus, if (x_i, y_i, δ_i) are independently and identically distributed, then $\eta_i(\beta^*, \phi^*)$ are independently and identically distributed, even though $\eta_i(\hat{\beta}^*, \hat{\phi}^*)$ are not necessarily independently and identically distributed. Because $\eta_i(\beta^*, \phi^*)$ are independently and identically distributed, we can apply the central limit theorem and the Slutsky theorem to get

$$\sqrt{n}\left(\hat{\theta}_I - \theta \right) \to^L N\left(0, \sigma^2\right), \tag{19}$$

where $\to^L$ denotes the convergence in distribution and $\sigma^2 = Var\left\{\eta_i(\beta^*, \phi^*)\right\}$. Furthermore, since $\eta_i(\beta^*, \phi^*)$ are independently and identically distributed with bounded fourth moments, we can apply the standard complete sample method to estimate the variance of $\tilde{\theta}_I = \sum_{i=1}^{n} w_i \eta_i(\beta^*, \phi^*)$. That is,

$$\hat{V}(\beta^*, \phi^*) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^{n} \left(\eta_i - \bar{\eta}_n \right)^2, \tag{20}$$

where $\eta_i = \eta_i(\beta^*, \phi^*)$ and $\bar{\eta}_n = n^{-1}\sum_{i=1}^n \hat{\eta}_i$, satisfies

$$p \lim_{n\to\infty} \hat{V}(\beta^*, \phi^*)/V = 1,$$

where $V = Var\left\{n^{-1}\sum_{i=1}^n \eta_i\left(\beta^*, \phi^*\right)\right\} = n^{-1}\sigma^2$. Therefore, by the Slutsky theorem again, we have

$$\frac{\hat{\theta}_I - \theta}{\sqrt{\hat{V}(\hat{\beta}^*, \hat{\phi}^*)}} \rightarrow^L N(0,1). \tag{21}$$

The asymptotic result in (21) can be used to construct the confidence interval for $\theta = E(Y)$. The reference distribution in (21) is either the outcome regression model or the nonresponse model.

4. Extension to finite population sampling

In this section, we consider the problem of doubly robust inference in the finite population sampling context. Consider a finite population U of size N. We are interested in estimating the mean of the finite population, $\theta_N = N^{-1}\sum_{i\in U} y_i$. To that end, a sample s, of size n is selected according to a given sampling design $p(s)$. In the complete data situation, a basic estimator is the expansion estimator given by (1) with $w_i = 1/(N\pi_i)$, where π_i denotes the first-order inclusion probability of unit i in the sample. In the presence of nonresponse to the y-variable, the imputed estimator $\hat{\theta}_I$ of θ_N is given by (3) with $w_i = 1/(N\pi_i)$. Note that $\hat{\theta}_I$ reduces to $\hat{\theta}_n$ in the complete data case (i.e., when $\delta_i = 1$ for all i).

In the finite population sampling, the set of respondents can be viewed as the result of a three-stage process. First, the finite population is generated from an infinite population according to a given model. Then, a sample s of size n, is selected from the finite population according to a given sampling design $p(s)$. Finally, the set of respondents is generated from s according to the unknown nonresponse mechanism. Therefore, we identify three sources of randomness: (i) the model m, which generates the vector of population values $Y_U = (y_1, ..., y_N)'$; (ii) the sampling design $p(s)$, which generates the vector of sample indicators $I_U = (I_1, ..., I_N)'$ such that $I_i = 1$ if unit i is selected in the sample and $I_i = 0$, otherwise; (iii) the nonresponse mechanism, which generates the vector of response indicators $\delta_U = (\delta_1, ..., \delta_N)'$. Here, the response indicator δ_i is defined for all the population units.

We discuss the asymptotic properties of the imputation estimator $\hat{\theta}_I$ of the form (3) using $(\hat{\beta}^*, \hat{\phi}^*)$, where $(\hat{\beta}^*, \hat{\phi}^*)$ is obtained by solving simultaneously (9) and (10). Again, under some regularity conditions, the asymptotic equivalence in (15) holds and the resulting imputed estimator is doubly robust.

Traditionally, the total variance of the imputed estimator $\hat{\theta}_I$ has been expressed as the sum of the sampling variance and the nonresponse variance. This decomposition of the total variance results from viewing nonresponse as a second-phase of selection. For this reason, this framework is often called the two-phase framework; e.g., Rao & Shao (1992), Särndal (1992) and Deville & Särndal (1994), among others. In this paper, we consider an alternative framework, which we call the *reverse framework*; e.g., Fay (1992), Shao & Steel (1999) and Kim & Rao (2009). It consists of viewing the situation prevailing in the presence of nonresponse as follows: first, applying the nonresponse mechanism, the finite population U is randomly divided into a population of respondents U_r and a population of nonrespondents U_m. Then, given (U_r, U_m), a sample s, containing both respondents and nonrespondents, is selected from U according to the given sampling design.

Under the nonresponse model approach, the total variance of $\hat{\theta}_I$ can be expressed as

$$V_T^{NM} = V_1^{NM} + V_2^{NM}, \tag{22}$$

where $V_1^{NM} = E\{V(\hat{\theta}_I \mid Y_U, X_U, \delta_U)|Y_U, X_U\}$ and $V_2^{NM} = V\{E(\hat{\theta}_I \mid Y_U, X_U, \delta_U)|Y_U, X_U\}$ with $X_U = (x_1, ..., x_N)'$. Under the outcome regression model approach, the total variance of $\hat{\theta}_I$ can be expressed as

$$V_T^{IM} = V_1^{IM} + V_2^{IM}, \tag{23}$$

where $V_1^{IM} = E\{V(\hat{\theta}_I - \theta_N \mid Y_U, X_U, \delta_U) \mid X_U, \delta_U\}$ and $V_2^{IM} = V\{E(\hat{\theta}_I - \theta_N \mid Y_U, X_U, \delta_U) \mid X_U, \delta_U\}$. An estimator of V_T^{NM} (respectively V_T^{IM}) is thus obtained by estimating separately V_1^{NM} and V_2^{NM} (respectively V_1^{IM} and V_2^{IM}). Under mild regularity conditions, the component V_1^{NM} (respectively V_1^{IM}) is of order $O\left(n^{-1}\right)$, whereas the components V_2^{NM} (respectively V_2^{IM}) is of order $O\left(N^{-1}\right)$. Therefore, the contribution of V_2^{NM} (respectively V_2^{IM}) to the total variance, V_2^{NM}/V_T^{NM} (respectively V_2^{IM}/V_T^{IM}) is of order $O\left(N^{-1}n\right)$ and is negligible when the sampling fraction n/N is negligible.

In order to estimate either V_1^{NM} or V_1^{IM}, it suffices to estimate $V(\hat{\theta}_I|Y_U, X_U, \delta_U)$, which represents the variance due to sampling conditional on Y_U, X_U and δ_U. Once again, we can apply Theorem 1, which states that $\hat{\theta}_I$ is asymptotically equivalent to $\tilde{\theta}_I$ given by (16). As a result, we can approximate $V(\hat{\theta}_I|Y_U, X_U, \delta_U)$ by $V(\tilde{\theta}_I|Y_U, X_U, \delta_U)$. For example, for a fixed size or random size without replacement sampling design, we have

$$V(\tilde{\theta}_I|Y_U, X_U, \delta_U) = \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij} - \pi_i\pi_j)\frac{\eta_i}{\pi_i}\frac{\eta_j}{\pi_j}, \tag{24}$$

where η_i is given by (18) and π_{ij} denotes the second order inclusion probability for units i and j. An estimator of V_1^{NM} (respectively V_1^{IM}), denoted by $\hat{V}_1$, is then given by

$$\hat{V}_1 = \frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}\frac{(\pi_{ij} - \pi_i\pi_j)}{\pi_{ij}}\frac{\hat{\eta}_i}{\pi_i}\frac{\hat{\eta}_j}{\pi_j},$$

where $\hat{\eta}_i$ is obtained from η_i by replacing (β_0, ϕ_0) with $(\hat{\beta}^*, \hat{\phi}^*)$. Note that $\hat{V}_1$ is obtained by applying a standard variance estimation methods to $\hat{\eta}_i$ in the sample. Under mild regularity conditions (e.g., Deville, 1999), the estimator $\hat{V}_1$ is consistent for either V_1^{NM} or V_1^{IM} regardless of the validity of the assumed nonresponse model and imputation model. Consistency of $\hat{V}_1$ follows from standard regularity conditions used in the complete data case. If the sampling fraction n/N is negligible, a consistent estimator of the total variance of $\hat{\theta}_I$ (under either the nonresponse model approach or the outcome regression model approach) is given by $\hat{V}_1$.

When the sampling fraction is not negligible, we must take the term V_2^{NM} into account (in the case of the nonresponse model approach) or V_2^{IM} (in the case of the outcome regression model approach). Once again, we use the asymptotic equivalence between $\hat{\theta}_I$ and $\tilde{\theta}_I$ established in Theorem 1. First, we have

$$E(\tilde{\theta}_I - \theta_N|Y_U, X_U, \delta_U) = \frac{1}{N}\sum_{i\in U}(\eta_i^* - y_i),$$

where $\eta_i^* = \eta_i(\beta^*, \phi^*)$ is defined in (18). Under the nonresponse model,

$$\begin{aligned} V_2^{NM} &= V\left\{E(\tilde{\theta}_I - \theta_N|Y_U, X_U, \delta_U)|Y_U, X_U\right\} \\ &= \frac{1}{N^2}\sum_{i\in U}\frac{p_i(1-p_i)}{p_i^2}\{y_i - m(x_i, \beta^*)\}^2. \end{aligned}$$

Thus, an estimator of V_2^{NM}, denoted by $\hat{V}_2$, is given by

$$\hat{V}_2 = \frac{1}{N^2}\sum_{i\in s}\pi_i^{-1}\delta_i\frac{(1 - p_i(\hat{\phi}))}{p_i(\hat{\phi})^2}\hat{e}_i^2, \tag{25}$$

where $\hat{e}_i = y_i - m(x_i, \hat{\beta})$. Because $(\hat{\beta}^*, \hat{\phi}^*)$ is a consistent estimator of $\left(\beta^*, \phi_0\right)$ under the nonresponse model, $\hat{V}_2$ in (25) is asymptotically unbiased and consistent for V_2^{NM} under the nonresponse model. Therefore, a consistent estimator of the total variance under the nonresponse model is given by

$$\hat{V}_T = \hat{V}_1 + \hat{V}_2. \tag{26}$$

To see if $\hat{V}_T$ in (26) is doubly robust, we need to check if $\hat{V}_2$ in (25) is consistent for V_2^{IM} under the outcome regression model. Note that

$$\begin{aligned} V_2^{IM} &= V\left\{E(\tilde{\theta}_I - \theta_N|Y_U, X_U, \delta_U) \mid X_U, \delta_U\right\} \\ &= \frac{1}{N^2}\sum_{i\in U}\left\{\frac{\delta_i}{p_i(\phi^*)} - 1\right\}^2 V(y_i \mid x_i) \\ &= \frac{1}{N^2}\sum_{i\in U}\left\{\frac{\delta_i}{p_i(\phi^*)^2} - \frac{2\delta_i}{p_i(\phi^*)} + 1\right\} V(y_i \mid x_i). \end{aligned} \tag{27}$$

Thus, the asymptotic bias of $\hat{V}_2$ in (25) as an estimator of V_2^{IM} under the outcome regression model is

$$E\left\{\hat{V}_2\right\} - V_2^{IM} \doteq \frac{1}{N^2}\sum_{i\in U}E\left\{\frac{\delta_i}{p_i(\phi^*)} - 1\right\} V(y_i \mid x_i). \tag{28}$$

Thus, under the outcome regression model, if we further assume that

$$V\left(y_i \mid x_i\right) = \psi(x_i; \alpha_0)$$

for some α_0 and a consistent estimator $\hat{\alpha_0}$ is available, then the right side of (28) can be estimated by

$$\hat{B}\left(\hat{V}_2\right) = \frac{1}{N^2}\sum_{i \in s} \pi_i^{-1}\left\{\frac{\delta_i}{p_i(\hat{\phi}^*)} - 1\right\}\psi(x_i; \hat{\alpha}). \tag{29}$$

Note that the expected value of the estimated bias term in (29) is asymptotically equal to zero under the nonresponse model because $p_i(\hat{\phi}^*)$ converges to the true response probability. Thus, a bias-adjusted estimator of the total variance

$$\hat{V}_T \quad = \quad \hat{V}_1 + \hat{V}_2 - \hat{B}\left(\hat{V}_2\right) \tag{30}$$

is doubly robust.

5. Simulation Study

To test our theory, we performed two limited simulation studies. The first simulation study, presented in Section5.1, compares the performance of several point and variance estimators in the infinite population set-up. In Section5.2, the case of finite population sampling is considered.

5.1 Infinite population set-up

In the first simulation, the simulation study can be described as a $2 \times 2 \times 5$ factorial design with B = 5,000 replication within each cell. The factors are two types of sampling distributions, two types of the nonresponse mechanisms, and five types of point estimators. For the sampling distributions, the first was generated from a linear regression model, whereas the second was generated according to a non-linear model. For the linear model, we used

$$y_i = 1 + x_{1i} + \epsilon_i, \tag{31}$$

where $x_{1i} \sim N(1, 1)$, $\epsilon_i \sim N(0, 1)$, and x_{1i} and ϵ_i are independent. For the non-linear model, we used the same x_{1i} and ϵ_i, but y_i was generated independently according to

$$y_i = 0.5(x_{1i} - 1.5)^2 + \epsilon_i. \tag{32}$$

Two sets of random sample of size $n = 500$ were separately generated from the two models. From each sample, we generated two types of the respondents from $Bernoulli(p_{1i})$ (Type A) and $Bernoulli(p_{2i})$ (Type B), respectively, where $\text{logit}\,(p_{1i}) = x_{2i}$ and $\text{logit}\,(p_{2i}) = -0.5 + 0.5(x_{2i} - 2)^2$, where $x_{2i} \sim \exp(1)$ and x_{2i} is independent of (x_{1i}, ϵ_i). The overall response rates were about 60% in both cases.

In each sample, we computed five estimators for $\theta = E(Y)$.

1. Complete sample estimator ($\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$).
2. The proposed doubly robust estimator. (New)
3. The doubly robust estimator proposed by Haziza & Rao (2006). (HR)
4. The doubly robust estimator proposed by Cao et al. (2009). (CTD)
5. The doubly robust estimator proposed by Tan (2006). (Tan)

We considered three scenarios at the estimation stage:

1. Scenario 1: Both models are correct. That is, the sample was generated from (31) and the respondents were generated from the Type A model. The "working" outcome regression model is $E(y_i \mid x_{1i}) = \beta_0 + \beta_1 x_{1i}$ and the "working" response model is $\delta_i \sim Bernoulli(p_i)$ with $\text{logit}(p_i) = \phi_0 + \phi_1 x_{2i}$.
2. Scenario 2: Only the outcome regression model is correct. That is, we used the same working models in Scenario 1 but the sample was generated from (31) and the respondents were generated from the Type B model.

Table 1: Monte Carlo average and variance of the point estimators, based on 5,000 Monte Carlo samples

Scenario	Method	Mean	Variance	Standardized Variance
1	Sample Mean	2.00	0.003925	100
	New	2.00	0.005524	141
	CTD	2.00	0.005907	150
	HR	2.00	0.005524	141
	Tan	2.00	0.005530	141
2	Sample Mean	2.00	0.003925	100
	New	2.00	0.005278	134
	CTD	2.00	0.005287	135
	HR	2.00	0.005360	137
	Tan	2.00	0.005623	143
3	Sample Mean	0.62	0.003466	100
	New	0.62	0.005936	171
	CTD	0.62	0.006540	189
	HR	0.62	0.005939	171
	Tan	0.62	0.005942	171

3. Scenario 3: Only the nonresponse model is correct. That is, we used the same working models in Scenario 1 but the sample was generated from (32) and the respondents were generated from the Type A model.

For the three estimators (HR, CTD, Tan), $\left(\hat{\phi}_0, \hat{\phi}_1\right)$ was computed by the maximum likelihood method but, whereas it was computed by solving

$$\sum_{i=1}^{n} \frac{\delta_i}{p_i(\phi)} (1, x_{2i}) = \sum_{i=1}^{n} (1, x_{2i}) \tag{33}$$

for the New estimator, where $\phi = (\phi_0, \phi_1)$. Once the $\hat{p}_i$'s were computed, both HR and the New methods used $(\hat{\beta}_0, \hat{\beta}_1)$ given by

$$(\hat{\beta}_0, \hat{\beta}_1)' = \left\{ \sum_{i=1}^{n} \delta_i \left(\hat{p}_i^{-1} - 1 \right) x_i x_i' \right\}^{-1} \sum_{i=1}^{n} \delta_i \left(\hat{p}_i^{-1} - 1 \right) x_i y_i, \tag{34}$$

where $x_i = (1, x_{1i})'$. For the CTD estimator, we used

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{c}_0, \hat{c}_1)' = \left\{ \sum_{i=1}^{n} \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1 \right) \tilde{x}_i \tilde{x}_i' \right\}^{-1} \sum_{i=1}^{n} \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1 \right) \tilde{x}_i y_i, \tag{35}$$

where $\tilde{x}_i = (1, x_{1i}, \hat{p}_i, \hat{p}_i x_{2i})'$. The doubly robust estimator of Tan (2006) is computed by

$$\hat{\theta}_{tan} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i y_i}{\hat{p}_i} - \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\delta_i}{\hat{p}_i} - 1 \right) \left(\hat{k}_0 + \hat{k}_1 \hat{m}_i \right)$$

where $\hat{m}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i}$ and

$$(\hat{k}_0, \hat{k}_1, \hat{c}_0, \hat{c}_1)' = \left\{ \sum_{i=1}^{n} \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1 \right) \tilde{z}_i \tilde{z}_i' \right\}^{-1} \sum_{i=1}^{n} \delta_i \hat{p}_i^{-1} \left(\hat{p}_i^{-1} - 1 \right) \tilde{z}_i y_i, \tag{36}$$

where $\tilde{z}_i = (1, \hat{m}_i, \hat{p}_i, \hat{p}_i x_{2i})'$.

Table 1 presents the Monte Carlo averages and variances of five estimators under three different scenarios. The four doubly robust estimators (New, HR, CTD, and Tan) were all approximately unbiased in all the scenarios, illustrating that they are doubly robust. Turning to relative efficiency, both the HR estimator and the New estimator showed similar performances and were more efficient than the CTD estimator and Tan's estimator in all the scenarios. In scenario 2, the New estimator performed the best since the calibration condition can be justified as the optimality condition when the outcome regression model is true. Tan's estimator showed slightly higher variance under scenario 2, whereas the CTD estimator had slightly higher variance under scenario 3.

Table 2: Monte Carlo percent relative bias of the two variance estimators, based on 5,000 Monte Carlo samples

Sample Size	Scenario	Method	Relative Bias (%)
500	1	New	2.27
		CTD	5.75
	2	New	4.69
		CTD	3.33
	3	New	-0.04
		CTD	8.27

We now turn to variance estimation. We computed variance estimators for both the CTD and the New estimators. The variance estimator proposed by Cao et al. (2009) was computed using (20) with

$$\eta_i = m(x_i;\hat{\beta}) + \frac{\delta_i}{\hat{p}_i}\left\{y_i - m(x_i;\hat{\beta})\right\} - \hat{\mathbf{c}}'\left(\delta_i - \hat{p}_i\right)(1, x_{2i})',$$

where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\mathbf{c}} = (\hat{c}_0, \hat{c}_1)$ were computed from (35). The variance estimator for the New estimator was computed using (20) with

$$\eta_i = m(x_i;\hat{\beta}) + \frac{\delta_i}{\hat{p}_i}\left\{y_i - m(x_i;\hat{\beta})\right\} \tag{37}$$

and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ is given by (34). In (37), we obtained $\hat{p}_i$ using the maximum likelihood method. Variance estimation in the context of Tan's estimator was not covered here as Tan (2006) did not discuss variance estimation.

Table 2 presents the Monte Carlo bias of the variance estimators of the CTD and the New estimators. The proposed variance estimator corresponding to the New estimator showed small relative biases (less than 5% in absolute values) in all the scenarios. Thus, the results from this study suggests that the variance estimator for the New estimator is doubly robust. The variance estimator for CTD method showed some modest bias (8.27%) under scenario 3.

5.2 Finite population set-up

We generated two finite populations of size $N = 5000$. In each population, we generated 4 variables: a variable of interest y and three auxiliary variables x_1, x_2 and x_3. First, the x_1 and x_3-values were generated from a Gamma distribution with parameters 2 and 2. The x_2-values were generated from a Gamma distribution with parameters 25 and 2. Given the x_1-values, the y-values were generated according to the linear model

$$y_i = 1 + x_{1i} + 1.8\epsilon_i,$$

for population 1 and according to the nonlinear model

$$y_i = 0.5\left(x_{1i} - 1.5\right)^2 + \epsilon_i,$$

for population 2. where the ϵ_i's were generated from a normal distribution with mean 0 and variance 1. Note that in each population, the model linking y and x_1 possesses an homoscedastic variance structure.

In both populations, the x_2-values were then sorted in ascending order and were partitioned into 4 strata U_1, U_2, U_3 and U_4 of size $N_1 = 2500$, $N_2 = 1000$, $N_3 = 1000$ and $N_4 = 500$, respectively. For population 1, the coefficient of determination (R^2) of the model linking y and x_1 varied from 0.75 to 0.81 across strata, whereas it varied from 0.72 to 0.76 for population 2.

The objective consisted in estimating the finite population mean $\theta_N = N^{-1}\sum_{i\in U} y_i$. From the population, we generated $R = 5,000$ samples according to stratified simple random sampling without replacement. That is, in each stratum, a simple random sample s_h of size n_h was selected from U_h, $h = 1, 2, 3, 4$. Equal allocation (i,.e. equal values of n_h) was used with $n_h = 125$ and $n_h = 250$, which correspond to an overall sampling fraction of 10% and 20%, respectively. In each selected sample, nonresponse to the study variable y was generated according to the nonresponse mechanism

$$\text{logit}\ (p_i) = 2 - x_{1i}$$

for population 1 and

$$\text{logit}\ (p_i) = (x_{1i} - 1.5)^2$$

Table 3: Working models used for estimation

Population	Scenario	Nonresponse working model	Outcome regression working model
1	1	logit $(p_i) = \phi_0 + \phi_1 x_{1i}$	$y_i = \lambda_0 + \lambda_1 x_{1i}$
	2	logit $(p_i) = \phi_0 + \phi_1 x_{1i}$	$y_i = \lambda_0 + \lambda_1 x_{3i}$
	3	logit $(p_i) = \phi_0 + \phi_1 x_{3i}$	$y_i = \lambda_0 + \lambda_1 x_{1i}$
2	4	logit $(p_i) = \phi_0 + \phi_1 x_{1i} + \phi_2 x_{1i}^2$	$y_i = \lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{1i}^2$
	5	logit $(p_i) = \phi_0 + \phi_1 x_{1i} + \phi_2 x_{1i}^2$	$y_i = \lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{3i}$
	6	logit $(p_i) = \phi_0 + \phi_1 x_{1i} + \phi_2 x_{3i}$	$y_i = \lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{1i}^2$

for population 2. The values of the parameters in the previous two expressions were chosen so that, within each stratum, the response rate was approximately equal to 70%.

We computed three estimators of the mean based on the working models presented in Table 3: (i) the complete sample estimator given by (1) with $w_i = 1/(N\pi_i)$; (ii) the estimator proposed by Haziza & Rao (HR) and (iii) the proposed estimator (New).

Finally, in each sample, we computed the estimator of the total variance given by (30). Note that, in order to compute (29), we used

$$\psi(x_i; \hat{\alpha}) = \frac{\sum_{i\in s} w_i \delta_i e_i^2}{\sum_{i\in s} w_i \delta_i},$$

where e_i denotes the residual attached to unit i obtained after fitting the working outcome regression model.

For each population, we considered three types of scenarios:

(i) Scenarios 1 and 4: Both the nonresponse model and the outcome regression model were correctly specified.

(ii) Scenario 2 and 5: Only the nonresponse model was correctly specified.

(iii) Scenario 3 and 6: Only the outcome regression model was correctly specified.

For each scenario, the working models are presented in Table 3.

Table 4 presents the Monte Carlo averages and variances of three estimators under six different scenarios. Both the HR and the New estimator showed negligible bias in all six scenarios, which is a clear indication that both estimators are robust to misspecification of either one model or the other. In terms of stability, the two estimators showed almost identical performances, with, in some cases, a slight advantage for the New estimator.

Table 5 show the Monte Carlo percent relative bias of the proposed variance estimator. We note that it performs relatively well in all the scenarios (with a relative absolute bias less than 6.02 %), which illustrates that it is doubly robust.

6. Concluding remarks

In this paper, we proposed a new doubly robust estimator that showed good finite sample performances in simulation studies. The resulting variance estimator is also doubly robust and can be readily implemented using complete data software, which is attractive from a data user's perspective. The proposed doubly robust estimator can be obtained by obtaining $(\hat{\beta}, \hat{\phi})$ from (9) and (10). Condition (10) is called the calibration condition and can often lead to an efficient estimator. In particular, if the "working" outcome regression model is a linear regression model $E(y_i) = x_i\beta$, then condition (10) is the typical calibration condition using x_i as the control variable. In this case, if the working regression model is good, then the resulting estimator is efficient. In the extreme case of $y_i = x_i\beta$, which means a perfect fit using x_i, the resulting estimator is algebraically equal to $\hat{\theta}_{DR} = \sum_{i=1}^n w_i y_i$, showing that the resulting doubly robust estimator is fully efficient when the outcome regression model is perfect. This type of consistency, so-called external consistency, does not hold for the other doubly robust estimators considered in this paper.

In the simulation studies, the new method showed better efficiency than the other doubly robust estimators in most cases, but there is no guarantee that it is optimal uniformly. Further investigation in this direction may be a topic of future research.

Acknowledgement

We are grateful to professor J.N.K. Rao for helpful comments. The research of the first author was partially supported by a Cooperative Agreement between the US Department of Agriculture Natural Resources Conservation Service

Table 4: Monte Carlo average and variance of the point estimators, based on 5,000 Monte Carlo samples

Scenario	Method	Mean	Variance	Standardized Variance	Mean	Variance	Standardized Variance
		$n_h = 125$			$n_h = 250$		
1	Complete	5.01	0.02751	100	5.01	0.0130	100
	New	5.01	0.03355	122	5.01	0.01586	122
	HR	5.01	0.03355	122	5.01	0.01586	122
2	Complete	5.01	0.02802	100	5.01	0.01275	100
	New	5.01	0.03446	123	5.01	0.01568	123
	HR	5.01	0.03502	125	5.01	0.01593	125
3	Complete	5.01	0.02851	100	5.01	0.01281	100
	New	5.00	0.03364	118	5.01	0.01524	119
	HR	5.00	0.03364	118	5.01	0.01524	119
4	Complete	0.62	0.00241	100	0.62	0.00115	100
	New	0.62	0.00279	116	0.62	0.00135	118
	HR	0.62	0.00279	116	0.62	0.00135	118
5	Complete	0.62	0.00229	100	0.62	0.00108	100
	New	0.62	0.00267	117	0.62	0.00126	117
	HR	0.62	0.00267	117	0.62	0.00126	117
6	Complete	0.62	0.00240	100	0.62	0.00105	100
	New	0.62	0.00261	109	0.62	0.00123	118
	HR	0.62	0.00261	109	0.62	0.00123	118

Table 5: Monte Carlo percent relative bias of the proposed variance estimator, based on 5,000 Monte Carlo samples

Scenario	$n_h = 125$	$n_h = 250$
1	-1.80	-2.51
2	3.85	5.44
3	-5.11	-3.55
4	-4.12	-2.15
5	5.24	6.02
6	-0.96	-0.33

and Iowa State University. The research of the second author was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

Appendix 1

Asymptotic properties

Before deriving the asymptotic properties of $\hat{\theta}_I$, we assume the following regularity conditions.

(C.1) The response probability is bounded below; that is, there is a fixed constant K_B such that $\pi_i^{-1} < K_B$ for all $i = 1, 2, \cdots, n$ uniformly.

(C.2) The assumed response probability function $p_i(\phi)$ is differentiable with continuous first order partial derivatives for all ϕ.

(C.3) The solution $\left(\hat{\beta}^*, \hat{\phi}^*\right)$ to (9) and (10) is uniquely determined and satisfies $(\hat{\beta}^*, \hat{\phi}^*) = (\beta^*, \phi^*) + o_p(1)$ for some (β^*, ϕ^*).

(C.4) The mean function $m(x_i; \beta)$ is twice differentiable with continuous second-order partial derivatives for all β.

(C.5) $W(\beta) = (X, Y, m(x; \beta), \dot{m}(x; \beta))$ has finite fourth moment for all β.

To prove Theorem 1, first write the imputed estimator as $\hat{\theta}_I = \hat{\theta}_I(\hat{\beta}^*, \hat{\phi}^*)$, where $(\hat{\beta}^*, \hat{\phi}^*)$ is the solution to (9) and (10). Now, if we define

$$U(\beta, \phi) = \sum_{i=1}^{n} w_i \left(\frac{\delta_i}{p_i(\phi)} - 1 \right) \{y_i - m(x_i; \beta)\},$$

we can express $\hat{\theta}_I(\hat{\beta}^*, \hat{\phi}^*)$ as

$$\hat{\theta}_I(\hat{\beta}^*, \hat{\phi}^*) = \hat{\theta}_n + U(\hat{\beta}^*, \hat{\phi}^*). \tag{38}$$

Note that $U(\beta, \phi)$ satisfies

$$\frac{\partial}{\partial \phi} U(\beta, \phi) = -\sum_{i=1}^{n} w_i \delta_i \left\{ \frac{1 - p_i(\phi)}{p_i(\phi)} \right\} \{y_i - m(x_i; \beta)\} x_i$$

and

$$\frac{\partial}{\partial \beta} U(\beta, \phi) = -\sum_{i=1}^{n} w_i \left\{ \frac{\delta_i}{p_i(\phi)} - 1 \right\} \dot{m}(x_i; \beta).$$

Thus, conditions (9) and (10), are equivalent to

$$\frac{\partial}{\partial(\beta, \phi)} U(\beta, \phi) = \mathbf{0}. \tag{39}$$

Because of the existence of the second moment of the partial derivatives in (39), standard arguments for the asymptotic normality of $(\hat{\beta}^*, \hat{\phi}^*)$ can be used to show that

$$(\hat{\beta}^*, \hat{\phi}^*) - (\beta^*, \phi^*) = O_p\left(n^{-1/2}\right). \tag{40}$$

Because $(\hat{\beta}^*, \hat{\phi}^*)$ satisfies (39), its probability limit (β^*, ϕ^*) satisfies

$$E\left\{ \frac{\partial}{\partial(\beta, \phi)} U(\beta, \phi) \mid \beta = \beta^*, \phi = \phi^* \right\} = \mathbf{0}. \tag{41}$$

Condition (41) implies that the contribution due to estimating the parameters (β, ϕ) is negligible in the asymptotic distribution of $U(\beta, \phi)$. Condition (41) is often called Randles (1982) condition. Results (40) and (41) implies that

$$U(\hat{\beta}^*, \hat{\phi}^*) = U(\beta^*, \phi^*) + o_p(n^{-1/2}). \tag{42}$$

Therefore, combining (38) and (42), we prove (15).

REFERENCES

Bang, H. and Robins, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–73.

Chang, T. and Kott. P. S. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika* **95**, 555–71.

Cao, W., Tsiatis, A. A. and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, **96**, 723–34.

Cochran, W.G. (1977). *Sampling Techniques.* Wiley: New-York.

Deville, J.-C. Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology* **25**, 193-203.

Deville, J.C. and Särndal, C.E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. *Journal of Official Statistics* **10**, 381-94.

Fay, R. E. (1992). When are inferences from multiple imputation valid ? In *Proceeding of the Survey Research Methods Section,* pp. 227-32. Washington, DC: American Statistical Association.

Haziza, D. and Rao, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology* **32**, 53-64.

Kang, J.D.Y. and Schafer, J.L. (2008). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523-39.

Kim, J.K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal Statistics* **35**, 501–14.

Kim, J.K. and Park, H.A. (2006). Imputation using response probability. *Canadian Journal Statistics* **34**, 171–82.

Kim, J.K. and Rao, J.N.K. (2009). Unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* **96**, 917–32.

Kott, P.S. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association* **89**, 693–6.

Lannacchione, V.G., Milne, J.G. and Folsom, R.E. (1991). Response probability weight adjustments using logistic regression. *Proceedings of the Survey Research Methods Section,* pp. 637–42. American Statistical Association.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Annals of Statistics* **10**, 462–74.

Rao, J. N. K. and Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811-22.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficient when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–66.

Särndal, C. E. (1992). Method for estimating the precision of survey estimates when imputation has been used. *Survey Methodology* **18**, 241–52.

Scharfstein, D.O., Rotnitzky, A. and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion and rejoinder). *Journal of the American Statistical Association* **94**, 1096–146.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association* **94**, 254-65.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association* **101**, 1619-37.