JIRSS (2026)
Vol. 25, No. 1, pp 104-126
DOI: 10.22034/jirss.2026.1999131.1006

Handling non-response in the presence of extreme values

Syed Abdul Rehman[1], **Javid Shabbir**[2]

[1]Department of Mathematical Sciences, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan.
[2]Department of Statistics, University of Wah, Wah Cantt, Pakistan.

Received: 2023-03-28, Accepted: 2026-07-11, Published online: 2026-07-16

Abstract. Non-response and extreme values are two major issues faced by surveyors at the survey and estimation stage; they lead to biased and inefficient estimators. This study proposes four unbiased estimators of the finite population mean in the presence of extreme values when non-response occurs in surveys. We consider two robust ranked set sampling (RSS) designs, called improved paired RSS (IPRSS) and median RSS (MRSS) for the derivation of the proposed estimators. Expressions are provided for the variance of the proposed estimator, and it is mathematically proved that these estimators are more efficient than the Hansen naïve model. Additionally, the efficiency conditions are provided in comparison with the Bouza model of handling non-response in RSS. We conduct a comprehensive simulation study to observe the relative efficiency of the proposed estimators based on hypothetical normal data. For applications in real life, we also consider heart failure data for simulations.

Keywords. Extreme value theory, concomitant variable, order statistics, Monte Carlo simulation, relative efficiency.

MSC: 62D05, 62G05, 62F12.

Corresponding Author: Syed Abdul Rehman (arehman@stat.qau.edu.pk).
Javid Shabbir (javidshabbir@gmail.com)

1 Introduction

In survey research, one of the biggest difficulties faced by a surveyor is non-response. Non-response refers to when one or more units from the selected sampling frame do not respond to the survey. Since it violates the basic and necessary assumptions of the sampling method being used, a less representative sample and biased measures are caused by non-response, which are two big problems that lead to imprecise and unreliable statistical inference. When data is missing at random (MAR), a lower response rate has no effect on the non-response error. However, a lower response rate raises the risk of larger non-response bias. The problem of non-response was first addressed by Hansen and Hurwitz (1946) who proposed a pioneer model which is explained later in this paper.

In statistical analysis, the reliability and precision of an estimate are both seriously affected by the extreme values if present in a population. The theory of extreme values (EVT) measures the frequency and size of such events. The Weibull distribution is an example of an EVT model which is part of the generalised extreme value (GEV) distributions that fits the tails of a distribution by using the shape, location, and scale factors. EVT illustrates maxima (or minima) from large samples, in which each extreme value is considered independent. Maxima can be the highest daily discharge for the river flow rate in a year from decades. In public health, extreme values could include a big heatwave, a pandemic, or an unintentional toxic exposure. In engineering, EVT is used for dams, bridges, highways, nuclear power plants and other applications. Additional ecological examples in application are excessive pesticide use, heavy metal intake in the diet and pollution deposition in surface soils. These observations, on the other hand, could be due to instrument fault, coding error, defect, or other problems. Outliers have been detected and minimized by using a range of strategies, including the use of median-like robust statistics and lognormal distribution. The specific threshold or percentile extrema are used to identify extrema when we have a small sample size. This method ensures statistical validity with the identification of "actual" extreme values. The two highest 5% to 10% observations are chosen in practice. EVT is being used to forecast the occurrence of an extensive range of disasters, including droughts, flooding, and financial collapses.

Estimation of population parameters is thought of as imprecise under simple random sampling (SRS) when extreme values exist in a population. In case of non-response it becomes more problematic, especially when the non-responding units have a pattern other than that followed by the responding units. This appeals to the fact that a sample selection technique must be chosen wisely since it affects the precision of estimates. Sampling methods other than SRS are preferred by experts in such cases, among which the ranked set sampling (RSS) procedure is believed to be more efficient. RSS increases the precision of estimates by reducing the sampling errors, and it has the advantage that it can utilize the ranks of the auxiliary variable to collect a sample. The RSS technique was first introduced by McIntyre (1952) where units of the study variable are ordered by either visual judgement or by some cheap quantitative measures. Takahasi and Wakimoto (1968) developed an unbiased estimator of the population mean under the RSS technique. Stokes (1977) exposed an important use of the auxiliary variable by using the ranks of the auxiliary variable to order the units of the study variable. Rehman and Shabbir (2021) proposed an improved paired RSS (IPRSS) sampling procedure which is robust to extreme values in the data. The double RSS (DRSS) procedure was proposed byAl-Saleh and Al-Kadiri (2000), while the double MRSS (DMRSS) procedure was proposed by Samawi and Tawalbeh (2002). Haq et al. (2016) proposed the double PRSS (DPRSS). Singh et al. (2014), Abu-Dayyeh et al. (2003), Kadilar and Cingi (2005),Malik and Singh (2013), Singh and Sharma (2015), Fatima et al. (2022) and Muneer et al. (2017) can also be seen for the use of the auxiliary variable. Several authors have claimed that use of the auxiliary variable at the designing stage may increase the precision of estimates.

Since non-response combined with the presence of outliers can substantially reduce the efficiency of classical RSS estimators. This motivates us to develop estimators based on robust RSS designs that limit the impact of outliers and extreme observations, thereby improving stability and reliability. In his study, we investigate the estimation of the finite population mean in the presence of non-response under MRSS and IPRSS, which are robust RSS designs for extreme values. We assume that the proposed estimators will enhance the precision of estimates since these estimators are relying on robust data collection methods.

2 Methodology

We consider the naive model of Hansen and Hurwitz (1946), according to which we draw a sample s of size n_s by using the SRSWOR method from a finite population Ω of size N. Let n_1 units respond to the survey at the first attempt, called the response group. The units that do not respond at the first attempt ($n_2 = n - n_1$) are called the non-response group. Special efforts are made to approach the non-responding units, and a part of them $\left(n_2' = n_2/k; k > 1\right)$ is included in the sample. Thus, we get a sample of size $n = n_1 + n_2'$ for estimation purposes. This allows the entire population to be separated into two complementary groups.

Hansen and Hurwitz (1946) proposed the following unbiased estimator for population mean:

$$\bar{y}^*_{HH} = w_1\bar{y}_{1srs} + w_2\bar{y}'_{2srs} \tag{2.1}$$

where

$$\bar{y}_{1\text{srs}} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i, \qquad \bar{y}'_{2\text{srs}} = \frac{1}{n_2'}\sum_{i=1}^{n_2'} y_i', \qquad w_i = \frac{n_i}{n}; \quad i = 1, 2.$$

The variance of $\bar{y}^*_{HH}$ is given by,

$$V\left(\bar{y}^*_{HH}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma_y^2 + \frac{W_2\left(K-1\right)}{n}\sigma_{y2}^2 \tag{2.2}$$

where $(Y_i)_{i=1}^{N}$ shows the characteristics under study, K shows the non-response rate and some other notations are defined as,

$$\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_i,\ \bar{Y}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} Y_i,\ W_i = \frac{N_i}{N},\ \bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,\ \sigma_{y2}^2 = \frac{1}{N_2-1}\sum_{i=1}^{N}(Y_{2i} - \bar{Y}_2)^2$$

2.1 Assumptions and Limitations

The estimators proposed in this study are developed considering some standard regularity assumptions. First, we assume that the ranking criteria for ranking the units are imperfect. This implies that the order of values of a concomitant variable is used to rank the values of the study variable. Additionally, since the efficiency of the estimators depends on a moderate to strong association between the study variable and the auxiliary variable, we assumed low, moderate and high correlation coefficients between the study variable and the auxiliary variable to know their impacts on relative efficiency. The non-response mechanism is assumed to be Missing at Random (MAR), implying that the probability of non-response depends only on observed information and not on unobserved values of the study variable. In addition, the underlying population is assumed to be approximately symmetric with a finite variance, which is necessary for the validity of the variance expressions and asymptotic properties of the estimators.

We also assume small sample sizes in our study; this is due to the fact that the RSS theory is fundamentally designed for efficient estimation with small sample sizes. The proposed estimators assume that non-response is not extreme because under very high non-response rates ($K > 2.5$), the efficiency of RSS may not be gained, and inference may become less reliable.

The gain in efficiency of estimators using RSS samples depends on the quality of ranking. Our study and simulation in Section 6 assume imperfect ranking; practitioners should assess the strength of the association between the study variable and the auxiliary before applying these methods. In case of imperfect ranking, reduction in efficiency may arise from a weak correlation between the study and auxiliary variables. In practice, experts recommend a stronger correlation, particularly greater than 0.5 for effective ranking. Additionally, we have provided Table 9 at the end of the simulation study, where readers can find the practical guidance for choosing the best estimator with respect to the correlation coefficient between the study variable and the concomitant variable.

3 Handling Non-response under RSS

Bouza-Herrera (2013) employed the RSS technique to collect the sample from the non-responding group and proved that it can provide more precise estimates of population mean than using the classical SRS method. proposed the following estimator of population mean:

$$\bar{y}^*_{Bouza} = w_1 \bar{y}_{srs} + w_2 \bar{y}'_{2rss} \tag{3.1}$$

where $\bar{y}_{srs}$ is the sample mean of the response group depending on n_1 units collected via the SRS method, while $\bar{y}'_{2rss} = \frac{1}{n'_2} \sum_{j=1}^{r} \sum_{i=1}^{\nu} Y_{i(i)j}$ is the sample mean of the non-response group depending on n'_2 units collected via the RSS method. The RSS method is described as:

Initially identify ν^2 units from a population and distribute them in ν independent sets, each consisting of ν units. Rank the units in each set by using ranks of a closely related auxiliary variable, and select the lowest judgement order statistics from the first set, the second lowest judgement order statistics from the second set, and keep selecting units until ν^{th} judgement order statistics are selected from the ν^{th} set. The selected units in the sample can be represented as $\left(Y_{i[i]j}\right)_{i=1}^{\nu}$, where $j = 1, 2, ..., r$ represents the number of times the RSS procedure is repeated to obtain a sample of size $r\nu$.

Bouza-Herrera (2013) showed that $\bar{y}^*_{Bouza}$ is an unbiased estimator of population mean with the following variance,

$$V\left(\bar{y}^*_{Bouza}\right) = V\left(\bar{y}^*_{srs}\right) - \frac{W_2 K}{r\nu^2} \sum_{i=1}^{\nu} \Delta^2_{2y[i]} \tag{3.2}$$

where $\Delta_{2y[i]} = \mu_{2y[i]} - \mu_{2y}$, μ_{2y} is the true mean of the study variable of the non-responding group, and $\mu_{2y[i]}$ shows the mean of the i^{th} judgement order of the study variable in the same group.

It can be observed that the variance of $\bar{y}^*_{Bouza}$ is smaller than the variance of $\bar{y}^*_{srs}$ by a factor $W_2 \left(\frac{K}{r\nu^2} \sum_{i=1}^{\nu} \Delta^2_{2y[i]}\right)$. which follows that, if $\mu_{2[i]} \neq \mu_2$ so, it is obvious that $V\left(\bar{y}^*_{Bouza}\right) < V\left(\bar{y}^*_{srs}\right)$.

4 Robust RSS designs to extreme values

The theory of survey sampling provides several robust data collection procedures for the situations in which a population being studied is suspected of containing extreme values. Here we consider the following two robust RSS sampling designs:

4.1 Improved Paired RSS (IPRSS)

This method is developed by Rehman and Shabbir (2021) according to which a sample is selected by the following method:

1. If it ν is even, then identify $(\nu/2)$ independent sets each of size $(\nu + 2)$ from the population. The units are ranked in each set by using ranks of the concomitant variable, and the smallest and second largest ranked units are selected from the

first set, and the third smallest and third-largest units are selected from the second set. Keep selecting the units in this way till $\{(\nu+2)/2\}$ and $\{(\nu+4)/2\}$ units are selected from the final set. Samples collected by this method can be represented as $\left(Y_{i[i+1]j}, Y_{i[m+2-i]j}\right)_{i=1}^{\nu+2}$.

2. If ν is odd, then identify $(\nu+1)/2$ sets each of size $(\nu+2)$. The units in each set are ranked by using the ranks of a concomitant variable. The second smallest and second largest ranked units are selected from the first set, and the third smallest and third largest units are selected from the second set. The paired units are chosen in this way until $((\nu-1)/2)$ set. Samples collected by this method can be represented as $\left(Y_{i[i+1]j}, Y_{i[\nu+2-i]j}, Y_{i[\nu+1/2]j}\right)_{i=1}^{\nu+2}$,

where $j = 1, 2, ..., r$ shows the number of times this procedure is repeated to obtain a sample of size $r\nu$. The corresponding mean estimator and their variances for even and odd sample sizes are given below as:

$$\bar{y}_{IPRSS}^{E} = \frac{1}{r\nu}\sum_{j=1}^{r}\sum_{i=1}^{\nu/2}\left(Y_{i[i+1]j} + Y_{i[\nu+2-i]j}\right) \tag{4.1}$$

$$V\left(\bar{y}_{IPRSS}^{E}\right) = \frac{\sigma_y^2}{r\nu} - \frac{1}{r\nu^2}\sum_{i=1}^{\nu/2}\left(\Delta_{y[i+1]}^2 + \Delta_{y[\nu+2-i]}^2 + 2\Delta_{y(i+1)}\Delta_{y[\nu+2-i]}\right) \tag{4.2}$$

and

$$\bar{y}_{IPRSS}^{O} = \frac{1}{r\nu}\sum_{j=1}^{r}\left(\sum_{i=1}^{\nu+1/2} Y_{i[i+1]j} + \sum_{i=1}^{\nu-1/2} Y_{i[\nu+2-i]j}\right) \tag{4.3}$$

$$V\left(\bar{y}_{IPRSS}^{O}\right) = \frac{\sigma_y^2}{r\nu} - \frac{1}{r\nu^2}\left(\sum_{i=1}^{\nu+1/2}\left(\Delta_{y[i+1]}^2 + \Delta_{y[\nu+2-i]}^2\right) + 2\sum_{i=1}^{\nu-1/2}\Delta_{y[i+1]}\Delta_{y[\nu+2-i]}\right) \tag{4.4}$$

4.2 Median RSS (MRSS)

The MRSS was developed by Muttlak (1998) which select sample units in the following method;
Initially identify ν^2 units from a population and distribute them into ν independent sets each having ν units. The units in each set are ordered by using ranks of a closely related concomitant variable, and then:

1. If ν is even, then select the $(\nu/2)^{th}$ units from the first $(\nu/2)$ sets, while $((\nu+2)/2)^{th}$ units are selected from the remaining sets. Samples collected by this method can be represented as $\left(\left(Y_{i[\nu/2]j}\right)_{i=1}^{\nu/2}, \left(Y_{i[\nu+2/2]j}\right)_{i=\nu+2/2}^{\nu}\right)$.

2. If ν is odd, select $(\nu+1/2)^{th}$ unit from each set which provides a sample represented by $\left(Y_{i[\nu+1/2]j}\right)_{i=1}^{\nu}$.

where $j = 1, 2, ..., r$ shows the number of times this procedure is repeated to obtain a sample of size $r\nu$. For even and odd sample sizes, the corresponding mean estimators and their variances are as follows:

$$\bar{y}^{E}_{MRSS} = \frac{1}{r\nu}\sum_{j=1}^{r}\left(\sum_{i=1}^{\nu/2} Y_{i[\nu/2]j} + \sum_{i=\nu+2/2}^{\nu} Y_{i[\nu+2/2]j}\right) \tag{4.5}$$

$$V\left(\bar{y}^{E}_{MRSS}\right) = \frac{\sigma^2_y}{r\nu} - \frac{1}{r\nu^2}\left(\sum_{i=1}^{\nu/2}\Delta^2_{y(\nu/2)} + \sum_{i=\nu+2/2}^{\nu}\Delta^2_{y(\nu+2/2)}\right) \tag{4.6}$$

and

$$\bar{y}^{O}_{MRSS} = \frac{1}{r\nu}\sum_{j=1}^{r}\sum_{i=1}^{\nu} Y_{i[\nu+1/2]j} \tag{4.7}$$

$$V\left(\bar{y}^{O}_{MRSS}\right) = \frac{\sigma^2_y}{r\nu} - \frac{1}{r\nu^2}\sum_{i=1}^{\nu}\Delta^2_{y[\nu+2/2]} \tag{4.8}$$

5 Proposed Estimators under robust RSS designs

In this section we propose four unbiased estimators of population mean under non-response by utilizing the robust RSS designs discussed above. Expressions for the variances are given in each case and the condition under which the proposed estimators perform better than the Bouza-Herrera (2013) estimator, are also given.
Note that in our study we will use the fixed set size ν for the execution of RSS designs from both the response and non-response groups. While the number of cycles r varies in the two groups. Additional detail of notations is given below.
$n_1 = r_1\nu,\ n_2 = r_2\nu,\ n'_2 = r'_2\nu,\ r = r_1 + r'_2,\ n = r\nu,\ k = \frac{n_2}{n'_2} = \frac{r_2}{r'_2}.$

5.1 Proposed estimator I

Consider we collect data from the non-response group at the second attempt by using the method of the IPRSS technique; then the proposed estimator-I for even and odd sizes can be written as:

$$\bar{y}^{*E}_{P1} = w_1\bar{y}_{1srs} + w_2\bar{y}'^{E}_{2iprss} \tag{5.1}$$

and

$$\bar{y}^{*O}_{P1} = w_1\bar{y}_{1srs} + w_2\bar{y}'^{O}_{2iprss} \tag{5.2}$$

Lemma 5.1. *The variance of $\bar{y}^{*E}_{P1}$ is more precise than $\bar{y}^{*}_{srs}$.*

Proof: *Eq (13) can also be written as,*

$$\bar{y}^{*E}_{P1} = \left[(w_1\bar{y}_{1srs} + w_2\bar{y}_{2srs})/s\right] + \left[w_2\left(\bar{y}'^{E}_{2iprss} - \bar{y}_{2srs}\right)\Big/s_2\right] \tag{5.3}$$

The term $(w_1\bar{y}_{1srs} + w_2\bar{y}_{2srs})/s$ represents the sample mean based on n units selected by the SRS method, which has the variance $\sigma_y^2\big/n$. Thus, the conditional variance of $\bar{y}_{P1}^{*E}$ can be written as,

$$V\left(\bar{y}_{P1}^{*E}\right) = \frac{\sigma_y^2}{n} + w_2^2 V\left(\frac{\bar{y}'_{2,\text{iprss}} - \bar{y}_{2,\text{srs}}}{s_2}\right). \tag{5.4}$$

Variance for the explicit function of the second term in the R.H.S., relying on units of the non-response group, is

$$\begin{aligned} V\Big(\bar{y}_{2iprss}^{\prime E} - \bar{y}_{2srs}\Big)\Big/s_2 &= E\Big(\bar{y}_{2iprss}^{\prime E} - \bar{y}_{2srs}\Big)^2 \\ &= E\Big(\Big(\bar{y}_{2iprss}^{\prime E} - \mu_2\Big) - (\bar{y}_{2srs} - \mu_2)\Big)^2 \\ &= E\Big(\Big(\bar{y}_{2iprss}^{\prime E} - \mu_2\Big)^2 + (\bar{y}_{2srs} - \mu_2)^2 - 2\left(\bar{y}_{2iprss}^{\prime E} - \mu_2\right)(\bar{y}_{2srs} - \mu_2)\Big)^2 \end{aligned}$$

Conditioning third term to sample of size n_2, we get

$$\begin{aligned} V\left(\bar{y}_{2iprss}^{\prime E} - \bar{y}_{2srs}\right) &= E\left(\left(\bar{y}_{2iprss}^{\prime E} - \mu_2\right)^2 - (\bar{y}_{2srs} - \mu_2)^2\right) \\ &= \left(V\left(\bar{y}_{2iprss}^{\prime E}\right) - V\left(\bar{y}_{2srs}\right)\right) \\ &= \frac{\sigma_{2y}^2}{n'_2} - \frac{1}{n'_2\nu}\sum_{i=1}^{\nu/2}\left(\Delta_{2y[i+1]}^2 + \Delta_{2y[\nu+2-i]}^2 + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right) - \frac{\sigma_{2y}^2}{n_2} \\ &= \frac{k\sigma_{2y}^2}{n_2} - \frac{\sigma_{2y}^2}{n_2} - \frac{1}{n'_2\nu}\sum_{i=1}^{\nu/2}\left(\Delta_{2y[i+1]}^2 + \Delta_{2y[\nu+2-i]}^2 + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right) \end{aligned}$$

The expected value after multiplying both sides by w_2^2 is,

$$\begin{aligned} w_2^2 V\left(\bar{y}_{2iprss}^{\prime E} - \bar{y}_{2srs}\right) = {} & \frac{W_2(K-1)}{n}\sigma_{2y}^2 - \frac{KW_2}{n\nu}\sum_{i=1}^{\nu/2}\left(\Delta_{2y[i+1]}^2 + \Delta_{2y[\nu+2-i]}^2 + 2\Delta_{2y[i+1]}\right) \\ & - \frac{KW_2}{n\nu}\sum_{i=1}^{\nu/2}\left(\Delta_{2y[\nu+2-i]}\right) \end{aligned}$$

Thus, the variance of $\bar{y}_{P1}^{*}$ becomes,

$$V\left(\bar{y}_{P1}^{*E}\right) = \frac{\sigma_y^2}{r\nu} + \frac{W_2(K-1)}{n}\sigma_{2y}^2 - \frac{W_2K}{n\nu}\sum_{i=1}^{\nu/2}\left(\Delta_{2y[i+1]}^2 + \Delta_{2y[\nu+2-i]}^2 + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right)$$

or

$$V\left(\bar{y}_{P1}^{*E}\right) = V\left(\bar{y}_{srs}^{*}\right) - \frac{W_2K}{n\nu}\sum_{i=1}^{\nu/2}\left(\Delta_{2y[i+1]}^2 + \Delta_{2y[\nu+2-i]}^2 + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right). \tag{5.5}$$

Eq (5.5) shows that $V\left(\bar{y}^{*E}_{P1}\right) \leq V\left(\bar{y}^{*}_{srs}\right)$. Also, if we compare Eq (4) and Eq (17), then $V\left(\bar{y}^{*E}_{P1}\right) < V\left(\bar{y}^{*}_{Bouza}\right)$ if the following condition holds,

$$\sum_{i=1}^{\nu/2}\left(\Delta^2_{2y[i+1]} + \Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right) > \sum_{i=1}^{\nu}\Delta^2_{2y[i]}$$

Similarly, we can derive the variance of $\bar{y}^{*O}_{P1}$ as,

$$V\left(\bar{y}^{*O}_{P1}\right) = V\left(\bar{y}^{*}_{srs}\right) - \frac{W_2K}{r\nu^2}\left(\sum_{i=1}^{\nu+1/2}\Delta^2_{2y[i+1]} + \left(\sum_{i=1}^{\nu-1/2}\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right)\right), \quad (5.6)$$

which also shows that $V\left(\bar{y}^{*O}_{P4}\right) \leq V\left(\bar{y}^{*}_{srs}\right)$. Also, if we compare Eq (4) and Eq (18), then $V\left(\bar{y}^{*O}_{P1}\right) < V\left(\bar{y}^{*}_{Bouza}\right)$ if the following condition holds,

$$\sum_{i=1}^{\nu+1/2}\Delta^2_{2y[i+1]} + \left(\sum_{i=1}^{\nu-1/2}\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right) > \sum_{i=1}^{\nu}\Delta^2_{2y[i]}$$

5.2 Proposed estimator II

It is an extension to the proposed estimator-I i.e., we collect the data from both groups by the method of IPRSS. The proposed estimator for even and odd sizes can be written as,

$$\bar{y}^{*E}_{P2} = w_1\bar{y}_{1iprss} + w_2\bar{y}'^{E}_{2iprss}, \quad (5.7)$$

and

$$\bar{y}^{*O}_{P2} = w_1\bar{y}_{1iprss} + w_2\bar{y}'^{O}_{2iprss}. \quad (5.8)$$

Both estimators $\bar{y}^{*E}_{P2}$ and $\bar{y}^{*O}_{P2}$ are unbiased with the following variances correspondingly.

$$\begin{aligned} V\left(\bar{y}^{*E}_{P2}\right) = V\left(\bar{y}^{*}_{srs}\right) &- \frac{1}{r\nu^2}\left[\sum_{i=1}^{\nu+1/2}\left(\Delta^2_{y[i+1]} + \Delta^2_{y[\nu+2-i]}\right) + \sum_{i=1}^{\nu-1/2}2\Delta_{y[i+1]}\Delta_{y[\nu+2-i]}\right] \\ &- \frac{W_2(K-1)}{r\nu^2}\left[\sum_{i=1}^{\nu+1/2}\left(\Delta^2_{2y[i+1]} + \Delta^2_{2y[\nu+2-i]}\right) + \sum_{i=1}^{\nu-1/2}2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right] \end{aligned} \quad (5.9)$$

and

$$\begin{aligned} V\left(\bar{y}^{*O}_{P2}\right) = V\left(\bar{y}^{*}_{srs}\right) &- \frac{1}{r\nu^2}\left[\sum_{i=1}^{\nu+1/2}\Delta^2_{y[i+1]} + \sum_{i=1}^{\nu-1/2}\left(\Delta^2_{y[\nu+2-i]} + 2\Delta_{y[i+1]}\Delta_{y[\nu+2-i]}\right)\right] \\ &- \frac{W_2(K-1)}{r\nu^2}\left[\sum_{i=1}^{\nu+1/2}\Delta^2_{2y[i+1]} + \sum_{i=1}^{\nu-1/2}\left(\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right] \end{aligned} \quad (5.10)$$

The variance of estimators can be derived by following the steps involved in lemma derived above in Section 5.1. Eq (21) and Eq (22) shows that $\bar{y}^{*E}_{P2}$ and $\bar{y}^{*O}_{P2}$ both are more efficient than $\bar{y}^{*}_{srs}$, while they will be more precise than $\bar{y}^{*}_{Bouza}$ if the following conditions holds,

$$\left[\begin{array}{c} \sum_{i=1}^{\nu+1/2} \Delta^2_{y[i+1]} + \sum_{i=1}^{\nu-1/2} \left(\Delta^2_{y[\nu+2-i]} + 2\Delta_{y[i+1]}\Delta_{y[\nu+2-i]}\right) \\ +W_2\left(K-1\right)\left[\sum_{i=1}^{\nu+1/2} \Delta^2_{2y[i+1]} + \sum_{i=1}^{\nu-1/2} \left(\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right] \end{array}\right] > \sum_{i=1}^{\nu} \Delta_{2y[i]}$$

if ν is even, and

$$\left[\begin{array}{c} \sum_{i=1}^{\nu+1/2} \Delta^2_{y[i+1]} + \Delta^2_{y[\nu+2-i]} + \sum_{i=1}^{\nu-1/2} 2\Delta_{y[i+1]} + \Delta_{y[\nu+2-i]} \\ +W_2\left(K-1\right)\left[\sum_{i=1}^{\nu+1/2} \Delta^2_{2y[i+1]} + \Delta^2_{2y[\nu+2-i]} + \sum_{i=1}^{\nu-1/2} 2\Delta_{2y[i+1]} + \Delta_{2y[\nu+2-i]}\right] \end{array}\right] > \sum_{i=1}^{\nu} \Delta_{2y[i]}$$

if ν is odd.

5.3 Proposed estimator III

Now we consider the two robust RSS designs together for the estimation of population mean under non-response. Consider we use the MRSS method for the collecting data from response group, while IPRSS is used for the collection data from sub sample of non-respondents, then the proposed estimator for even and odd set size ν can be written as,

$$\bar{y}^{*E}_{P3} = w_1\bar{y}^{E}_{1mrss} + w_2\bar{y}'^{E}_{2iprss} \tag{5.11}$$

and

$$\bar{y}^{*O}_{P3} = w_1\bar{y}^{O}_{1mrss} + w_2\bar{y}'^{O}_{2iprss} \tag{5.12}$$

both estimators $\bar{y}^{*E}_{P3}$ and $\bar{y}^{*O}_{P3}$ are unbiased while their variances are derived by following the steps of above proven lemma which are given as,

$$\begin{aligned} V\left(\bar{y}^{*E}_{P3}\right) = V\left(\bar{y}^{*}_{srs}\right) - \frac{1}{r\nu^2}&\left[\sum_{i=1}^{\nu/2} \Delta^2_{y[\nu/2]} + \sum_{i=\nu+2/2}^{\nu} \Delta^2_{y[\nu+/2]} + W_2\left(\sum_{i=1}^{\nu/2} \Delta^2_{2y[\nu/2]} + \sum_{i=\nu+2/2}^{\nu} \Delta^2_{2y[\nu+2/2]}\right)\right] \\ - \frac{W_2}{r\nu^2}&\left[K\left\{\sum_{i=1}^{\nu+1/2} \Delta^2_{2y[i+1]} + \sum_{i=1}^{\nu-1/2} \left(\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right\}\right] \end{aligned} \tag{5.13}$$

and

$$\begin{aligned} V\left(\bar{y}^{*O}_{P3}\right) = V\left(\bar{y}^{*}_{srs}\right) - \frac{1}{r\nu^2}&\left[\sum_{i=1}^{\nu} \Delta^2_{y[\nu+1/2]} + W_2\left(\sum_{i=1}^{\nu/2} \Delta^2_{2y[\nu+1/2]}\right)\right] \\ - \frac{W_2 K}{r\nu^2}&\sum_{i=1}^{\nu/2} \left(\Delta^2_{2y[i+1]} + \Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right) \end{aligned} \tag{5.14}$$

Eq (25) and Eq (26) shows that $\bar{y}^{*E}_{P3}$ and $\bar{y}^{*O}_{P3}$ both are more precise than $\bar{y}^{*}_{srs}$, while they are more efficient than $\bar{y}^{*}_{Bouza}$ if the following condition holds,

$$\left[\begin{array}{l} \sum_{i=1}^{\nu/2}\Delta^2_{y[\nu/2]}+\sum_{i=\nu+2/2}^{\nu}\Delta^2_{y[\nu+/2]}+W_2\left\{\sum_{i=1}^{\nu/2}\Delta^2_{2y[\nu/2]}+\sum_{i=\nu+2/2}^{\nu}\Delta^2_{2y[\nu+2/2]}\right\} \\ +W_2\left[K\left\{\sum_{i=1}^{\nu+1/2}\Delta^2_{2y[i+1]}+\sum_{i=1}^{\nu-1/2}\left(\Delta^2_{2y[\nu+2-i]}+2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right\}\right] \end{array}\right] > K\sum_{i=1}^{\nu}\Delta^2_{2y[i]}$$

if ν is even, and

$$\left[\begin{array}{c} \sum_{i=1}^{\nu}\Delta^2_{y[\nu+1/2]}+W_2\left\{\sum_{i=1}^{\nu/2}\Delta^2_{2y[\nu+1/2]}\right\} \\ +W_2K\sum_{i=1}^{\nu/2}\left(\Delta^2_{2y[i+1]}+\Delta^2_{2y[\nu+2-i]}+2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right) \end{array}\right] > K\sum_{i=1}^{\nu}\Delta^2_{2y[i]}$$

if ν is odd.

5.4 Proposed estimator IV

Let we use the IPRSS method for the collection of data from the response group, while MRSS is used for collecting data from a sub-sample of non-respondents. Then the proposed estimator for even and odd set sizes ν can be written as,

$$\bar{y}^{*E}_{P4} = w_1\bar{y}^{E}_{1iprss} + w_2\bar{y}'^{E}_{2mrss} \tag{5.15}$$

and

$$\bar{y}^{*O}_{P4} = w_1\bar{y}^{O}_{1iprss} + w_2\bar{y}'^{O}_{2mrss} \tag{5.16}$$

Both estimators $\bar{y}^{*E}_{P4}$ and $\bar{y}^{*O}_{P4}$ are unbiased, for which the variances are derived by following the steps involved in lemma in Section 5.1. Their variances are given as,

$$\begin{aligned} V\left(\bar{y}^{*E}_{P4}\right) = V\left(\bar{y}^{*}_{srs}\right) &- \frac{1}{r\nu^2}\left[\sum_{i=1}^{\nu/2}\left(\Delta^2_{y[i+1]}+\Delta^2_{y[\nu+2-i]}+2\Delta_{y[i+1]}\Delta_{y[\nu+2-i]}\right)\right] \\ &- \frac{W_2}{r\nu^2}\left\{\sum_{i=1}^{\nu/2}\left(\Delta^2_{2y[i+1]}+\Delta^2_{2y[\nu+2-i]}+2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right\} \\ &- \frac{W_2K}{r\nu^2}\left(\sum_{i=1}^{\nu/2}\Delta^2_{2y[\nu/2]}+\sum_{i=\nu+2/2}^{\nu}\Delta^2_{2y[\nu+1/2]}\right) \end{aligned} \tag{5.17}$$

and

$$V\left(\bar{y}^{*O}_{P4}\right) = V\left(\bar{y}^{*}_{srs}\right) - \frac{1}{r\nu^2}\left\{\sum_{i=1}^{\nu+1/2}\Delta^2_{y[i+1]} + \sum_{i=1}^{\nu-1/2}\left(\Delta^2_{y[\nu+2-i]} + 2\Delta_{y[i+1]}\Delta_{2[\nu+2-i]}\right)\right\}$$
$$- \frac{W_2}{r\nu^2}\sum_{i=1}^{\nu-1/2}\left(\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right) \tag{5.18}$$
$$- \frac{W_2}{r\nu^2}\left[K\sum_{i=1}^{\nu/2}\Delta^2_{2y[\nu+1/2]} + \sum_{i=1}^{\nu+1/2}\Delta^2_{2y[i+1]}\right]$$

Eq (29) and Eq (30) shows that $\bar{y}^{*E}_{P4}$ and $\bar{y}^{*O}_{P4}$ both are at least as precise as $\bar{y}^{*}_{srs}$, while they are more precise than $\bar{y}^{*}_{Bouza}$ if the following conditions holds,

$$\left[\begin{array}{l} \sum_{i=1}^{\nu/2}\left(\Delta^2_{y[i+1]} + \Delta^2_{y[\nu+2-i]} + 2\Delta_{y[i+1]}\Delta_{y[\nu+2-i]}\right) \\ +W_2\left\{\sum_{i=1}^{\nu/2}\left(\Delta^2_{2y[i+1]} + \Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right\} \\ +W_2K\left(\sum_{i=1}^{\nu/2}\Delta^2_{2y[\nu/2]} + \sum_{i=\nu+2/2}^{\nu}\Delta^2_{2y[\nu+1/2]}\right) \end{array}\right] > K\sum_{i=1}^{\nu}\Delta^2_{2y[i]}$$

for even set size ν, and

$$\left[\begin{array}{l} \sum_{i=1}^{\nu+1/2}\Delta^2_{y[i+1]} + \sum_{i=1}^{\nu-1/2}\left(\Delta^2_{y[\nu+2-i]} + 2\Delta_{y[i+1]}\Delta_{2[\nu+2-i]}\right) \\ +W_2\left\{\sum_{i=1}^{\nu+1/2}\Delta^2_{2y[i+1]} + \sum_{i=1}^{\nu-1/2}\left(\Delta^2_{2y[\nu+2-i]} + 2\Delta_{2y[i+1]}\Delta_{2y[\nu+2-i]}\right)\right\} \\ +W_2K\sum_{i=1}^{\nu/2}\Delta^2_{2y[\nu+1/2]} \end{array}\right] > K\sum_{i=1}^{\nu}\Delta^2_{2y[i]}$$

for odd set size ν.

If the regularity conditions, such as finite fourth moments and consistent ranking are satisfied by the auxiliary variable then the proposed estimator $\bar{y}^{*}_{ph}; h = 1,2,3,4$ converges to the distribution of a normal random variable. For small samples, the expressions of variance provided for each proposed estimator can be used directly for inferences, as supported by our simulation results. The choice of the most appropriate estimator for the given data is explained in the flowchart shown in Figure 1.

5.5 Variance Estimation in Practice

The variance expressions derived in Sections 5.1–5.4 involve population parameters $\Delta^2_{y[i]} = (\mu_{y[i]} - \mu_y)^2$, which are unknown in practice. For implementation with sample data, we propose the following sample based estimators:

For a ranked set sample of size $r\nu$, the term $\Delta^2_{y[i]}$ can be estimated by:

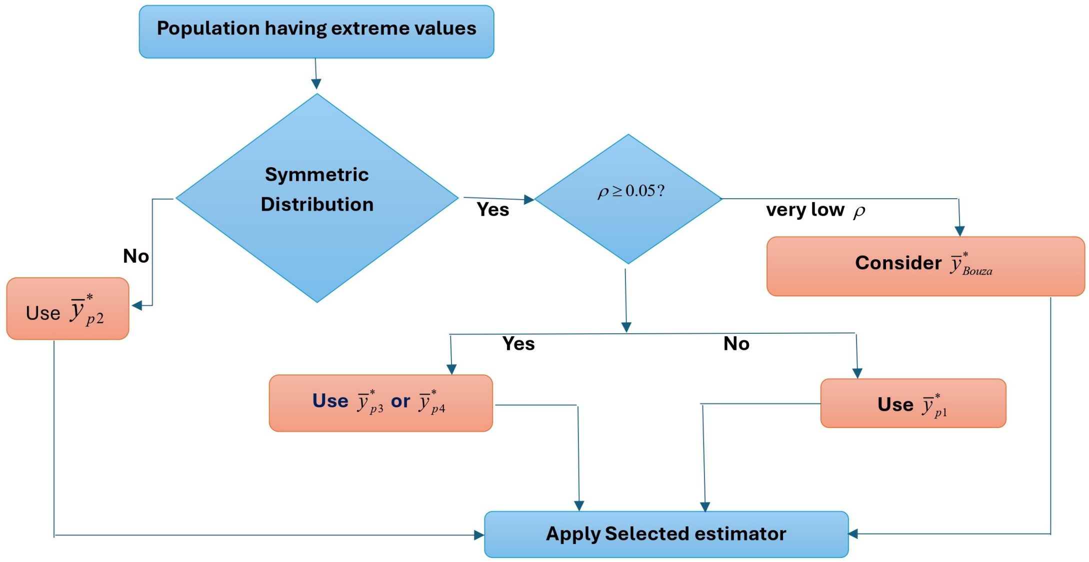

Figure 1: Flowchart for selecting the best proposed estimator under non-response with extreme values.

$$\hat{\Delta}^2_{y[i]} = \frac{1}{r-1}\sum_{j=1}^{r}(y_{i[j]} - \bar{y}_{[i]})^2,$$

where $\bar{y}_{[i]} = \frac{1}{r}\sum_{j=1}^{r} y_{i[j]}$ is the sample mean of the ith judgment order statistic across r cycles. This provides an unbiased estimator of $\Delta^2_{y[i]}$ under perfect ranking. For the proposed estimators under non-response, the variance estimator for $\bar{y}^E_{P1}$ becomes:

$$\hat{V}(\bar{y}^E_{P1}) = \frac{s^2_y}{rv} + \frac{W_2(K-1)}{n}s^2_{2y} - \frac{W_2K}{nv}\sum_{i=1}^{v/2}\left(\hat{\Delta}^2_{2y[i+1]} + \hat{\Delta}^2_{2y[v+2-i]} + 2\hat{\Delta}_{2y[i+1]}\hat{\Delta}_{2y[v+2-i]}\right),$$

where s^2_y and s^2_{2y} are sample variances from the response and non-response groups, respectively. Similar expressions follow for $\bar{y}^E_{P2}$, $\bar{y}^E_{P3}$, and $\bar{y}^E_{P4}$. These sample based variance estimators are consistent and can be used directly for constructing confidence intervals and hypothesis tests in practice.

5.6 Asymptotic Properties

The proposed estimators $\bar{y}^*_{Ph}$ are linear combinations of means from the response and non-response groups. Under standard regularity conditions of finite moments, independent sampling, and consistent ranking, each estimator is asymptotically normal by the central limit theorem. This permits large sample inference using the sample based variance estimators developed above together with normal quantiles. Thus, an approximate $100(1-\alpha)\%$ confidence interval for the population mean is $\bar{y}^*_{Ph} \pm z_{\alpha/2}\sqrt{\hat{V}(\bar{y}^*_{Ph})}$, where $\hat{V}(\bar{y}^*_{Ph})$ is computed from sample data and $z_{\alpha/2}$ is the standard normal critical value.

6 Simulation study

Simulation study is carried out in view of the Hansen and Hurwitz (1946) assumptions. We generated the auxiliary variable for the response and non-response groups of population correspondingly using $X_j \sim Poisson\left(N_j, \mu_{yj}\right)$, $j = 1, 2$. These two groups together make an entire population. The corresponding study variable is produced using the relationship $Y_j = \rho_{yx}X_j + e_j\sqrt{1-\rho^2_{yx}}$, where $e_j \sim Normal\left(N_j, 0, 1\right)$ and ρ_{yx} is fix coefficient of correlation between Y and X. A sample of size $n\,(= n_1 + n'_2)$ is selected from population by using the method of SRS, RSS, MRSS, DRSS, DMRSS (for Situation-I and Situation-II), and the sample mean is estimated using the corresponding mean estimators under non-response. This procedure is repeated $20,000$ times to calculate the bias and variance of the estimates by using the following expression,

$$Bias(\bar{y}^*_{\bullet}) = \frac{1}{20,000}\sum_{j=1}^{20,000}\left(\bar{y}^*_{\bullet} - \mu_y\right) \tag{6.1}$$

$$MSE(\bar{y}^*_{\bullet}) = \frac{1}{20,000}\sum_{j=1}^{20,000}\left(\bar{y}^*_{\bullet} - \mu_y\right)^2 \tag{6.2}$$

where $\bar{y}^*_{\bullet}$ represents the corresponding mean estimator as discussed above. The relative efficiency of the proposed estimators $y^*_{Ph};\ \ h = 1, 2, 3, 4$, for different values of non-response rate K is calculated as:

$$RE = \frac{MSE\left(\bar{y}^*_{Bouza}\right)}{MSE(\bar{y}^*_{Ph})} \tag{6.3}$$

The scatter plot for the hypothetically generated population is shown in Figure 2. The bias results are presented in Tables 1–3, while the relative efficiency results are presented in Tables 4–6.

Tables 4–6 show that the overall RE of the proposed estimators increases with the increase in the coefficient of correlation, while a decline is observed in RE when set size ν is increased. Further, we can achieve more precise estimates with a smaller value of set size with a higher correlation coefficient. A minor reduction in RE is observed with the increase in overall sample size n . It is also observed that the RE of proposed estimators increases as the value of the non-response rate K increases.

6.1 Sensitivity to Ranking Quality

To assess the robustness of the proposed estimators to imperfect ranking, we conducted a sensitivity analysis examining how relative efficiency (RE) varies with the correlation

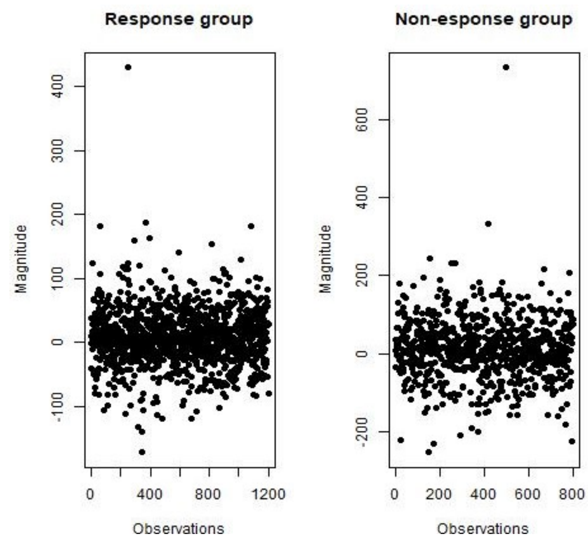

Figure 2: Hypothetically generated response and non-response groups.

Table 1: Bias of estimators for $K = 1.5$

(r_1, r_2')	Estimator	$\rho = 0.30$			$\rho = 0.60$			$\rho = 0.90$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
(2, 1)	$\bar{y}^*_{Bouza}$	0.1726	0.0542	0.0456	0.1692	0.0745	0.0279	0.1703	0.0661	0.0373
	$\bar{y}^*_{P1}$	3.6308	0.5198	3.3701	1.1860	0.1956	0.8317	2.4030	0.3597	2.0848
	$\bar{y}^*_{P2}$	3.9985	0.9643	3.8146	1.2411	0.3722	1.0473	2.6121	0.6680	2.4122
	$\bar{y}^*_{P3}$	3.9775	1.3627	4.3454	1.2261	0.6048	1.2932	2.5966	0.9866	2.8096
	$\bar{y}^*_{P4}$	1.3157	1.2354	1.2462	0.6495	0.4198	0.3891	0.9818	0.8273	0.8147
(4, 2)	$\bar{y}^*_{Bouza}$	0.1686	0.1552	0.0194	0.1886	0.1580	0.0015	0.1915	0.1559	0.0101
	$\bar{y}^*_{P1}$	3.3383	1.0674	4.4655	0.8740	0.8142	2.0244	3.0471	0.9464	3.2503
	$\bar{y}^*_{P2}$	3.9825	1.6948	4.9791	1.2214	1.1365	2.2909	3.4447	1.4271	3.6422
	$\bar{y}^*_{P3}$	3.8732	1.9739	5.5863	1.0866	1.2390	2.5855	3.4059	1.6231	4.1063
	$\bar{y}^*_{P4}$	1.3396	2.1509	2.0531	0.6578	1.3884	1.1626	1.0044	1.7848	1.6259
(6, 3)	$\bar{y}^*_{Bouza}$	0.0020	0.0787	0.0878	0.0209	0.0778	0.1067	0.0213	0.0777	0.0981
	$\bar{y}^*_{P1}$	4.2890	1.0483	4.2876	1.9082	0.7872	1.8753	2.4384	0.9252	3.0772
	$\bar{y}^*_{P2}$	4.9250	1.9117	4.7403	2.2520	1.3761	2.0979	2.9052	1.6555	3.4143
	$\bar{y}^*_{P3}$	5.0170	2.2280	5.2030	2.3260	1.4816	2.3113	3.0234	1.8672	3.7564
	$\bar{y}^*_{P4}$	0.9707	2.4395	1.6506	0.2662	1.7051	0.8654	1.0202	2.0923	1.2646

ρ between study and auxiliary variables. Figure 3 presents line plots of RE against ρ for all four estimators under symmetric population distribution with $\nu = 5$ and $K = 2$.

The results in Figure 3 show an important pattern. For $\rho \geq 0.6$, all estimators maintain RE > 1.5, with $\bar{y}^*_{P2}$ and $\bar{y}^*_{P3}$ exceeding RE > 2.0 under favourable conditions. However, efficiency gains deteriorate rapidly for $\rho < 0.4$, with RE approaching 1.0 (equivalent to SRS) at $\rho \approx 0.2$. Notably, estimators incorporating median-based RSS ($\bar{y}^*_{P3}$ and $\bar{y}^*_{P4}$) demonstrate greater resilience to weak correlation in asymmetric populations. The threshold $\rho = 0.5$, recommended in Section 4, serves as a practical boundary:

Table 2: Bias of estimators for $K = 2$

(r_1, r_2')	Estimator	$\rho = 0.30$			$\rho = 0.60$			$\rho = 0.90$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
(3,1)	$\bar{y}^*_{Bouza}$	0.0122	0.0807	0.0079	0.0273	0.0781	0.1692	0.0437	0.0717	0.1821
	$\bar{y}^*_{P1}$	3.6115	0.5313	3.6924	2.3878	0.3788	2.1651	1.1676	0.2255	0.8990
	$\bar{y}^*_{P2}$	3.9867	0.8785	4.0785	2.6105	0.6007	2.6080	1.2428	0.3240	1.2336
	$\bar{y}^*_{P3}$	4.1168	1.3618	4.4774	2.7341	0.9986	2.9133	1.3567	0.6297	1.3919
	$\bar{y}^*_{P4}$	1.0757	1.1842	2.1624	0.7466	0.7909	0.9128	0.4177	0.3968	0.4819
(6,2)	$\bar{y}^*_{Bouza}$	0.0888	0.0210	0.2174	0.0935	0.0170	0.0789	0.0987	0.0103	0.0831
	$\bar{y}^*_{P1}$	4.8786	0.8135	6.7656	3.6847	0.6934	4.5654	2.4692	0.5650	3.2173
	$\bar{y}^*_{P2}$	5.1964	1.3631	7.2014	3.8383	1.1053	4.5846	2.4658	0.8419	3.1186
	$\bar{y}^*_{P3}$	5.2837	1.8912	7.6803	3.9241	1.5457	4.5879	2.5451	1.1957	2.9631
	$\bar{y}^*_{P4}$	1.4067	1.7820	2.0378	1.0822	1.4070	0.4379	0.7550	1.0301	0.0354
(9,3)	$\bar{y}^*_{Bouza}$	0.2146	0.0093	0.0117	0.2193	0.0084	0.0476	0.2206	0.0048	0.0515
	$\bar{y}^*_{P1}$	4.3442	1.2415	4.3541	3.1014	1.1167	3.3009	1.8501	0.9823	1.9561
	$\bar{y}^*_{P2}$	5.1691	1.8875	5.1267	3.7721	1.6505	4.0631	2.3689	1.4031	2.5901
	$\bar{y}^*_{P3}$	5.1716	2.1946	5.6552	3.7622	1.8864	4.7306	2.3529	1.5646	3.0938
	$\bar{y}^*_{P4}$	1.7360	2.3388	0.9704	1.4202	2.0057	1.3475	1.0999	1.6606	0.9224

Table 3: Bias of estimators for $K = 2.5$

(r_1, r_2')	Estimator	$\rho = 0.30$			$\rho = 0.60$			$\rho = 0.90$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
(4,1)	$\bar{y}^*_{Bouza}$	0.1282	0.0444	0.0719	0.1353	0.0408	0.0219	0.1407	0.0326	0.0256
	$\bar{y}^*_{P1}$	3.4540	0.3490	3.4307	2.2322	0.1788	2.2725	1.0224	0.0107	1.0081
	$\bar{y}^*_{P2}$	4.0895	0.8612	3.9036	2.7142	0.5646	2.5810	1.3495	0.2706	1.2081
	$\bar{y}^*_{P3}$	3.9956	1.1804	4.3480	2.6171	0.7969	2.8971	1.2487	0.4079	1.3720
	$\bar{y}^*_{P4}$	1.0331	1.0933	0.9665	0.6852	0.6981	0.8225	0.3389	0.3073	0.4053
(8,2)	$\bar{y}^*_{Bouza}$	0.0062	0.0051	0.0680	0.0017	0.0076	0.0634	0.0140	0.0077	0.0482
	$\bar{y}^*_{P1}$	4.0097	0.7954	3.2449	2.8102	0.6623	0.7839	1.6033	0.5254	0.5581
	$\bar{y}^*_{P2}$	4.4197	0.9952	4.1346	3.0530	0.7353	1.1320	1.6803	0.4732	0.3383
	$\bar{y}^*_{P3}$	4.4925	1.0910	4.7590	3.1122	0.7438	1.3081	1.7290	0.3930	0.2901
	$\bar{y}^*_{P4}$	0.7058	1.2426	1.6387	0.3636	0.8916	1.6447	0.0215	0.5390	1.2142
(12,3)	$\bar{y}^*_{Bouza}$	0.1625	0.0620	0.0509	0.1614	0.0619	0.0691	0.1541	0.0581	0.0263
	$\bar{y}^*_{P1}$	3.4909	0.6790	3.4052	2.2484	0.5562	4.3480	1.0111	0.4226	2.9772
	$\bar{y}^*_{P2}$	3.8608	1.0935	4.1314	2.4527	0.8334	4.8005	1.0538	0.5601	3.3449
	$\bar{y}^*_{P3}$	3.9146	1.3007	4.7029	2.5064	0.9527	5.2289	1.1014	0.5832	3.6038
	$\bar{y}^*_{P4}$	1.1630	1.3730	2.2950	0.8160	1.0120	1.6030	0.4675	0.6356	1.2653

above this value, RSS consistently outperforms SRS, while below it gains are modest and may not justify additional ranking costs. These findings establish clear efficiency boundaries, suggesting that practitioners should employ the proposed methods only when auxiliary correlation exceeds 0.4–0.5, depending on distributional characteristics.

Table 4: RE of proposed estimators for $K = 1.5$

(r_1, r_2')	Estimator	$\rho = 0.30$			$\rho = 0.60$			$\rho = 0.90$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
(2,1)	$\bar{y}^*_{P1}$	1.602	1.243	1.328	1.675	1.249	1.487	1.716	1.252	1.582
	$\bar{y}^*_{P2}$	2.627	2.207	1.930	2.900	2.233	2.393	3.072	2.249	2.724
	$\bar{y}^*_{P3}$	2.602	2.263	2.588	2.866	2.310	3.853	3.032	2.341	5.206
	$\bar{y}^*_{P4}$	2.055	2.695	3.196	2.083	2.763	3.333	2.102	2.796	3.414
(4,2)	$\bar{y}^*_{P1}$	1.466	1.215	1.214	1.575	1.350	1.089	1.651	1.357	1.269
	$\bar{y}^*_{P2}$	2.263	2.252	1.034	2.417	2.314	1.407	2.971	2.367	1.830
	$\bar{y}^*_{P3}$	2.256	2.352	1.096	2.376	2.461	1.687	2.938	2.556	2.604
	$\bar{y}^*_{P4}$	2.047	2.389	2.558	2.008	2.521	2.843	2.120	2.635	3.101
(6,3)	$\bar{y}^*_{P1}$	1.206	1.236	1.196	1.429	1.242	1.215	1.538	1.247	1.138
	$\bar{y}^*_{P2}$	1.625	1.886	1.780	2.302	1.962	1.091	2.589	2.027	1.465
	$\bar{y}^*_{P3}$	1.592	1.880	1.699	2.233	2.010	1.234	2.570	2.123	1.900
	$\bar{y}^*_{P4}$	1.960	2.009	2.491	1.916	2.157	2.794	2.039	2.290	3.047

Table 5: RE of proposed estimators for $K = 2$

(r_1, r_2')	Estimator	$\rho = 0.30$			$\rho = 0.60$			$\rho = 0.90$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
(3,1)	$\bar{y}^*_{P1}$	1.739	1.355	1.517	1.847	1.359	1.773	1.914	1.361	1.937
	$\bar{y}^*_{P2}$	2.598	2.067	1.928	2.906	2.081	2.494	3.111	2.087	2.988
	$\bar{y}^*_{P3}$	2.508	2.149	2.393	2.806	2.193	3.567	3.009	2.222	5.050
	$\bar{y}^*_{P4}$	1.997	2.545	2.629	2.025	2.612	3.333	2.045	2.649	3.434
(6,2)	$\bar{y}^*_{P1}$	1.466	1.369	1.319	1.684	1.378	0.904	1.870	1.385	1.179
	$\bar{y}^*_{P2}$	1.933	2.056	1.912	2.427	2.109	1.079	2.911	2.148	1.533
	$\bar{y}^*_{P3}$	1.867	1.998	1.858	2.337	2.097	1.283	2.798	2.173	2.078
	$\bar{y}^*_{P4}$	1.931	2.199	2.203	1.975	2.321	3.101	2.005	2.416	3.185
(9,3)	$\bar{y}^*_{P1}$	1.309	1.240	1.187	1.560	1.257	1.187	1.770	1.272	1.110
	$\bar{y}^*_{P2}$	1.485	1.773	1.657	1.967	1.857	1.765	2.483	1.931	1.275
	$\bar{y}^*_{P3}$	1.476	1.696	1.569	1.958	1.809	1.641	2.470	1.910	1.378
	$\bar{y}^*_{P4}$	1.713	1.814	3.162	1.762	1.974	2.270	1.801	2.121	2.483

7 Application on real-life data

We consider heart-failure data collected by Muneer et al. (2017). Our goal is to calculate the population mean of the three variables that enhance the risk of heart failure: (1) level of creatinine in the blood, (2) platelets, and (3) level of serum sodium in the blood. Note that ejection fraction is used as a concomitant variable for selecting a sample of serum sodium and platelets, while age is considered as a concomitant variable for collecting a sample of serum creatinine. The scatter plots for the three data sets are shown in Figures 4–6, while Tables 7–8 present the bias and relative efficiency of the proposed estimators.

Table 8 shows that the RE of the proposed estimators is generally higher at moderate set size ν, while it increases with the increase in the value of the non-response rate K. Table 7 shows that the overall bias is increased with the increase in the value of the

Table 6: RE of proposed estimators for $K = 1.5$

(r_1, r_2')	Estimator	$\rho = 0.30$			$\rho = 0.60$			$\rho = 0.90$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
(4,1)	$\bar{y}^*_{P1}$	1.813	1.527	1.492	1.930	1.539	1.815	1.996	1.547	2.034
	$\bar{y}^*_{P2}$	2.423	2.149	1.818	2.724	2.181	2.519	2.920	2.199	3.097
	$\bar{y}^*_{P3}$	2.385	2.207	2.130	2.668	2.258	3.369	2.848	2.294	4.918
	$\bar{y}^*_{P4}$	1.914	2.670	2.874	1.938	2.746	3.095	1.950	2.789	3.180
(8,2)	$\bar{y}^*_{P1}$	1.580	1.421	1.011	1.818	1.426	1.543	2.003	1.427	1.566
	$\bar{y}^*_{P2}$	1.910	2.014	1.000	2.374	2.045	2.130	2.775	2.062	2.298
	$\bar{y}^*_{P3}$	1.846	2.103	0.978	2.291	2.157	2.770	2.675	2.183	3.187
	$\bar{y}^*_{P4}$	1.750	2.332	2.442	1.778	2.429	2.349	1.796	2.494	2.569
(12,3)	$\bar{y}^*_{P1}$	1.524	1.317	1.259	1.806	1.328	1.248	2.001	1.337	1.287
	$\bar{y}^*_{P2}$	1.863	1.845	1.764	2.447	1.905	1.814	2.917	1.951	1.812
	$\bar{y}^*_{P3}$	1.876	1.872	1.710	2.482	1.963	1.836	2.980	2.029	1.635
	$\bar{y}^*_{P4}$	1.903	2.125	1.844	1.938	2.275	2.320	1.955	2.384	2.509

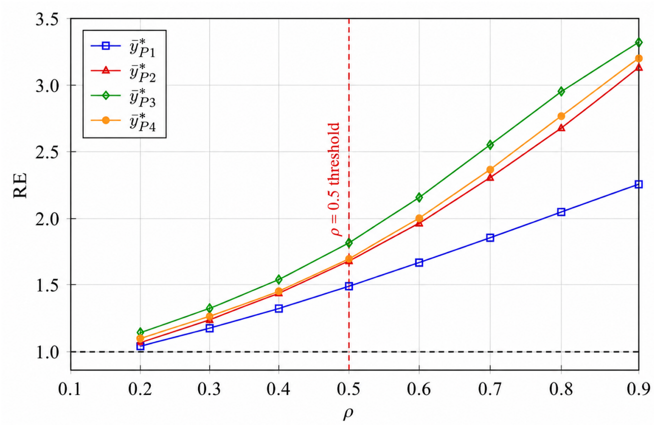

Figure 3: Sensitivity of RE to ρ for all four proposed estimators ($\nu = 5$, $K = 2$, symmetric population).

non-response rate K as well as the set size ν.

Based on our simulation findings, we present a decision table to facilitate practical application. Table 9 provides guidance for practitioners in selecting the best suitable estimator based on the distributional characteristics and the strength of the auxiliary correlation.

8 Conclusions

In this study we proposed four estimators of population mean under non-response by utilizing two robust RSS designs called IPRSS and MRSS. We showed that the proposed estimators estimate the mean of a finite population more precisely as compared to the Bouza-Herrera (2013) estimator.

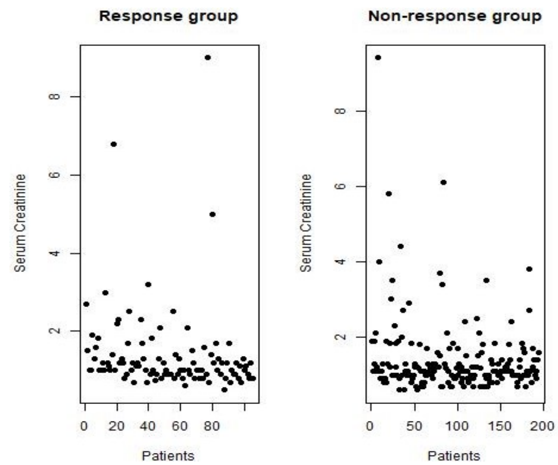

Figure 4: Scatter plots for the response and non-response groups of Serum Creatinine.

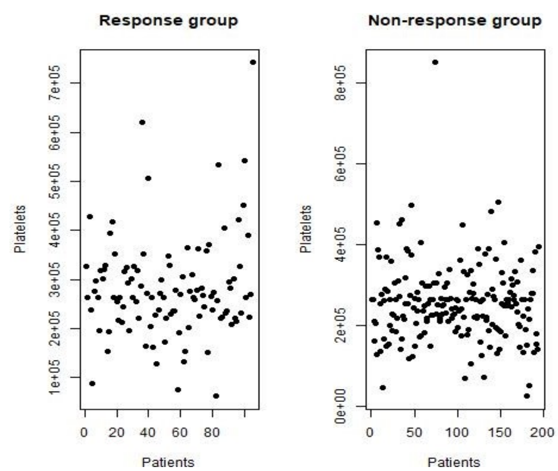

Figure 5: Scatter plots for the response and non-response groups of Platelets.

From the results of a simulation study based on hypothetical and heart-failure data, we recommend the proposed estimators for the estimation of population mean when the population is suspected of containing extreme values. Also, we advise the use of robust RSS designs for data collection from both groups, as it provides more precise estimates. Since the proposed estimators come with explicit expressions for their variances. This allows practitioners, with the data in hand, to compare the efficiency of different estimators and choose the most suitable one for their study. Thus, the derived variance formulas provide a practical tool for guiding estimator selection in real applications, even when the underlying population distributions are unknown.

While SRS offers simplicity, RSS provides improved precision and outlier robustness at the cost of ranking complexity. Our results show RSS estimators achieve 1.2–3.2 times greater efficiency than SRS counterparts when correlation is moderate to strong ($\rho \geqslant 0.5$). This makes RSS particularly valuable in populations with extreme values or

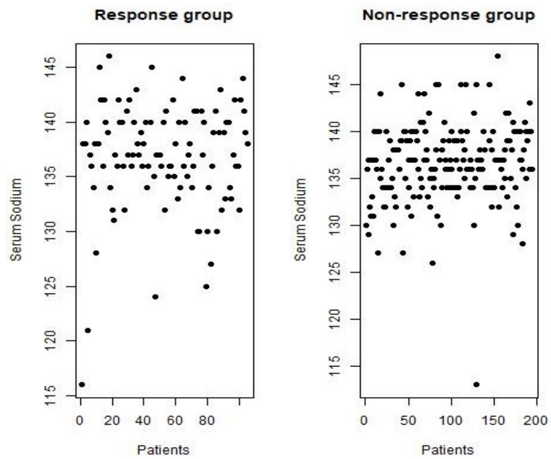

Figure 6: Scatter plots for the response and non-response groups of Serum Sodium.

Table 7: Bias of estimators for different values of K

Data	Estimator	$K = 1.5$			$K = 2$			$K = 2.5$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
1	$\bar{y}^*_{Bouza}$	0.0032	0.0070	0.0003	0.0017	0.0004	0.0016	0.0024	0.0013	0.0002
	$\bar{y}^*_{P1}$	0.1447	0.1224	0.1092	0.1448	0.1213	0.1102	0.1426	0.1219	0.1090
	$\bar{y}^*_{P2}$	0.2164	0.1937	0.1791	0.2168	0.1900	0.1794	0.2147	0.1921	0.1784
	$\bar{y}^*_{P3}$	0.2201	0.1986	0.2015	0.2212	0.1971	0.2024	0.2191	0.1980	0.2021
	$\bar{y}^*_{P4}$	0.1927	0.1950	0.2188	0.1930	0.1905	0.2199	0.1939	0.1921	0.2178
2	$\bar{y}^*_{Bouza}$	168.1	259.2	286.1	168.6	245.9	407.9	18.4	85.4	129.3
	$\bar{y}^*_{P1}$	11152.2	85.3	2943.2	11284.5	265.7	3033.7	11250.6	116.1	2914.0
	$\bar{y}^*_{P2}$	8528.5	2786.4	5657.4	8384.6	2477.6	5658.0	8287.9	2580.6	5638.0
	$\bar{y}^*_{P3}$	8304.0	3072.7	6754.8	8232.8	2765.8	6801.0	8105.6	2943.1	6825.6
	$\bar{y}^*_{P4}$	6264.6	5835.9	6411.3	6249.2	6173.0	6548.0	6333.9	5822.6	6417.6
3	$\bar{y}^*_{Bouza}$	0.0305	0.0001	0.0239	0.0003	0.0136	0.0077	0.0102	0.0014	0.0023
	$\bar{y}^*_{P1}$	0.4384	0.1695	0.0235	0.4076	0.1823	0.0407	0.4369	0.1741	0.0481
	$\bar{y}^*_{P2}$	0.5711	0.3002	0.1660	0.5546	0.3261	0.1809	0.5758	0.3111	0.1808
	$\bar{y}^*_{P3}$	0.5733	0.3154	0.2085	0.5600	0.3266	0.2263	0.5826	0.3227	0.2313
	$\bar{y}^*_{P4}$	0.1888	0.1792	0.1922	0.1941	0.1898	0.1823	0.1920	0.1846	0.1900

high non-response, where SRS performance declines.

Although all four estimators perform better than competing estimators in situations where the population under consideration may possess extreme values. However, we recommend that the second proposed estimator should be considered when the population is symmetrically distributed, as it leverages paired RSS in both respondent and non-respondent groups. While in the case of asymmetric distribution, one is free to choose among proposed estimators 3 or 4 since these estimators consider median values and leverage the effects of asymmetry on representative sample selection.

The present work assumes a Missing at Random (MAR) non-response mechanism since this assumption is commonly adopted in survey sampling literature. Future re-

Table 8: RE of proposed estimators for different values of K

Data	Estimator	$K = 1.5$			$K = 2$			$K = 2.5$		
		$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$	$\nu = 3$	$\nu = 5$	$\nu = 7$
1	$\bar{y}^*_{P1}$	1.779	1.727	1.647	1.957	1.989	1.802	1.896	2.012	1.822
	$\bar{y}^*_{P2}$	1.414	1.475	1.408	1.470	1.568	1.364	1.382	1.473	1.294
	$\bar{y}^*_{P3}$	1.654	1.490	1.218	1.610	1.539	1.162	1.488	1.444	1.095
	$\bar{y}^*_{P4}$	2.040	1.557	1.195	1.958	1.625	1.136	1.841	1.516	1.097
2	$\bar{y}^*_{P1}$	1.305	1.585	1.438	1.315	1.666	1.480	1.302	1.713	1.505
	$\bar{y}^*_{P2}$	1.268	1.945	1.865	1.344	1.950	1.732	1.351	1.927	1.700
	$\bar{y}^*_{P3}$	1.813	2.276	1.871	1.726	2.150	1.705	1.663	2.084	1.653
	$\bar{y}^*_{P4}$	2.045	2.034	2.410	1.936	1.942	2.218	1.806	1.926	2.229
3	$\bar{y}^*_{P1}$	1.064	1.359	1.356	1.043	1.392	1.356	1.530	1.490	1.442
	$\bar{y}^*_{P2}$	1.259	1.517	1.886	1.356	1.452	1.675	1.807	1.527	1.700
	$\bar{y}^*_{P3}$	1.183	1.760	1.935	1.096	1.653	1.660	1.504	1.709	1.679
	$\bar{y}^*_{P4}$	1.868	1.909	2.250	1.764	1.815	2.122	1.714	1.905	2.108

Table 9: Estimator selection with expected efficiency gains over SRS

Case	**Distribution**	ρ **range**	**Best Estimator**	**Expected RE**
1	Symmetric	0.6–0.9	$\bar{y}^*_{P2}$	1.8–2.5
2	Symmetric	0.3–0.5	$\bar{y}^*_{P1}$	1.3–1.7
3	Asymmetric	0.6–0.9	$\bar{y}^*_{P3}$	2.0–3.2
4	Asymmetric	0.3–0.5	$\bar{y}^*_{P4}$	1.5–2.0
5	Heavy-tailed	Any	$\bar{y}^*_{P3}$ or $\bar{y}^*_{P4}$	1.7–2.8
6	Unknown	< 0.3	$\bar{y}^*_{Bouza}$	1.0–1.2

search may examine the performance of the proposed estimators assuming Missing Not at Random (MNAR) mechanisms, particularly when extreme values directly affect the probability of response. Such an expansion would increase the practical applicability of the proposed methodology.

Acknowledgment

The authors express their sincere gratitude to the anonymous reviewers and the editor for their valuable comments and constructive suggestions, which significantly improved the quality and clarity of this manuscript.

Funding

This research received no specific grant from any funding agency.

Conflict of Interest

The authors declare that they have no competing interests.

References

Abu-Dayyeh WA, Ahmed MS, Ahmed RA, Muttlak HA. Some estimators of a finite population mean using auxiliary information. Applied Mathematics and Computation. 2003;139(2–3):287–298.

Al-Saleh MF, Al-Kadiri MA. Double-ranked set sampling. Statistics and Probability Letters. 2000;48(2):205–212.

Bouza-Herrera CN. Handling Missing Data in Ranked Set Sampling. Springer; 2013.

Fatima M, Shahbaz SH, Hanif M, Shahbaz MQ. A modified regression-cum-ratio estimator for finite population mean in presence of nonresponse using ranked set sampling. AIMS Mathematics. 2022;7(4):6478–6488.

Hansen MH, Hurwitz WN. The problem of non-response in sample surveys. Journal of the American Statistical Association. 1946;41(236):517–529.

Haq A, Brown J, Moltchanova E, Al-Omari AI. Paired double-ranked set sampling. Communications in Statistics - Theory and Methods. 2016;45(10):2873–2889.

Kadilar C, Cingi H. A new estimator using two auxiliary variables. Applied Mathematics and Computation. 2005;162(2):901–908.

Malik S, Singh R. An improved estimator using two auxiliary attributes. Applied Mathematics and Computation. 2013;219(23):10983–10986.

McIntyre GA. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research. 1952;3(4):385–390.

Muneer S, Shabbir J, Khalil A. Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. Communications in Statistics - Theory and Methods. 2017;46(5):2181–2192.

Muttlak HA. Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. Environmetrics. 1998;9(3):255–267.

Rehman SA, Shabbir J. On the improvement of paired ranked set sampling to estimate population mean. Statistics in Transition New Series. 2021;22(3):145–161.

Samawi HM, Tawalbeh EM. Double median ranked set sample: Comparing to other double ranked samples for mean and ratio estimators. Journal of Modern Applied Statistical Methods. 2002;1(2):428–438.

Singh HP, Tailor R, Singh S. General procedure for estimating the population mean using ranked set sampling. Journal of Statistical Computation and Simulation. 2014;84(5):931–945.

Singh R, Sharma P. A class of exponential ratio estimators of finite population mean using two auxiliary variables. Pakistan Journal of Statistics and Operation Research. 2015;11(2):221–229.

Stokes SL. Ranked set sampling with concomitant variables. Communications in Statistics - Theory and Methods. 1977;6(12):1207–1211.

Takahasi K, Wakimoto K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. Annals of the Institute of Statistical Mathematics. 1968;20(1):1–31.