



OPEN Novel logarithmic imputation procedures using multi auxiliary information under ranked set sampling

Anoop Kumar¹, Shashi Bhushan^{2✉}, Walid Emam³, Yusra Tashkandy³ & M. J. S. Khan⁴

Ranked set sampling (RSS) is known to increase the efficiency of the estimators while comparing it with simple random sampling. The problem of missingness creates a gap in the information that needs to be addressed before proceeding for estimation. Negligible amount of work has been carried out to deal with missingness utilizing RSS. This paper proposes some logarithmic type methods of imputation for the estimation of population mean under RSS using auxiliary information. The properties of the suggested imputation procedures are examined. A simulation study is accomplished to show that the proposed imputation procedures exhibit better results in comparison to some of the existing imputation procedures. Few real applications of the proposed imputation procedures is also provided to generalize the simulation study.

Keywords Missing data, Imputation, Ranked set sampling, Simulation study

Survey practitioners are always faced with the issue of missing data. The presence of missing data may be the reason in obtaining more efficient estimates of population parameters because of reduced sample size, but also conventional sampling techniques can not be directly enforced to examine data which is available in hand. To date, a number of imputation techniques have been proposed in previous research to use the auxiliary data on the non-response to imputation the missing value of the research variable y .

The problem of missing data may occur either at the estimate stage of parameters or at the stage of sampling design or at both the stages. Hansen and Hurwitz¹ was the first who sort out the problem of missing data at the sampling design stage and subsequently^{2,34} mooted and extended this technique in a mailed questionnaire survey. Several writers have emphasised the significance of auxiliary variables in the process of imputed missing data. The foundational ideas of missing at random (MAR), observed at random (OAR), and parameter distribution (PD) were envisioned by Rubin³ in his seminal research. These concepts provided a basis for drawing information from the data ridden with missing observations. The difference between MAR and missing completely at random (MCAR) approaches was made by Heitjan and Basu⁴ in their earlier work. To tackle the missing value, various improvements have been proposed over existing imputation procedures using auxiliary information. Bhushan and Pandey⁷ advocated some optimal imputation procedures to sort out the missing data employing multi auxiliary information. An optimal second call imputation under probability proportional to size sampling was introduced by Sohil et al.⁶. Bhushan et al.⁵ employed SRS to estimate population mean in the presence of missing data. Prasad and Yadav⁸ developed product type exponential imputation procedures in survey sampling, while, Yadav and Prasad⁹ suggested generalized class of factor type exponential imputation procedures through simulation approach. Bhushan et al.³⁵ proposed design based synthetic imputation procedures for domain mean. A few notable studies in this field include those by Lee et al.³⁶, Singh and Horn¹², Ahmed et al.¹³, and Diana and Perri¹⁴.

Statistics professionals frequently face the challenge of measuring research variables, which can be expensive or complicated. However, unit ranking can be accomplished visually or through other cost-free methods. In these situations, McIntyre¹¹ proposed the idea of RSS as a practical substitute for simple random sampling (SRS); however, he did not offer a strong mathematical foundation for the theory. The necessary mathematical basis for RSS was supplied by Takahasi and Wakimoto¹⁵, who examined the idea of McIntyre¹¹. Shahzad et al.¹⁶ advocated a new population mean estimator under RSS along with the sensitive issues, while, to deal with sensitivity issues,

¹Department of Statistics, Central University of Haryana, Mahendergarh 123031, India. ²Department of Statistics, Lucknow University, Lucknow 226007, India. ³Department of Statistics and Operations Research, Faculty of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia. ⁴Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh 202002, India. ✉email: bhushan_s@lkouniv.ac.in

Shahzad et al.¹⁷ suggested regression type estimator for mean estimation under RSS. Some optimal classes of estimators and log type estimators were also suggested for the mean estimation under RSS by Bhushan and Kumar¹⁸ and Bhushan and Kumar¹⁹, respectively. Kocyigit and Kadilar²² investigated ratio estimators for RSS in the availability of tie information. Khalid et al.²¹ proposed a new exponential ratio estimator under RSS. Kocyigit and Rather²⁰ developed the new sub-regression type estimator under RSS. Bhushan and Kumar^{23,24} utilized RSS for the predictive estimation of population mean by proposing difference, ratio, and logarithmic type estimators. Some multi-auxiliary information based logarithmic estimators were suggested by Bhushan and Kumar³¹ for estimating population mean. Kocyigit²⁶ suggested a novel sub-type mean estimator under RSS employing dual auxiliary variables.

In practice, it is possible for there to be instances where every data set has a minuscule number of observations, and each observation proves to be crucial for reaching an appropriate conclusion. Moreover, utilising such sets of information that contain missing values might taint the outcome and make the estimation process less effective. Bouza and Al-Omari²⁷ developed imputation of missing data under RSS, while handling of missing data under RSS was investigated by Bouza²⁸. Sohail et al.²⁹ suggested a class of ratio type estimators for imputing missing data under RSS. Bhushan et al.³⁰ proposed efficient difference and ratio type imputation procedures under RSS. Bhushan and Kumar²⁵ dealt with the imputation of missing data under RSS utilizing multi auxiliary information, whereas¹⁰ suggested RSS imputation procedures in the presence of correlated measurement errors. This study uses the MCAR approach by default for imputation of missing data.

Methodological foundation and notations

In RSS, a parent population of size N units serves as the source of quantification for m simple random samples, each of size m . Now, the units of every randomly selected sample are arranged according to the selected variable. Suppose x_1 and x_2 are the two auxiliary variables and y is the research variable. The population's m^2 trivariate samples are quantified at random and divided into m sets, with a total size m . At this point, every sample is ordered in relation to either x_1 or x_2 , the auxiliary variables. The unit with the lowest rank of x_1 , variable y , and x_2 related to the lowest rank of x_1 are now quantified based on the initial sample. The variables y and x_2 linked to the second smallest rank of x_1 are measured from the second sample of size m . Until the y and x_2 values linked to the highest rank of x_1 are quantified from the sample m , the aforementioned process is maintained. One cycle is finished by this technique. The required ranked set sample of size $n = mk$ units is obtained by repeating this process k times.

Out of the selected m units, let r units render the answer for the study variable such that $m \geq r$ from the first cycle. Repeat the same process until k such cycles are completed, at which point we will get replies from $n_1 = mkr$ units from the $n = mk$ chosen units, provided $n \geq n_1$. Since it is expected that there is one missing value in the m slashed units, which is easily apparent from the figuration representation provided in Appendix A, let the ranking be done over the additional variable x_1 in this case.

Let a finite population U consist of N distinguishable units with mean $\bar{Y} = N^{-1} \sum_{i=1}^N y_i$. To get the population mean $\bar{Y}$, let's quantify a ranked set sample s of size $n = mk$ from U . Let r possess the responding units out of sampled n units. Let p and $(1 - p)$ be the probability that the respondent i either belongs to the responding group A or to the non-responding group $\bar{A}$ such that $s = A + \bar{A}$, where, $+$ denotes disjoint union. Every unit possesses an observed value of y_i , $i \in A$; however, the values for the units $i \in \bar{A}$ are missing and must be imputed to provide the complete data structure required for a successful inference. Also, the auxiliary variable x helps impute the values that are missing.

The remainder of the paper is set up as follows. The traditional imputation techniques and their characteristics under bivariate and multiple auxiliary information setup utilizing RSS were covered in detail in Section “Review of prominent existing imputation procedures”. We presented a few logarithmic imputation techniques and their characteristics in Section “Proposed imputation procedures”. The efficiency criteria are obtained in Section “Efficiency conditions”. A simulation study is conducted in Section “Simulation study”, while a real data application is provided in Section Real data applications. Finally, Section “Conclusion” provides a conclusion.

Review of prominent existing imputation procedures

Under RSS, the typical mean imputation technique may be given as

$$y_{.im} = \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} & \text{if } i \in \bar{A} \end{cases} \quad (1)$$

with the resultant estimator as

$$T_m = \bar{y}_{r,rss}$$

Bivariate auxiliary information

In order to have comprehensive understanding, the imputation procedures are divided into three strategies under the deliberation of availability of bivariate auxiliary information.

Strategy I: When $\bar{X}$ is known and $\bar{x}_{1n}$ and $\bar{x}_{2n}$ are used.

Strategy II: When $\bar{X}$ is known and $\bar{x}_{1r}$ and $\bar{x}_{2r}$ are used.

Strategy III: When $\bar{X}$ is unknown and $\bar{x}_{1n}$, $\bar{x}_{1r}$ and $\bar{x}_{2r}$ are used.

We define the conventional difference and ratio type imputation procedures in presence of bivariate auxiliary information under RSS as

Strategy I

$$\begin{aligned}
 y_{.id_1} &= \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} + \frac{n}{n-r} [\beta_1(\bar{x}_{1n,rss} - \bar{X}_1) + \beta_2(\bar{x}_{2n,rss} - \bar{X}_2)] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.iu_1} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\bar{x}_{1n,rss}} \right)^{\theta_1} \left(\frac{\bar{X}_2}{\bar{x}_{2n,rss}} \right)^{\theta_2} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.iu_4} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\phi_1\bar{x}_{1n,rss} + (1-\phi_1)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\phi_2\bar{x}_{2n,rss} + (1-\phi_2)\bar{X}_2} \right) - r\bar{y}_{r,rss} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

Strategy II

$$\begin{aligned}
 y_{.id_2} &= \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} + \frac{n}{n-r} [\beta_3(\bar{x}_{1n,rss} - \bar{X}_1) + \beta_4(\bar{x}_{2n,rss} - \bar{X}_2)] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.iu_2} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\bar{x}_{1r,rss}} \right)^{\theta_3} \left(\frac{\bar{X}_2}{\bar{x}_{2r,rss}} \right)^{\theta_4} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.iu_5} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\phi_3\bar{x}_{1r,rss} + (1-\phi_3)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\phi_4\bar{x}_{2r,rss} + (1-\phi_4)\bar{X}_2} \right) - r\bar{y}_{r,rss} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

Strategy III

$$\begin{aligned}
 y_{.id_3} &= \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} + \frac{n}{n-r} [\beta_5(\bar{x}_{1r,rss} - \bar{x}_{1n,rss}) + \beta_6(\bar{x}_{2r,rss} - \bar{x}_{2n,rss})] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.iu_3} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_{r,rss} \left(\frac{\bar{x}_{1n,rss}}{\bar{x}_{1r,rss}} \right)^{\theta_5} \left(\frac{\bar{x}_{2n,rss}}{\bar{x}_{2r,rss}} \right)^{\theta_6} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.iu_6} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\phi_5\bar{x}_{1r,rss} + (1-\phi_5)\bar{x}_{1n,rss}} \right) \left(\frac{\bar{X}_2}{\phi_6\bar{x}_{2r,rss} + (1-\phi_6)\bar{x}_{2n,rss}} \right) - r\bar{y}_{r,rss} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

The consequent estimators are under strategies I, II and III are given by

$$\begin{aligned}
 T_{d_1} &= y_{r,rss} + \beta_1(\bar{x}_{1n,rss} - \bar{X}_1) + \beta_2(\bar{x}_{2n,rss} - \bar{X}_2) \\
 T_{d_2} &= y_{r,rss} + \beta_3(\bar{x}_{1r,rss} - \bar{X}_1) + \beta_4(\bar{x}_{2r,rss} - \bar{X}_2) \\
 T_{d_3} &= y_{r,rss} + \beta_5(\bar{x}_{1r,rss} - \bar{x}_{1n,rss}) + \beta_6(\bar{x}_{2r,rss} - \bar{x}_{2n,rss}) \\
 T_{u_1} &= y_{r,rss} \left(\frac{\bar{X}_1}{\bar{x}_{1n,rss}} \right)^{\theta_1} \left(\frac{\bar{X}_2}{\bar{x}_{2n,rss}} \right)^{\theta_2} \\
 T_{u_2} &= y_{r,rss} \left(\frac{\bar{X}_1}{\bar{x}_{1r,rss}} \right)^{\theta_3} \left(\frac{\bar{X}_2}{\bar{x}_{2r,rss}} \right)^{\theta_4} \\
 T_{u_3} &= y_{r,rss} \left(\frac{\bar{x}_{1n,rss}}{\bar{x}_{1r,rss}} \right)^{\theta_5} \left(\frac{\bar{x}_{2n,rss}}{\bar{x}_{2r,rss}} \right)^{\theta_6} \\
 T_{u_4} &= \bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\phi_1\bar{x}_{1n,rss} + (1-\phi_1)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\phi_2\bar{x}_{2n,rss} + (1-\phi_2)\bar{X}_2} \right) \\
 T_{u_5} &= \bar{y}_{r,rss} \left(\frac{\bar{X}_1}{\phi_3\bar{x}_{1r,rss} + (1-\phi_3)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\phi_4\bar{x}_{2r,rss} + (1-\phi_4)\bar{X}_2} \right) \\
 T_{u_6} &= \bar{y}_{r,rss} \left(\frac{\bar{x}_{1n,rss}}{\phi_5\bar{x}_{1r,rss} + (1-\phi_5)\bar{x}_{1n,rss}} \right) \left(\frac{\bar{x}_{2n,rss}}{\phi_6\bar{x}_{2r,rss} + (1-\phi_6)\bar{x}_{2n,rss}} \right)
 \end{aligned}$$

where β_i , $i = 1, 2, \dots, 6$, θ_i , and ϕ_i are well chosen optimising scalars.

Bhushan and Kumar²⁵ suggested nine new imputation procedures using bivariate auxiliary information under RSS as

Strategy I

$$\begin{aligned}
 y_{.is_1} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{(n\alpha_1-r)}{n-r} \bar{y}_{r,rs} + \eta_1 \left(\bar{x}_{1n,rs} - \frac{(n\bar{X}_1-r\bar{x}_{1n,rs})}{n-r} \right) + \eta_2 \left(\bar{x}_{2n,rs} - \frac{(n\bar{X}_2-r\bar{x}_{2n,rs})}{n-r} \right) & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_4} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\alpha_4 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\bar{x}_{1n,rs}} \right)^{\delta_1} \left(\frac{\bar{X}_2}{\bar{x}_{2n,rs}} \right)^{\delta_2} - r\bar{y}_{r,rs} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_7} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\alpha_7 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\Theta_1 \bar{x}_{1n,rs} + (1-\Theta_1)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\Theta_2 \bar{x}_{2n,rs} + (1-\Theta_2)\bar{X}_2} \right) - r\bar{y}_{r,rs} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

Strategy II

$$\begin{aligned}
 y_{.is_2} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{(n\alpha_2-r)}{n-r} \bar{y}_{r,rs} + \eta_3 \left(\bar{x}_{1r,rs} - \frac{(n\bar{X}_1-r\bar{x}_{1r,rs})}{n-r} \right) + \eta_4 \left(\bar{x}_{2r,rs} - \frac{(n\bar{X}_2-r\bar{x}_{2r,rs})}{n-r} \right) & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_5} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\alpha_5 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\bar{x}_{1r,rs}} \right)^{\delta_3} \left(\frac{\bar{X}_2}{\bar{x}_{2r,rs}} \right)^{\delta_4} - r\bar{y}_{r,rs} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_8} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\alpha_8 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\Theta_3 \bar{x}_{1r,rs} + (1-\Theta_3)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\Theta_4 \bar{x}_{2r,rs} + (1-\Theta_4)\bar{X}_2} \right) - r\bar{y}_{r,rs} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

Strategy III

$$\begin{aligned}
 y_{.is_3} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{(n\alpha_3-r)}{n-r} \bar{y}_{r,rs} + \eta_5 \left(\bar{x}_{1r,rs} - \frac{(n\bar{x}_{1n,rs}-r\bar{x}_{1r,rs})}{n-r} \right) + \eta_6 \left(\bar{x}_{2r,rs} - \frac{(n\bar{x}_{2n,rs}-r\bar{x}_{2r,rs})}{n-r} \right) & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_6} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\alpha_6 \bar{y}_{r,rs} \left(\frac{\bar{x}_{1n,rs}}{\bar{x}_{1r,rs}} \right)^{\delta_5} \left(\frac{\bar{x}_{2n,rs}}{\bar{x}_{2r,rs}} \right)^{\delta_6} - r\bar{y}_{r,rs} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_9} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\alpha_9 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\Theta_5 \bar{x}_{1r,rs} + (1-\Theta_5)\bar{x}_{1n,rs}} \right) \left(\frac{\bar{X}_2}{\Theta_6 \bar{x}_{2r,rs} + (1-\Theta_6)\bar{x}_{2n,rs}} \right) - r\bar{y}_{r,rs} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

The consequent estimators under strategies I, II, and III are given by

$$\begin{aligned}
 T_{s_1} &= \alpha_1 y_{r,rs} + \eta_1 (\bar{x}_{1n,rs} - \bar{X}_1) + \eta_2 (\bar{x}_{2n,rs} - \bar{X}_2) \\
 T_{s_2} &= \alpha_2 y_{r,rs} + \eta_3 (\bar{x}_{1r,rs} - \bar{X}_1) + \eta_4 (\bar{x}_{2r,rs} - \bar{X}_2) \\
 T_{s_3} &= \alpha_3 y_{r,rs} + \eta_5 (\bar{x}_{1r,rs} - \bar{x}_{2n,rs}) + \eta_6 (\bar{x}_{2r,rs} - \bar{x}_{2n,rs}) \\
 T_{s_4} &= \alpha_4 y_{r,rs} \left(\frac{\bar{X}_1}{\bar{x}_{1n,rs}} \right)^{\delta_1} \left(\frac{\bar{X}_2}{\bar{x}_{2n,rs}} \right)^{\delta_2} \\
 T_{s_5} &= \alpha_5 y_{r,rs} \left(\frac{\bar{X}_1}{\bar{x}_{1r,rs}} \right)^{\delta_3} \left(\frac{\bar{X}_2}{\bar{x}_{2r,rs}} \right)^{\delta_4} \\
 T_{s_6} &= \alpha_6 y_{r,rs} \left(\frac{\bar{x}_{1n,rs}}{\bar{x}_{1r,rs}} \right)^{\delta_5} \left(\frac{\bar{x}_{2n,rs}}{\bar{x}_{2r,rs}} \right)^{\delta_6} \\
 T_{s_7} &= \alpha_7 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\Theta_1 \bar{x}_{1n,rs} + (1-\Theta_1)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\Theta_2 \bar{x}_{2n,rs} + (1-\Theta_2)\bar{X}_2} \right) \\
 T_{s_8} &= \alpha_8 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\Theta_3 \bar{x}_{1r,rs} + (1-\Theta_3)\bar{X}_1} \right) \left(\frac{\bar{X}_2}{\Theta_4 \bar{x}_{2r,rs} + (1-\Theta_4)\bar{X}_2} \right) \\
 T_{s_9} &= \alpha_9 \bar{y}_{r,rs} \left(\frac{\bar{X}_1}{\Theta_5 \bar{x}_{1r,rs} + (1-\Theta_5)\bar{x}_{1n,rs}} \right) \left(\frac{\bar{X}_2}{\Theta_6 \bar{x}_{2r,rs} + (1-\Theta_6)\bar{x}_{2n,rs}} \right)
 \end{aligned}$$

where α_i , $i = 1, 2, \dots, 9$, δ_i and Θ_i are well chosen optimising scalars.

Multi-auxiliary information

With $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p$, let $X_{n \times p} = x_{ij}$ represent the $n \times p$ data matrix based on p additional variables connected with variable of interest y . While the study variable y has missing values, it is supposed that all information on the additional variable vector $\tilde{x}$ is available.

When several auxiliary information are present, the imputation techniques are further split into three strategies, which are as follows:

Strategy I: When $\bar{X}$ is known and estimates $\bar{x}_{1n_1}, \bar{x}_{2n_2}, \dots, \bar{x}_{pn_k}$ are used.

Strategy II: When $\bar{X}$ is known and estimates $\bar{x}_{1r_1}, \bar{x}_{2r_2}, \dots, \bar{x}_{pr_k}$ are used.

Strategy III: When $\bar{X}$ is unknown and estimates $\bar{x}_{1n_1}, \bar{x}_{2n_2}, \dots, \bar{x}_{pn_k}$ and $\bar{x}_{1r_1}, \bar{x}_{2r_2}, \dots, \bar{x}_{pr_k}$ are used.

Strategy I

$$y_{id1multi} = \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} + \frac{n}{n-r} \left[\sum_{j=1}^p \beta_{lj} (\bar{x}_{jn,rss} - \bar{X}_j) \right] & \text{if } i \in \bar{A} \end{cases}$$

$$y_{iu4multi} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_{r,rss} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jn,rss}} \right)^{\theta_{lj}} - r \bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

$$y_{iu7multi} = \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_{r,rss} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\phi_{lj} \bar{x}_{jn,rss} + (1-\phi_{lj}) \bar{X}_j} \right) - r \bar{y}_{r,rss} \right] & \text{for } i \in \bar{A} \end{cases}$$

Strategy II

$$y_{is2multi} = \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} + \frac{n}{n-r} \left[\sum_{j=1}^p \beta_{llj} (\bar{x}_{jr,rss} - \bar{X}_j) \right] & \text{if } i \in \bar{A} \end{cases}$$

$$y_{iu5multi} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_{r,rss} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jr,rss}} \right)^{\theta_{llj}} - r \bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

$$y_{iu8multi} = \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_{r,rss} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\phi_{llj} \bar{x}_{jr,rss} + (1-\phi_{llj}) \bar{X}_j} \right) - r \bar{y}_{r,rss} \right] & \text{for } i \in \bar{A} \end{cases}$$

Strategy III

$$y_{id3multi} = \begin{cases} y_i & \text{if } i \in A \\ \bar{y}_{r,rss} + \frac{n}{n-r} \left[\sum_{j=1}^p \beta_{lllj} (\bar{x}_{jr,rss} - \bar{x}_{jn,rss}) \right] & \text{if } i \in \bar{A} \end{cases}$$

$$y_{iu6multi} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_{r,rss} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rss}}{\bar{x}_{jr,rss}} \right)^{\theta_{lllj}} - r \bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

$$y_{is9multi} = \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n \bar{y}_{r,rss} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rss}}{\phi_{lllj} \bar{x}_{jr,rss} + (1-\phi_{lllj}) \bar{x}_{jn,rss}} \right) - r \bar{y}_{r,rss} \right] & \text{for } i \in \bar{A} \end{cases}$$

The consequent estimators under strategies I, II and III are given by

$$\begin{aligned}
T_{d1multi} &= \bar{y}_{r,rs} + \sum_{j=1}^p \beta_{Ij} (\bar{x}_{jn,rs} - \bar{X}_j) \\
T_{d2multi} &= \bar{y}_{r,rs} + \sum_{j=1}^p \beta_{IIj} (\bar{x}_{jr,rs} - \bar{X}_j) \\
T_{d3multi} &= \bar{y}_{r,rs} + \sum_{j=1}^p \beta_{IIIj} (\bar{x}_{jr,rs} - \bar{x}_{jn,rs}) \\
T_{u4multi} &= \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jn,rs}} \right)^{\theta_{Ij}} \\
T_{u5multi} &= \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jr,rs}} \right)^{\theta_{IIj}} \\
T_{u6multi} &= \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rs}}{\bar{x}_{jr,rs}} \right)^{\theta_{IIIj}} \\
T_{u7multi} &= \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\phi_{Ij} \bar{x}_{jn,rs} + (1 - \phi_{Ij}) \bar{X}_j} \right) \\
T_{u8multi} &= \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\phi_{IIj} \bar{x}_{jr,rs} + (1 - \phi_{IIj}) \bar{X}_j} \right) \\
T_{u9multi} &= \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rs}}{\phi_{IIIj} \bar{x}_{jr,rs} + (1 - \phi_{IIIj}) \bar{x}_{jn,rs}} \right)
\end{aligned}$$

where β_{Ij} , $j = 1, 2, \dots, p$, β_{IIj} , β_{IIIj} , θ_{Ij} , θ_{IIj} , θ_{IIIj} , ϕ_{Ij} , ϕ_{IIj} , and ϕ_{IIIj} are well chosen optimising scalars.

Utilizing multi auxiliary information, Bhushan and Kumar²⁵ presented the generalized form of the imputation procedures using RSS under three different strategies as

Strategy I

$$\begin{aligned}
y_{is10} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{(n\alpha_{10}-r)}{(n-r)} \bar{y}_{r,rs} + \frac{n}{n-r} \left[\sum_{j=1}^p \eta_{Ij} (\bar{x}_{jn,rs} - \bar{X}_j) \right] & \text{if } i \in \bar{A} \end{cases} \\
y_{is13} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\alpha_{13} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jn,rs}} \right)^{\delta_{Ij}} - r \bar{y}_{r,rs} \right] & \text{if } i \in \bar{A} \end{cases} \\
y_{is16} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\alpha_{16} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\Theta_{Ij} \bar{x}_{jn,rs} + (1 - \Theta_{Ij}) \bar{X}_j} \right) - r \bar{y}_{r,rs} \right] & \text{for } i \in \bar{A} \end{cases}
\end{aligned}$$

Strategy II

$$\begin{aligned}
y_{is11} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{(n\alpha_{11}-r)}{(n-r)} \bar{y}_{r,rs} + \frac{n}{n-r} \left[\sum_{j=1}^p \eta_{IIj} (\bar{x}_{jr,rs} - \bar{X}_j) \right] & \text{if } i \in \bar{A} \end{cases} \\
y_{is14} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\alpha_{14} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jr,rs}} \right)^{\delta_{IIj}} - r \bar{y}_{r,rs} \right] & \text{if } i \in \bar{A} \end{cases} \\
y_{is17} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\alpha_{17} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\Theta_{IIj} \bar{x}_{jr,rs} + (1 - \Theta_{IIj}) \bar{X}_j} \right) - r \bar{y}_{r,rs} \right] & \text{for } i \in \bar{A} \end{cases}
\end{aligned}$$

Strategy III

$$\begin{aligned}
 y_{.is_{12}} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{(n\alpha_{12}-r)}{(n-r)} \bar{y}_{r,rs} + \frac{n}{n-r} \left[\sum_{j=1}^p \eta_{IIIj} (\bar{x}_{jr,rs} - \bar{x}_{jn,rs}) \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_{15}} &= \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\alpha_{15} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rs}}{\bar{x}_{jr,rs}} \right)^{\delta_{IIIj}} - r \bar{y}_{r,rs} \right] & \text{if } i \in \bar{A} \end{cases} \\
 y_{.is_{18}} &= \begin{cases} y_i & \text{for } i \in A \\ \frac{1}{n-r} \left[n\alpha_{18} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rs}}{\Theta_{IIIj} \bar{x}_{jr,rs} + (1-\Theta_{IIIj}) \bar{x}_{jn,rs}} \right) - r \bar{y}_{r,rs} \right] & \text{for } i \in \bar{A} \end{cases}
 \end{aligned}$$

The consequent estimators are

$$T_{s_{10}} = \alpha_{10} \bar{y}_{r,rs} + \sum_{j=1}^p \eta_{Ij} (\bar{x}_{jn,rs} - \bar{X}_j) \quad (2)$$

$$T_{s_{11}} = \alpha_{11} \bar{y}_{r,rs} + \sum_{j=1}^p \eta_{IIj} (\bar{x}_{jr,rs} - \bar{X}_j) \quad (3)$$

$$T_{s_{12}} = \alpha_{12} \bar{y}_{r,rs} + \sum_{j=1}^p \eta_{IIIj} (\bar{x}_{jr,rs} - \bar{x}_{jn,rs}) \quad (4)$$

$$T_{s_{13}} = \alpha_{13} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jn,rs}} \right)^{\delta_{Ij}} \quad (5)$$

$$T_{s_{14}} = \alpha_{14} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\bar{x}_{jr,rs}} \right)^{\delta_{IIj}} \quad (6)$$

$$T_{s_{15}} = \alpha_{15} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rs}}{\bar{x}_{jr,rs}} \right)^{\delta_{IIIj}} \quad (7)$$

$$T_{s_{16}} = \alpha_{16} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\Theta_{Ij} \bar{x}_{jn,rs} + (1 - \Theta_{Ij}) \bar{X}_j} \right) \quad (8)$$

$$T_{s_{17}} = \alpha_{17} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{X}_j}{\Theta_{IIj} \bar{x}_{jr,rs} + (1 - \Theta_{IIj}) \bar{X}_j} \right) \quad (9)$$

$$T_{s_{18}} = \alpha_{18} \bar{y}_{r,rs} \prod_{j=1}^p \left(\frac{\bar{x}_{jn,rs}}{\Theta_{IIIj} \bar{x}_{jr,rs} + (1 - \Theta_{IIIj}) \bar{x}_{jn,rs}} \right) \quad (10)$$

where α_i , $i = 10, 11, \dots, 18$. Also, η_{Ij} , $j = 1, 2, \dots, p$, η_{IIj} , η_{IIIj} , δ_{Ij} , δ_{IIj} , δ_{IIIj} , Θ_{Ij} , Θ_{IIj} , and Θ_{IIIj} are the well chosen optimizing scalars. The mean square error (MSE) of these resulting estimators is given in Appendix B.

Proposed imputation procedures

Imputation methods provide a method to effectively tackle the problem of missing data by making the data set structure complete so that further inferences can be made. It is also desirable to have certain optimal criterion governing imputation. In this paper, the following objectives were laid down while proposing the imputations.

- Providing an effective imputation for the purpose of estimating the population mean;
- While imputing the missing observations, keep the original replies unaltered.

In view of the above, we suggest some logarithmic type imputation procedures under MCAR assumption. The motivation to propose logarithmic type imputation comes from the works of Bhushan and Kumar³¹ and Bhushan

et al.⁵ where the logarithmic type estimators perform extremely well in both symmetric as well as asymmetric populations. The reader is referred to the works for a detailed exposition.

Bivariate auxiliary information

We propose imputation procedures under three distinct strategies, each utilizing bivariate auxiliary data as

Strategy I

$$y_{iw_1} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\zeta_1 \bar{y}_{r,rss} \left\{ 1 + \log \left(\frac{\bar{x}_{1n,rss}}{\bar{X}_1} \right) \right\}^{\psi_1} \left\{ 1 + \log \left(\frac{\bar{x}_{2n,rss}}{\bar{X}_2} \right) \right\}^{\psi_2} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

Strategy II

$$y_{iw_2} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\zeta_2 \bar{y}_{r,rss} \left\{ 1 + \log \left(\frac{\bar{x}_{1r,rss}}{\bar{X}_1} \right) \right\}^{\psi_3} \left\{ 1 + \log \left(\frac{\bar{x}_{2r,rss}}{\bar{X}_2} \right) \right\}^{\psi_4} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

Strategy III

$$y_{iw_3} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\zeta_3 \bar{y}_{r,rss} \left\{ 1 + \log \left(\frac{\bar{x}_{1n,rss}}{\bar{X}_1} \right) \right\}^{\psi_5} \left\{ 1 + \log \left(\frac{\bar{x}_{2n,rss}}{\bar{X}_2} \right) \right\}^{\psi_6} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

The consequent estimators are given by

$$\begin{aligned} T_{w_1} &= \zeta_1 \bar{y}_{r,rss} \left\{ 1 + \log \left(\frac{\bar{x}_{1n,rss}}{\bar{X}_1} \right) \right\}^{\psi_1} \left\{ 1 + \log \left(\frac{\bar{x}_{2n,rss}}{\bar{X}_2} \right) \right\}^{\psi_2} \\ T_{w_2} &= \zeta_2 \bar{y}_{r,rss} \left\{ 1 + \log \left(\frac{\bar{x}_{1r,rss}}{\bar{X}_1} \right) \right\}^{\psi_3} \left\{ 1 + \log \left(\frac{\bar{x}_{2r,rss}}{\bar{X}_2} \right) \right\}^{\psi_4} \\ T_{w_3} &= \zeta_3 \bar{y}_{r,rss} \left\{ 1 + \log \left(\frac{\bar{x}_{1n,rss}}{\bar{X}_1} \right) \right\}^{\psi_5} \left\{ 1 + \log \left(\frac{\bar{x}_{2n,rss}}{\bar{X}_2} \right) \right\}^{\psi_6} \end{aligned}$$

where ζ_i , $i = 1, 2, 3$ and ψ_j , $j = 1, 2, \dots, 6$ are suitably chosen scalars.

Multi auxiliary information

We propose imputation procedures under three distinct strategies, each utilizing multi auxiliary data as

Strategy I

$$y_{iw_4} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\zeta_4 \bar{y}_{r,rss} \prod_{j=1}^p \left\{ 1 + \log \left(\frac{\bar{x}_{jn,rss}}{\bar{X}_j} \right) \right\}^{\psi_{Ij}} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

Strategy II

$$y_{iw_5} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\zeta_5 \bar{y}_{r,rss} \prod_{j=1}^p \left\{ 1 + \log \left(\frac{\bar{x}_{jr,rss}}{\bar{X}_j} \right) \right\}^{\psi_{IIj}} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

Strategy III

$$y_{iw_6} = \begin{cases} y_i & \text{if } i \in A \\ \frac{1}{n-r} \left[n\zeta_6 \bar{y}_{r,rss} \prod_{j=1}^p \left\{ 1 + \log \left(\frac{\bar{x}_{jn,rss}}{\bar{X}_j} \right) \right\}^{\psi_{IIIj}} - r\bar{y}_{r,rss} \right] & \text{if } i \in \bar{A} \end{cases}$$

The consequent estimators are given by

$$T_{w_4} = \zeta_4 \bar{y}_{r,rss} \prod_{j=1}^p \left\{ 1 + \log \left(\frac{\bar{x}_{jn,rss}}{\bar{X}_j} \right) \right\}^{\psi_{Ij}}$$

$$T_{w_5} = \zeta_5 \bar{y}_{r,rss} \prod_{j=1}^p \left\{ 1 + \log \left(\frac{\bar{x}_{jr,rss}}{\bar{X}_j} \right) \right\}^{\psi_{IIj}}$$

$$T_{w_6} = \zeta_6 \bar{y}_{r,rss} \prod_{j=1}^p \left\{ 1 + \log \left(\frac{\bar{x}_{jn,rss}}{\bar{X}_j} \right) \right\}^{\psi_{IIIj}}$$

where ζ_i , $i = 4, 5, 6$ and ψ_{Ij} , $j = 1, 2, \dots, p$; ψ_{IIj} , and ψ_{IIIj} are suitably chosen scalars.

Theorem 1 The suggested difference kind of imputation techniques comprise the resulting estimators T_{w_i} , $i = 1, 2, \dots, 6$, whose bias and minimum MSE are

$$\text{Bias}(T_{w_i}) = \bar{Y}(\zeta_i Q_i - 1) \quad (11)$$

$$\text{minMSE}(T_{w_i}) = \bar{Y}^2 \left(1 - \frac{Q_i^2}{P_i} \right) \quad (12)$$

Proof Appendix C provides a concise description and derivation of parametric functions P_i and Q_i . $\square$

Efficiency conditions

On comparing the minimum MSE of the proposed consequent estimators with the minimum MSE of the existing consequent estimators, we get the following efficiency conditions.

- (i) According to strategy I, the suggested imputation technique y_{w_1} outperforms the traditional difference imputation procedure y_{d_1} and the power transformation imputation procedures y_{u_1} , y_{u_4} .

$$\frac{Q_1^2}{P_1} > 1 - V_0^* + \frac{(V_1 V_{02}^2 + V_2 V_{01}^2 - 2V_{01} V_{02} V_{12})}{(V_1 V_2 - V_{12}^2)} \quad (13)$$

- (ii) According to strategy II, the suggested imputation technique y_{w_2} outperforms the traditional difference imputation procedure y_{d_2} and the power transformation imputation procedures y_{u_2} , y_{u_5} .

$$\frac{Q_2^2}{P_2} > 1 - V_0^* + \frac{(V_1^* V_{02}^2 + V_2^* V_{01}^2 - 2V_{01}^* V_{02}^* V_{12}^*)}{(V_1^* V_2^* - V_{12}^{*2})} \quad (14)$$

- (iii) According to strategy I, the suggested imputation technique y_{w_3} outperforms the traditional difference imputation procedure y_{d_3} and the power transformation imputation procedures y_{u_3} , y_{u_6} .

$$\frac{Q_3^2}{P_3} > 1 - V_0^* + \frac{\left((V_1 - V_1^*)(V_{02}^* - V_{02})^2 + (V_2^* - V_2)(V_{01}^* - V_{01})^2 \right)}{(V_1^* - V_1^*)(V_2^* - V_2) - (V_{12}^* - V_{12})^2} \quad (15)$$

- (iv) Under the corresponding strategies, the suggested imputation approaches y_{w_i} , $i = 1, 2, 3$ outperform those of Bhushan and Kumar²⁵ approaches y_{s_i} , $i = 1, 2, 3$.

$$\frac{Q_i^2}{P_i} > \alpha_{i(opt)} \quad (16)$$

- (v) Under the corresponding strategies, the suggested imputation approaches y_{w_i} , $i = 1, 2, 3$ outperform those of Bhushan and Kumar²⁵ approaches y_{s_j} , $j = 4, 5, \dots, 9$.

$$\frac{Q_i^2}{P_i} > \frac{B_j^2}{A_j} \quad (17)$$

The above conditions are the limitations under which the suggested imputation techniques outperform the well-known imputation techniques now in use. Furthermore, a simulation study is used to determine these conditions.

Simulation study

In order to assess our primary objective i.e., to draw an efficient imputation for the population mean estimation, we perform a comprehensive simulation under different correlations. The simulation steps are described below:

- (i) Generate artificially a trivariate normal population $N(\underline{\mu}, \Sigma)$ of size $N=500$ with mean vector $\underline{\mu}$ and dispersion matrix Σ given by

$$\underline{\mu} = \begin{pmatrix} \mu_y \\ \mu_{x_1} \\ \mu_{x_2} \end{pmatrix} = \begin{pmatrix} 20 \\ 15 \\ 10 \end{pmatrix}$$

$$\text{and } \Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{x_1 y} & \sigma_{x_2 y} \\ \sigma_{x_1 y} & \sigma_{x_1}^2 & \sigma_{x_1 x_2} \\ \sigma_{x_2 y} & \sigma_{x_1 x_2} & \sigma_{x_2}^2 \end{pmatrix} = \begin{pmatrix} 25 & 22.36 \times \rho_{yx_1} & 75 \times \rho_{x_2 y} \\ 22.36 \times \rho_{x_1 y} & 20 & 67.08 \times \rho_{x_1 x_2} \\ 75 \times \rho_{x_2 y} & 67.08 \times \rho_{x_1 x_2} & 15 \end{pmatrix}.$$

The covariances are addressed to ensure different values of correlation coefficients $\rho_{x_1 y}$, $\rho_{x_2 y}$ and $\rho_{x_1 x_2}$. The values of $\rho_{x_1 y}$, $\rho_{x_2 y}$, $\rho_{x_1 x_2}$ seems to be a reasonable selection to notice the behaviour of the proposed imputation procedures.

- (ii) Select a sample of size $n = 12$ units with set size $m = 3$ and number of cycles $r = 4$ from the above artificially generated population.
- (iii) Drop down $(n - n_1)$ units with the responding probability $p = 0.67$ from the above sample.
- (iv) Using 15000 iterations, we have calculated the MSE and percent relative efficiency (PRE) of proposed consequent estimators w.r.t. the conventional mean estimator using the following expression.

$$MSE = \frac{1}{15,000} \sum_{i=1}^{15,000} (t^* - \bar{Y})^2$$

$$PRE = \frac{MSE(\bar{y}_{[n]})}{MSE(T)} \times 100$$

- (v) The simulation results are cited in Table 1 in terms of MSE and PRE of the existing and proposed resulting estimators for different choices of correlation coefficients.

Discussion of simulation results

The simulation findings cited in Table 1 are interpreted as follows:

- (i) Under strategy I, the proposed logarithmic imputation procedure T_{w_1} dominates the mean per unit imputation procedure t_m , difference type imputation procedure T_{d_1} , ratio type imputation procedures T_{u_1} and T_{u_4} , improved difference and ratio type imputation procedures T_{s_i} , $i = 1, 4, 7$ envisaged by Bhushan and Kumar³⁵ with least MSE and highest PRE. The similar dominance can also be observed from strategy II and strategy III.
- (ii) Under strategy I, the MSE of the existing and proposed imputation procedures decreases as the correlation coefficients $\rho_{x_1 y}$, $\rho_{x_2 y}$ and $\rho_{x_1 x_2}$ increases. This effect can be seen from Fig. 1. The similar effect can also be seen from strategies II and III and the figures for the same can be drawn.
- (iii) Under strategy I, the PRE of the existing and proposed imputation procedures increases as the correlation coefficients $\rho_{x_1 y}$, $\rho_{x_2 y}$ and $\rho_{x_1 x_2}$ increases. This effect can be seen from Fig. 2. The similar effect can also be seen from strategies II and III and the figures for the same can be drawn.
- (iv) The MSE and PRE of the proposed estimator T_{w_2} are least and maximum under strategy II as compare to the MSE and PRE of the proposed estimators T_{w_1} and T_{w_3} under strategies I and III, respectively.

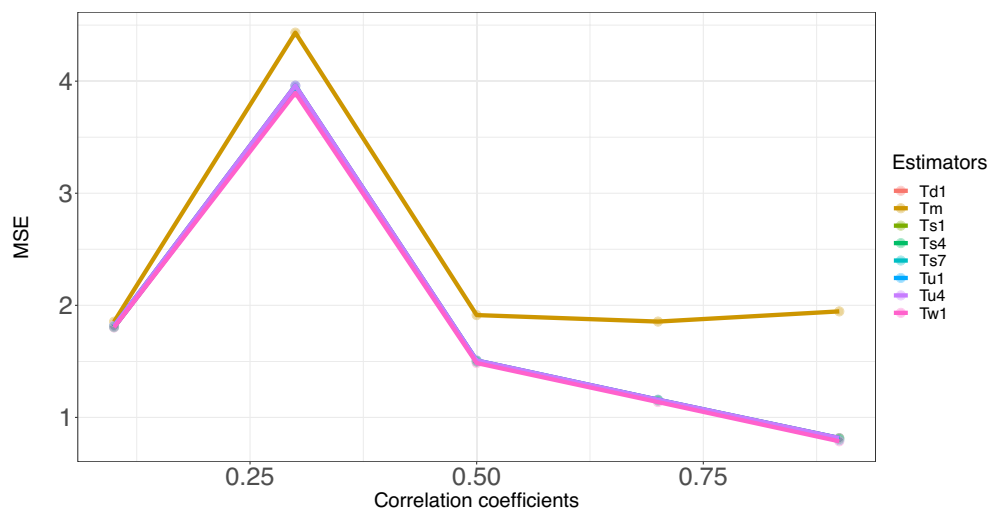


Figure 1. Graph of MSE of resulting estimators for simulated population under Strategy I.

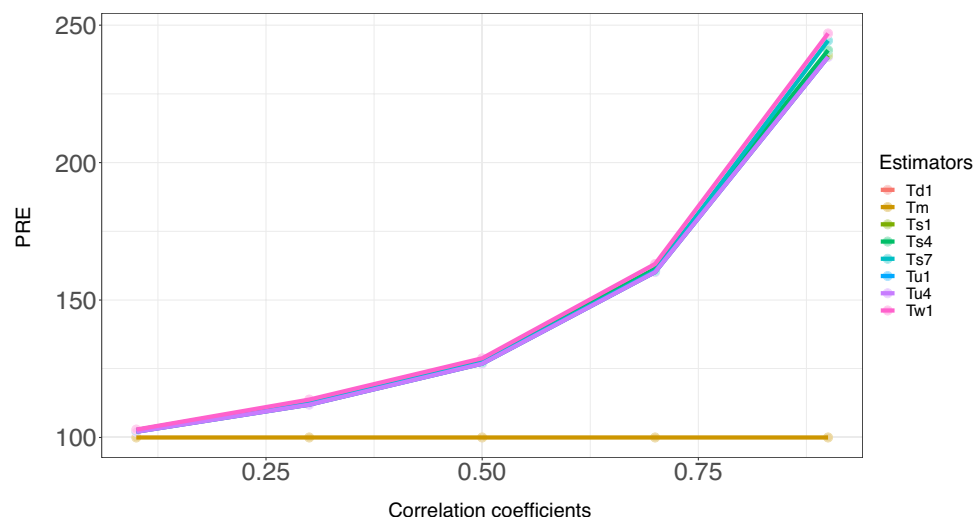


Figure 2. Graph of PRE of resulting estimators for simulated population under Strategy I.

ρ_{x_1y}	0.1		0.3		0.5		0.7		0.9	
ρ_{x_2y}	0.1		0.3		0.5		0.7		0.9	
$\rho_{x_1x_2}$	0.1		0.3		0.5		0.7		0.9	
Estimators	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
T_m	1.854	100.00	4.433	100.00	1.913	100.00	1.855	100.00	1.946	100.00
Strategy I										
T_{d1}	1.816	102.091	3.960	111.941	1.508	126.855	1.158	160.173	0.816	238.464
T_{u1}	1.816	102.091	3.960	111.941	1.508	126.855	1.158	160.173	0.816	238.464
T_{u4}	1.816	102.091	3.960	111.941	1.508	126.855	1.158	160.173	0.816	238.464
T_{s1}	1.807	102.589	3.924	112.983	1.502	127.368	1.154	160.672	0.814	238.988
T_{s4}	1.805	102.718	3.913	113.282	1.496	127.871	1.149	161.392	0.807	240.892
T_{s7}	1.807	102.588	3.924	112.959	1.502	127.325	1.155	160.563	0.796	244.383
T_{w1}	1.802	102.875	3.896	113.782	1.485	128.810	1.137	163.129	0.787	246.979
Strategy II										
T_{d2}	1.797	103.153	3.727	118.936	1.308	146.193	0.815	227.642	0.259	749.857
T_{u2}	1.797	103.153	3.727	118.936	1.308	146.193	0.815	227.642	0.259	749.857
T_{u5}	1.797	103.153	3.727	118.936	1.308	146.193	0.815	227.642	0.259	749.857
T_{s2}	1.789	103.651	3.695	119.978	1.304	146.705	0.813	228.140	0.259	750.381
T_{s5}	1.785	103.851	3.680	120.470	1.295	147.678	0.806	230.050	0.252	771.725
T_{s8}	1.789	103.651	3.696	119.940	1.305	146.634	0.814	227.925	0.233	832.059
T_{w2}	1.781	104.108	3.653	121.351	1.277	149.780	0.786	235.893	0.219	887.281
Strategy III										
T_{d3}	1.835	101.019	4.200	105.545	1.714	111.641	1.512	122.704	1.389	140.054
T_{u3}	1.835	101.019	4.200	105.545	1.714	111.641	1.512	122.704	1.389	140.054
T_{u6}	1.835	101.019	4.200	105.545	1.714	111.641	1.512	122.704	1.389	140.054
T_{s3}	1.826	101.517	4.159	106.588	1.706	112.153	1.505	123.202	1.384	140.578
T_{s6}	1.825	101.578	4.154	106.722	1.703	112.350	1.503	123.437	1.380	140.985
T_{s9}	1.826	101.517	4.159	106.576	1.706	112.134	1.506	123.160	1.374	141.531
T_{w3}	1.824	101.650	4.146	106.930	1.698	112.662	1.497	123.885	1.371	141.883

Table 1. MSE and PRE of resultant estimators for normal population. The bold values are representing the least MSE and highest PRE of the estimators.

Real data applications

In this part, we use four real populations made up of bivariate auxiliary data to analyze the theoretical outcomes of the suggested and contemporary estimators. Bhushan and Kumar²⁵ have recently utilized populations 1 and 2. The following is a description of these populations' data sets:

Population 1: The data set was taken from Singh⁽³²⁾, pp. 1116), where the study variable y is the quantity of fish captured in 1995, and the auxiliary variables x_1 and x_2 are the quantity of fish caught in 1994 and 1993, respectively. The required parameter values are given in the following manner: $N=69$, $n = 12$, $m = 3$, $k = 4$, $P = 0.67$, $\bar{Y} = 4514.89$, $\bar{X}_1 = 4954.43$, $\bar{X}_2 = 4591.07$, $C_y = 1.35$, $C_{x_1} = 1.42$, $C_{x_2} = 1.37$, $\rho_{x_1y} = 0.96$, $\rho_{x_2y} = 0.95$, and $\rho_{x_1x_2} = 0.97$.

Population 2: The data set for population 2 is derived from Sarndal⁽³³⁾, pp. 652–659), where the variables y , x_1 , and x_2 represent the 1983 population (in millions), import (in millions of U.S. dollars), and export (in millions of U.S. dollars), respectively. The required parameter values are given in the following manner: $N=124$, $n = 12$, $m = 3$, $k = 4$, $P = 0.67$, $\bar{Y} = 36.65$, $\bar{X}_1 = 14604.49$, $\bar{X}_2 = 14276.03$, $C_y = 3.18$, $C_{x_1} = 2.33$, $C_{x_2} = 2.20$, $\rho_{x_1y} = 0.23$, $\rho_{x_2y} = 0.22$, and $\rho_{x_1x_2} = 0.98$.

Population 3: Sarndal⁽³³⁾, pp. 652–659) is another source of data for population 3. There, the total number of seats (S82) in the municipal council in 1982 is taken as y , the number of conservative seats (CS82) in the municipal council in 1982 is taken as variable x_1 , and the number of social-democratic seats (SS82) in the municipal council in 1982 is taken as variable x_2 . The required parameter values are given in the following manner: $N=284$, $n = 12$, $m = 3$, $k = 4$, $P = 0.67$, $\bar{Y} = 47.50$, $\bar{X}_1 = 9.05$, $\bar{X}_2 = 22.11$, $C_y = 0.23$, $C_{x_1} = 0.54$, $C_{x_2} = 0.33$, $\rho_{x_1y} = 0.65$, $\rho_{x_2y} = 0.75$, and $\rho_{x_1x_2} = 0.20$.

The MSE and PRE of the suggested imputation techniques in relation to the traditional mean imputation procedure are computed using the following formula, which takes into account the characteristics of the aforementioned populations:

$$PRE = \frac{MSE(t_m)}{MSE(T)} \times 100 \quad (18)$$

Table 2 presents an overview of the empirical findings derived from the aforementioned populations using MSE and PRE. The findings of Table 2 under strategy I show that, with the lowest MSE and maximum PRE, the proposed imputation techniques outperform the difference and ratio type imputation procedures as well as the imputation procedures recommended by Bhushan and Kumar²⁵. One can observe this tendency in Figs. 3 and 4, respectively. Additionally, strategies II and III have a similar trend. For strategies II and III, the MSE and PRE graphs can also be created.

Conclusion

This paper considered some novel logarithmic type imputation procedures and the consequent estimators for the estimation of population mean in situation of missing data under RSS. These imputation procedures are suggested under the assumption of MCAR. The MSE of the proposed consequent estimators has obtained to the first order approximation. The limitations of the proposed imputation procedures have been obtained in terms of efficiency conditions under which the proposed imputation procedures dominate the existing imputation procedures. These efficiency conditions have been evaluated by an extensive simulation study performed over an artificially drawn population for various prespecified values of correlation coefficients. The simulation results are summarized in Table 1 by MSE and PRE. It has been observed from the simulation results that the suggested logarithmic imputation techniques y_{iwj} , $j = 1, 2, 3$ outperform their conventional counterparts such as mean imputation y_{im} , regression imputations y_{idj} , $j = 1, 2, 3$, power transformation imputations y_{iuj} , $j = 1, 2, \dots, 6$, and the imputation procedures y_{isj} , $j = 1, 2, \dots, 9$ recommended by Bhushan and Kumar²⁵ using multi auxiliary information under RSS. As the correlation coefficient values increase, the PRE of the suggested imputation techniques y_{iwj} , $j = 1, 2, 3$ steadily increases in every strategy, but it still outperforms its traditional equivalents.

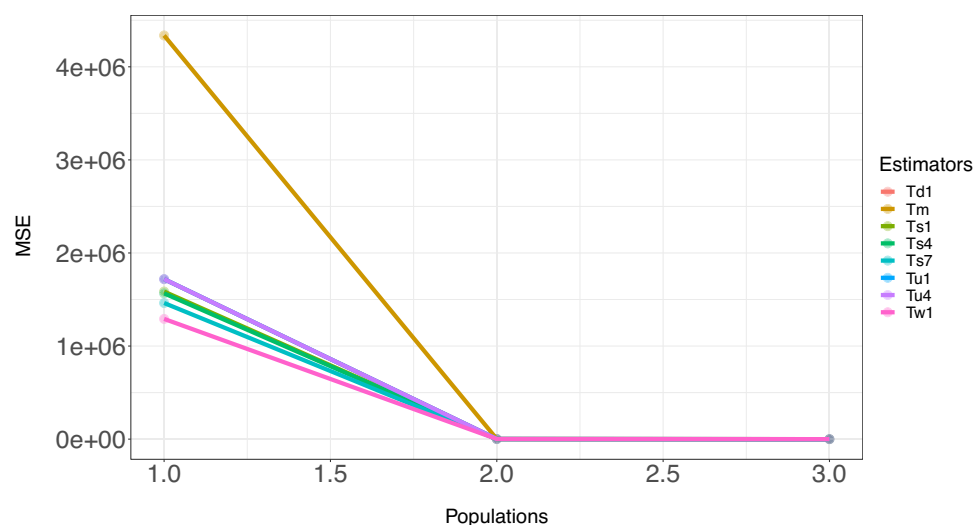


Figure 3. Graph of MSE of resulting estimators for real populations under Strategy I.

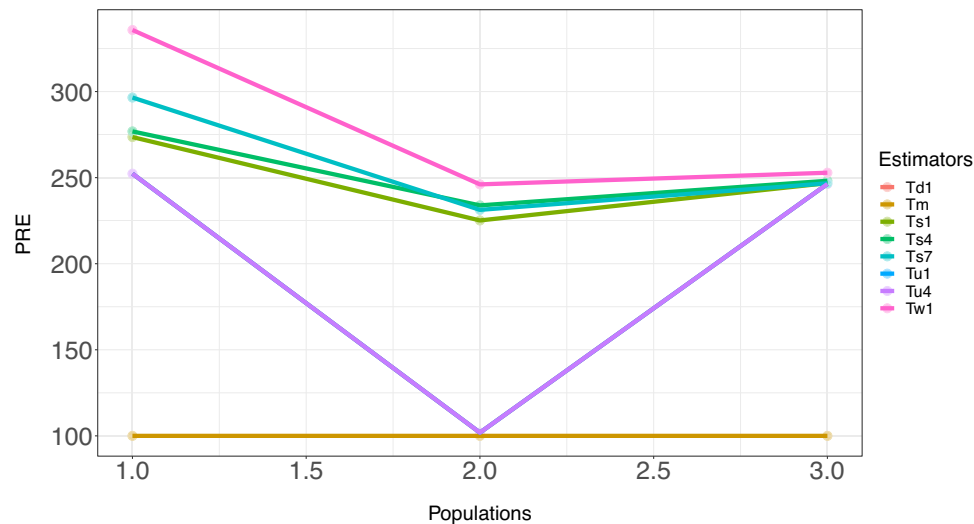


Figure 4. Graph of PRE of resulting estimators for real populations under Strategy I.

Estimators	Population 1		Population 2		Population 3	
	MSE	PRE	MSE	PRE	MSE	PRE
T_m	4,336,620.00	100	1655.81	100	11.07	100
Strategy I						
T_{d1}	1,718,834.00	252.30	1625.27	101.87	4.49	246.32
T_{u1}	1,718,834.00	252.30	1625.27	101.87	4.49	246.32
T_{u4}	1,718,834.00	252.30	1625.27	101.87	4.49	246.32
T_{s1}	1,585,170.00	273.57	735.46	225.13	4.48	246.81
T_{s4}	1,566,310.00	276.86	707.96	233.88	4.46	248.18
T_{s7}	1,462,490.00	296.52	715.91	231.28	4.48	246.72
T_{w1}	1,291,567.00	335.76	672.86	246.08	4.38	252.85
Strategy II						
T_{d2}	429,476.00	1009.74	1610.23	102.83	1.25	882.02
T_{u2}	429,476.00	1009.74	1610.23	102.83	1.25	882.02
T_{u5}	429,476.00	1009.74	1610.23	102.83	1.25	882.02
T_{s2}	420,614.10	1031.02	732.36	226.09	1.25	882.51
T_{s5}	411,322.50	1054.31	691.28	239.52	1.23	900.22
T_{s8}	327,292.30	1324.99	703.16	235.48	1.25	882.11
T_{w2}	129,543.40	3347.61	638.84	259.18	1.10	1006.05
Strategy III						
T_{d3}	3,047,263.00	142.31	1640.77	100.91	7.83	141.35
T_{u3}	3,047,263.00	142.31	1640.77	100.91	7.83	141.35
T_{u6}	3,047,263.00	142.31	1640.77	100.91	7.83	141.35
T_{s3}	2,650,968.00	163.58	738.61	224.17	7.80	141.85
T_{s6}	2,636,627.00	164.47	725.08	228.36	7.79	142.14
T_{s9}	2,565,499.00	169.03	728.99	227.13	7.81	141.82
T_{w3}	2,463,578.00	176.02	707.82	233.92	7.75	142.85

Table 2. MSE and PRE of consequent estimators for real populations. The bold values are representing the least MSE and highest PRE of the estimators.

The simulation results have also indicated that the maximum efficiency for the discussed imputation procedures have found with respect to strategy II for various correlation coefficient values. The results of real data application have shown that the proposed imputation procedures obtained least MSE and highest PRE in comparison to the existing imputation procedures. Therefore, the proposed imputation procedures may be opted to estimate the population mean in presence of missing data under RSS.

In the future studies, to deal with the missing data, utilizing multi-auxiliary information, we hope to develop these imputation procedures under stratified ranked set sampling.

Data availability

all data generated or analysed during this study are included in this article.

Received: 9 May 2024; Accepted: 30 July 2024

Published online: 04 August 2024

References

- Hansen, M. H. & Hurwitz, W. N. The problem of non-response in sample surveys. *J. Am. Stat. Assoc.* **41**, 517–529 (1946).
- Srinath, K. P. Multiphase sampling in non-response problems. *J. Am. Stat. Assoc.* **66**, 583–586 (1971).
- Rubin, R. B. Inference and missing data. *Biometrika* **63**(3), 581–592 (1976).
- Heitjan, D. F. & Basu, S. Distinguishing ‘missing at random’ and ‘missing completely at random’. *Am. Stat.* **50**, 207–213 (1996).
- Bhushan, S., Kumar, A., Pandey, A. P. & Singh, S. Estimation of population mean in presence of missing data under simple random sampling. *Commun. Stat. Simul. Comput.* **52**(12), 6048–6069 (2022).
- Sohil, F., Sohail, M. U. & Shabbir, J. Optimum second call imputation in PPS sampling. *PLoS ONE* **17**(1), e0261834 (2022).
- Bhushan, S. & Pandey, A. P. Optimal imputation of the missing data using multi auxiliary information. *Comput. Stat.* **36**(1), 449–477 (2020).
- Prasad, S. & Yadav, V. K. Imputation of missing data through product type exponential methods in sampling theory. *Rev. Colomb. Estad.* **46**(1), 111–127 (2023).
- Yadav, V. K. & Prasad, S. Generalized class of factor type exponential imputation techniques for population mean using simulation approach. *J. Stat. Comput. Simul.* **94**(9), 1997–2039 (2024).
- Bhushan, S. & Kumar, A. Ranked set sampling imputation methods in presence of correlated measurement errors. *Commun. Stat. Theory Methods* <https://doi.org/10.1080/03610926.2024.2352031> (2024).
- McIntyre, G. A. A method of unbiased selective sampling using ranked set. *Aust. J. Agric. Res.* **3**, 385–390 (1952).
- Singh, S. & Horn, S. Compromised imputation in survey sampling. *Metrika* **51**, 267–276 (2000).
- Ahmed, M. S., Al-Titi, O., Al-Rawi, Z. & Abu-Dayyeh, W. Estimation of a population mean using different imputation methods. *Stat. Transit.* **7**(6), 1247–1264 (2006).
- Diana, G. & Perri, P. F. Improved estimators of the population mean for missing data. *Commun. Stat. Theory Methods* **39**, 3245–3251 (2010).
- Takahasi, K. & Wakimoto, K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Stat. Math.* **20**, 1–31 (1968).
- Shahzad, U., Hanif, M., Koyuncu, N. & Luengo, A. V. G. A new estimator of mean under ranked set sampling alongside the sensitivity issue. *J. Stat. Manag. Syst.* **21**(8), 1553–1564 (2018).
- Shahzad, U., Hanif, M., Koyuncu, N. & Luengo, A. V. G. A regression type estimator for mean estimation under ranked set sampling alongside the sensitivity issue. *Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.* **68**(2), 2037–2049 (2019).
- Bhushan, S. & Kumar, A. On optimal classes of estimators under ranked set sampling. *Commun. Stat. Theory Methods* **51**(8), 2610–2639 (2022).
- Bhushan, S. & Kumar, A. Novel log type class of estimators under ranked set sampling. *Sankhya B* **84**, 421–447 (2022).
- Kocyyigit, E. G. & Rather, K. U. I. The new sub-regression type estimator in ranked set sampling. *J. Stat. Theory Pract.* **17**(2), 27 (2023).
- Khalid, R., Kocyyigit, E. G. & Unal, C. New exponential ratio estimator in Ranked set sampling. *Pak. J. Stat. Oper. Res.* <https://doi.org/10.18187/pjsor.v18i2.3921> (2022).
- Kocyyigit, E. G. & Kadilar, C. Ratio estimators for ranked set sampling in the presence of tie information. *Commun. Stat. Simul. Comput.* **51**(11), 6826–6839 (2022).
- Bhushan, S. & Kumar, A. Predictive estimation approach using difference and ratio type estimators in ranked set sampling. *J. Comput. Appl. Math.* **410**, 114214 (2022).
- Bhushan, S. & Kumar, A. Novel predictive estimators using ranked set sampling. *Concur. Comput. Pract. Exp.* **35**(3), e7435 (2022).
- Bhushan, S. & Kumar, A. Imputation of missing data using multi auxiliary information under ranked set sampling. *Commun. Stat. Simul. Comput.* <https://doi.org/10.1080/03610918.2023.2288796> (2023).
- Kocyyigit, E. G. A novel sub-type mean estimator for ranked set sampling with dual auxiliary variables. *J. New Theory* **44**, 79–86 (2023).
- Bouza Herrera, C. N. & Al-Omari, A. I. Ranked set estimation with imputation of the missing observations: The median estimator. *Rev. Investig. Oper.* **32**(1), 30–37 (2011).
- Bouza, C. N. *Handling Missing Data in Ranked Set Sampling* (Springer, 2013).
- Sohail, M. U., Shabbir, J. & Ahmed, S. A class of ratio type estimators for imputing the missing values under rank set sampling. *J. Stat. Theory Pract.* **12**, 704–717 (2018).
- Bhushan, S., Kumar, A., Zaman, T. & Al Mutairi, A. Efficient difference and ratio-type imputation methods under ranked set sampling. *Axioms* **12**, 558 (2023).
- Bhushan, S. & Kumar, A. New efficient logarithmic estimators using multi-auxiliary information under ranked set sampling. *Concur. Comput. Pract. Exp.* **34**(27), e7337 (2022).
- Singh, S. *Advanced Sampling Theory with Applications: How Michael Selected Amy* (Kluwer, 2003).
- Sarndal, C. E., Swensson, B. & Wretman, J. *Model Assisted Survey Sampling* (Springer-Verlag, 2003).
- El-Badry, M. A. A sampling procedure for mailed questionnaires. *J. Am. Stat. Assoc.* **51**, 209–227 (1956).
- Bhushan, S., Kumar, A., Pokhrel, R., Bakr, M. E. & Mekiso, G. T. Design based synthetic imputation methods for domain mean. *Sci. Rep.* **14**(1), 4250 (2024).
- Lee, H., Rancourt, E. & Sarndal, C. E. Experiments with variance estimation from survey data with imputed values. *J. Off. Stat.* **10**(3), 231–243 (1994).

Acknowledgements

The authors are extremely grateful to the panel of reviewers and the Editor-in-Chief for their careful insights and constructive suggestions which led to considerable improvement of the paper. The study was funded by Researchers Supporting Project number (RSP2024R488), King Saud University, Riyadh, Saudi Arabia.

Author contributions

A.K.: writing original manuscript, methodology, simulation study, software; S.B.: methodology, review and editing; W.E. and Y.T.: Project administration, financial support; M.J.S.K: review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68940-4>.

Correspondence and requests for materials should be addressed to S.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024