

Estimation of Finite Population Mean and Superpopulation Parameters when the Sampling Design is Informative and Nonresponse Mechanism is Nonignorable

Abdulhakeem A.H. Eideh
Department of Mathematics, Al-Quds University
Abu-Dies Campus, Palestine, P.O. Box 20002
msabdul@staff.alquds.edu

Abstract

In this paper we study the joint treatment of not missing at random response mechanism and informative sampling for survey data. This is the most general situation in surveys and other combinations of sampling informativeness and response mechanisms can be considered as special cases. The proposed method combines two methodologies used in the analysis of sample surveys for the treatment of informative sampling and the nonignorable nonresponse mechanism. One incorporates the dependence of the first order inclusion probabilities on the study variable, while the other incorporates the dependence of the probability of nonresponse on unobserved or missing observations. The main purpose here is the estimation of finite population mean and superpopulation parameters when the sampling design is informative and nonresponse mechanism is nonignorable. Under four scenarios of sampling design and nonresponse mechanism, we obtained the method of moment estimators of finite population mean, with their biases and mean square errors. Furthermore, a four-step estimation method is introduced for the estimation of superpopulation parameters under informative sampling and nonignorable nonresponse mechanism. New relationships between moments of response, nonresponse, sample, sample-complement and population distributions were derived. Most estimators for finite population mean known from sampling surveys can be derived as a special case of the results derived in this paper.

Keywords: Response distribution, Nonignorable nonresponse, Informative sampling design, Poststratification.

1. Introduction

Data collected by sample surveys are used extensively to make inferences on assumed population models. Often, survey design features (clustering, stratification, unequal probability selection, etc.) are ignored and the sample data are then analyzed using classical methods based on simple random sampling. This approach can, however, lead to erroneous inference because of sample selection bias implied by informative sampling - the sample selection probabilities depend on the values of the model outcome variable (or the model outcome variable is correlated with design variables not included in the model). See Pfeffermann et. al (1998) and Eideh and Nathan (2006). In addition to the effect of complex sample design, one of the major problems in the analysis of survey data is that of missing values. Rubin (1976) and Little and Rubin (2002) consider three types of nonresponse mechanism or missing data mechanism:

(a) Missing completely at random (MCAR): if the response probability does not depend on the study variable, or the auxiliary population variable, the missing data are MCAR.

(b) Missing at random (MAR) given auxiliary population variable: if the response probability depends on the auxiliary population variable but not on the study variable, the missing data are MAR.

(c) Not missing at random (NMAR): if the response probability depends on the value of a missing study variable, the missing data are NMAR.

So, the cross-classification of sampling design and response mechanism is summarized in the following table:

Table 1

Sampling Design	Response Mechanism		
	MCAR	MAR	NMAR
Informative-INF	INFMCAR	INFMAR	INFNMAR
Noninformative-NINF	NINFMCAR	NINFMAR	NINFNMAR

The literature dealing with the treatment of nonresponse in surveys, deals with NINFMCAR, NINFMAR, NINFNMAR, INFMCAR, and INFMAR. See for example, Little and Rubin 2002, Schafer 1997, Little 1982, Rubin 1976, Särndal and Swensson 1987, Cobben 2009, Chambers and Skinner 2003, Pfeiffermann and Sikov (2011), Little (1993; 1994), Tang et al. (2003) Qin et al. (2002) Chang and Kott (2008). The methods used in these papers are summarized by Pfeiffermann and Sikov (2011) and Eideh (2012).

For inference problem, Little (1982) classify the nonresponse mechanism as ignorable (MAR and MCAR) and nonignorable (NMAR). For this sense, the cross classification of sampling design and nonresponse mechanism is:

Table 2

Sampling Design	Nonresponse Mechanism	
	Ignorable	Nonignorable
Informative	II	IN
Noninformative	NI	NN

Pfeiffermann and Sikov (2011), and Eideh (2012) consider estimation of superpopulation parameters and prediction of finite population parameters (census parameters) under nonignorable nonresponse via response and nonresponse distributions when the sampling design is noninformative.

None of the above studies consider simultaneously the problem of informative sampling and the problem of nonignorable nonresponse when analyzing survey data.

In this paper, we study, within a modeling framework, the joint treatment of nonignorable nonresponse mechanism and informative sampling for survey data, by specifying the probability distribution of the observed measurements when the sampling design is informative. This is the most general situation in surveys and other combinations of sampling informativeness and response mechanisms can be considered as special cases.

It should be pointed here that, according to Sarndal (2011) “Nonresponse causes both bias and increased variance. Its square is typically the dominant portion of the Mean Squared Error (MSE). We address primarily surveys on individuals and households with

quite large sample sizes, as is typical for Journal of Official Statistics for government surveys; consequently, the variance contribution to MSE is low by comparison. Increased variance due to nonresponse is nevertheless an issue; striking a balance between variance increase and bias reduction is considered, for example, in Little and Vartivarian (2005).” Furthermore, Brick (2013) mentioned that “Model assumptions and adjustments are made in an attempt to compensate for missing data. Because the mechanisms that cause unit nonresponse are almost never adequately reflected in the model assumptions, survey estimates may be biased even after the model based adjustments. Nonresponse also causes a loss in the precision of survey estimates, primarily due to reduced sample size and secondarily as the result of increased variation of the survey weights. However, bias is the dominant component of the nonresponse-related error in the estimates, and nonresponse bias generally does not decrease as the sample size increases. Thus, bias is often the largest component of mean square error of the estimates even for subdomains when the sample size is large”. In we focus here on the bias, variance and MSE.

The paper is structured as follow. Section 2 reviews the definition of response distribution and estimation of response probabilities. Section 3 introduces new relationships between moments of response, nonresponse, sample, sample-complement and population distributions. Section 4 describes the estimation of finite population total $\bar{Y}_U = N^{-1} \sum_{i=1}^N y_i$, under the four scenarios mentioned in Table 2. Also the main purpose in this section is the computation of the biases and mean square errors of the estimators. Section 5 is devoted to the estimation of superpopulation parameters under informative sampling and nonignorable nonresponse mechanism. Section 6 provides the conclusions.

2. Response and Nonresponse Distributions

Let $U = \{1, \dots, N\}$ denote a finite population consisting of N units. Let y be the study variable of interest and let y_i be the value of y for the i th population unit. A probability sample s is drawn from U according to a specified sampling design. The sample size is denoted by n . Let $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$, $i \in U$ be the values of a vector of auxiliary variables, x_1, \dots, x_p , and $\mathbf{z} = \{z_1, \dots, z_N\}$ be the values of known design variables, used for the sample selection process not included in the model under consideration. In what follows, we consider a sampling design with selection probabilities $\pi_i = \Pr(i \in s) > 0$, and sampling weight $w_i = 1/\pi_i$; $i = 1, \dots, N$. In practice, the π_i 's may depend on the population values $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. We express this dependence by writing: $\pi_i = \Pr(i \in s | \mathbf{x}, \mathbf{y}, \mathbf{z})$ for all units $i \in U$. Denote by $\mathbf{I} = (I_1, \dots, I_N)'$ the N by 1 sample indicator (vector) variable, such that $I_i = 1$ if unit $i \in U$ is selected to the sample and $I_i = 0$ if otherwise. So that $s = \{i | i \in U, I_i = 1\}$ and its complement is $\bar{s} = c = \{i | i \in U, I_i = 0\}$. We consider the population values y_1, \dots, y_N as random variables, which are independent realizations from a distribution with probability density functions (pdf) $f_p(y_i | \mathbf{x}_i; \theta)$, indexed by a vector of parameters θ .

In addition to the effect of complex sample design, one of the major problems in the analysis of survey data is that of missing values. In recent articles by Eideh (2009), Pfeffermann and Sikov (2011), and Eideh (2012), the authors defined and studied the problem of nonignorable nonresponse using the response and nonresponse distributions where the sampling design is noninformative. Following the notations, denote by $R = (R_1, \dots, R_N)'$ the N by 1 response indicator (vector) variable such that $R_i = 1$ if unit $i \in s$ is observed and $R_i = 0$ if otherwise. We assume that these random variables are independent of one another and of the sample selection mechanism (Oh and Scheuren 1983). The response set is defined accordingly as $r = \{i \in s \mid R_i = 1\}$ and the nonresponse set by $\bar{r} = \{i \in s \mid R_i = 0\}$. We assume probability sampling, so that $\pi_i = \Pr(i \in s) > 0$ for all units $i \in U$. Let the response probability $\psi_i = \Pr(i \in r \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) > 0$ for all units $i \in s$ and $\phi_i = 1/\psi_i$ be the response weight for $i \in s$.

Eideh (2009) defined and studies the properties of response and nonresponse distributions when the sampling design is informative and nonresponse mechanism is NMAR or nonignorable. According to Eideh (2009), the (marginal) response pdf of y_i is given by:

$$f_r(y_i \mid \mathbf{x}_i, \theta, \eta, \gamma) = \frac{E_s(\psi_i \mid \mathbf{x}_i, y_i, \gamma) f_s(y_i \mid \mathbf{x}_i, \theta, \eta)}{E_s(\psi_i \mid \mathbf{x}_i, \theta, \eta, \gamma)} \quad (1)$$

where, according to Pfeffermann *et al.* (1998), the sample pdf of y_i is:

$$f_s(y_i \mid \mathbf{x}_i, \theta, \gamma) = \frac{E_p(\pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{E_p(\pi_i \mid \mathbf{x}_i, \theta, \gamma)} \quad (2)$$

Furthermore, Sverchkov and Pfeffermann (2001) define the sample-complement pdf of y_i as:

$$f_{\bar{r}}(y_i \mid \mathbf{x}_i, \theta, \gamma) = \frac{E_p(1 - \pi_i \mid \mathbf{x}_i, y_i, \gamma) f_p(y_i \mid \mathbf{x}_i, \theta)}{E_p(1 - \pi_i \mid \mathbf{x}_i, \theta, \gamma)} \quad (3)$$

Note that $f_r(y_i \mid \mathbf{x}_i, \theta, \eta, \gamma)$ is completely specified by $f_p(y_i \mid \mathbf{x}_i, \theta)$, $E_p(\pi_i \mid \mathbf{x}_i, y_i, \eta)$ and $E_s(\psi_i \mid \mathbf{x}_i, y_i, \gamma)$.

Similarly, the (marginal) nonresponse pdf of y_i is given by:

$$f_{\bar{r}}(y_i \mid \mathbf{x}_i) = \frac{\{1 - E_s(\psi_i \mid \mathbf{x}_i, y_i)\} f_s(y_i \mid \mathbf{x}_i)}{\{1 - E_s(\psi_i \mid \mathbf{x}_i)\}} \quad (4)$$

It should be noted here that, the parameters γ and η that index mechanisms of response and sample models, respectively, are characteristics of the data collection but are not generally of scientific interest. For more discussion on the use of response and nonresponse distributions for analytic inference in survey sampling, see Pfeffermann and Sverchkov (2004), Pfeffermann and Sikov (2011), and Eideh (2009, 2012).

Estimation of response probabilities ψ_i for all $i \in s$:

If the nonresponse mechanism is not missing at random, then the classical methods for estimating the response probabilities using auxiliary variables, available for respondents and nonrespondents, is logistic or probit models. If we use the logistic model, then

$$\psi_i = \Pr(R_i = 1 | i \in s, x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \quad (5)$$

We can fit this model using maximum likelihood approach. Thus the estimate of ψ_i is:

$$\hat{\psi}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)} \quad (6)$$

If the nonresponse mechanism is NMAR, then values of y_i for $i \in r$ is available, but for $i \notin r$ are not available, so we cannot fit the following model:

$$\psi_i = \Pr(R_i = 1 | i \in s, x, y) = \frac{\exp(\beta_0 + \beta_1 x_i + \beta_2 y_i)}{1 + \exp(\beta_0 + \beta_1 x_i + \beta_2 y_i)} \quad (7)$$

directly using maximum likelihood method. A recent approach of estimation ψ_i under nonignorable nonresponse is discussed by Matei A. and Giovanna M.R. (2015), based on latent modeling approach.

3. Relationships between Moments of Response, Nonresponse, Sample, Sample-Complement and Population distributions.

Following the definitions of response, and nonresponse distributions, we notice that, given the population distribution, these distributions are completely determined by identifying $E_p(\pi_i | \mathbf{x}_i, y_i), E_s(\psi_i | \mathbf{x}_i, y_i)$. In practice, these conditional expectations are not known. Assuming that the available data to the analyst is $\{I_i, i \in U\}, \{R_i, x_i, w_i, \psi_i, i \in s\}$ and $\{y_i, x_i, i \in r\}$, which is the case in secondary analysis, the question that arises is: how can we identify and estimate, $E_p(\pi_i | y_i, \mathbf{x}_i)$ and $E_s(\psi_i | \mathbf{x}_i, y_i)$ based only on the theses data? The following relationships answers this question. Let (y_i, \mathbf{x}_i) be a vector of random variables.

According to Pfeffermann and Sverchkov (1999), the following relationship hold:

$$E_p(y_i | \mathbf{x}_i) = \{E_s(w_i | \mathbf{x}_i)\}^{-1} E_s(w_i y_i | \mathbf{x}_i) \quad (8)$$

Also, according to Sverchkov and Pfeffermann (2004), we have:

$$E_s(y_i | \mathbf{x}_i) = \frac{E_p\{(1 - \pi_i)y_i | \mathbf{x}_i\}}{E_p\{(1 - \pi_i) | \mathbf{x}_i\}} = \frac{E_s\{(w_i - 1)y_i | \mathbf{x}_i\}}{E_s\{(w_i - 1) | \mathbf{x}_i\}} \quad (9)$$

Furthermore, for vector of random variables (y_i, \mathbf{x}_i) , Eideh (2009), proved the following relationship:

$$E_s(y_i | \mathbf{x}_i) = \{E_r(\phi_i | \mathbf{x}_i)\}^{-1} E_r(\phi_i y_i | \mathbf{x}_i) \quad (10)$$

$$E_{\bar{r}}(y_i | \mathbf{x}_i) = \frac{E_s\{(1 - \psi_i)y_i | \mathbf{x}_i\}}{E_s\{(1 - \psi_i) | \mathbf{x}_i\}} = \frac{E_r\{(\phi_i - 1)y_i | \mathbf{x}_i\}}{E_r\{(\phi_i - 1) | \mathbf{x}_i\}} \quad (11)$$

The previous relationships and the following new relationships are fruitful in estimation of the parameters indexing superpopulation model, informative sampling design, nonresponse mechanism, and prediction of finite population parameters.

Lemma 1.

$$E_p(y_i | \mathbf{x}_i) = \frac{E_r(\phi_i w_i y_i | \mathbf{x}_i)}{E_r(\phi_i w_i | \mathbf{x}_i)} \quad (12)$$

Proof:

Using (8) and (10), we get:

$$\begin{aligned} E_p(y_i | \mathbf{x}_i) &= \frac{E_s(w_i y_i | \mathbf{x}_i)}{E_s(w_i | \mathbf{x}_i)} = \frac{E_r(\phi_i w_i y_i | \mathbf{x}_i)}{E_r(\phi_i w_i | \mathbf{x}_i)} \frac{E_s(w_i y_i | \mathbf{x}_i)}{E_s(w_i | \mathbf{x}_i)} \\ &= \frac{E_r(\phi_i w_i y_i | \mathbf{x}_i) E_r(\phi_i | \mathbf{x}_i)}{E_r(\phi_i | \mathbf{x}_i) E_r(\phi_i w_i | \mathbf{x}_i)} = \frac{E_r(\phi_i w_i y_i | \mathbf{x}_i)}{E_r(\phi_i w_i | \mathbf{x}_i)} \end{aligned}$$

Lemma 2.

$$E_{\bar{s}}(y_i | \mathbf{x}_i) = \frac{E_r\{\phi_i(w_i - 1)y_i | \mathbf{x}_i\}}{E_r\{\phi_i(w_i - 1) | \mathbf{x}_i\}} \quad (13)$$

Proof:

According to (9) and (10), we have:

$$\begin{aligned} E_{\bar{s}}(y_i | \mathbf{x}_i) &= \frac{E_s\{(w_i - 1)y_i | \mathbf{x}_i\}}{E_s\{(w_i - 1) | \mathbf{x}_i\}} = \frac{E_r\{\phi_i(w_i - 1)y_i | \mathbf{x}_i\} E_r\{\phi_i | \mathbf{x}_i\}}{E_r\{\phi_i | \mathbf{x}_i\} E_r\{\phi_i(w_i - 1) | \mathbf{x}_i\}} \\ &= \frac{E_r\{\phi_i(w_i - 1)y_i | \mathbf{x}_i\}}{E_r\{\phi_i(w_i - 1) | \mathbf{x}_i\}} \end{aligned}$$

Lemma 3.

$$E_{\bar{r}}(y_i) = E_s(y_i) - \frac{Cov_s(\psi_i, y_i)}{1 - E_s(\psi_i)} \quad (14)$$

$$E_s(y_i) - E_{\bar{r}}(y_i) = \frac{Cov_s(\psi_i, y_i)}{1 - E_s(\psi_i)} \quad (15)$$

where $Cov_s(\psi_i, y_i) = E_s(y_i \psi_i) - E_s(y_i) E_s(\psi_i)$.

Proof:

According to (11), we have

$$\begin{aligned} E_{\bar{r}}(y_i) &= \frac{E_s\{(1-\psi_i)y_i\}}{E_s(1-\psi_i)} = \frac{E_s(y_i) - E_s(y_i\psi_i)}{1 - E_s(\psi_i)} \\ &= \frac{E_s(y_i) - E_s(y_i\psi_i) - E_s(y_i)E_s(\psi_i) + E_s(y_i)E_s(\psi_i)}{1 - E_p(\psi_i)} \\ &= \frac{E_s(y_i)E_s(1-\psi_i) - Cov_s(\psi_i, y_i)}{1 - E_s(\psi_i)} = E_s(y_i) - \frac{Cov_s(\psi_i, y_i)}{1 - E_s(\psi_i)} \end{aligned}$$

Similar proof for equation (15).

Note that Section 3.3 of Beaumont (2002) is a special case of equation (14).

Lemma 4.

$$E_{\bar{s}}(y_i) = E_p(y_i) - \frac{Cov_p(\pi_i, y_i)}{1 - E_p(\pi_i)} \quad (16)$$

$$E_p(y_i) - E_{\bar{s}}(y_i) = \frac{Cov_p(\pi_i, y_i)}{1 - E_p(\pi_i)} \quad (17)$$

where $Cov_p(\pi_i, y_i) = E_p(\pi_i y_i) - E_p(y_i)E_p(\pi_i)$.

Proof:

According to (9), we have

$$\begin{aligned} E_{\bar{s}}(y_i) &= \frac{E_p\{(1-\pi_i)y_i\}}{E_p(1-\pi_i)} = \frac{E_p(y_i) - E_p(y_i\pi_i)}{1 - E_p(\pi_i)} \\ &= \frac{E_p(y_i) - E_p(y_i\pi_i) - E_p(y_i)E_p(\pi_i) + E_p(y_i)E_p(\pi_i)}{1 - E_p(\pi_i)} \\ &= \frac{E_p(y_i)E_p(1-\pi_i) - Cov_p(\pi_i, y_i)}{1 - E_p(\pi_i)} = E_p(y_i) - \frac{Cov_p(\pi_i, y_i)}{1 - E_p(\pi_i)} \end{aligned}$$

Lemma 5.

$$E_p(y_i) - E_{\bar{r}}(y_i) = -\frac{Cov_r(\phi_i w_i, y_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} + \frac{E_r(\phi_i w_i y_i)E_r(\phi_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} \quad (18)$$

where $Cov_r(\phi_i w_i, y_i) = E_r(\phi_i w_i y_i) - E_r(y_i)E_r(\phi_i w_i)$.

Proof:

According to Lemma 1, and (11), we get:

$$\begin{aligned}
 E_p(y_i) - E_r(y_i) &= \frac{E_s(w_i y_i)}{E_s(w_i)} - \frac{E_s\{(1 - \psi_i)y_i\}}{E_s\{(1 - \psi_i)\}} = \frac{E_r(\phi_i w_i y_i)}{E_r(\phi_i w_i)} - \frac{E_r\{(\phi_i - 1)y_i\}}{E_r(\phi_i - 1)} \\
 &= \frac{E_r(\phi_i w_i y_i)[E_r(\phi_i) - 1] - [E_r(\phi_i y_i) - E_r(y_i)]E_r(\phi_i w_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} \\
 &= \frac{E_r(\phi_i w_i y_i)E_r(\phi_i) - E_r(\phi_i w_i y_i) - E_r(\phi_i y_i)E_r(\phi_i w_i) + E_r(y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} \\
 &= \frac{-\{E_r(\phi_i w_i y_i) - E_r(\phi_i w_i)E_r(y_i)\}}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} + \frac{E_r(\phi_i w_i y_i)E_r(\phi_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} \\
 &= -\frac{Cov_r(\phi_i w_i, y_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}} + \frac{E_r(\phi_i w_i y_i)E_r(\phi_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)E_r\{(\phi_i - 1)\}}
 \end{aligned}$$

Lemma 6.

$$E_p(y_i) - E_s(y_i) = -\frac{E_r(\phi_i)E_r(\phi_i w_i y_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)[E_r(\phi_i w_i) - E_r(\phi_i)]} \quad (19)$$

Proof:

Using Lemmas 1 and 2, we get:

$$\begin{aligned}
 E_p(y_i) - E_s(y_i) &= \frac{E_s(w_i y_i)}{E_s(w_i)} - \frac{E_s\{(w_i - 1)y_i\}}{E_s\{(w_i - 1)\}} \\
 &= \frac{E_s(w_i y_i)[E_s(w_i) - 1] - [E_s(w_i y_i) - E_s(y_i)]E_s(w_i)}{E_r(w_i)E_r\{(w_i - 1)\}} \\
 &= \frac{E_s(w_i y_i)E_s(w_i) - E_s(w_i y_i) - E_s(w_i y_i)E_s(w_i) + E_s(y_i)E_s(w_i)}{E_r(\phi_i)E_r\{(\phi_i - 1)\}} \\
 &= \frac{-\{E_s(w_i y_i) - E_s(y_i)E_s(w_i)\}}{E_s(w_i)E_s\{(w_i - 1)\}} = -\frac{Cov_s(w_i, y_i)}{E_s(w_i)E_s\{(w_i - 1)\}} \\
 &= -\frac{E_r(\phi_i)E_r(\phi_i w_i y_i) - E_r(\phi_i y_i)E_r(\phi_i w_i)}{E_r(\phi_i w_i)[E_r(\phi_i w_i) - E_r(\phi_i)]}
 \end{aligned}$$

4. Method of Moments Estimators of Finite Population Mean

In this section we consider the estimation of finite population total $\bar{Y}_U = N^{-1} \sum_{i=1}^N y_i$, under the four scenarios mentioned in Table 2, namely: IN, II, NI, and NN. Also the main purpose of this section is the computation of the biases and mean square errors of these estimators.

Case 1: Informative sampling design and nonignorable nonresponse (IN).

According to Eideh (2009), we can show that the method of moments estimate (MME) of the finite population total, $\bar{Y}_U = N^{-1} \sum_{i \in U} y_i$ is given by

$$\bar{y}_{\phi w} = \frac{\sum_{i \in r} \phi_i w_i y_i}{\sum_{i \in r} \phi_i w_i} \quad (20)$$

which is the two-phase nonresponse adjusted estimator, see Sarndal and Lundstrom (2005, p 51).

Lemma 7. Statistical properties of $\bar{y}_{\phi w}$.

(a) $B(\bar{y}_{\phi w}) = E(\bar{y}_{\phi w}) - \bar{Y}_U \approx 0$

(b) If N is known,

$$V(\bar{y}_{\phi w}) = \frac{1}{t_2^2} \sum_{i \in r} \sum_{j \in r} \frac{\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j}{\pi_{ij} \psi_{ij}} (y_i - (\bar{y}_{\phi w})) (y_j - (\bar{y}_{\phi w})) \phi_i w_i \phi_j w_j$$

where $t_2 = N$.

If N is unknown

$$V(\bar{y}_{\phi w}) = \frac{1}{\left(\sum_{i \in r} \phi_i w_i \right)^2} \sum_{i \in r} \sum_{j \in r} \frac{\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j}{\pi_{ij} \psi_{ij}} (y_i - (\bar{y}_{\phi w})) (y_j - (\bar{y}_{\phi w})) \phi_i w_i \phi_j w_j$$

where $\psi_{ij} = \psi_i \psi_j$ for all units $i \neq j$ and $\psi_{ij} = \psi_i$ for $i = j$.

Proof:

Let $\hat{t}_1 = \sum_{i \in r} \phi_i w_i y_i$ and $\hat{t}_2 = \sum_{i \in r} \phi_i w_i$. Note that $\bar{y}_{\phi w} = \hat{t}_1 / \hat{t}_2$ is a ratio estimator of \bar{Y}_U .

Now,

$$\begin{aligned} E(\hat{t}_1) &= E\left(\sum_{i \in r} \phi_i w_i y_i\right) = E\left(\sum_{i \in U} \phi_i w_i y_i R_i\right) \\ &= \sum_{i \in U} \phi_i w_i y_i \pi_i \psi_i = \sum_{i \in U} y_i = Y_U = t_1 \end{aligned}$$

That is, $\hat{t}_1 = \sum_{i \in r} \phi_i w_i y_i$ is an unbiased estimator of t_1 .

Similarly,

$$\begin{aligned} E(\hat{t}_2) &= E\left(\sum_{i \in r} \phi_i w_i\right) = E\left(\sum_{i \in U} \phi_i w_i R_i\right) \\ &= \sum_{i \in U} \phi_i w_i \pi_i \psi_i = \sum_{i \in U} 1 = N = t_2 \end{aligned}$$

That is, $\hat{t}_2 = \sum_{i \in r} \phi_i w_i$ is an unbiased estimator of t_2 .

Expand $\bar{y}_{\phi_w} = \hat{t}_1 / \hat{t}_2$ in a Taylor series, around $\hat{t}_1 = t_1$ and $\hat{t}_2 = t_2$, we have

$$\begin{aligned}\bar{y}_{\phi_w} &= \frac{\sum_{i \in r} \phi_i w_i y_i}{\sum_{i \in r} \phi_i w_i} = \frac{\hat{t}_1}{\hat{t}_2} \\ &\approx \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2)\end{aligned}$$

So that $E(\bar{y}_{\phi_w}) \approx t_1 / t_2$. Hence $B(\bar{y}_{\phi_w}) = E(\bar{y}_{\phi_w}) - \bar{Y}_U \approx 0$.

(b) Since $B(\bar{y}_{\phi_w}) = E(\bar{y}_{\phi_w}) - \bar{Y}_U \approx 0$, therefore $MSE(\bar{y}_{\phi_w}) = V(\bar{y}_{\phi_w}) + B^2(\bar{y}_{\phi_w}) \approx V(\bar{y}_{\phi_w})$. Note that, \bar{y}_{ϕ_w} can be written as:

$$\begin{aligned}\bar{y}_{\phi_w} - \frac{t_1}{t_2} &= \bar{y}_{\phi_w} - \bar{Y}_U \approx \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2) \\ &= \frac{1}{t_2} \left(\hat{t}_1 - \frac{t_1}{t_2} \hat{t}_2 \right) \\ &= \frac{1}{t_2} (\hat{t}_1 - \bar{Y}_U \hat{t}_2) = \frac{1}{t_2} \left(\sum_{i \in U} \phi_i w_i y_i R_i - \bar{Y}_U \sum_{i \in U} \phi_i w_i R_i \right) \\ &= \frac{1}{t_2} \sum_{i \in U} (y_i - \bar{Y}_U) \phi_i w_i R_i\end{aligned}$$

So that,

$$\begin{aligned}V(\bar{y}_{\phi_w}) &= V(\bar{y}_{\phi_w} - \bar{Y}_U) = V\left(\frac{1}{t_2} \sum_{i \in U} (y_i - \bar{Y}_U) \phi_i w_i R_i\right) \\ &= \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}_U) (y_j - \bar{Y}_U) \phi_i w_i \phi_j w_j \text{Cov}(R_i, R_j)\end{aligned}$$

But,

$$\begin{aligned}\text{Cov}(R_i, R_j) &= E(R_i R_j) - E(R_i) E(R_j) \\ &= \pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j\end{aligned}$$

Hence,

$$V(\bar{y}_{\phi_w}) = \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j) (y_i - \bar{Y}_U) (y_j - \bar{Y}_U) \phi_i w_i \phi_j w_j$$

Estimation of $V(\bar{y}_{\phi_w})$

If N known,

$$\hat{V}(\bar{y}_{\phi_w}) = \frac{1}{N^2} \sum_{i \in r} \sum_{j \in r} \frac{\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j}{\pi_{ij} \psi_{ij}} (y_i - (\bar{y}_{\phi_w})) (y_j - (\bar{y}_{\phi_w})) \phi_i w_i \phi_j w_j \quad (21)$$

where $\psi_{ij} = \Pr(i, j \in r \mid \mathbf{x}, \mathbf{y}, \mathbf{z}) > 0$ for all units $i, j \in r$.

If N unknown,

$$\hat{V}(\bar{y}_{\phi_w}) = \frac{1}{\left(\sum_{i \in r} \phi_i w_i\right)^2} \sum_{i \in r} \sum_{j \in r} \frac{\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j}{\pi_{ij} \psi_{ij}} (y_i - (\bar{y}_{\phi_w})) (y_j - (\bar{y}_{\phi_w})) \phi_i w_i \phi_j w_j \quad (22)$$

Case 2: Informative sampling design and nonresponse mechanism is ignorable (II).

The MME of \bar{Y}_U becomes,

$$\bar{y}_w = \frac{\sum_{i \in r} w_i y_i}{\sum_{i \in r} w_i} \quad n \quad (23)$$

which is similar to the estimator given by Sarndal (1980) and discussed in details by Bethlehem (1988).

Lemma 8. Statistical properties of \bar{y}_w (Bethlehem (1988)).

$$(a) B(\bar{y}_w) \approx \frac{C(\psi, y)}{\bar{\psi}}$$

$$(b) V(\bar{y}_w) = \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) w_i w_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j), \sum_{i \in U} \psi_i = N \bar{\psi} = t_2$$

Proof:

I did not see the proof anywhere, I decided to show the reader the proof.

(a) Let $\hat{t}_1 = \sum_{i \in r} w_i y_i$ and $\hat{t}_2 = \sum_{i \in r} w_i$, so that $\bar{y}_w = \hat{t}_1 / \hat{t}_2$ is a ratio estimator of \bar{Y}_U . Also,

$$\begin{aligned} E(\hat{t}_1) &= E\left(\sum_{i \in r} w_i y_i\right) = E\left(\sum_{i \in U} w_i y_i R_i\right) \\ &= \sum_{i \in U} w_i y_i \pi_i \psi_i = \sum_{i \in U} \psi_i y_i = t_1 \end{aligned}$$

and

$$\begin{aligned} E(\hat{t}_2) &= E\left(\sum_{i \in r} w_i\right) = E\left(\sum_{i \in U} w_i R_i\right) \\ &= \sum_{i \in U} \psi_i = N \bar{\psi} = t_2 \end{aligned}$$

Using Taylor series, expand $\bar{y}_w = \hat{t}_1 / \hat{t}_2$ around $\hat{t}_1 = t_1$ and $\hat{t}_2 = t_2$, we get:

$$\bar{y}_w = \frac{\sum_{i \in r} w_i y_i}{\sum_{i \in r} w_i} = \frac{\hat{t}_1}{\hat{t}_2} \approx \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2)$$

So that,

$$E(\bar{y}_w) \approx \frac{t_1}{t_2} = \frac{\sum_{i \in U} \psi_i y_i}{N \bar{\psi}} = \bar{Y}^*$$

Now,

$$\begin{aligned} B(\bar{y}_w) &\approx E(\bar{y}_w) - \bar{Y}_U = \frac{\sum_{i \in U} \psi_i y_i}{N \bar{\psi}} - \frac{1}{N} \sum_{i \in U} y_i \\ &= \frac{1}{N} \left(\frac{\sum_{i \in U} \psi_i y_i}{\bar{\psi}} - \sum_{i \in U} y_i \right) = \frac{1}{N \bar{\psi}} \left(\sum_{i \in U} \psi_i y_i - \bar{\psi} \sum_{i \in U} y_i \right) \\ &= \frac{1}{N \bar{\psi}} \left(\sum_{i \in U} \psi_i y_i - N \bar{\psi} \bar{Y}_U \right) = \frac{C(\psi, y)}{\bar{\psi}} \end{aligned}$$

Note that, if the nonresponse mechanism is ignorable, that is the population covariance between the study variable and response probability is zero, $C(\psi, y) = 0$, then $B(\bar{y}_w) = 0$, and \bar{y}_w is an unbiased estimator of \bar{Y}_U . So, in order to reduce the bias, we can apply poststratification estimation, based on the estimated response probabilities $\hat{\psi}_i$ for all $i \in s$.

(b) **Computation of $V(\bar{y}_w)$** : Note that

$$\begin{aligned} \bar{y}_w &= \frac{\sum_{i \in r} w_i y_i}{\sum_{i \in r} w_i} = \frac{\hat{t}_1}{\hat{t}_2} \approx \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2) \\ \bar{y}_w &\approx -\frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2) = \frac{1}{t_2} \left(\hat{t}_1 - \frac{t_1}{t_2} \hat{t}_2 \right) \\ &= \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - \bar{Y}^* \hat{t}_2) = \frac{1}{t_2} \left(\sum_{i \in U} w_i y_i R_i - \bar{Y}^* \sum_{i \in U} w_i R_i \right) \\ &= \bar{Y}^* + \frac{1}{t_2} \sum_{i \in U} (y_i - \bar{Y}^*) w_i R_i \end{aligned}$$

So that,

$$\begin{aligned} V(\bar{y}_w) &= V(\bar{y}_w) = V \left(\bar{Y}^* + \frac{1}{t_2} \sum_{i \in U} (y_i - \bar{Y}^*) w_i R_i \right) = \frac{1}{t_2^2} V \left(\sum_{i \in U} (y_i - \bar{Y}^*) w_i R_i \right) \\ &= \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) w_i w_j \text{Cov}(R_i, R_j) \\ &= \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) w_i w_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j) \end{aligned}$$

Estimation of $MSE(\bar{y}_w)$

First need estimation of $V(\bar{y}_w)$:

$$\hat{V}(\bar{y}_w) = \frac{1}{\left(\sum_{i \in r} w_i \right)^2} \sum_{i \in r} \sum_{j \in r} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) w_i w_j \frac{(\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j)}{\pi_{ij} \psi_{ij}} \quad (24)$$

Estimation of $B(\bar{y}_w)$

$$\hat{B}(\bar{y}_w) \approx \frac{\hat{C}(\psi, y)}{\hat{\bar{\psi}}} \quad (25)$$

where

$$\hat{\bar{\psi}} = \bar{\psi}_w = \frac{\sum_{i \in r} w_i \psi_i}{\sum_{i \in r} w_i} \quad (26)$$

and

$$\hat{C}(\psi, y) = \frac{1}{\sum_{i \in r} w_i} \sum_{i \in r} w_i (y_i - \bar{y}_w)(\psi_i - \bar{\psi}_w) \quad (27)$$

Case 3: Noninformative sampling design and nonignorable nonresponse mechanism

We can show that the MME of \bar{Y}_U is,

$$\bar{y}_\phi = \frac{\sum_{i \in r} \phi_i y_i}{\sum_{i \in r} \phi_i} \quad (28)$$

Lemma 9. Statistical properties of \bar{y}_ϕ

- (a) $B(\bar{y}_\phi) \approx E(\bar{y}_\phi) - \bar{Y}_U = \frac{C(y, \pi)}{\bar{\pi}}$
 (b) $V(\bar{y}_\phi) = \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*)(y_j - \bar{Y}^*) \phi_i \phi_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j),$

where $\sum_{i \in U} \pi_i = N\bar{\pi} = t_2$.

Proof:

- (a) Let $\hat{t}_1 = \sum_{i \in r} \phi_i y_i$ and $\hat{t}_2 = \sum_{i \in r} \phi_i$, so that $\bar{y}_\phi = \hat{t}_1 / \hat{t}_2$ is a ratio estimator of \bar{Y}_U . Also,

$$\begin{aligned} E(\hat{t}_1) &= E\left(\sum_{i \in r} \phi_i y_i\right) = E\left(\sum_{i \in U} \phi_i y_i R_i\right) \\ &= \sum_{i \in U} \phi_i y_i \pi_i \psi_i = \sum_{i \in U} \pi_i y_i = t_1 \end{aligned}$$

and

$$\begin{aligned} E(\hat{t}_2) &= E\left(\sum_{i \in r} \phi_i\right) = E\left(\sum_{i \in U} \phi_i R_i\right) \\ &= \sum_{i \in U} \pi_i = N\bar{\pi} = t_2 \end{aligned}$$

Using Taylor series, expand $\bar{y}_\phi = \hat{t}_1/\hat{t}_2$ around $\hat{t}_1 = t_1$ and $\hat{t}_2 = t_2$, we get:

$$\bar{y}_\phi = \frac{\sum_{i \in r} \phi_i y_i}{\sum_{i \in r} \phi_i} = \frac{\hat{t}_1}{\hat{t}_2} \approx \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2)$$

So that

$$E(\bar{y}_\phi) \approx \frac{t_1}{t_2} = \frac{\sum_{i \in U} \pi_i y_i}{N\bar{\pi}} = \bar{Y}^*$$

Hence,

$$\begin{aligned} B(\bar{y}_\phi) &\approx E(\bar{y}_\phi) - \bar{Y}_U = \frac{\sum_{i \in U} \pi_i y_i}{N\bar{\pi}} - \frac{1}{N} \sum_{i \in U} y_i \\ &= \frac{1}{N} \left(\frac{\sum_{i \in U} \pi_i y_i}{\bar{\pi}} - \sum_{i \in U} y_i \right) = \frac{1}{N\bar{\pi}} \left(\sum_{i \in U} \pi_i y_i - \bar{\pi} \sum_{i \in U} y_i \right) \\ &= \frac{1}{N\bar{\pi}} \left(\sum_{i \in U} \pi_i y_i - N\bar{\pi} \bar{Y}_U \right) = \frac{C(y, \pi)}{\bar{\pi}} \end{aligned}$$

Note that, if the sampling design is noninformative, that is the population covariance between the study variable and first order inclusion probability is zero, $C(y, \pi)$, then $B(\bar{y}_\phi) = 0$, and \bar{y}_ϕ is an unbiased estimator of \bar{Y}_U . So, in order to reduce the bias, we can apply poststratification estimation, based on the inclusion probabilities π_i for all $i \in s$.

(b) Computation of $V(\bar{y}_\phi)$

We can write \bar{y}_ϕ as follows:

$$\begin{aligned} \bar{y}_\phi &= \frac{\sum_{i \in r} \phi_i y_i}{\sum_{i \in r} \phi_i} = \frac{\hat{t}_1}{\hat{t}_2} \approx \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2) \\ \bar{y}_\phi &\approx \frac{t_1}{t_2} + \frac{1}{t_2} (\hat{t}_1 - t_1) - \frac{t_1}{t_2^2} (\hat{t}_2 - t_2) = \frac{1}{t_2} \left(\hat{t}_1 - \frac{t_1}{t_2} \hat{t}_2 \right) \\ &= \bar{Y}^* + \frac{1}{t_2} (\hat{t}_1 - \bar{Y}^* \hat{t}_2) = \bar{Y}^* + \frac{1}{t_2} \left(\sum_{i \in U} \phi_i y_i R_i - \bar{Y}^* \sum_{i \in U} \phi_i R_i \right) \\ &= \bar{Y}^* + \frac{1}{t_2} \sum_{i \in U} (y_i - \bar{Y}^*) \phi_i R_i \end{aligned}$$

So that,

$$\begin{aligned} V(\bar{y}_\phi) &= V\left(\bar{Y}^* + \frac{1}{t_2} \sum_{i \in U} (y_i - \bar{Y}^*) \phi_i R_i\right) \\ &= \frac{1}{t_2^2} V\left(\sum_{i \in U} (y_i - \bar{Y}^*) \phi_i R_i\right) \\ &= \frac{1}{t_2^2} \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) \phi_i \phi_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j) \end{aligned}$$

Estimation of $V(\bar{y}_\phi)$

$$\hat{V}(\bar{y}_\phi) = \frac{1}{\left(\sum_{i \in r} \phi_i\right)^2} \sum_{i \in r} \sum_{j \in r} \frac{\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j}{\pi_{ij} \psi_{ij}} (y_i - \bar{y}_\phi) (y_j - \bar{y}_\phi) \phi_i \phi_j \quad (29)$$

Estimation of $B(\bar{y}_\phi)$

$$\hat{B}(\bar{y}_\phi) \approx \frac{\hat{C}(\pi, y)}{\hat{\pi}} \quad (30)$$

where

$$\hat{\pi} = \bar{\pi}_\phi = \frac{\sum_{i \in r} \phi_i \pi_i}{\sum_{i \in r} \phi_i} \quad (31)$$

and

$$\hat{C}(\pi, y) = \frac{1}{\sum_{i \in r} \phi_i} \sum_{i \in r} \phi_i (y_i - \bar{y}_\phi) (\pi_i - \bar{\pi}_\phi) \quad (32)$$

Case 4: Sampling design is noninformative and nonresponse mechanism is ignorable

Here, we can show that the MME of \bar{Y}_U is given by:

$$\bar{y}_r = \frac{\sum_{i \in r} y_i}{\sum_{i \in r} 1} = \frac{\sum_{i \in r} y_i}{m} \quad (33)$$

Lemma 10. Statistical properties of \bar{y}_r

$$(a) B(\bar{y}_r) = \frac{C(y, \pi \psi)}{(\pi \psi)}$$

$$(b) V(\bar{y}_r) = \frac{1}{m^2} \sum_{i \in U} \sum_{j \in U} y_i y_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j)$$

Proof:

(a) Since

$$E(\bar{y}_r) = E\left(\frac{\sum_{i \in r} y_i}{m}\right) = \frac{1}{m} E\left(\sum_{i \in U} y_i R_i\right) = \frac{1}{m} \sum_{i \in U} y_i \pi_i \psi_i = \frac{1}{m} Y_{\pi\psi}$$

Therefore,

$$B(\bar{y}_r) = E(\bar{y}_r) - \bar{Y}_U = \frac{1}{m} \sum_{i \in U} y_i \pi_i \psi_i - \frac{1}{N} \sum_{i \in U} y_i$$

Assume $m = E(m) = \sum_{i \in U} \pi_i \psi_i = N(\overline{\pi\psi})$, then

$$\begin{aligned} B(\bar{y}_r) &= \frac{1}{N(\overline{\pi\psi})} \sum_{i \in U} y_i \pi_i \psi_i - \frac{1}{N} \sum_{i \in U} y_i = \frac{1}{N(\overline{\pi\psi})} \left(\sum_{i \in U} y_i \pi_i \psi_i - (\overline{\pi\psi}) \sum_{i \in U} y_i \right) \\ &= \frac{1}{N(\overline{\pi\psi})} \left(\sum_{i \in U} y_i \pi_i \psi_i - N(\overline{\pi\psi}) \bar{Y}_U \right) \\ &= \frac{1}{(\overline{\pi\psi})} \frac{1}{N} \sum_{i \in U} (y_i - \bar{Y}_U) (\pi_i \psi_i - \overline{\pi\psi}) = \frac{C(y, \pi\psi)}{(\overline{\pi\psi})} \end{aligned}$$

Note that, if the sampling design is noninformative and nonresponse mechanism is ignorable, that is the population covariance between the study variable and inclusion probability is zero, $C(y, \pi\psi) = 0$, then $B(\bar{y}_r) = 0$, and \bar{y}_r is an unbiased estimator of \bar{Y}_U . So, in order to reduce the bias, we can apply poststratification estimation, based on the product of inclusion probabilities π_i and estimated response probabilities $\hat{\psi}_i$, $(\pi_i \hat{\psi}_i)$ for all $i \in s$.

(b) **Computation of** $V(\bar{y}_r)$

$$\begin{aligned} V(\bar{y}_r) &= \frac{1}{m^2} V\left(\sum_{i \in r} y_i\right) = \frac{1}{m^2} V\left(\sum_{i \in r} y_i R_i\right) \\ &= \frac{1}{m^2} V\left(\sum_{i \in r} y_i R_i\right) = \frac{1}{m^2} \sum_{i \in U} \sum_{j \in U} y_i y_j \text{Cov}(R_i, R_j) \\ &= \frac{1}{m^2} \sum_{i \in U} \sum_{j \in U} y_i y_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j) \end{aligned}$$

Estimation of $V(\bar{y}_r)$

$$\hat{V}(\bar{y}_r) = \frac{1}{m^2} \sum_{i \in r} \sum_{j \in r} y_i y_j \frac{(\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j)}{\pi_{ij} \psi_{ij}} \quad (34)$$

Estimation of $B(\bar{y}_r)$

$$\hat{B}(\bar{y}_r) = \frac{\hat{C}(y, \pi\psi)}{(\pi\psi)} = \frac{1}{\sum_{i \in r} \pi_i \psi_i} \sum_{i \in r} (y_i - \bar{y}_r) (\pi_i \psi_i - (\pi\psi)_r) \quad (35)$$

where $m(\pi\psi)_r = \sum_{i \in r} \pi_i \psi_i$.

The four cases can be summarized in Table 3.

Table 3: Method of Moments Estimators of \bar{Y}_U Bias and Variance, $MSE = VAR + (BIAS)^2$

Sampling Design and Nonresponse Mechanism	Estimator	Bias (\approx)	Variance
IN	$\bar{y}_{\phi w} = \frac{\sum_{i \in r} \phi_i w_i y_i}{\sum_{i \in r} \phi_i w_i}$	0	$\frac{1}{N^2} \left\{ \sum_{i \in r} \sum_{j \in r} \frac{\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j}{\pi_i \psi_i \pi_j \psi_j} \times \right.$ $\left. (y_i - (\bar{y}_{\phi w})) (y_j - (\bar{y}_{\phi w})) \phi_i w_i \phi_j w_j \right\}$
II	$\bar{y}_w = \frac{\sum_{i \in r} w_i y_i}{\sum_{i \in r} w_i}$	$\frac{C(\psi, y)}{\bar{\psi}}$	$\frac{1}{t_2^2} \left\{ \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) w_i w_j \times \right.$ $\left. \frac{(\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j)}{\pi_i \psi_i \pi_j \psi_j} \right\}$ $\sum_{i \in U} \psi_i = N \bar{\psi} = t_2$
NN	$\bar{y}_{\phi} = \frac{\sum_{i \in r} \phi_i y_i}{\sum_{i \in r} \phi_i}$	$\frac{C(y, \pi)}{\bar{\pi}}$	$\frac{1}{t_2^2} \left\{ \sum_{i \in U} \sum_{j \in U} (y_i - \bar{Y}^*) (y_j - \bar{Y}^*) \phi_i \phi_j \times \right.$ $\left. \frac{(\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j)}{\pi_i \psi_i \pi_j \psi_j} \right\}$ $\sum_{i \in U} \pi_i = N \bar{\pi} = t_2$
NI	$\bar{y}_r = \frac{\sum_{i \in r} y_i}{m}$	$\frac{C(y, \pi\psi)}{(\pi\psi)}$	$\frac{1}{m^2} \sum_{i \in U} \sum_{j \in U} y_i y_j (\pi_{ij} \psi_{ij} - \pi_i \psi_i \pi_j \psi_j)$

An interesting feature of the theses results is that several classical estimators in common use, within randomization theory (design-based school) of survey sampling, are shown to be special cases of the proposed approach, thus providing them a new justification.

5. Estimation under Informative Sampling and Nonignorable Nonresponse Mechanism

One of the main advantages of basing the inference on the response distribution is that it permits the use of standard inference procedures like those based on the likelihood principle. Having derived the response distribution when the sampling design is informative and the nonresponse mechanism in nonignorable (NMAR) and if the response measurements are independent, then the response likelihood for θ (the parameter indexing the superpopulation model), η (the parameter indexing the sampling design) and γ (the parameter indexing the nonresponse mechanism), is given by:

$$\begin{aligned} L_{r.in}(\theta, \eta, \gamma) &= \prod_{i=1}^m f_r(y_i | \mathbf{x}_i, \theta, \eta, \gamma) \\ &= \prod_{i=1}^m \frac{E_s(\psi_i | \mathbf{x}_i, y_i, \gamma) f_s(y_i | \mathbf{x}_i, \theta, \eta)}{E_s(\psi_i | \mathbf{x}_i, \theta, \eta, \gamma)} \\ &= \prod_{i=1}^m \frac{E_s(\psi_i | \mathbf{x}_i, y_i, \gamma) E_p(\pi_i | \mathbf{x}_i, y_i, \eta) f_p(y_i | \mathbf{x}_i, \theta)}{E_s(\psi_i | \mathbf{x}_i, \theta, \eta, \gamma) E_p(\pi_i | \mathbf{x}_i, \theta, \eta)} \end{aligned} \quad (36)$$

and the logarithm of the response likelihood for (θ, η, γ) is:

$$\begin{aligned} l_{r.in}(\theta, \eta, \gamma) &= \sum_{i=1}^m \log f_r(y_i | \mathbf{x}_i, \theta, \eta, \gamma) \\ &= \sum_{i=1}^m \log f_p(y_i | \mathbf{x}_i, \theta) + \sum_{i=1}^m \log E_p(\pi_i | \mathbf{x}_i, y_i, \eta) - \sum_{i=1}^m \log E_p(\pi_i | \mathbf{x}_i, \theta, \eta) \\ &\quad + \sum_{i=1}^m \log E_s(\psi_i | \mathbf{x}_i, y_i, \gamma) - \sum_{i=1}^m \log E_s(\psi_i | \mathbf{x}_i, \theta, \eta, \gamma) \end{aligned} \quad (37)$$

The function given in equation (37) can be maximized with respect to (θ, η, γ) to obtain the maximum response likelihood estimates of these parameters. Maximum response likelihood estimators of other parameters, which are the parameters of interest, (e.g. the parameter θ characterizing the population distribution of y) are defined using the invariance properties of the maximum likelihood (ML) approach.

The response likelihood function, $L_{r.in}(\theta, \eta, \gamma)$, can be interpreted as a weighted likelihood, where the weight is the product of the two ratios, the first one is $E_p(\pi_i | \mathbf{x}_i, y_i, \eta) / E_p(\pi_i | \mathbf{x}_i, \theta, \eta)$, which characterize the sampling design, and the second ratio is $E_s(\psi_i | \mathbf{x}_i, y_i, \gamma) / E_s(\psi_i | \mathbf{x}_i, \theta, \eta, \gamma)$, that characterise the missing data mechanism.

It should emphasize here that,

- (a) If $E_p(\pi_i | \mathbf{x}_i, y_i) = E_p(\pi_i | \mathbf{x}_i)$ for all values of y_i , then the sampling design is noninformative.
- (b) If $E_s(\psi_i | \mathbf{x}_i, y_i) = E_s(\psi_i | \mathbf{x}_i)$ for all values of y_i , then the nonresponse process is ignorable.

Particular Cases:

Case 1: Sampling design is noninformative and nonresponse process is nonignorable. Then (40) becomes:

$$l_{r.nn}(\theta, \gamma) = \sum_{i=1}^m \log f_p(y_i | \mathbf{x}_i, \theta) + \sum_{i=1}^m \log E_s(\psi_i | \mathbf{x}_i, y_i, \gamma) - \sum_{i=1}^m \log E_s(\psi_i | \mathbf{x}_i, \theta, \gamma) \quad (38)$$

Case 2: Sampling design is noninformative and nonresponse process is ignorable. Then (37) becomes:

$$l_{r.ni}(\theta) = \sum_{i=1}^m \log f_p(y_i | \mathbf{x}_i, \theta) \quad (38)$$

which is the standard estimation processes, where the missing value mechanism and process of sampling design are ignored and base the inference on the classical log-likelihood function. However, analysis using standard estimation methods, which ignores the last four terms of (22), leads to inconsistent estimates of θ . Thus the effect of the nonignorable missing value mechanism and informative sampling design must be taken into account.

Case 3: Sampling design is informative and nonresponse process is ignorable. Then (37) becomes:

$$l_{r.ii}(\theta, \eta) = \sum_{i=1}^m \log f_p(y_i | \mathbf{x}_i, \theta) + \sum_{i=1}^m \log E_p(\pi_i | \mathbf{x}_i, y_i, \eta) - \sum_{i=1}^m \log E_p(\pi_i | \mathbf{x}_i, \theta, \eta) \quad (40)$$

Now, assuming $f_p(y_i | \mathbf{x}_i, \theta)$, $E_p(\pi_i | \mathbf{x}_i, y_i, \eta)$ and $E_s(\psi_i | \mathbf{x}_i, y_i, \gamma)$ are completely specified, then the maximum likelihood (ML) estimator of (θ, η, γ) can be obtained by maximizing the log likelihood function given in (40) with respect to (θ, η, γ) simultaneously, or in four-step method. For modeling of $E_p(\pi_i | \mathbf{x}_i, y_i, \eta)$, Pfeffermann et al. (1998) introduced exponential and polynomial function of (\mathbf{x}_i, y_i) , later Eideh (2003) considered logit and probit functions. Furthermore, Eideh (2012) adopted the exponential, linear, logit and probit functions for modeling $E_s(\psi_i | \mathbf{x}_i, y_i, \gamma)$.

In practice the response probabilities are theoretical quantities and they are unknown. For estimation of ψ_i , see Section 2.

Four steps method

Step 1: Estimation of ψ_i . See Section 2. Denote the estimate by $\hat{\psi}_i$, so that $\hat{\phi}_i = 1/\hat{\psi}_i$. We refer to $\hat{\psi}_i$ as the response propensity.

Step 2: Estimation of the effect of nonresponse mechanism. Estimate the parameter γ using the relationship given in (10), namely:

$$E_s(\psi_i | \mathbf{x}_i, y_i, \gamma) = \frac{1}{E_r(\phi_i | \mathbf{x}_i, y_i, \gamma)} \quad (41)$$

Thus the parameter γ can be estimated by regressing $\hat{\phi}_i$ on (\mathbf{x}_i, y_i) using the data set $\{\hat{\phi}_i, y_i, \mathbf{x}_i, i \in r\}$. Denoting the resulting estimate of γ by $\tilde{\gamma}$.

Step 3: Estimation of the effect of sampling design. Estimate the parameter η using the relationship given in (12), namely:

$$E_p(y_i | \mathbf{x}_i, \eta) = \frac{E_r(\phi_i w_i y_i | \mathbf{x}_i)}{E_r(\phi_i w_i | \mathbf{x}_i)} \quad (42)$$

Thus the parameter η can be estimated using regression analysis. This can be proceed as follows:

(a) Write (42) as:

$$E_p(y_i | \mathbf{x}_i, \eta) = \frac{E_r(\phi_i w_i y_i | \mathbf{x}_i)}{E_r(\phi_i w_i | \mathbf{x}_i)} = E_r(l_i y_i | \mathbf{x}_i) \quad (43)$$

where

$$l_i = \frac{\phi_i w_i}{E_r(\phi_i w_i | \mathbf{x}_i)} \quad (44)$$

(b) Estimate $E_r(\phi_i w_i | \mathbf{x}_i)$ by regressing $\hat{\phi}_i w_i$ on \mathbf{x}_i using the data set $\{\mathbf{x}_i, \hat{\phi}_i, w_i, i \in r\}$.

(c) Let $\hat{l}_i = \hat{\phi}_i w_i / \hat{E}_r(\phi_i w_i | \mathbf{x}_i)$, and then regress $\hat{l}_i y_i$ on \mathbf{x}_i using the response data set $\{\hat{l}_i y_i, \mathbf{x}_i, i \in r\}$. Denoting the resulting estimate of η by $\tilde{\eta}$.

Step 4: Estimation of the superpopulation model parameter. Substitute $\tilde{\gamma}$ and $\tilde{\eta}$ in the response log-likelihood function, (40), and since $E_p(\pi_i | \mathbf{x}_i, y_i, \tilde{\eta})$ and $E_s(\psi_i | \mathbf{x}_i, y_i, \tilde{\gamma})$ do not contain θ , then the ML estimator of θ is obtained by maximizing the resulting response log-likelihood function with respect to the population parameter θ , namely:

$$\tilde{l}_{r.in}(\theta) = \sum_{i=1}^m \log f_p(y_i | \mathbf{x}_i, \theta) - \sum_{i=1}^m \log E_p(\pi_i | \mathbf{x}_i, \theta, \tilde{\eta}) - \sum_{i=1}^m \log E_s(\psi_i | \mathbf{x}_i, \theta, \tilde{\eta}, \tilde{\gamma}) \quad (45)$$

In this paper we study the joint treatment of not missing at random response mechanism and informative sampling for survey data. This is the most general situation in surveys and other combinations of sampling informativeness and response mechanisms can be considered as special cases. The proposed method combines two methodologies used in the analysis of sample surveys for the treatment of informative sampling and the nonignorable nonresponse mechanism. One incorporates the dependence of the first order inclusion probabilities on the study variable, while the other incorporates the dependence of the probability of nonresponse on unobserved or missing observations. The main purpose here is the estimation of finite population mean and superpopulation parameters when the sampling design is informative and nonresponse mechanism is nonignorable. Under four scenarios of sampling design and nonresponse mechanism, we obtained the method of moment estimators of finite population mean, with their biases and mean

square errors. Furthermore, a four-step estimation method is introduced for the estimation of superpopulation parameters under informative sampling and nonignorable nonresponse mechanism. New relationships between moments of response, nonresponse, sample, sample-complement and population distributions were derived. Most estimators for finite population mean known from sampling surveys can be derived as a special case of the results derived in this paper. This paper can be considered as generalization and extension of Bethlehem paper (1988).

6. Conclusions

In this article we use two methodologies used in the analysis of sample surveys for the treatment of informative sampling and the nonignorable nonresponse mechanism. One incorporates the dependence of the first order inclusion probabilities on the study variable, while the other incorporates the dependence of the probability of nonresponse on unobserved or missing observations. Using the new relationships, derived in the present study, between moments of response, nonresponse, sample, sample-complement and population distributions, we develop four estimators of finite population mean under classification of sampling design and nonresponse mechanism. Known estimators in common use in official statistics are shown to be special cases of the present theory, so provide new justification of these estimators as method of moments estimators. Further experimentation (simulation and real data problem) with this kind of estimators and is therefore highly recommended.

Furthermore, in this paper, we show the role of informative sampling design and nonignorable nonresponse in adjusting various estimators for bias reduction. In addition to the estimation of finite population mean, we introduce a new method for the estimation of superpopulation parameters under informative sampling and nonignorable nonresponse mechanism.

In brief, ignoring informativeness of sampling design and nonignorable nonresponse, will yield biased estimators of finite population total. To reduce the bias, we propose the use of poststratification based on first order inclusion probabilities (in case of informative sampling design and ignorable nonresponse mechanism), or estimated response probabilities (for noninformative sampling design and nonignorable nonresponse mechanism), or product of them (if the sampling design is noninformative and the nonresponse mechanism is ignorable).

I hope that the new mathematical results obtained in the present article will encourage further theoretical, simulation, real data problem, empirical and practical research in these directions.

References

1. Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, pp 251-260.
2. Chang, T. and Kott, P.S. (2008). Using Calibration Weighting to Adjust for Nonresponse under a Plausible Model. *Biometrika*, 95, 555–571.

3. Chambers, R. and Skinner, C. (2003). *Analysis of survey data*. New York: John Wiley.
4. Cobben, F. (2009). *Nonresponse in Sample Survey – Method for Analysis and Adjustment*. Statistics Netherlands.
5. Eideh A.H. (2009). On the use of the sample distribution and sample likelihood for inference under informative probability sampling. *DIRASAT (Natural Science)*, Volume 36 (2009), Number 1, pp18-29.
6. Eideh, A. H. and Nathan, G. (2006). Fitting time series models for longitudinal survey data under informative sampling. *Journal of Statistical Planning and Inference* 136, 9, pp 3052-306. [Corrigendum, 137 (2007), p 628].
7. Eideh, A. H. and Nathan, G. (2009). Two-stage informative cluster sampling with application in small area estimation. *Journal of Statistical Planning and Inference*.139, pp 3088-3101.
8. Eideh A.H. (2012). Estimation and Prediction under Nonignorable Nonresponse via Response and Nonresponse Distributions. *Journal of Indian Society of Agriculture Statistics*. 66(3)2012, pp 359-380.
9. Little, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77, pp 237-250.
10. Little, R.J.A. (1993). Pattern-mixture Models for Multivariate Incomplete Data. *Journal of the American Statistical Association*, 88, pp 125–134.
11. Little, R.J.A. (1994). A Class of Pattern-mixture Models for Normal Incomplete Data. *Biometrika*, 81, 471–483.
12. Little, R.J.A. and Rubin, D.B. (2002). *Statistical analysis with missing data*. New York: Wiley.
13. Pfeiffermann, D., Krieger, A.M. and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling'. *Statistica Sinica*, 8, pp 1087-1114.
14. Pfeiffermann, D. and Sverchkov, M.(1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya*, 61, B, pp 166-186.
15. Pfeiffermann, D. and Sikov, A. (2011). Imputation and Estimation under Nonignorable Nonresponse in Household Surveys with Missing Covariate Information. *Journal of Official Statistics*, Vol. 27, No. 2, 2011, pp. 181–209.
16. Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya*, 61, B, pp 166-186.
17. Qin, J., Leung, D., and Shao, J. (2002). Estimation with Survey Data under Nonignorable Nonresponse or Informative Sampling. *Journal of the American Statistical Association*, 97, pp 193–200.
18. Matei, A. and Ranalli, M.G. (2015). Dealing with non-ignorable nonresponse in survey sampling: A latent modeling approach. *Survey Methodology*, 41, No. 1, pp. 145-164.
19. Rubin, D.B. (1976). Inference and Missing Data. *Biometrika*, 63, pp 581–590.

20. Särndal C.E. and Swensson B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *International Statistical Review*, 55, pp 279-294.
21. Särndal, C.E. and Lundstorm, S. (2005). *Estimation in surveys with nonresponse*. New York: John Wiley.
22. Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall.
23. Skinner, C.J., Holt, D., and Smith, T.M.F (eds.) (1989). *Analysis of complex surveys*, New York: Wiley.
24. Sverchkov, M., and Pfeffermann, D. (2004). Prediction of finite population totals based on the sample distribution. *Survey Methodology*, 30, pp 79-92.