https://doi.org/10.56093/JISAS.V79I3.6
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 79(3) 2025 255–266

Estimation of Domain Mean under Two Stage Sampling with Units Missing at Random

Kaustav Aditya[1], Bharti[1], Neha Purwar[1], Safeela Nasrin[1] and Shrila Das[2]
[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[2]*ICAR-Indian Agricultural Research Institute, New Delhi*

Received 12 November 2025; Revised 10 December 2025; Accepted 11 January 2026

SUMMARY

Estimation of domain means is crucial in large-scale surveys; however, it becomes difficult when some of the selected units fail to respond. Sub-sampling of non-respondents serves as a practical and methodologically sound approach to address this issue. This study considers the estimation of domain mean under a random response mechanism using a two-stage sampling design with two phases at the second stage. Three estimators were developed based on sub-sampling of non-respondents, where additional effort was made to collect information from a selected sub-sample. Variance expressions were obtained for each estimator, and the optimum sample sizes were determined by minimizing the expected cost for a given level of precision. The empirical analysis shows a considerable reduction in expected cost for all proposed estimators, with the highest reduction observed when nonresponse is minimum. The study concludes that the proposed estimators are more efficient and cost-effective for domain mean estimation when nonresponse is present.

Keywords: Random response; Missing at random; Sub-sampling; Cost function, Domain estimation, Two stage sampling.

1. INTRODUCTION

For large or medium-scale surveys, the sampling frame of ultimate stage units is not available and the cost of construction of such frame is very high. Moreover, when the population elements are widely dispersed geographically, the resulting sample also becomes scattered, leading to high costs of enumeration and difficulties in supervising fieldwork. For such situations, two-stage or multi-stage sampling designs are very effective. However, in many human surveys, another practical difficulty arises when data are not obtained from all selected units. Nonresponse often persists even after repeated call-backs, leading to incomplete data. Since the respondents may differ systematically from the non-respondents, the resulting estimates can be biased. To mitigate this issue, Hansen and Hurwitz (1946) proposed a technique for adjusting for nonresponse, providing a framework to reduce such bias and improve the reliability of survey estimates. The technique consists of selecting a sub-sample of non-respondents. Through specialized efforts data are collected from the non-respondents so as to obtain an estimate of non-responding units in the population. Foradori (1961) studied the sub-sampling of the non-respondents to estimate the population total in two stages using unequal probability sampling. Srinath (1971) used a different procedure for selecting the sub-sample of respondents where the sub-sampling procedure varied according to the nonresponse rates. Oh and Schereun (1983) attempted to compensate for nonresponse by weighing adjustment. Kalton and Karsprzyk (1986) tried the imputation technique. Tripathi and Khare (1997) extended the sub-sampling of non-respondents' approach to multivariate cases. Okafor and Lee (2000) extended the approach to double sampling for ratio and regression estimation. Okafor (2001, 2005)

Corresponding author: Bharti
E-mail address: bhartibhardwaj69@gmail.com

further extended the approach in the context of element sampling and two-stage sampling, respectively, on two successive occasions. Chhikara and Sud (2009) used the sub-sampling of non-respondents' approach for estimation of population and domain totals in the context of item nonresponse. Sud *et al.* (2012) addresses the problem of estimation of finite population mean in the presence of nonresponse under two stage sampling design assuming a deterministic response mechanism. Subsequently, Sud *et al.* (2013) examined the same problem under a two-stage sampling design when the response mechanism was assumed to be random. Aditya *et al.* (2014, 2015) developed estimators for the domain mean in the presence of nonresponse under a two-stage sampling design, assuming a deterministic response mechanism. They extended the methodology to situations where population level auxiliary information was not available at the earlier stages of selection.

It may be noted that weighting and imputation procedures aim at elimination of bias introduced by nonresponse. However, these procedures are based on certain assumptions on the response mechanism. When these assumptions do not hold good the resulting estimate may be seriously biased. Further, when the nonresponse is confounded i.e. the response probability is dependent on the survey character; it becomes difficult to eliminate the bias entirely. Rancourt *et al.* (1994) provided a partial correction for the situation. Hansen and Hurwitz's sub-sampling approach, although costly, is free from any assumptions. When the bias caused by nonresponse is serious this technique is very effective i.e. one does not have to go for 100 percent response, which can be very expensive.

In many cases, estimates for specific subgroups or domains of the population are required alongside the overall estimates (Sarndal *et al.,* 1992). In the context of estimation of the domain parameters (mean/total), Agrawal and Midha (2007) proposed a two-phase sampling design when the size of the domain was not known. Chhikara and Sud (2009) considered the problem of estimation of finite population mean of a domain in the presence of item nonresponse under a random response mechanism. Sud *et al.* (2010) considered the problem of estimation of finite population mean of a domain in the presence of nonresponse under a deterministic response mechanism. Recent studies on estimation in the presence of nonresponse are Aditya *et al.* (2013,14), Raman *et al.* (2013 and 2016), Audu *et al.* (2022), Chaudhury *et al.* (2021, 2023) and Singh *et al.* (2024), Pandey *et al.* (2024), Alamri & Aloraini (2025) *etc.*

In this study, estimators of domain mean under two-stage sampling design were developed using the technique of sub-sampling of the non-respondents, assuming a random response mechanism. The variances of the proposed estimators were derived, and the optimum sample sizes were determined by minimizing the expected cost for a specified level of precision.

2. THEORETICAL DEVELOPMENTS

Let the finite population $\boldsymbol{U}$ under consideration consists of N known primary stage units (psus) labeled 1 through N. Let the i^{th} psu comprise M second stage units (ssus). Let us consider a population $U = (1,\ldots,k,\ldots,N)$ of size N partitioned into D sub-sets $U_1 \ldots U_d \ldots U_D$ (hereafter we refer them as domains) and let N_d (which is assumed large) be the size of $U_d\,(d = 1,\ldots,D)$ such that $U = \bigcup_{d=1}^{D} U_d$ and $N = \sum_{d=1}^{D} N_d$. It is assumed that N_d is known. Let a sample s of n psus is drawn from the population using simple random sampling without replacement (srswor) sampling design. Let s_d denote the part of sample s that happens to fall in U_d, that is, $s_d = s \cap U_d$. Let us denote by n_d the size of s_d such that $s = \bigcup_{d=1}^{D} s_d$ and $n = \sum_{d=1}^{D} n_d$. When the domain sizes are small, n_d may turn out to be very small or it may be equal to '0' in some cases. In such cases small area estimation techniques are needed for reliable estimation at the domain level. However, we do not consider this case here. Let M_d be the size of the units in each psu belonging to the d^{th} domain and from each selected psu m_d ssus are selected by srswor and letters/mails containing questionnaires are sent to each unit in the sample. With the random sample of observations, the statistician's task is to make the best possible estimate for the domain. Let y_{dkj} be the value of study character pertaining to j^{th} ssu in the k^{th} psu in d^{th} domain, k=1,2,…,N_d, j=1,2,…,M_d, d=1, 2,..., D. The objective this study is to estimate the domain mean $\overline{Y}_d = \frac{1}{N_d}\sum_{k=1}^{N_d}\frac{1}{M_d}\sum_{j=1}^{M_d} y_{dkj}$.

Considering the estimator developed by Aditya (2016), let n psus be selected from N psus by srswor where n_d out of n psus fall in the d^{th} domain and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Let, a sample of m_d ssus selected from M_d ssus, m_{dk_1} units responds and m_{dk_2} units do not respond, $m_{dk_1} + m_{dk_2} = m_d$. From the m_{dk_2} nonresponding units a subsample of h_{dk_2} units is selected by srswor, $m_{dk_2} = h_{dk_2} f_{dk_2}$, k=1, 2,..., n_d. Let $\bar{y}_{m_{dk_1}}$ denote the mean of the sample from the response class for the d^{th} domain while $\bar{y}_{h_{dk_2}}$ denote the mean of the sample for the nonresponse class, where $\bar{y}_{m_{dk_1}} = \frac{1}{m_{dk_1}} \sum_{j=1}^{m_{dk_1}} y_{dkj}$ and $\bar{y}_{h_{dk_2}} = \frac{1}{h_{dk_2}} \sum_{k=1}^{h_{dk_2}} y_{dkj}$. Here at the second stage m_{dk_1} responding and m_{dk_2} nonresponding units are being generated as a result of m_d independent Bernoulli trials, one for each element k in m_d with constant probability θ_{dk} of "success", i.e. the response. Therefore

$$\Pr(k \in m_{dk_1} \mid m_d) = \theta_{k|m_d} = \theta_{dk} \text{ and}$$

$$\Pr(k \,\&\, l \in m_{dk_1} \mid m_d) = \theta_{kl|m_d} = \theta_{dk}^2$$

An unbiased estimator of $\bar{Y}_d$ given by Aditya (2016) is as,

$$\bar{y}_{1dr} = \frac{N}{nN_d} \sum_{k=1}^{n_d} \frac{1}{m_d} (m_{dk_1} \bar{y}_{m_{dk_1}} + m_{dk_2} \bar{y}_{h_{dk_2}}) \qquad (1)$$

with variance

$$V(\bar{y}_{1dr}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)} S_{bd}^2 + \frac{N(N-n)}{nN_d^2(N-1)} N_d Q_d \bar{Y}_d^2 + \frac{N}{nN_d^2} \sum_{k=1}^{N_d} \left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2 + \frac{N}{nN_d^2} \sum_{k=1}^{N_d} \frac{(1-\theta_{dk})}{m_d} (f_{dk_2} - 1) S_{dk}^2 \qquad (2)$$

where,

$$S_{bd}^2 = \frac{1}{(N_d - 1)} \sum_{k=1}^{N_d} (\bar{Y}_{dk} - \bar{Y}_d)^2, \quad \bar{Y}_{dk} = \frac{1}{M_d} \sum_{j=1}^{M_d} Y_{dk_j}$$

and $\bar{Y}_d = \frac{1}{N_d} \sum_{k=1}^{N_d} \bar{Y}_{dk} = \frac{1}{N_d} \sum_{k=1}^{N} Z_{dk} \bar{Y}_k$.

$$S_{dk}^2 = \frac{1}{(M_d - 1)} \sum_{j=1}^{M_d} (Y_{dk_j} - \bar{Y}_{dk})^2, \quad P_d = \frac{N_d}{N}, Q_d = 1 - P_d,$$

Now, determining the optimum values of n, m_d and f_{dk_2} by minimizing the expected cost for a fixed variance as given in Aditya (2016). The following cost functionwas considered,

$$C = C_{1d} n_d + C_{2d} \sum_{k=1}^{n_d} m_{dk_1} + C_{3d} \sum_{k=1}^{n_d} h_{dk_2}$$

where,C: Total cost; C_{1d}: Per unit travel and miscellaneous cost in the d^{th} domain; C_{2d}: Cost per unit of collecting the information on the study character in the first attempt in the d^{th} domain; C_{3d}: Cost per unit of collecting the information by expensive method after the first attempt of collecting information failed in the d^{th} domain.

The expected cost in this case is,

$$E(C) = \frac{n}{N} [C_{1d} N_d + C_{2d} m_d \sum_{k=1}^{N_d} \theta_{dk} + C_{3d} m_d \sum_{k=1}^{N_d} \frac{(1-\theta_{dk})}{f_{dk_2}}]$$

The optimum values were obtained as,

$$n_{opt} = \frac{K_{24}}{K_{23}}, \; m_{dopt} = \frac{-b_{12} \pm \sqrt{b_{12}^2 - 4a_{12}e_{12}}}{2a_{12}} \quad \text{and}$$

$$f_{2dopt} = \pm \sqrt{\frac{C_{3d} \sum_{k=1}^{N_d} (1-\theta_{dk}) \left(\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{i=1}^{N_d} (1-\theta_{dk}) S_{dk}^2 \right)}{C_{2d} \sum_{k=1}^{N_{2d}} \theta_{dk} \sum_{k=1}^{N_d} (1-\theta_{dk}) S_{dk}^2}}$$

To avoid negative values, the following values were considered

$$m_{dopt} = \frac{-b_{12} + \sqrt{b_{12}^2 - 4a_{12}e_{12}}}{2a_{12}} \quad \text{and}$$

$$f_{2dopt} = \sqrt{\frac{C_{3d} \sum_{k=1}^{N_d} (1-\theta_{dk}) \left(\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{i=1}^{N_d} (1-\theta_{dk}) S_{dk}^2 \right)}{C_{2d} \sum_{k=1}^{N_{2d}} \theta_{dk} \sum_{k=1}^{N_d} (1-\theta_{dk}) S_{dk}^2}}$$

where,

$$K_{23} = V_0 + \frac{N}{N_d^2} \left\{ \frac{N_d Q_d \bar{Y}_d^2}{(N-1)} \right\} + \frac{k_{12}}{N_d} S_{bd}^2,$$

$$\left[k_{12} = \left[\frac{N(N_d - 1)}{N_d(N-1)}\right]\right],$$

$$K_{24} = \left[\frac{N}{N_d}k_{12}S_{bd}^2 + \frac{N^2}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} + \frac{N}{N_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d} - \frac{1}{M_d}\right)S_{dk}^2 + \frac{N}{N_d^2}\sum_{k=1}^{N_d}\frac{(1-\theta_{dk})}{m_d}(f_{2d}-1)S_{dk}^2\right],$$

$$a_{12} = \left(C_{3d}\sum_{k=1}^{N_d}\frac{(1-\theta_{dk})}{f^2{}_{2d}}\right)\left[N_d k_{12} S_{bd}^2 + N\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} - \sum_{k=1}^{N_d}\frac{1}{M_d}S_{dk}^2\right],$$

$$b_{12} = -\left(C_{2d}\sum_{k=1}^{N_d}\theta_{dk}\sum_{k=1}^{N_d}(1-\theta_{dk})S_{dk}^2 - C_{3d}\sum_{k=1}^{N_d}\frac{(1-\theta_{dk})}{f^2{}_{2d}}\left\{\sum_{k=1}^{N_d}S_{dk}^2 - \sum_{k=1}^{N_d}(1-\theta_{dk})S_{dk}^2\right\}\right),$$

$$e_{12} = -C_{1d}N_d\sum_{k=1}^{N_d}(1-\theta_{dk})S_{dk}^2$$

and $V_0 = 0.0025 \times \bar{Y}_d^2$.

2.1 Proposed Estimator 1 ($\bar{y}_{2dr}$)

Let n psus be selected from N psus by srswor design where n_d out of n psus falls in the d-th domain and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Let there be no nonresponse in n_{1d}psus. In the remaining n_{2d} psus m_{dk_1} units responds and m_{dk_2} units do not respond, $m_{dk_1} + m_{dk_2} = m_d$. From the m_{dk_2} nonresponding units a subsample of h_{dk_2} units is selected by srswor and data was collected through specialized effort, $m_{dk_2} = h_{dk_2} f_{dk_2}$, k=1, 2,...,n_{2d}. In this context, the proposed estimator was

$$\bar{y}_{2dr} = \frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}})\right\} \tag{3}$$

which is unbiased for $\bar{Y}_d$.

$$E\left(\bar{y}_{2dr}\right) = E_1E_2E_3E_4E_5E_6E_7E_8\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}\left(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{h_{dk_2}}\right)\right\}$$

$$= E_1E_2E_3E_4E_5E_6E_7\frac{N}{nN_d}\left\{\left(\sum_{k=1}^{n_{dd}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}\left(m_{dk_1}\bar{y}_{m_{dk_1}} + m_{dk_2}\bar{y}_{m_{d_2}}\right)\right)\right\}$$

$$= E_1E_2E_3E_4E_5E_6\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}\left(m_{dk_1}\bar{y}_{dk} + m_{dk_2}\bar{y}_{dk}\right)\right\}$$

$$= E_1E_2E_3E_4E_5\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\frac{1}{m_d}\left(\theta_{dk}m_d\bar{y}_{dk} + \left(1-\theta_{dk}\right)m_d\bar{y}_{dk}\right)\right\}$$

$$= E_1E_2E_3E_4E_5\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk} + \sum_{k=1}^{n_{2d}}\bar{y}_{dk}\right\}$$

$$= E_1E_2E_3E_4\frac{N}{nN_d}\left\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}$$

$$= E_1E_2E_3\frac{N}{nN_d}\left\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\} := E_1E_2\frac{N}{nN_d}\left\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}$$

$$= E_1\frac{N}{nN_d}\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk} = \frac{1}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk} = \bar{Y}_d n$$

Thus, $\bar{y}_{2dr}$ is an unbiased estimator of $\bar{Y}_d$.Where E_8represents conditional expectations of all possible samples of size h_{dk_2} drawn from m_{dk_2}, E_7 is the conditional expectation of all possible samples of size m_{dk_1}, m_{dk_2} respectively drawn from M_{dk_1} and M_{dk_2} by keeping m_{dk_1}, m_{dk_2} fixed, E_6 refers to conditional expectation arising out of m_d independent Bernoulli trials resulting in m_{dk_1} success and m_{dk_2} failures, $m_{dk_1} + m_{dk_2} = m_d$, E_5 is the conditional expectation of all possible samples of size m_d drawn from M_d, E_4 arising out of selection of all possible samples of size n_{1d}, n_{2d} from N_{1d}, N_{2d} keeping n_{1d}, n_{2d} fixed, E_3 refers to conditional expectation arising out of randomness of n_{1d}, n_{2d}, E_2 arising out of selection of all possible samples of size n_d drawn from N_d keeping n_d fixed whereas E_1 is the expectation arising out of randomness of n_d. The variance of estimator was obtained as follows:

$$V(\bar{y}_{2dr}) = V_1E_2E_3E_4E_5E_6E_7E_8(\bar{y}_{2dr}) + E_1V_2E_3E_4E_5 E_6E_7E_8(\bar{y}_{2dr}) + E_1E_2V_3E_4E_5E_6E_7E_8(\bar{y}_{2dr}) + E_1E_2E_3V_4E_5E_6E_7E_8(\bar{y}_{2dr}) + E_1E_2E_3E_4V_5 E_6E_7E_8(\bar{y}_{2dr}) + E_1E_2E_3E_4E_5V_6E_7E_8(\bar{y}_{2dr}) + E_1E_2E_3E_4E_5E_6V_7E_8(\bar{y}_{2dr}) + E_1E_2E_3E_4E_5 E_6E_7V_8(\bar{y}_{2dr})$$

($V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8$ are defined similarly as $E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8$).

Here

$$V_1E_2E_3E_4E_5E_6E_7E_8(\bar{y}_{2dr}) = \frac{N(N-n)}{nN_d^2(N-1)} N_d Q_d \bar{Y}^2$$

$$E_1V_2E_3E_4E_5E_6E_7E_8(\bar{y}_{2dr}) = \frac{(N-n)}{nN_d}\left[1 - \frac{(N-N_d)}{N_d(N-1)}\right] S_{bd}^2,$$

$$E_1E_2V_3E_4E_5E_6E_7E_8(\bar{y}_{2dr}) = 0,$$
$$E_1E_2E_3V_4E_5E_6E_7E_8(\bar{y}_{2dr}) = 0,$$

$$E_1E_2E_3E_4V_5E_6E_7E_8(\bar{y}_{2dr}) = \frac{N}{nN_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2,$$

$$E_1E_2E_3E_4E_5V_6E_7E_8(\bar{y}_{2dr}) = 0,$$
$$E_1E_2E_3E_4E_5E_6V_7E_8(\bar{y}_{2dr}) = 0 \text{ and}$$

$$E_1E_2E_3E_4E_5E_6E_7V_8(\bar{y}_{2dr}) = \frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{m_d}(f_{dk_2} - 1) S_{dk}^2$$

By adding the above three terms, the variance of estimator was obtained as

$$V(\bar{y}_{2dr}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)} S_{bd}^2 + \frac{N(N-n)}{nN_d^2(N-1)} N_d Q_d \bar{Y}_d^2 + \frac{N}{nN_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2 + \frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{m_d}(f_{dk_2} - 1) S_{dk}^2 \quad (4)$$

where $\bar{y}_{m_{dk_1}}, \bar{y}_{h_{dk_2}}$, S_{bd}^2, S_{dk}^2 etc. have been defined earlier.

The optimum values of n, m_d and f_{dk_2} were determined by minimizing the expected cost for a fixed variance. The relevant cost function in this case is,

$$C = C_{1d} n_{2d} + C_{2d} n_{1d} m_{dk} + C_{2d}\sum_{k=1}^{n_{2d}} m_{dk_1} + C_{3d}\sum_{k=1}^{n_{2d}} h_{dk_2}$$

C, C_{1d}, C_{2d} and C_{3d} are same as defined earlier. The expected cost was,

$$E(C) = \frac{n}{N}\left[C_{1d}N_{2d} + C_{2d}N_{1d}m_d + C_{2d}m_d\sum_{k=1}^{N_{2d}}\theta_{dk} + C_{3d}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})m_d}{f_{dk_2}}\right]$$

Consider the function $\phi = E(C) + \lambda\{V(\bar{y}_{2dr}) - V_0\}$, where λ is the Lagrangian multiplier. To get closed form expression of the optimum values, it was assumed that $m_{dk_2} = h_{dk_2} f_{2d}$, k=1, 2,..., n_{2d} in place of $m_{dk_2} = h_{dk_2} f_{dk_2}$, k=1, 2,..., n_{2d}.

The optimum values obtained through minimization were as follows:

$$n_{opt} = \frac{K_{25}}{K_{26}}, \; m_{dopt} = \frac{-b_{13} \pm \sqrt{b_{13}^{\,2} - 4a_{13}e_{13}}}{2a_{13}} \quad \text{and}$$

$$f_{2dopt} = \pm\sqrt{\frac{C_{3d}\sum_{k=1}^{N_{2d}}(1-\theta_{dk})\left(\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{i=1}^{N_{2d}}(1-\theta_{dk})S_{dk}^2\right)}{C_{2d}\left(N_{1d} + \sum_{k=1}^{N_{2d}}\theta_{dk}\right)\sum_{k=1}^{N_{2d}}(1-\theta_{dk})S_{dk}^2}}$$

Only positive values were considered, therefore,

$$m_{dopt} = \frac{-b_{13} + \sqrt{b_{13}^{\,2} - 4a_{13}e_{13}}}{2a_{13}} \quad \text{and}$$

$$f_{2dopt} = \sqrt{\frac{C_{3d}\sum_{k=1}^{N_d}(1-\theta_{dk})\left(\sum_{k=1}^{N_d} S_{dk}^2 - \sum_{i=1}^{N_d}(1-\theta_{dk})S_{dk}^2\right)}{C_{2d}\left(N_{1d} + \sum_{k=1}^{N_{2d}}\theta_{dk}\right)\sum_{k=1}^{N_d}(1-\theta_{dk})S_{dk}^2}}$$

where,

$$K_{26} = V_0 + \frac{N}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} + \frac{k_{12}}{N_d} S_{bd}^2,$$

$$\left[k_{12}=\left[\frac{N(N_d-1)}{N_d(N-1)}\right]\right],$$

$$K_{25}=\left[\begin{array}{l}\frac{N}{N_d}k_{12}S_{bd}^2+\frac{N^2}{N_d^2}\left\{\frac{N_dQ_d\bar{Y}_d^2}{N-1}\right\}+\frac{N}{N_d^2}\sum_{k=1}^{N_d}(\frac{1}{m_d}-\\ \frac{1}{M_d})S_{dk}^2+\frac{N}{N_d^2}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{m_d}(f_{2d}-1)S_{dk}^2\end{array}\right],$$

$$a_{12}=\left(C_{3d}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{f^2{}_{2d}}\right)\left[\begin{array}{l}N_dk_{12}S_{bd}^2+\\ N\left\{\frac{N_dQ_d\bar{Y}_d^2}{N-1}\right\}-\sum_{k=1}^{N_d}\frac{1}{M_d}S_{dk}^2\end{array}\right],$$

$$b_{12}=-\left(\begin{array}{l}C_{2d}(N_{1d}+\sum_{k=1}^{N_{2d}}\theta_{dk})\sum_{k=1}^{N_{2d}}(1-\theta_{dk})S_{dk}^2-\\ C_{3d}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{f^2{}_{2d}}\left\{\sum_{k=1}^{N_d}S_{dk}^2-\sum_{k=1}^{N_{2d}}(1-\theta_{dk})S_{dk}^2\right\}\end{array}\right),$$

$$e_{12}=-C_{1d}N_{2d}\sum_{k=1}^{N_{2d}}(1-\theta_{dk})S_{dk}^2$$

and $V_0=0.0025\times\bar{Y}_d^2$.

2.2 Proposed Estimator 2 ($\bar{y}_{3dr}$)

Let n psus be selected from Npsus by srswor where n_d out of n psusfall in the d-th domain and within each selected psu, m_d ssusare also selected from M_d ssus by srswor. Let there be no nonresponse in n_{1d} psus. In the remaining n_{2d} psus m_{dk_1} units responds and m_{dk_2} units do not respond, $m_{dk_1}+m_{dk_2}=m_d$. From the m_{dk_2} nonresponding units a sub-sample of h_{dk_2} units is selected by srswor, $m_{dk_2}=h_{dk_2}f_{dk_2}$, k=1, 2,...,n_{2d}. Let there be complete nonresponse in the n_{3d} psus, $n_{1d}+n_{2d}+n_{3d}=n_d$. Further a sub-sample of h_{3d} psus is drawn out of n_{3d} psus and data are collected through specialized efforts on each of m_d ssus in the selected h_{3d} psus, n_{3d}=$f_{3d}h_{3d}$. In this context, the proposed estimator isas

$$\bar{y}_{3dr}=\frac{N}{nN_d}\left\{\begin{array}{l}\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}}+m_{dk_2}\bar{y}_{h_{dk_2}})+\\ \frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\end{array}\right\} \tag{5}$$

which is unbiased for $\bar{Y}_d$.

$$E\left(\bar{y}_{3dr}\right)=E_1E_2E_3E_4E_5E_6E_7E_8E_9\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}}+m_{dk_2}\bar{y}_{h_{dk_2}})+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}$$

$$=E_1E_2E_3E_4E_5E_6E_7E_8\frac{N}{nN_d}\left\{\left(\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{m_{dk_1}}+m_{dk_2}\bar{y}_{m_{dk_2}})+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right)\right\}$$

$$=E_1E_2E_3E_4E_5E_6E_7\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(m_{dk_1}\bar{y}_{dk}+m_{dk_2}\bar{y}_{dk})+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}$$

$$=E_1E_2E_3E_4E_5E_6\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\frac{1}{m_d}(\theta_{dk}m_d\bar{y}_{dk}+(1-\theta_{dk})m_d\bar{y}_{dk})+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}$$

$$=E_1E_2E_3E_4E_5E_6\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{y}_{dk}+\sum_{k=1}^{n_{2d}}\bar{y}_{dk}+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{y}_{dk}\right\}$$

$$=E_1E_2E_3E_4E_5\frac{N}{nN_d}\left\{\sum_{k=1}^{n_{1d}}\bar{Y}_{dk}+\sum_{k=1}^{n_{2d}}\bar{Y}_{dk}+\frac{n_{3d}}{h_{3d}}\sum_{k=1}^{h_{3d}}\bar{Y}_{dk}\right\}$$

$$=E_1E_2E_3E_4\frac{N}{nN_d}\left[\left\{\sum_{k=1}^{n_{1d}}\bar{Y}_{dk}+\sum_{k=1}^{n_{2d}}\bar{Y}_{dk}+\sum_{k=1}^{n_{3d}}\bar{Y}_{dk}\right\}\right]$$

$$=E_1E_2E_3E_4\frac{N}{nN_d}\left\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}$$

$$=E_1E_2\frac{N}{nN_d}\left\{\sum_{k=1}^{n_d}\bar{Y}_{dk}\right\}=E_1\frac{N}{nN_d}\frac{n_d}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}$$

$$=\frac{1}{N_d}\sum_{k=1}^{N_d}\bar{Y}_{dk}=\bar{Y}_d$$

Thus, $\bar{y}_{3dr}$ is an unbiased estimator of $\bar{Y}_d$. Where E_9 represents conditional expectations of all possible samples of size h_{dk_2} drawn from m_{dk_2}, E_8 is the conditional expectation of all possible samples of size m_{dk_1}, m_{dk_2} respectively drawn from M_{dk_1} and M_{dk_2} by keeping m_{dk_1},m_{dk_2} fixed, E_7 refers to conditional expectation arising out of m_d independent Bernoulli trials resulting in m_{dk_1} success and m_{dk_2} failures, $m_{dk_1}+m_{dk_2}=m_d$, E_6 is the conditional expectation of all possible samples of size m_d drawn from M_d, E_5 represents the conditional expectation of selection of h_{3d} from n_{3d}, E_4 arising out of selection of all possible samples of size n_{1d}, n_{2d}, n_{3d} from N_{1d}, N_{2d}, N_{3d} keeping n_{1d}, n_{2d}, n_{3d} fixed, E_3 refers to conditional expectation arising out of randomness of n_{1d}, n_{2d}, n_{3d}, E_2 arising out of selection of all possible samples of size n_d drawn from N_d keeping n_d fixed whereas E_1 is the expectation arising out of randomness of n_d. The variance of the proposed estimator was obtained as follows:

$$\begin{aligned}V(\bar{y}_{3dr}) &= V_1E_2E_3E_4E_5E_6E_7E_8E_9(\bar{y}_{3dr}) + E_1V_2E_3E_4\\&E_5E_6E_7E_8E_9(\bar{y}_{3dr}) + E_1E_2V_3E_4E_5E_6E_7E_8\\&E_9(\bar{y}_{3dr}) + E_1E_2E_3V_4E_5E_6E_7E_8E_9(\bar{y}_{3dr})+\\&E_1E_2E_3E_4V_5E_6E_7E_8E_9(\bar{y}_{3dr}) + E_1E_2E_3E_4\\&E_5V_6E_7E_8E_9(\bar{y}_{3dr}) + E_1E_2E_3E_4E_5E_6V_7E_8\\&E_9(\bar{y}_{3dr}) + E_1E_2E_3E_4E_5E_6E_7V_8E_9(\bar{y}_{3dr})+\\&E_1E_2E_3E_4E_5E_6E_7E_8V_9(\bar{y}_{3dr})\end{aligned}$$

($V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9$ are defined similarly as $E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8, E_9$)

$$V_1E_2E_3E_4E_5E_6E_7E_8E_9(\bar{y}_{3dr}) = \frac{N(N-n)}{nN^2_d}\frac{N_d}{(N-1)}Q_d\bar{Y}_d^2,$$

$$E_1V_2E_3E_4E_5E_6E_7E_8E_9(\bar{y}_{3dr}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2,$$

$$E_1E_2V_3E_4E_5E_6E_7E_8E_9(\bar{y}_{3dr}) = 0,$$

$$E_1E_2E_3V_4E_5E_6E_7E_8E_9(\bar{y}_{3dr}) = 0,$$

$$E_1E_2E_3E_4V_5E_6E_7E_8E_9(\bar{y}_{3dr}) = \frac{NN_{3d}}{nN_d^2}(f_{3d}-1)S^2_{b_{N_{3d}}},$$

$$E_1E_2E_3E_4E_5V_6E_7E_8E_9(\bar{y}_{3dr}) = \frac{N}{nN_d^2}\left[\sum_{k=1}^{N_{1d}+N_{2d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2 + f_{3d}\sum_{k=1}^{N_{3d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2\right],$$

$$E_1E_2E_3E_4E_5E_6V_7E_8E_9(\bar{y}_{3dr}) = 0,$$

$$E_1E_2E_3E_4E_5E_6E_7V_8E_9(\bar{y}_{3dr}) = 0 \text{ and}$$

$$E_1E_2E_3E_4E_5E_6E_7E_8V_9(\bar{y}_{3dr}) = \frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{m_d}(f_{dk_2}-1)S_{dk}^2$$

Thus, by adding all the terms, the variance of proposed estimator was obtained as

$$\begin{aligned}V(\bar{y}_{3dr}) &= \frac{N(N-n)}{nN^2_d}\frac{N_d}{N-1}Q_d\bar{Y}_d^2 + \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)}S_{bd}^2 +\\&\frac{N}{nN_d^2}\left[\sum_{k=1}^{N_{1d}+N_{2d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2 + f_{3d}\sum_{k=1}^{N_{3d}}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2\right]\\&+\frac{NN_{3d}}{nN_d^2}(f_{3d}-1)S^2_{b_{N_{3d}}} + \frac{N}{nN_d^2}\sum_{k=1}^{N_{2d}}\frac{(1-\theta_{dk})}{m_d}(f_{dk_2}-1)S_{dk}^2\end{aligned} \quad (6)$$

$$S^2_{bd_{N_{3d}}} = \frac{1}{N_{3d}-1}\sum_{k=1}^{N_{3d}}(\bar{Y}_k - \bar{Y}_{N_{3d}})^2, \text{ where,}$$

$$\bar{Y}_{N_{3d}} = \frac{1}{N_{3d}}\sum_{k=1}^{N_{3d}}\bar{Y}_k$$

To determine the optimum values of n, m_d, f_{dk_2} and f_{3d}, following steps were proceed as follows,

The cost function in this case is,

$$C = C_{1d}(n_{2d}+h_{3d}) + C_{2d}n_{1d}m_d + C_{2d}\sum_{k=1}^{n_{2d}}m_{dk_1} + C_{3d}\left(\sum_{k=1}^{n_{2d}}h_{dk_2} + h_{3d}m_d\right)$$

where C, C_{1d}, C_{2d} and C_{3d} are same as defined earlier. The expected cost was,

$$E(C) = \frac{n}{N}[C_{1d}N_{2d} + C_{1d}\frac{N_{3d}}{f_{3d}} + C_{2d}N_{1d}m_d + C_{2d}\sum_{k=1}^{N_{2d}}\theta_{dk}m$$

To minimize the expected cost subject to fixed variance,following function was considered,

$\phi = E(C) + \lambda\{V(\bar{y}_{3dr}) - V_0\}$. To get close form expression of the optimum value, it was assumed that $m_{dk_2} = h_{dk_2}f_{2d}$, k=1, 2,..., n_{2d} in place of $m_{dk_2} = h_{dk_2}f_{dk_2}$, k=1, 2,...,n_{2d}. To overcome the problem arising due to simultaneous minimization of n, m_d, f_{2d}, f_{3d} it was assumed that n_{3d}=$f_{2d}h_{3d}$.

Thus, minimization gives the optimum values as,

$$n_{opt} = \frac{K_{27}}{K_{28}}, \quad m_{dopt} = \frac{-b_{21} \pm \sqrt{b_{21}{}^2 - 4a_{21}e_{21}}}{2a_{21}} \quad \text{and}$$

$$f_{2ddopt} = \frac{-b_{31} \pm \sqrt{b_{31}{}^2 - 4a_{31}e_{31}}}{2a_{31}}$$

To avoid the negative values, following values were considered

$$m_{dopt} = \frac{-b_{21} + \sqrt{b_{21}{}^2 - 4a_{21}e_{21}}}{2a_{21}} \quad \text{and}$$

$$f_{2dopt} = \frac{-b_{31} + \sqrt{b_{31}{}^2 - 4a_{31}e_{31}}}{2a_{31}}$$

where,

$$K_{28} = V_0 + \frac{N}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} + \frac{k_{12}}{N_d} S_{bd}^2,$$

$$\left[k_{12} = \left[\frac{N(N_d - 1)}{N_d(N-1)}\right]\right],$$

$$K_{27} = \frac{N}{N_d} k_{12} S_{bd}^2 + \frac{N^2}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{(N-1)}\right\} + \frac{N}{N_d^2}\left(\sum_{k=1}^{N_{1d}+N_{2d}} \left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2 + f_{3d} \sum_{k=1}^{N_{3d}} \left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2\right) + \frac{N}{N_d^2} \sum_{k=1}^{N_{2d}} \frac{(1-\theta_{dk})}{m_d} (f_{2d} - 1) S_{dk}^2 + \frac{NN_{3d}}{N_d^2} (f_{3d} - 1) S_{b_{N_{3d}}}^2,$$

$$a_{21} = C_{3d} \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) \left(N_{3d} S_{bN_{3d}}^2 - \sum_{k=1}^{N_{3d}} \frac{1}{M_d} S_{dk}^2\right),$$

$$b_{21} = C_{3d} \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) \sum_{k=1}^{N_{3d}} S_{dk}^2 - C_{3d} N_{3d} \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) S_{dk}^2,$$

$$e_{21} = -C_{1d} N_{3d} \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) S_{dk}^2,$$

$$e_{31} = -C_{3d} \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) \left(\sum_{k=1}^{N_{1d}+N_{2d}} S_{dk}^2 - \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) S_{dk}^2\right),$$

$$a_{31} = \left(C_{2d} N_{1d} + C_{2d} \sum_{k=1}^{N_{2d}} \theta_{dk}\right) \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) S_{dk}^2,$$

$$b_{31} = C_{3d} \{N_{3d} \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) S_{dk}^2 - \sum_{k=1}^{N_{2d}} (1-\theta_{dk}) \sum_{k=1}^{N_{3d}} S_{dk}^2\}$$

2.3 Control Estimator

This estimator was also considered for efficiency comparison purpose. Let n psus be selected from N psus by srswor design where n_d out of $n$psus fall in the d^{th} domain and within each selected psu, m_d ssus are also selected from M_d ssus by srswor. Data is collected through specialized efforts i.e. there is no nonresponse. Then unbiased estimator for $\bar{Y}_d$ is given as,

$$\bar{y}_{4dr} = \frac{N}{nN_d} \sum_{k=1}^{N_d} \bar{y}_{dk} \tag{7}$$

Proof: By definition,

$$E(\bar{y}_{4dr}) = E_1 E_2 E_3 \left(\frac{N}{nN_d} \sum_{k=1}^{n_d} \bar{y}_{dk}\right) = E_1 E_2 \left(\frac{N}{nN_d} \sum_{k=1}^{n_d} \bar{Y}_{dk}\right)$$

$$= E_1 \frac{N}{nN_d} \frac{n_d}{N_d} \sum_{k=1}^{N_d} \bar{Y}_{dk} \qquad \left[\because E_1(n_d) = \frac{nN_d}{N}\right]$$

$$= \frac{1}{N_d} \sum_{k=1}^{N_d} \bar{Y}_{dk} = \bar{Y}_d$$

Thus, $\bar{y}_{4dr}$ is an unbiased estimator of $\bar{Y}_d$.

where E_3 is conditional expectation pertaining to all possible samples of size m_d drawn from M_d and E_2 is expectation pertaining to all possible samples of size n_d drawn from N_d keeping n_d fixed and E_1 is the expectations arising out of randomness of n_d.

The variance of estimator was obtained as

$$V(\bar{y}_{4dr}) = V_1 E_2 E_3(\bar{y}_{4dr}) + E_1 V_2 E_3(\bar{y}_{4dr}) + E_1 E_2 V_3(\bar{y}_{4dr})$$

$$V_1 E_2 E_3(\bar{y}_{4dr}) = \frac{N(N-n)}{nN_d^2(N-1)} N_d Q_d \bar{Y}^2,$$

$$E_1 V_2 E_3(\bar{y}_{4dr}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)} S_{bd}^2,$$

$$E_1 E_2 V_3(\bar{y}_{4dr}) = \frac{N}{nN_d^2} \sum_{k=1}^{N_d} \left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2$$

Thus,

$$V(\bar{y}_{4dr}) = \frac{N(N-n)(N_d-1)}{nN_d^2(N-1)} S_{bd}^2 + \frac{N(N-n)}{nN_d^2(N-1)} N_d Q_d \bar{Y}_d^2 + \frac{N}{nN_d^2} \sum_{k=1}^{N_d} \left(\frac{1}{m_d} - \frac{1}{M_d}\right) S_{dk}^2, \tag{8}$$

S_{bd}^2, S_{dk}^2, Q_d etc. are defined earlier.

For optimization the relevant cost function in this case is,

$$C = C_{1d} n_d + C_{3d} n_d m_d$$

The expected cost $E(C) = \frac{n}{N}\left(C_{1d} N_d + C_{3d} N_d m_d\right)$

where, C, C_{1d}, C_{3d} have been defined earlier.

To obtain optimum values of n and m_d the expected cost was minimized by fixing the variance. The optimum values were obtained in the same way as earlier,

$$n_{opt} = \frac{\left[\frac{N^2}{N_d^2}\frac{(N_d-1)}{(N-1)}S_{bd}^2 + \frac{N^2}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{N-1}\right\} + \frac{N}{N_d^2}\sum_{k=1}^{N_d}\left(\frac{1}{m_d}-\frac{1}{M_d}\right)S_{dk}^2\right]}{\left(V_0 + \frac{N}{N_d^2}\left\{\frac{N_d Q_d \bar{Y}_d^2}{N-1}\right\} + \frac{N(N_d-1)}{N_d^2(N-1)}S_{bd}^2\right)} \text{ and}$$

$$m_{dopt} = \pm\sqrt{\frac{C_{1d}\sum_{k=1}^{N_d} S_{dk}^2}{C_{3d}\left(\frac{N(N_d-1)}{(N-1)}S_{bd}^2 + \left\{\frac{N N_d Q_d \bar{Y}_d^2}{N-1}\right\} - \frac{1}{M_d}\sum_{k=1}^{N_d} S_{dk}^2\right)}}$$

Negative value was avoided by considering

$$m_{dopt} = \sqrt{\frac{C_{1d}\sum_{k=1}^{N_d} S_{dk}^2}{C_{3d}\left(\frac{N(N_d-1)}{(N-1)}S_{bd}^2 + \left\{\frac{N N_d Q_d \bar{Y}_d^2}{N-1}\right\} - \frac{1}{M_d}\sum_{k=1}^{N_d} S_{dk}^2\right)}}$$

3. EMPIRICAL ILLUSTRATION

For empirical illustration data pertaining to MU 284 population given in Sarndal *et al.* (1992) was used and analysis was performed in SAS software. The variable considered was P85 which was 1985 human

Table 1. The optimum values of sample sizes along with percentage reduction in expected cost of $\bar{y}_{11r}, \bar{y}_{21r}, \bar{y}_{31r}$ over $\bar{y}_{41r}$ in domain 1

cost			Control ($\bar{y}_{41r}$)		Existing Estimator ($\bar{y}_{11r}$)					Proposed Estimator($\bar{y}_{21r}$)					Proposed Estimator($\bar{y}_{31r}$)				
								% RIEC					% RIEC					% RIEC	
C_{11}	C_{21}	C_{31}	n	m_1	n	m_1	f_{21}	5% CV	10% CV	n	m_1	f_{21}	5% CV	10% CV	n	m_1	f_{21}	5% CV	10% CV
25	1	45	46	4	46	9	4.28	74.32	74.32	63	4	1.95	97.42	97.41	63	5	2.69	94.51	94.50
25	1	50	46	4	46	9	4.51	74.99	74.99	63	4	2.06	97.49	97.48	63	5	2.80	94.87	94.85
25	1	55	46	4	46	9	4.73	75.57	75.57	63	4	2.16	97.55	97.54	63	5	2.91	95.17	95.16
25	2	45	46	4	46	8	3.02	67.98	67.98	63	3	1.38	96.96	96.95	63	5	2.01	92.84	92.82
25	2	50	46	4	46	8	3.19	68.77	68.77	63	3	1.45	97.04	97.03	63	5	2.11	93.34	93.32
25	2	55	46	4	46	8	3.34	69.46	69.46	63	3	1.52	97.10	97.09	63	5	2.19	93.76	93.74
30	1	45	46	4	46	9	4.28	74.31	74.31	63	4	1.95	97.38	97.37	63	6	2.69	94.24	94.22
30	1	50	46	4	46	9	4.51	75.01	75.01	63	4	2.06	97.46	97.45	63	6	2.80	94.60	94.58
30	1	55	46	4	46	9	4.73	75.61	75.61	63	4	2.16	97.52	97.51	63	5	2.91	94.91	94.89
30	2	45	46	4	46	9	3.02	68.63	68.63	63	3	1.38	96.96	96.95	63	6	2.01	92.49	92.47
30	2	50	46	4	46	9	3.19	69.43	69.43	63	3	1.45	97.05	97.04	63	6	2.11	93.00	92.98
30	2	55	46	4	46	9	3.34	70.13	70.13	63	3	1.52	97.12	97.11	63	5	2.19	93.43	93.41
35	1	45	46	5	46	9	4.28	74.18	74.18	63	4	1.95	97.34	97.33	63	7	2.69	94.00	93.98
35	1	50	46	4	46	9	4.51	74.89	74.89	63	4	2.06	97.41	97.41	63	6	2.80	94.37	94.35
35	1	55	46	4	46	9	4.73	75.51	75.51	63	4	2.16	97.48	97.47	63	6	2.91	94.69	94.67
35	2	45	46	5	46	9	3.02	68.99	68.99	63	3	1.38	96.96	96.95	63	7	2.01	92.20	92.17
35	2	50	46	4	46	9	3.19	69.81	69.81	63	3	1.45	97.04	97.03	63	6	2.11	92.72	92.69
35	2	55	46	4	46	9	3.34	70.52	70.52	63	3	1.52	97.11	97.10	63	6	2.19	93.16	93.13

Table 2. The optimum values of sample sizes along with percentage reduction in expected cost of $\bar{y}_{12r}, \bar{y}_{22r}, \bar{y}_{32r}$ over $\bar{y}_{42r}$ in domain 2

cost			Control ($\bar{y}_{41r}$)		Existing Estimator ($\bar{y}_{11r}$)					Proposed Estimator($\bar{y}_{21r}$)					Proposed Estimator($\bar{y}_{31r}$)				
								% RIEC					% RIEC					% RIEC	
C_{12}	C_{22}	C_{32}	n	m_2	n	m_2	f_{22}	5% CV	10% CV	n	m_2	f_{22}	5% CV	10% CV	n	m_2	f_{22}	5% CV	10% CV
25	1	45	52	9	52	3	1.52	77.96	77.96	64	6	1.91	97.88	97.88	64	3	1.91	93.90	93.88
25	1	50	52	8	52	3	1.60	78.83	78.83	64	6	2.01	97.93	97.93	64	2	1.96	94.30	94.28
25	1	55	52	8	52	3	1.68	79.58	79.58	64	6	2.11	97.98	97.98	64	2	2.00	94.63	94.62
25	2	45	52	9	52	2	1.07	76.63	76.63	64	4	1.35	97.62	97.62	64	3	1.56	93.01	93.00
25	2	50	52	8	52	2	1.13	77.51	77.51	64	4	1.42	97.68	97.68	64	2	1.62	93.52	93.50
25	2	55	52	8	52	2	1.18	78.27	78.27	64	4	1.49	97.73	97.73	64	2	1.66	93.93	93.92
30	1	45	52	10	52	3	1.52	76.63	76.63	64	6	1.91	97.83	97.83	64	3	1.91	93.33	93.32
30	1	50	52	9	52	3	1.60	77.56	77.56	64	6	2.01	97.89	97.89	64	3	1.96	93.77	93.76
30	1	55	52	9	52	3	1.68	78.37	78.37	64	6	2.11	97.94	97.93	64	3	2.00	94.14	94.12
30	2	45	52	10	52	3	1.07	75.43	75.43	64	5	1.35	97.59	97.59	64	3	1.56	92.37	92.35
30	2	50	52	9	52	3	1.13	76.38	76.38	64	5	1.42	97.66	97.65	64	3	1.62	92.92	92.90
30	2	55	52	9	52	3	1.18	77.20	77.20	64	5	1.49	97.71	97.71	64	3	1.66	93.37	93.36
35	1	45	52	11	52	3	1.52	75.38	75.38	64	6	1.91	97.79	97.78	64	4	1.91	92.81	92.80
35	1	50	52	10	52	3	1.60	76.38	76.38	64	6	2.01	97.85	97.84	64	3	1.96	93.29	93.27
35	1	55	52	10	52	3	1.68	77.24	77.24	64	6	2.11	97.90	97.89	64	3	2.00	93.68	93.67
35	2	45	52	11	52	3	1.07	74.28	74.28	64	5	1.35	97.56	97.56	64	4	1.56	91.77	91.75
35	2	50	52	10	52	3	1.13	75.29	75.29	64	5	1.42	97.63	97.62	64	3	1.62	92.36	92.35
35	2	55	52	10	52	3	1.18	76.17	76.17	64	5	1.49	97.69	97.68	64	3	1.66	92.86	92.84

Table 3. The optimum values of sample sizes along with percentage reduction in expected cost of $\bar{y}_{1dr}, \bar{y}_{2dr}, \bar{y}_{3dr}$ over $\bar{y}_{4dr}$ in domain 3

cost			Control ($\bar{y}_{41r}$)		Existing Estimator ($\bar{y}_{11r}$)					Proposed Estimator($\bar{y}_{21r}$)					Proposed Estimator($\bar{y}_{31r}$)				
								% RIEC					% RIEC					% RIEC	
C_{12}	C_{22}	C_{32}	n	m_3	n	m_3	f_{23}	5% CV	10% CV	n	m_3	f_{23}	5% CV	10% CV	n	m_3	f_{23}	5% CV	10% CV
25	1	45	50	12	50	5	1.52	81.54	81.54	64	10	1.56	97.43	97.42	65	36	8.89	90.12	90.10
25	1	50	50	12	50	5	1.60	82.20	82.20	64	10	1.64	97.47	97.47	65	36	9.81	90.59	90.56
25	1	55	50	11	50	5	1.68	82.78	82.78	64	10	1.72	97.52	97.51	65	36	10.72	90.99	90.96
25	2	45	50	12	50	4	1.07	79.63	79.63	64	8	1.10	96.88	96.87	64	36	4.73	83.06	83.01
25	2	50	50	12	50	4	1.13	80.32	80.32	64	8	1.16	96.94	96.93	64	36	5.20	83.79	83.75
25	2	55	50	11	50	4	1.18	80.91	80.91	64	8	1.22	96.99	96.98	64	36	5.66	84.44	84.40
30	1	45	50	13	50	5	1.52	80.69	80.69	64	10	1.56	97.45	97.44	65	36	8.89	90.62	90.59
30	1	50	50	13	50	5	1.60	81.41	81.41	64	10	1.64	97.50	97.49	65	36	9.81	91.06	91.03
30	1	55	50	12	50	5	1.68	82.03	82.03	64	10	1.72	97.54	97.53	65	36	10.72	91.44	91.42
30	2	45	50	13	50	4	1.07	78.98	78.98	64	9	1.10	96.95	96.94	64	36	4.73	84.17	84.12
30	2	50	50	13	50	4	1.13	79.71	79.71	64	9	1.16	97.01	97.00	64	36	5.20	84.86	84.82
30	2	55	50	12	50	4	1.18	80.35	80.35	64	9	1.22	97.06	97.05	64	36	5.66	85.46	85.42
35	1	45	50	14	50	5	1.52	79.87	79.87	64	11	1.56	97.45	97.45	65	36	8.89	90.97	90.95
35	1	50	50	14	50	5	1.60	80.63	80.63	64	11	1.64	97.50	97.50	65	36	9.81	91.40	91.37
35	1	55	50	13	50	5	1.68	81.29	81.29	64	11	1.72	97.55	97.54	65	36	10.72	91.77	91.74
35	2	45	50	14	50	4	1.07	78.30	78.30	64	9	1.10	97.00	96.99	64	36	4.73	85.00	84.96
35	2	50	50	14	50	4	1.13	79.08	79.08	64	9	1.16	97.06	97.05	64	36	5.20	85.66	85.62
35	2	55	50	13	50	4	1.18	79.76	79.76	64	9	1.22	97.11	97.10	64	36	5.66	86.23	86.20

population (in thousands) of 284 municipalities of Sweden. Using this data a population with N=27 was generated by combining the adjacent 10 units and allocating them to the respective psus. From the Npsus a sample of n=21 psus each of size M_d=10 was drawn using srswor. The whole population was divided into three domains. Different values of θ_{dk} were considered for each psu in each domain and various combinations of C_{1d}, C_{2d}, and C_{3d} were considered. The percentage reduction in expected cost of $\bar{y}_{1dr}, \bar{y}_{2dr}, \bar{y}_{3dr}$ over $\bar{y}_{4dr}$ along with optimum values of sample sizes for different values of C_{1d}, C_{2d}, and C_{3d} are given in Table 1 for domain 1, in Table 2for domain 2 and in Table 3for domain 3. Two values of CV i.e. 5% and 10% were considered.

4. RESULT AND DISCUSSION

A close perusal of Tables1 to Table 3reveals that the %RIEC is maximum in estimator $\bar{y}_{21r}$ followed by the estimator $\bar{y}_{31r}$, and it is least in the estimator $\bar{y}_{11r}$.

In almost all the three estimators the %RIEC increases with the increase in the cost per unit of collecting the information by expensive method after the first attempt to obtain information failed. A close perusal of all the tables shows that all the estimators of domain mean in the presence of nonresponse, based on sub-sampling of the non-respondents, when domain size is known in advance, are better than an estimator based only on the interview method and having 100% response at the second stage. Among all three estimators, for all the domains, %RIEC is maximum in estimator $\bar{y}_{21r}$ followed by the estimator $\bar{y}_{31r}$, and it is least in the estimator $\bar{y}_{11r}$. The %RIEC is maximum for the estimator $\bar{y}_{21r}$ because the nonresponse is minimum in this estimator.

5. CONCLUSION

It is visible from the results from all the table that proposed estimator performs better than the control estimator as well as the estimator developed by Aditya (2016) with respect to %RIEC. It is worthwhile to mention that the results agree with Durbin (1954) i.e. sub-sampling will show a marked profit if C_{3d} is not large as compared to ($C_{1d} + C_{2d}W_{1d}$), where W_{1d} is the response rate. Hence,it can be concluded that all the proposed estimators are more efficient than the existing estimators in the presence of nonresponse under two stage sampling design when the objective of the study is to estimate the domain mean.

Practical Applicability and Limitation

The proposed estimator is suitable for real world surveys where limited follow-up is preferred due to budget, time, or logistical constraints. However, its effectiveness relies on the practical feasibility of follow-up for the selected sub-sample and the availability of reliable cost information for accurate optimization.

REFERENCES

Aditya, K. (2016). Estimation of domain mean using two-stage sampling in the presence of non-response. *International Journal for Research in Social Sciences,* **6**(5), 496-511.

Aditya, K., Sud, U.C., and Chandra, H. (2012). Estimation of domain total for unknown domain size in the presence of nonresponse. *Statistics and Applications (New Series),* **10**(1-2), 13-25.

Agrawal, M.C., and Midha, C.K. (2007). Some efficient estimators of the domain parameters. *Statistics & Probability Letters,* **77,** 704-709.

Chaudhary, M.K., Prajapati, A., and Ray, B.K. (2021). Some convex-type classes of estimators of finite population mean under random non-response. *Journal of Information and Optimization Sciences,* **42**(8), 1951-1965.

Chaudhary, M.K., Ray, B.K., Vishwakarma, G.K., and Kadilar, C. (2023). A calibration-based approach on estimation of mean of a stratified population in the presence of non-response. *Communications in Statistics – Theory and Methods,* **77**(3). https://doi.org/10.1080/03610926.2023.2257818

Chhikara, R.S., and Sud, U.C. (2009). Estimation of population and domain totals under two-phase sampling in the presence of non-response. *Journal of the Indian Society of Agricultural Statistics,* **63(**3), 297-304.

Durbin, J. (1954). Nonresponse and call-backs in surveys. *Bulletin of the International Statistical Institute,* **34**, 72-86.

Foradori, G.T. (1961). Some non-response sampling theory for two-stage designs. Institute of Statistics, North Carolina State College.

Hansen, M.H., and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association,* **41,** 517-529.

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology,* **12,** 1-16.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustment for unit non-response. In W.G. Madow, I. Olkin, & B. Rubin (Eds.), *Incomplete data in sample surveys* (Vol. 2, pp. 143-184). New York: Academic Press.

Okafor, F.C. (2001). Treatment of non-response in successive sampling. *Statistica,* **61**(2), 195-204.

Okafor, F.C. (2005). Sub-sampling the non-respondents in two-stage sampling over successive occasions. *Journal of the Indian Statistical Association,* **43**(1), 33-49.

Okafor, F.C., and Lee, H. (2000). Double sampling for ratio and regression estimation with sub-sampling the non-respondents. *Survey Methodology,* **26**(2), 183-188.

Raman, R.K., Sud, U.C., Chandra, H., and Gupta, V.K. (2013). Calibration estimator of population total with sub-sampling of non-respondents. *Journal of the Indian Society of Agricultural Statistics,* **67**(3), 329-337.

Rancourt, E., Lee, H., and Särndal, C.E. (1994). Bias corrections for survey estimates from data with ratio imputed values for confounded non-response. *Survey Methodology,* **20**, 137-147.

Sarndal, C.E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Singh, G., Jaiswal, A. and Khan, M. (2024). Some improved estimation procedures of population mean under non-response in sample survey. *Authorea.*https://doi.org/10.22541/au.170669774.44050980/v1

Srinath, K.P. (1971). Multiphase sampling in non-response problems. *Journal of the American Statistical Association,* **66**, 583-586.

Sud, U.C., Chandra, H., and Chhikara, R.S. (2010). Domain estimation in the presence of non-response. *Journal of the Indian Society of Agricultural Statistics,* **64**, 343-347.

Tripathi, T.P., and Khare, B.B. (1997). Estimation of mean vector in presence of non-response. *Communications in Statistics – Theory and Methods,* **26**(9), 2255-2269.

Vyshna, I.C., Aditya, K., Nasrin, S., Bharti, B., Singh, D., Biswas, A., and Ahmed, T. (2025). Calibration estimation of population total by double use of auxiliary information in two-stage sampling design. *Journal of the Indian Society for Probability and Statistics.* https://doi.org/10.1007/s41096-025-00242-y