The Rao, Hartley and Cochran Scheme with Dubious Random Non-Response in Survey Sampling

María del Mar Rueda[1], Antonio Arcos[1], Raghunath Arnab[2], Sarjinder Singh[3]

[1]Department of Statistics and O.R., University of Granada, Avd. Fuentenueva s/n. Granada. Spain **E-mail**: mrueda@ugr.es, arcos@ugr.es

[2]Department of Statistics, University of Botswana, Private Bag UB 00705, Gaborone, Botswana. **E-mail:** arnab@mopipi.ub.bw

[3]Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363, USA. **E-mail:** sarjinder@yahoo.com

Abstract

In the present investigation, we consider the problem of estimation of population total using the well known Rao, Hartley and Cochran (1962), say RHC scheme, in the presence of dubious random non-response. The proposed estimator has been compared with the usual estimators of the population total in the presence of random non-response. A new idea of "Dubious Random Non-response (DRN)" through transformations on the response probabilities has been introduced and studied.

Key Words: Rao, Hartley and Cochran's scheme, Estimation of population total, Random non-response.

1. Introduction

In the presence of random non-response, a huge amount of literature is available in the field of survey sampling as one can refer to Rubin (1976). To our knowledge, no one has paid attention to study the Rao, Hartley and Cochran (1962) scheme in the presence of random non-response that motivated the authors to think and study on these lines. Note that the Rao, Hartley and Cochran (1962) scheme has very good reputation and image among the survey statisticians from the last four-five decades, and nobody could challenge it by now because of its simplicity and practicability in real surveys. Before going further, let us first discuss the Rao, Hartley and Cochran (1962) scheme. Suppose a population consists of N units and we wish to draw a sample of n units. First of all, divide randomly the N units into n groups as shown below:

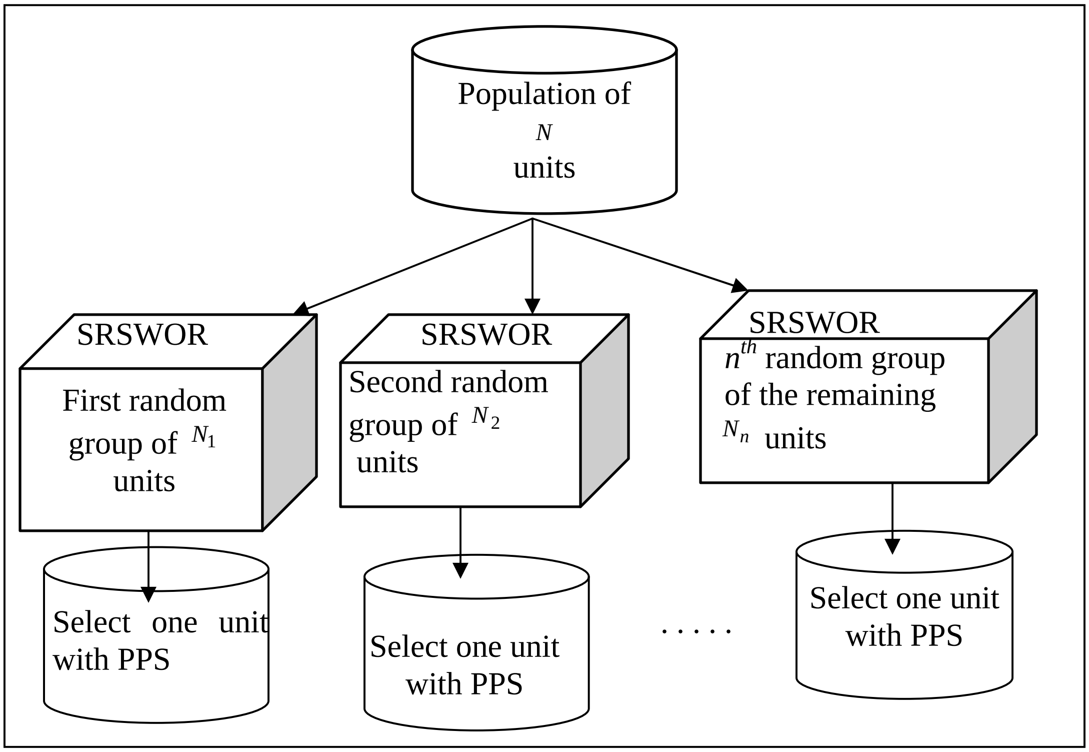

Figure 1. Pictorial representation of the Rao, Hartley and Cochran (1962).

First random group: Out of N units, select N_1 units by using SRSWOR sampling.

Second random group: Out of $(N - N_1)$ units, select N_2 units by SRSWOR sampling and so on such that

$$\sum_{i=1}^{n} N_i = N. \tag{1.1}$$

The allocation of units to different groups is done randomly and we select one unit from each of the n groups with probability proportional to size (PPS) and thus we obtain a sample of size n. Suppose $p_{1,}$ p_2, ..., p_N are the probabilities associated with the N units in the population and $\sum_{i=1}^{N} p_i = 1$. Further suppose that p_{ij} denotes the probability corresponding to the j^{th} unit in the i^{th} group, G_i, $\forall\ i = 1, 2, ..., n$. Thus the Rao, Hartley, and Cochran (1962) mechanism can be better understood from the following table, which gives the structure of population units after making random groups, as follows:

Structure of data in RHC-Sampling Strategy											
1st Group (G_1)		2nd Group (G_2)			i^{th} Group (G_i)			n^{th} Group (G_n)	
Value	Prob.	Value	Prob.		Value	Prob.		Value	Prob.
Y_{11}	p_{11}	Y_{21}	p_{21}		Y_{i1}	p_{i1}		Y_{n1}	p_{n1}
Y_{12}	p_{12}	Y_{22}	p_{22}		Y_{i2}	p_{i2}		Y_{n2}	p_{n2}
.	.	.	.		.	.		.	.
.	.	.	.		.	.		.	.
Y_{1N_1}	p_{1N_1}	Y_{2N_2}	p_{2N_2}		Y_{iN_i}	p_{iN_i}		Y_{nN_n}	p_{nN_n}
	τ_1		τ_2			τ_i			τ_n

where $\tau_i = \sum_{j \in G_i} p_{ij}$, $i = 1, 2, ..., n$,denotes the sum of selection probabilities of the i^{th} random group. Let s be a sample of size n selected using the RHC scheme $p(s)$.

Then, an unbiased estimator of population total Y is given by:

$$\hat{Y}_{\text{RHC}} = \sum_{i \in s} \frac{y_i}{p_i^*} . \tag{1.2}$$

with $p_i^* = p_i / \tau_i$ and the variance of the estimator $\hat{Y}_{\text{RHC}}$ is given by:

$$V\left(\hat{Y}_{\text{RHC}}\right) = \alpha \left[\sum_{j=1}^{N} \frac{Y_j^2}{p_j} - Y^2 \right]. \tag{1.3}$$

where

$$\alpha = \frac{\sum_{i \in s} N_i^2 - N}{N(N-1)} \tag{1.4}$$

An unbiased estimator of the variance $V\left(\hat{Y}_{\text{RHC}}\right)$ is given by:

$$\hat{v}\left(\hat{Y}_{\text{RHC}}\right) = \frac{\left(\sum_{i \in s} N_i^2 - N\right)}{\left(N^2 - \sum_{i \in s} N_i^2\right)} \left[\sum_{i \in s} \frac{y_i^2}{\left(p_i^2 / \tau_i\right)} - \hat{Y}_{\text{RHC}}^2 \right]. \tag{1.5}$$

In the next section, we consider a new situation when some of the respondents selected using the RHC scheme either fails to respond or are unavailable in a completely random way called missing completely at random (MCAR).

2. RHC with Dubious Random Non-Response

Consider that response on the study variable y_i is available only on the G_i, $i \in s_r$ random groups, while that is not available from the remaining G_i , $i = (s - s_r)$ random groups. Following Särndal (1992), let

$$\delta_i = \begin{cases} 1 & \text{if the ith unit responds} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

such that

$$E_r(\delta_i) = \Pr ob(\delta_i = 1) = \delta_i^* = \phi(x_i) \text{, say} \tag{2.2}$$

and

$$V_r(\delta_i) = \delta_i^*(1 - \delta_i^*) \tag{2.3}$$

where E_r and V_r denote the expected value over the response mechanism and δ_i is a Bernoulli variable with probability of success δ_i^*.

For example, we consider with the following transformations on the response probabilities $\phi(x_i)$ as:

$$\delta_{i0}^* = \phi(x_i) \tag{2.4}$$

$$\delta_{i1}^* = [\phi(x_i)]^{(1-r/n)} \tag{2.5}$$

$$\delta_{i2}^* = \left(1 - \frac{r}{n}\right)\phi(x_i) + \frac{r}{n} \tag{2.6}$$

$$\delta_{i3}^* = \left(1 + \frac{r}{n}\right)^{r/n} (1 + \phi(x_i))^{(1-r/n)} - 1 \tag{2.7}$$

$$\delta_{i4}^* = \frac{1}{\left(1 - \frac{r}{n}\right)\frac{1}{\phi(x_i)} + \frac{r}{n}} \tag{2.8}$$

and

$$\delta_{i5}^* = \frac{1}{\left(1 + \frac{r}{n}\right)^{r/n} \left(1 + \frac{1}{\phi(x_i)}\right)^{(1-r/n)} - 1} \tag{2.9}$$

Note that if $r \to n$ then $\delta_i^* \to 1$ and if $r \to 0$ then $\delta_i^* \to \phi(x_i)$.

Now we consider transformations on response probabilities in which a coefficient of judgment λ is being used to compromise between MAR and MCAR leading to new dubious non-response (DNR) cases as follows:

$$\delta_{i6}^{*} = \lambda\,\phi(x_i) + (1-\lambda)\frac{r}{n} \tag{2.10}$$

and

$$\delta_{i7}^{*} = \left(1+\frac{r}{n}\right)^{(1-\lambda)}\left(1+\phi(x_i)\right)^{\lambda} - 1 \tag{2.11}$$

Note that if $\lambda \to 1$ then $\delta_{i6}^{*} \to \phi(x_i)$ and $\delta_{i7}^{*} \to \phi(x_i)$; and if $\lambda \to 0$ then $\delta_{i6}^{*} \to r/n$ and $\delta_{i7}^{*} \to r/n$. A natural good guess of λ could be a known value of the positive population correlation coefficient ρ_{xy} between x and y, but a better choice of λ based on an investigator's judgment may differ from ρ_{xy} for survey to survey. Thus, an optimum value of λ may be investigated through simulation study.

Under such a response mechanism, we define a new estimator of the population total as:

$$\hat{Y}_{\text{RHC(DNR)}} = \sum_{i\in s}\left(\frac{y_i}{p_i^{*}}\right)\left(\frac{\delta_i}{\delta_i^{*}}\right) \tag{2.12}$$

Then we have the following theorems:

Theorem 2.1.The estimator $\hat{Y}_{\text{RHC(DNR)}}$ is an unbiased estimator of the population total Y .

Proof. Let E_p denote the expected value over the used sample selection design $p(s)$, then taking the expected value on both sides of (2.5), we have

$$E\left[\hat{Y}_{\text{RHC(DNR)}}\right] = E_p E_r\left[\sum_{i\in s}\left(\frac{y_i}{p_i^{*}}\right)\left(\frac{\delta_i}{\delta_i^{*}}\right)\right] = E_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^{*}}\right)\left(\frac{E_r(\delta_i)}{\delta_i^{*}}\right)\right]$$

$$= E_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^{*}}\right)\left(\frac{\delta_i^{*}}{\delta_i^{*}}\right)\right] == E_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^{*}}\right)\right] = Y$$

Hence the theorem.

Theorem 2.2.The variance of the estimator $\hat{Y}_{RHC(DNR)}$ is given by

$$V\left[\hat{Y}_{\mathrm{RHC(DNR)}}\right]=V\left(\hat{Y}_{RHC}\right)+\sum_{i=1}^{N}\frac{(1-\delta_i^*)}{\delta_i^*}y_i^2+\left[\frac{\sum\limits_{i\in s}N_i^2-N}{N^2}\right]\sum_{i=1}^{N}\frac{(1-\delta_i^*)(1-p_i)}{\delta_i^* p_i}y_i^2 \quad (2.13)$$

where $V\left(\hat{Y}_{RHC}\right)$ is same as given in (1.3).

Proof. Let V_p denote the variance operator over the sampling design $p(s)$, then we have

$$V\left[\hat{Y}_{RHC(DNR)}\right]=V_pE_r\left[\hat{Y}_{RHC(DNR)}\right]+E_pV_r\left[\hat{Y}_{RHC(DNR)}\right] \quad (2.14)$$

Now

$$V_pE_r\left[\hat{Y}_{RHC(DNR)}\right]=V_pE_r\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)\left(\frac{\delta_i}{\delta_i^*}\right)\right]=V_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)\left(\frac{E_r(\delta_i)}{\delta_i^*}\right)\right]$$

$$=V_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)\left(\frac{\delta_i^*}{\delta_i^*}\right)\right]=V_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)\right]$$

$$=\frac{\left(\sum\limits_{i\in s}N_i^2-N\right)}{N(N-1)}\left[\sum_{j=1}^{N}\frac{Y_j^2}{p_j}-Y^2\right]=V\left(\hat{Y}_{\mathrm{RHC}}\right). \quad (2.15)$$

Let E_G be the expected value over all possible random groups and G_i denote the ith random group, then

$$E_pV_r\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)\left(\frac{\delta_i}{\delta_i^*}\right)\right]=E_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)^2\left(\frac{V_r(\delta_i)}{(\delta_i^*)^2}\right)\right]=E_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)^2\left(\frac{\delta_i^*(1-\delta_i^*)}{(\delta_i^*)^2}\right)\right]$$

$$=E_p\left[\sum_{i\in s}\left(\frac{y_i}{p_i^*}\right)^2\left(\frac{(1-\delta_i^*)}{\delta_i^*}\right)\right]=E_G\left[\sum_{i\in s}E\left\{\frac{y_i^2(1-\delta_i^*)}{(p_i^*)^2\delta_i^*}\,|\,G_i\right\}\right]$$

$$=E_G\left[\sum_{j\in s}\sum_{j\in G_i}\left\{\frac{y_j^2(1-\delta_j^*)}{(p_j^*)\delta_j^*}\right\}\right]=E_G\left[\sum_{j\in s}\sum_{j\in G_i}\left\{\frac{y_j^2(1-\delta_j^*)}{p_j\delta_j^*}\sum_{k\in G_i}p_k\right\}\right]$$

$$=E_G\left[\sum_{j\in s}\sum_{j\in G_i}\frac{y_j^2(1-\delta_j^*)}{\delta_j^*}+\sum_{i\in s}\sum_{j\neq k\in G_i}\sum\frac{y_j^2(1-\delta_j^*)p_k}{p_j\delta_j^*}\right]$$

$$=\sum_{i=1}^{N}\frac{y_i^2(1-\delta_i^*)}{\delta_i^*}+E_G\left[\sum_{i\in s}\sum_{j\neq k\in G_i}\sum\frac{y_j^2(1-\delta_j^*)p_k}{p_j\delta_j^*}\right]$$

$$= \sum_{i=1}^{N} \frac{y_i^2(1-\delta_i^*)}{\delta_i^*} + \left[\frac{\sum_{i\in s} N_i^2 - N}{N^2}\right] \sum_{j\neq k=1}^{N} \sum^{N} \frac{y_j^2(1-\delta_j^*)p_k}{p_j \delta_j^*}$$

$$= \sum_{i=1}^{N} \frac{y_i^2(1-\delta_i^*)}{\delta_i^*} + \left[\frac{\sum_{i\in s} N_i^2 - N}{N^2}\right] \sum_{i=1}^{N} \frac{y_i^2(1-\delta_i^*)(1-p_i)}{p_i \delta_i^*} \quad (2.16)$$

Using (2.16) and (2.15) in (2.14), we have the theorem.

Theorem 2.3. An unbiased estimator of the variance of the estimator $\hat{Y}_{RHC(DNR)}$ is given by

$$\hat{v}\left(\hat{Y}_{\mathrm{RHC(DNR)}}\right) = \frac{\alpha}{1+\alpha}\left[\sum_{i\in s} \frac{y_i^2 \delta_i}{p_i^* \delta_i^*} - \left\{\hat{Y}_{\mathrm{RHC(DNR)}}\right\}^2\right] + \frac{1}{1+\alpha} \sum_{i\in s} \left(\frac{y_i}{p_i^*}\right)^2 \frac{(1-\delta_i^*)}{\delta_i^*} \quad (2.17)$$

Proof. It follows from the facts that:

(a) An unbiased estimator of $\sum_{i=1}^{N} \frac{y_i^2}{p_i}$ is $\sum_{i\in s} \frac{y_i^2 \delta_i}{p_i^* \delta_i^*}$

(b) An unbiased estimator of Y^2 is $\left[\hat{Y}_{RHC(DNR)}\right]^2 - \hat{v}\left[\hat{Y}_{RHC(DNR)}\right]$

Hence an unbiased estimator of $V\left(\hat{Y}_{RHC}\right)$ in the presence of doubtful non-response is given by

$$\hat{v}\left(\hat{Y}_{\mathrm{RHC}}\right) = \alpha\left[\sum_{i\in s} \frac{y_i^2 \delta_i}{p_i^* \delta_i^*} - \left\{\left(\hat{Y}_{\mathrm{RHC(DNR)}}\right)^2 - \hat{v}\left(\hat{Y}_{\mathrm{RHC(DNR)}}\right)\right\}\right]$$

An unbiased estimator of the second term on the right hand side of (2.13):

$$\sum_{i=1}^{N} \frac{(1-\delta_i^*)}{\delta_i^*} y_i^2 + \left[\frac{\sum_{i\in s} N_i^2 - N}{N^2}\right] \sum_{i=1}^{N} \frac{(1-\delta_i^*)(1-p_i)}{\delta_i^* p_i} y_i^2$$

is given by

$$\sum_{i\in s} \left(\frac{y_i}{p_i^*}\right)^2 \frac{V_r(\delta_i)}{(\delta_i^*)^2} = \sum_{i\in s} \left(\frac{y_i}{p_i^*}\right)^2 \frac{(1-\delta_i^*)}{\delta_i^*}$$

Hence the unbiased estimator of $V\left[\hat{Y}_{\mathrm{RHC(DNR)}}\right]$ is obtained by solving the equation

$$\hat{v}\left[\hat{Y}_{\mathrm{RHC(DNR)}}\right] = \alpha\left[\sum_{i\in s} \frac{y_i^2 \delta_i}{p_i^* \delta_i^*} - \left\{\left(\hat{Y}_{\mathrm{RHC(DNR)}}\right)^2 - \hat{v}\left(\hat{Y}_{\mathrm{RHC(DNR)}}\right)\right\}\right] + \sum_{i\in s} \left(\frac{y_i}{p_i^*}\right)^2 \left(\frac{1-\delta_i^*}{\delta_i^*}\right)$$

which proves the theorem.

3. Numerical Comparison of the Estimators

In this section, we present a comparison of the proposed estimator with the estimator

$$\hat{Y}_{\text{SRSWOR(MCAR)}} = \frac{N}{r}\sum_{i=1}^{r} y_i \tag{3.1}$$

whose variance will under MCAR is given by:

$$V\left(\hat{Y}_{\text{SRSWOR(MCAR)}}\right) = N^2\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 \tag{3.2}$$

We can also compare it with the Rao and Sitter (1995) ratio estimator in the presence of MCAR non-response defined as:

$$\hat{Y}_{\text{RS(MCAR)}} = N\,\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) \tag{3.3}$$

with variance given by

$$V\left(\hat{Y}_{\text{RS(MCAR)}}\right) = N^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(S_y^2 + R^2 S_x^2 - 2RS_{xy}\right)\right] \tag{3.4}$$

where $R = \bar{Y}/\bar{X}$.

For this comparison we use the data of a real population. The population considered (called Cancer) consists on N = 301 counties in North Carolina, South Carolina and Georgia with the white female population in 1960; this population was studied by Royall and Cumberland (1981). The auxiliary variable x is the adult female population in 1960 and the main variable y is breast cancer mortality in 1950-1969. For each estimator e we calculate the relative efficiency respect to the estimator $\hat{Y}_{\text{SRSWOR(MCAR)}}$ as:

$$\text{RE}(e) = \frac{V(e)}{V(\hat{Y}_{SRSWOR(MCAR)})} \tag{3.5}$$

The population is divided randomly into 30 groups (29 groups of size 10 and the last group of sample 11). For each δ_{ij}^* (j=0 to 7) we calculate the estimator

$$\hat{Y}^{j}_{\text{RHC(DNR)}} = \sum_{i \in s} \left(\frac{y_i}{p_i^*} \right) \left(\frac{\delta_{ij}}{\delta_{ij}^*} \right). \qquad (3.6)$$

We use a logistic model for the response probabilities $\phi(x_i) = \dfrac{1}{1+e^{-x_i}}$.

Table 1 shows the relative efficiency of the estimator $\hat{Y}_{\text{RS(MCAR)}}$ and the proposed estimators $\hat{Y}^{j}_{\text{RHC(DNR)}}$ for *j*=0 to 5 for all values of *r*. Table 2 and table 3 show the relative efficiency of the proposed estimators $\hat{Y}^{6}_{\text{RHC(DNR)}}$ and $\hat{Y}^{7}_{\text{RHC(DNR)}}$ for some values of λ.

Table 1. Relative efficiency for $\hat{Y}_{\text{RS(MCAR)}}$ and $\hat{Y}^{j}_{\text{RHC(DNR)}}$ estimators (j=0,…,5)

n	*r*	$\hat{Y}_{\text{RS(MCAR)}}$	$\hat{Y}^{0}_{\text{RHC(DNR)}}$	$\hat{Y}^{1}_{\text{RHC(DNR)}}$	$\hat{Y}^{2}_{\text{RHC(DNR)}}$	$\hat{Y}^{3}_{\text{RHC(DNR)}}$	$\hat{Y}^{4}_{\text{RHC(DNR)}}$	$\hat{Y}^{5}_{\text{RHC(DNR)}}$
30	1	1.06	0.03	0.02	0.02	0.03	0.02	0.02
30	2	1.06	0.05	0.05	0.05	0.06	0.05	0.04
30	3	1.06	0.08	0.07	0.06	0.09	0.07	0.06
30	4	1.06	0.1	0.09	0.08	0.13	0.09	0.07
30	5	1.05	0.13	0.1	0.09	0.16	0.11	0.08
30	6	1.05	0.16	0.12	0.11	0.2	0.13	0.08
30	7	1.05	0.18	0.13	0.12	0.24	0.14	0.08
30	8	1.05	0.21	0.14	0.13	0.27	0.16	0.09
30	9	1.05	0.24	0.15	0.13	0.3	0.17	0.09
30	10	1.04	0.26	0.16	0.14	0.33	0.18	0.08
30	11	1.04	0.29	0.17	0.14	0.36	0.19	0.08
30	12	1.04	0.32	0.17	0.15	0.37	0.2	0.08
30	13	1.04	0.35	0.18	0.15	0.39	0.2	0.07
30	14	1.04	0.37	0.18	0.15	0.4	0.21	0.07
30	15	1.03	0.4	0.18	0.15	0.41	0.21	0.06
30	16	1.03	0.43	0.18	0.15	0.41	0.21	0.06
30	17	1.03	0.46	0.18	0.14	0.4	0.21	0.05
30	18	1.03	0.49	0.17	0.14	0.39	0.21	0.05
30	19	1.03	0.52	0.17	0.14	0.38	0.2	0.04
30	20	1.02	0.55	0.16	0.13	0.36	0.2	0.04
30	21	1.02	0.58	0.15	0.12	0.34	0.19	0.03
30	22	1.02	0.61	0.15	0.12	0.32	0.18	0.03
30	23	1.02	0.64	0.14	0.11	0.29	0.17	0.03
30	24	1.01	0.67	0.13	0.1	0.26	0.16	0.03
30	25	1.01	0.7	0.11	0.09	0.23	0.14	0.02
30	26	1.01	0.73	0.1	0.08	0.19	0.13	0.02
30	27	1.01	0.76	0.09	0.07	0.16	0.11	0.03
30	28	1	0.79	0.07	0.06	0.12	0.09	0.03
30	29	1	0.82	0.06	0.05	0.08	0.06	0.03

Table 2. Relative efficiency of $\hat{Y}^{6}_{\text{RHC(DNR)}}$ for λ =0.1,…,0.9

n	*r*	$\hat{Y}^{6}_{\text{RHC(DNR)}}$								
		0.1	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*
30	1	0.28	0.17	0.12	0.09	0.07	0.05	0.04	0.04	0.03
30	2	0.4	0.27	0.2	0.16	0.13	0.1	0.09	0.07	0.06
30	3	0.46	0.34	0.27	0.21	0.18	0.15	0.12	0.11	0.09
30	4	0.49	0.39	0.31	0.26	0.22	0.19	0.16	0.14	0.12
30	5	0.51	0.42	0.35	0.3	0.25	0.22	0.19	0.17	0.15
30	6	0.51	0.43	0.37	0.33	0.29	0.25	0.22	0.2	0.18
30	7	0.51	0.45	0.39	0.35	0.31	0.28	0.25	0.23	0.2
30	8	0.5	0.45	0.41	0.37	0.33	0.3	0.28	0.25	0.23
30	9	0.5	0.45	0.42	0.38	0.35	0.32	0.3	0.28	0.26
30	10	0.48	0.45	0.42	0.39	0.37	0.34	0.32	0.3	0.28
30	11	0.47	0.45	0.42	0.4	0.38	0.36	0.34	0.32	0.31
30	12	0.46	0.44	0.42	0.4	0.39	0.37	0.36	0.35	0.33
30	13	0.44	0.43	0.42	0.41	0.4	0.39	0.38	0.37	0.36
30	14	0.42	0.42	0.41	0.41	0.4	0.4	0.39	0.39	0.38
30	15	0.41	0.41	0.41	0.41	0.41	0.4	0.4	0.4	0.4
30	16	0.39	0.39	0.4	0.4	0.41	0.41	0.42	0.42	0.43
30	17	0.37	0.38	0.39	0.4	0.41	0.42	0.43	0.44	0.45
30	18	0.35	0.36	0.38	0.39	0.41	0.42	0.44	0.45	0.47
30	19	0.33	0.35	0.36	0.38	0.4	0.42	0.45	0.47	0.49
30	20	0.31	0.33	0.35	0.38	0.4	0.43	0.45	0.48	0.51
30	21	0.29	0.31	0.34	0.37	0.39	0.43	0.46	0.5	0.53
30	22	0.27	0.29	0.32	0.35	0.39	0.43	0.47	0.51	0.56
30	23	0.24	0.27	0.31	0.34	0.38	0.42	0.47	0.52	0.58
30	24	0.22	0.26	0.29	0.33	0.37	0.42	0.47	0.53	0.59
30	25	0.2	0.24	0.27	0.32	0.37	0.42	0.48	0.54	0.61
30	26	0.18	0.22	0.26	0.3	0.36	0.41	0.48	0.55	0.63
30	27	0.15	0.19	0.24	0.29	0.35	0.41	0.48	0.56	0.65
30	28	0.13	0.17	0.22	0.27	0.33	0.4	0.48	0.57	0.67
30	29	0.11	0.15	0.2	0.26	0.32	0.4	0.48	0.57	0.69

Table 3. Relative efficiency of $\hat{Y}^{7}_{\text{RHC(DNR)}}$ for λ =0.1,…,0.9

n	*r*	$\hat{Y}^{7}_{\text{RHC(DNR)}}$								
		0.1	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*
30	1	0.32	0.19	0.13	0.1	0.08	0.06	0.05	0.04	0.03
30	2	0.43	0.3	0.22	0.17	0.14	0.11	0.09	0.08	0.06
30	3	0.48	0.37	0.29	0.23	0.19	0.16	0.13	0.11	0.09
30	4	0.51	0.41	0.33	0.28	0.23	0.2	0.17	0.14	0.12
30	5	0.52	0.43	0.36	0.31	0.27	0.23	0.2	0.17	0.15
30	6	0.52	0.45	0.39	0.34	0.3	0.26	0.23	0.2	0.18
30	7	0.52	0.46	0.4	0.36	0.32	0.29	0.26	0.23	0.2
30	8	0.51	0.46	0.41	0.38	0.34	0.31	0.28	0.25	0.23
30	9	0.5	0.46	0.42	0.39	0.36	0.33	0.3	0.28	0.26
30	10	0.49	0.45	0.42	0.4	0.37	0.35	0.32	0.3	0.28
30	11	0.47	0.45	0.42	0.4	0.38	0.36	0.34	0.32	0.31
30	12	0.46	0.44	0.42	0.41	0.39	0.37	0.36	0.35	0.33
30	13	0.44	0.43	0.42	0.41	0.4	0.39	0.38	0.37	0.36
30	14	0.42	0.42	0.41	0.41	0.4	0.4	0.39	0.39	0.38
30	15	0.41	0.41	0.41	0.41	0.41	0.4	0.4	0.4	0.4

30	16	0.39	0.39	0.4	0.4	0.41	0.41	0.42	0.42	0.43
30	17	0.37	0.38	0.39	0.4	0.41	0.42	0.43	0.44	0.45
30	18	0.35	0.36	0.38	0.39	0.41	0.42	0.44	0.45	0.47
30	19	0.33	0.35	0.37	0.39	0.41	0.43	0.45	0.47	0.49
30	20	0.31	0.33	0.35	0.38	0.4	0.43	0.46	0.49	0.52
30	21	0.29	0.31	0.34	0.37	0.4	0.43	0.46	0.5	0.54
30	22	0.27	0.3	0.33	0.36	0.4	0.43	0.47	0.51	0.56
30	23	0.25	0.28	0.31	0.35	0.39	0.43	0.48	0.53	0.58
30	24	0.22	0.26	0.3	0.34	0.38	0.43	0.48	0.54	0.6
30	25	0.2	0.24	0.28	0.33	0.38	0.43	0.49	0.55	0.62
30	26	0.18	0.22	0.27	0.32	0.37	0.43	0.49	0.56	0.64
30	27	0.16	0.2	0.25	0.3	0.36	0.43	0.5	0.57	0.66
30	28	0.13	0.18	0.23	0.29	0.35	0.42	0.5	0.59	0.68
30	29	0.11	0.16	0.22	0.28	0.35	0.42	0.5	0.6	0.7

It is clear that our proposed estimators $\hat{Y}^{j}_{\text{RHC(DNR)}}$ for j=0,…,7 fare better than the alternative estimators independently of the value of r. Respect to $\hat{Y}^{6}_{\text{RHC(DNR)}}$ and $\hat{Y}^{7}_{\text{RHC(DNR)}}$ estimators, we also observed that if r is small, we should use large α-values and reciprocally, for large values of r we should use small α-values.

It is interesting to note that the $\hat{Y}^{5}_{\text{RHC(DNR)}}$ has a very good behaviour: the relative efficiency is always less than 0.1, that is, the estimator produces a gain in accuracy to the $\hat{Y}_{\text{SRSWOR(MCAR)}}$ estimator higher than 90% in all cases.

Finally noted that we have tried with other functions for the response probabilities and we have seen that the behaviour is very dependent on this choice.

4. Simulation Study

We conducted a small simulation study to investigate the finite sample performance of the proposed estimators. We considered the same population (Cancer). The coefficient of correlation between variables is 0.967094. For each unit i of this population we generated a Bernoulli variable with probability of success $\phi(x_i) = \dfrac{1}{1+e^{-x_i}}$ and we assumed the y_i-value as missing if the results of this variable was 0. At each simulation run, a sample of size 25 was taken using the Rao, Hartley and Cochran (1962) scheme (twenty four groups of size 12 and one group of size 13) and the considered estimators of the mean were computed. The process was repeated B=1000 times. The average response over all simulations run is 15.380. Table 4 shows Relative Efficiency x 100 , Relative Bias x 100 and the gain in efficiency over SRSWORMCAR (1/RE) over all simulations runs.

Table 4. Relative efficiency and relative bias in % of considered estimators with non-response

Estimator	RE	RB	1/RE
SRSWORMCAR	100.000	14667.8163	100
RSMCAR	67.983	12198.1858	147,1
RHCDNR0	3.272	1975.9157	3056,23
RHCDNR1	2.422	-2213.7086	4128,82
RHCDNR2	3.201	-2596.0284	3124,02
RHCDNR3	0.401	364.7691	24937,66
RHCDNR4	1.689	-1766.5721	5920,66
RHCDNR5	6.085	-3686.2522	1643,39
RHCDNR601	0.256	-55.0479	39062,5
RHCDNR602	0.289	120.3309	34602,08
RHCDNR603	0.370	305.7192	27027,03
RHCDNR604	0.508	501.9871	19685,04
RHCDNR605	0.712	710.1290	14044,94
RHCDNR606	0.995	931.2865	10050,25
RHCDNR607	1.373	1166.7774	7283,32
RHCDNR608	1.865	1418.1323	5361,93
RHCDNR609	2.495	1687.1412	4008,02
RHCDNR701	0.255	-46.2185	39215,69
RHCDNR702	0.293	136.8093	34129,69
RHCDNR703	0.382	328.5172	26178,01
RHCDNR704	0.530	529.5182	18867,92
RHCDNR705	0.745	740.4957	13422,82
RHCDNR706	1.038	962.2130	9633,91
RHCDNR707	1.420	1195.5270	7042,25
RHCDNR708	1.908	1441.4018	5241,09
RHCDNR709	2.518	1700.9279	3971,41

Table 4 can be summarized as follows: (i.) the model of non-response used introduces a serious problem of bias in all estimators, but specially in the estimators based on the assumption of MCAR non-response: $\hat{Y}_{\text{SRSWOR(MCAR)}}$ and $\hat{Y}_{\text{RS(MCAR)}}$; (ii) the proposed estimators $\hat{Y}^{6}_{\text{RHC(DNR)}}$ and $\hat{Y}^{7}_{\text{RHC(DNR)}}$ with λ =0.1 are the smallest bias; (iii) these estimators are the most efficient estimators for the mean; (iv) the gain in efficiency for these estimators decreases as α increases.

References

Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962). A simple procedure of unequal probability sampling without replacement. *J. R. Statist. Soc.*, B, 24, 482-491.

Rao, J.N.K. and Sitter, R.R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82, 453-460.

Royal, R.M. and Cumberland, W.G. (1981) The finite population linear regression and estimators of its variance. An empirical study. *J. Am. Stat. Assoc*. 76, 924-930

Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

Särndal, C.E. (1992). Methods for estimating the precision of survey estimates when imputation is used. *Survey Methodology*, 18, 241-252.