Comparative Study of Reproducibility of Ranked Set Sampling Methods using Predictive Inference

Syed Abdul Rehman[1*], Tahani Coolen-Maturi[2], Frank P. A. Coolen[3], Javid Shabbir[4]

[1*]Department of Mathematical Sciences, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, 87300, Pakistan.
[2,3]Department of Mathematical Sciences, Durham University, Durham, United Kingdom.
[4]Department of Statistics, University of Wah, Wah Cantt, Pakistan.

*Corresponding author(s). E-mail(s): arehman@stat.qau.edu.pk; Contributing authors: tahani.maturi@durham.ac.uk; frank.coolen@durham.ac.uk; javidshabbir@gmail.com;

Abstract

Ranked set sampling (RSS) is an important survey technique aimed at efficient estimation of population characteristics. Various RSS methods can be used to collect an RSS sample, but the reproducibility of estimates using these methods remains unexplored. Reproducibility refers to obtaining similar estimates when the survey is repeated under identical conditions. This study compares the reproducibility of population mean estimates using four basic RSS methods: classical RSS (RSS), Median RSS (MRSS), Extreme RSS (ERSS), and Paired RSS (PRSS). We assess the reproducibility of these methods using Nonparametric Predictive Inference (NPI) bootstrapping. Simulations are conducted for varying sample sizes with both perfect and imperfect rankings, and results are compared for weak and strong associations between the study and concomitant variables. Additionally, we apply these methods to agricultural data from Punjab, India. Our findings indicate that MRSS provides the best reproducibility for population mean estimates, while ERSS performs the worst in this regard.

Keywords: Imprecise probabilities, order statistics, nonparametric predictive inference, statistical reproducibility

1 Introduction

Estimation of population parameters using sample data is a fundamental aspect of classical statistical inference. Various sampling techniques are available to select a sample from a population. Ranked Set Sampling (RSS) is one such method that offers more precise estimates of population characteristics compared to Simple Random Sampling (SRS). McIntyre (1952) introduced the RSS method, and it is considered appropriate when the characteristic under study is time-consuming and costly. This technique enhances estimation accuracy by reducing sampling error. It requires the utilization of a closely related concomitant variable to collect data for the study variable; for example, the soil fertility level can be used to collect data on crop yield. RSS is especially useful in the environmental and biological sciences where obtaining a large sample may be challenging. Takahasi and Wakimoto (1968) developed an unbiased estimator of the population mean under the RSS technique. Muttlak (1998) introduced the Median RSS (MRSS) method and proposed an unbiased estimator of the population mean. Lynne Stokes (1977) used ranks of the concomitant variable to order units of the study variable. The extreme RSS (ERSS) method was suggested by Samawi et al. (1996), it uses extreme order statistics to estimate the population mean. Muttlak (1996) suggested a method of paired RSS (PRSS) and showed that it also allows unbiased estimation of population mean. These four methods are regarded as basic RSS methods; however, other RSS approaches that are modifications of these basic methods have been provided by Al-Saleh and Al-Kadiri (2000), Abu-Dayyeh et al. (2003), and Bhushan et al. (2024).

Recently, researchers have shown an increasing interest in the concept of statistical reproducibility. In estimation theory, it represents the extent to which a statistical method or technique consistently yields similar estimates when the procedure is repeated under the same conditions. Reproducing previous estimates enhances the evidence for any statistical methods. Ensuring the validity and reliability of statistical procedures and their results is a fundamental component of scientific inquiry. Initially, the topic of reproducibility was discussed by Goodman (1992) in statistical tests. He also addressed the misunderstanding between statistical p-value and reproducibility. Further distinctions between the p-value and reproducibility were discussed by Senn (2002). Miller (2009) highlighted the challenges in drawing meaningful conclusions from a single preliminary experiment, particularly when the power of a test is unknown due to the lack of knowledge about effective sample size.The reproducibility was estimated by Posavac (2002) by comparing the value of a test statistic obtained from the actual test results with the corresponding critical value. Shao and Chow (2002) suggested three methods, including the Bayesian approach, to assess reproducibility. The predictive nature of reproducibility and its association with the effective sample size was studied by Killeen (2005). De Martini (2008) proposed several definitions of the reproducibility of statistically significant findings. He also suggested various reproducibility estimators for the Wilcoxon rank sum test and examined their efficiency. Begley and Ellis (2012) highlighted concerns about experiment reproducibility by demonstrating that around 25% of significant results in pre-clinical cancer trials could be reproduced, underscoring the critical issue of reproducibility in scientific research.

Nonparametric Predictive Inference (NPI) is a statistical approach that enables predictions about future observations without relying on strong assumptions regarding the underlying data distribution. The idea is to make probabilistic predictions about future observations based on the observed data. One example of NPI is the construction of predictive probability intervals using the concepts of lower and upper probabilities, which are based on order statistics and provide a way to express uncertainty without assuming a specific distribution. As a result of using this methodology, the concept of Reproducibility Probability (RP) was developed by Coolen and Himd (2020) within the frequentist statistical framework. It is possible to draw logical inferences about RP, given the explicitly predictive nature of NPI. A bootstrap method based on NPI was developed by Coolen and Bin Himd (2014), they used it to estimate the RP. Recently, Simkus et al. (2022) applied the NPI bootstrapping technique to examine the RP of the t-test, demonstrating the practical applicability of NPI in assessing reproducibility.

The current study compares the reproducibility of four basic RSS methods for estimating the population mean using NPI bootstrapping, following the approach of Rehman et al. (2024). In line with Coolen and Alqifari (2018), we describe reproducibility of an RSS method as the probability that the estimate based on a future sample will be similar to the estimate based on the initial sample if sampling is done under the same conditions. We examine the reproducibility of these methods for different ranking criteria, different sample sizes, and the correlation coefficient between the concomitant variable and the study variable using a simulation study. Our objective is to contribute to the RSS literature by investigating the reproducibility of these methods, identifying their potential limitations and advantages. In agricultural research, precise crop yield estimates are essential for efficient farming and food security. Our study comparing RSS methods offers insights for selecting reliable sampling methods, aiding farmers and policymakers in decision-making. Similarly, precise evaluation of the distribution of pollutants and rare species is essential for conservation in environmental studies. Our study guides on choosing more reproducible and reliable sampling strategies, enhancing environmental monitoring, and conservation planning effectiveness. This study commences with a review of RSS methods in Section 2, while Section 3 defines NPI reproducibility and explains its significance. Building on this, Section 4 explains the integration of NPI into RSS estimates and presents an algorithm to investigate it. A detailed simulation study is presented in Section 5 to examine the RP of the RSS methods for hypothetically generated normal and exponential populations. The application of this study to real data is presented in Section 6, while the key findings and conclusions are summarized in Section 7.

2 RSS Methods

This section presents the methodological framework for sample selection under four basic RSS approaches and provides their respective mean estimators. In addition, the imperfect RSS method is also explained in this section.

2.1 The Classical Ranked Set Sampling (RSS)

This method was introduced by McIntyre (1952) and efficiently estimated the mean pasture yield. This method has proven to be an effective approach for estimating the mean and various other population characteristics. The procedure for selecting a sample using this method entails the following steps:

Initially, identify m^2 units from the population and assign them into m independent sets of size m. In each set, the units are ranked using some visual judgments or by using the ranks of any closely related concomitant variable. The lowest order statistic from the first set is selected, and the second lowest order statistic is selected from the second set. Continue selecting units in this manner until the m^{th} order statistic is selected from the m^{th} set. The selected units can be represented as $\left\{Y_{i(i)}\right\}_{i=1}^{m}$. This procedure can be repeated r times to obtain a final sample of size $n = rm$. An estimator of the population mean based on this method was developed by Takahasi and Wakimoto (1968) as

$$\bar{y}_{rss} = \frac{1}{n}\sum_{j=1}^{r}\sum_{i=1}^{m} Y_{i(i)j}. \tag{1}$$

where $Y_{i(i)j}$ shows i^{th} order statistics in the i^{th} set of j^{th} cycle. The estimator $\bar{y}_{rss}$ is unbiased and has less variance than the mean estimator under SRS i.e.,

$$V\left(\bar{y}_{rss}\right) = \frac{\sigma^2}{n} - \frac{1}{rm^2}\sum_{i=1}^{m}\Delta_{(i)}^2. \tag{2}$$

Here, $\frac{\sigma_y^2}{n}$ represents the variance of the mean estimator obtained from an SRSWR sample of size n, denoted by $\bar{y}_{srs}$. The quantity $\Delta_{(i)} = \mu_{(i)} - \mu$ highlights the magnitude of deviation of the i^{th} order statistic mean from the overall population mean μ. Accordingly, Equation (2) indicates that $\bar{y}_{rss}$ provides a more precise estimate than $\bar{y}_{srs}$ whenever $\mu_{(i)} \neq \mu$.

2.2 Median Ranked Set Sampling (MRSS)

This method was proposed by Muttlak (1998). By selecting the median units, the MRSS reduces the sampling variation and subsequently offers robust, efficient, and precise estimates of population mean. This method is particularly useful in situations when dealing with asymmetric populations or outliers. This novel method contributes to improving the accuracy of population parameter estimation in various fields of study. The sample selection procedure through MRSS involves examining m^2 units of the population. These units are distributed into m independent sets of size m. In each set, the units are ranked using visual judgments or ranks of a closely related concomitant variable, and

- If m is even, select $(m/2)^{th}$ order statistic from first $(m/2)$ sets and $((m+2)/2)^{th}$ order statistic is selected from the remaining sets. The sample selected by this method is represented as $\left\{\left(Y_{i(m/2)}\right)_{i=1}^{m/2}, \left(Y_{i((m+2)/2)}\right)_{i=(m+2)/2}^{m}\right\}$.

- If m is odd, select the middle order statistics i.e., $((m+1)/2)^{th}$ unit is selected from all sets. The selected sample can be represented as $\left(Y_{i((m+1)/2)}\right)_{i=1}^{m}$.

The mean estimator for odd and even set sizes m is given by

$$\bar{y}_{mrss} = \frac{1}{n}\sum_{j=1}^{r}\begin{cases}\sum_{i=1}^{\frac{m}{2}} Y_{i\left(\frac{m}{2}\right)} + \sum_{i=\frac{m+2}{2}}^{m} Y_{i\left(\frac{m+2}{2}\right)} & \text{if } m \text{ is even}\\ \sum_{i=1}^{m} Y_{i\left(\frac{m+1}{2}\right)} & \text{if } m \text{ is odd}\end{cases} \tag{3}$$

The estimator $\bar{y}_{mrss}$ is unbiased, whereas its variance is given by

$$V\left(\bar{y}_{mrss}\right) = \frac{1}{nm}\begin{cases}\sum_{i=1}^{m} \sigma^2_{\left(\frac{m}{2}\right)} + \sum_{i=\frac{m}{2}+1}^{m} \sigma^2_{\left(\frac{m}{2}+1\right)} & \text{if } m \text{ is even}\\ \sum_{i=1}^{m} \sigma^2_{\left(\frac{m+1}{2}\right)} & \text{if } m \text{ is odd}\end{cases} \tag{4}$$

The estimator $\bar{y}_{mrss}$ will be more efficient than $\bar{y}_{rss}$ if the condition $V\left(\bar{y}_{mrss}\right) < V\left(\bar{y}_{rss}\right)$ is satisfied by the underlying population; Muttlak (1998) has obtained an expression for this.

2.3 Extreme Ranked Set Sampling (ERSS)

This method was suggested by Samawi et al. (1996) and it is seen as a valuable alternative in situations where the underlying population exhibits a fat-tailed distribution. By focusing on extreme observations, it effectively accommodates populations with high variability, thereby yielding more robust and reliable estimates of population characteristics. This method is advantageous in fields like finance, where asset returns frequently exhibit fat-tailed distributions, and in environmental studies, where extreme events exert a substantial influence on ecological processes. Sample selection procedure for this method involves examining m^2 units from population, assigning them randomly to m independent sets each of size m. Rank the units within each set and

- If m is even, select the lowest ranked units from the first $\frac{m}{2}$ sets and the highest ranked units from the last $\frac{m}{2}$ sets. The selected sample is expressed as $\left\{\left(Y_{i(1)}\right)_{i=1}^{m/2}, \left(Y_{i(m)}\right)_{i=(m+2)/2}^{m}\right\}$.
- If m is odd, the lowest ranked units are selected from the first $((m-1)/2)$ sets, and the highest ranked units are selected from set $\frac{m+1}{2}$ to set $(m-1)$. The median unit, i.e., $(\frac{m+1}{2})^{th}$ is selected from the last set. The selected sample can be expressed as $\left\{\left(Y_{i(1)}\right)_{i=1}^{n/m}, \left(Y_{i(m)}\right)_{i=(m+1)/2}^{m}, Y_{m((m+1)/2)}\right\}$.

Repetition of this procedure r times yields a final sample of size $n = rm$. The estimator of population mean is given by

$$\bar{y}_{erss} = \frac{1}{n}\sum_{j=1}^{r}\begin{cases}\sum_{i=1}^{\frac{m}{2}} Y_{i(1)} + \sum_{i=\frac{m+2}{2}}^{m} Y_{i(m)} & \text{if } m \text{ is even}\\ \sum_{i=1}^{\frac{m-1}{2}} Y_{i(1)} + \sum_{i=\frac{m+1}{2}}^{m-1} Y_{i(m)} + Y_{\left(\frac{m(m+1)}{2}\right)} & \text{if } m \text{ is odd}\end{cases} \tag{5}$$

The estimator $\bar{y}_{erss}$ is unbiased whereas its variance is given by

$$V\left(\bar{y}_{erss}\right) = \frac{\sigma^2}{n} - \frac{1}{rm^2}\begin{cases}\sum_{i=1}^{\frac{m}{2}} \Delta^2_{(1)} + \sum_{i=\frac{m+2}{2}}^{m} \Delta^2_{(m)} & \text{if } m \text{ is even}\\ \sum_{i=1}^{\frac{m-1}{2}} \Delta^2_{(1)} + \sum_{i=\frac{m+1}{2}}^{m-1} \Delta^2_{(m)} + \Delta^2_{\left(\frac{m+1}{2}\right)} & \text{if } m \text{ is odd}\end{cases} \tag{6}$$

Equation (6) shows that $\bar{y}_{erss}$ is more efficient than $\bar{y}_{srs}$. Furthermore, if the underlying data meets specific conditions, as shown by Samawi et al. (1996), then $\bar{y}_{erss}$ will be more efficient than $\bar{y}_{rss}$ and $\bar{y}_{mrss}$.

2.4 Paired Ranked Set Sampling (PRSS)

Muttlak (1996) proposed this method, arguing that it draws a sample of size m by examining fewer population units than the other RSS methods; thus, it is cost- and time-efficient. PRSS offers a unique advantage by pairing items based on their ranks, which helps reduce variability and enhance the precision of population parameter estimates. This approach is particularly valuable in studies where the relationship between paired observations is important for understanding the underlying population characteristics. The sampling procedure is described below.

- If m is even, identify $\frac{m^2}{2}$ units of the population and assign them randomly to $\frac{m}{2}$ independent sets, each of size m. Rank the units in each set and select the lowest and largest ranked units from the first set. Select the second lowest and the second largest ranked units from the second set and continue selecting paired units in this manner until $(\frac{m}{2})^{th}$ and $(\frac{m}{2})^{th}$ ranked units are selected from the last set. Sample selected through this method can be represented as $\left\{\left(Y_{i(i)}\right)_{i=1}^{m}, \left(Y_{i(m+1-i)}\right)_{i=1}^{m}\right\}$.
- If m is odd, examine $\frac{m(m+1)}{2}$ units from the population, and assign them to $\frac{m+1}{2}$ independent sets, each of size m. Rank the units within sets and select the lowest and the largest units from the first set. Select the second lowest and the second largest units from the second set. Selection of paired units continues until the $(\frac{m+1}{2})^{th}$ ranked unit is selected from the last set. The selected sample through this method can be expressed as $\left\{\left(Y_{i(i)}\right)_{i=1}^{(m+1)/2}, \left(Y_{i(m+1-i)}\right)_{i=1}^{(m-1)/2}\right\}$.

The mean estimator based on this method is given by

$$\bar{y}_{prss} = \frac{1}{n} \sum_{j=1}^{r} \begin{cases} \sum_{i=1}^{\frac{m}{2}} \left(Y_{i(i)} + Y_{i(m+1-i)}\right) & \text{if } m \text{ is even} \\ \sum_{i=1}^{\frac{m+1}{2}} Y_{i(i)} + \sum_{i=1}^{\frac{m-1}{2}} Y_{i(m+1-i)} & \text{if } m \text{ is odd} \end{cases} \tag{7}$$

where r shows number of times a PRSS method is repeated. The estimator $\bar{y}_{prss}$ is unbiased with variance given as

$$V\left(\bar{y}_{prss}\right) = \frac{\sigma^2}{n} - \frac{1}{rm^2} \begin{cases} \sum_{i=1}^{\frac{m}{2}} \left(\Delta^2_{(i)} + \Delta^2_{(m+1-i)}\right) + 2\sum_{i=1}^{\frac{m}{2}} \sigma_{(i,m+1-i)} & \text{if } m \text{ is even} \\ \sum_{i=1}^{\frac{m+1}{2}} \Delta^2_{(i)} + \sum_{i=1}^{\frac{m-1}{2}} \Delta^2_{(m+1-i)} + 2\sum_{i=1}^{\frac{m-1}{2}} \sigma_{(i,m+1-i)} & \text{if } m \text{ is odd.} \end{cases} \tag{8}$$

Here $\sigma_{(i,m+1-i)}$ is the covariance between the i^{th} and $(m+1-i)^{th}$ order statistics. Equation (8) shows that $\bar{y}_{prss}$ is more efficient than $\bar{y}_{srs}$, however, its efficiency comparison with $\bar{y}_{rss}$, $\bar{y}_{mrss}$, and $\bar{y}_{erss}$ can be seen from Muttlak (1996) and references therein.

2.5 RSS with Perfect and Imperfect Ranking

When the ranking is carried out directly on the basis of the study variable itself, it is referred to as perfect ranking. This approach can be implemented through the visual judgement of the surveyor and, despite being simple and cost-effective, is generally assumed to be error-free. Perfect ranking is typically applied when sampling units exhibit a natural order or hierarchy. However, when such ranking is not feasible, Lynne Stokes (1977) introduced an alternative procedure in which the ranking of study units is performed using a closely associated concomitant variable, provided that information on the auxiliary variable is available or easily obtainable. This approach, termed imperfect ranking, has also been described by some researchers as ranking with errors, since larger units may occasionally be placed before smaller ones due to observational limitations. Weak correlation between the study and concomitant variables increases the likelihood of such misplacement, potentially leading to reduced efficiency in the estimates. This study considers both ranking criteria for comparing the reproducibility of population mean estimation for basic RSS methods.

3 NPI reproducibility

NPI relies on Hill's assumption $A_{(n)}$ which is used for prediction when there is no prior information about an underlying distribution (Hill 1968). It is used to predict direct conditional probabilities for one or more future values. To explain the assumption $A_{(n)}$, we consider n real-valued exchangeable random quantities $Y_1, ..., Y_n$. Our aim is to predict future value Y_{n+1} based on n observed values. Let the ordered observed values be denoted as $y_{(1)}, ..., y_{(n)}$, with $y_{(0)} = -\infty$ and $y_{(n+1)} = \infty$, or using boundaries, either known or assumed, to support the random variables such as $y_{(0)} = L$ and $y_{(n+1)} = R$. Thus, for the future observation $Y_{(n+1)}$, the assumption $A_{(n)}$ derived from n observed values is

$$P\left(Y_{n+1} \in \left(y_{(k-1)}, y_{(k)}\right)\right) = \frac{1}{n+1}, \quad \text{for } k = 1, 2, ..., n+1. \tag{9}$$

$A_{(n)}$ assumes that there are no ties, however, ties in data can be managed by breaking them by a tiny amount. This post-data assumption and statistical inference based on it is predictive and nonparametric. It can provide lower and upper bounds for probabilities, known as interval probabilities or imprecise probabilities in probability theory. Augustin and Coolen (2004) referred to this statistical approach as nonparametric predictive inference (NPI). Augustin et al. (2014) provide a comprehensive overview of imprecise probabilities, including both the theoretical foundations and practical implementations. The lower probability $\underline{P}(A)$ in NPI is the maximum lower bound for the probability for A, while the upper probability $\overline{P}(A)$ is the minimum upper bound for the probability for A, where $0 \leqslant \underline{P}(A) \leqslant \overline{P}(A) \leqslant 1$. For the event $Y_{n+1} \in B$, where $B \subset \mathbb{R}$, the NPI lower and upper probabilities, respectively are

$$\underline{P}(Y_{n+1} \in B) = \frac{1}{n+1} \left|\{k : I_k \subseteq B\}\right| \tag{10}$$

and

$$\overline{P}(Y_{n+1} \in B) = \frac{1}{n+1} \left|\{k : I_k \cap B \neq \emptyset\}\right|. \tag{11}$$

Coolen and Alqifari (2018) used NPI for multiple future observations by consecutively applying Hill's assumption $A_{(n)}, A_{(n+1)}, ..., A_{(n+m-1)}$, jointly denoted by $A_{(\cdot)}$. Given n data observations, an ordering O_i represents a potential ordering of the $m > 1$ future observations. For n data observations and m future observations, there are $\binom{n+m}{n}$ possible orderings, and all orderings have equal probability under $A_{(\cdot)}$. Coolen and Bin Himd (2014) developed bootstrap method based on NPI, called the Nonparametric Predictive Inference Bootstrap (NPI-B) which predicts future values from the intervals between data points, adding it to the data set before drawing another value. We employ this bootstrapping technique to investigate the reproducibility of mean estimates for basic RSS methods, thus, we describe the procedure. For n real-valued data in one dimension on an interval $[L, R]$, the NPI-B algorithm proceeds as:

1. Form $n + 1$ intervals using the n ordered observations.
2. Select an interval randomly with probability $\frac{1}{n+1}$, and draw a future value from it using an assumed probability distribution over the interval.
3. Data is updated by adding the drawn value, increasing the sample size to $n + 1$ observations.
4. For updated data, Steps 1–3 are repeated to generate another future value, which is then added to the data.
5. In total, iterate this procedure n times to obtain n future values that form an NPI-B sample of size n.

The probability distributions in Step 2 of this algorithm are different for finite and infinite intervals. For finite intervals, a future unit is uniformly selected from any of the selected intervals. However, in the case of infinite intervals, the future value is sampled from tail intervals differently. The first and last intervals, that is, $I_1 = \left(-\infty, y_{(1)}\right)$ and $I_{n+1} = \left(y_{(n)}, +\infty\right)$ are assumed to be tails of the Normal distribution with parameters $\mu_y = \left(y_{(1)} + y_{(n)}\right)/2$ and $\sigma_y = \left(y_{(n)} - \mu_y\right)/\left(\Phi^{-1}\left(n/(n+1)\right)\right)$, where Φ^{-1} is cumulative distribution function. If the selected interval is I_1 then a value is sampled from a normal distribution with parameters $\left(\mu_y, \sigma_y^2\right)$, and accepted if it is less than $y_{(1)}$. Similarly, accept the sampled value for I_{n+1} if it is greater than $y_{(n)}$. If the random quantities are known to be positive, then exponential distribution can be used in the interval $\left(y_{(n)}, \infty\right)$ with rate parameter $\lambda = \log(n+1)/y_{(n)}$. These distributions are fitted such that the intervals have probability $\frac{1}{n+1}$.

Recently, the reproducibility of estimates using NPI-B bootstrapping was studied by Alghamdi (2022). In this study, the sample estimate derived from a sample $y_1, ..., y_n$ is denoted by $\hat{\theta}$. The NPI-B algorithm is used to create a future sample $b_1, ..., b_n$, let $\hat{\theta}_B$ denote the estimate based on the future sample. On reproducing future samples many time (say n_B), and estimating $\hat{\theta}$ enables us to assess the ε-reproducibility of estimates as

$$\widehat{RP}\left(\varepsilon\right) = \frac{1}{n_B}\sum_{i=1}^{n_B} \mathbf{1}\left\{\left|\hat{\theta} - \hat{\theta}_{B_i}\right| \leqslant \varepsilon\right\}, \tag{12}$$

where $\varepsilon \geqslant 0$. The function $RP(\varepsilon)$ quantifies the likelihood of deviation between the two estimates within a specific margin of ε. The term $\mathbf{1}\left\{\mathbf{A}\right\}$ is an indicator function such that $\mathbf{1}\left\{\mathbf{A}\right\}$=1 if event A is true and zero otherwise.

4 NPI reproducibility for RSS estimates

In this section, we describe the mathematical approach designed to evaluate and compare the reproducibility of basic RSS methods within the framework of NPI. Following Alghamdi (2022), $RP\left(\varepsilon\right)$ is the probability that the mean estimated from reproduced samples will fall within the ε deviation of the mean estimated based on the original sample, assuming that the sampling process is repeated under identical conditions.
Consider a set of sampling units obtained using any of the basic RSS methods, where n original observations $y_i; i = 1, \ldots, n$ are assumed to be independently and identically distributed, and their ordered statistics are denoted by $y_{(1)} < ... < y_{(n)}$. We determine the lower and upper bound $\left[y_{(0)}, y_{(n+1)}\right]$ and make $n+1$ intervals on the real line. Using the steps discussed in Section 3, a set of n NPI-B future values is produced and replaced with the original set. Repeat this procedure for all sets and obtain the setup of NPI-B-RSS for reproducing the original sample. We produce a large number of NPI-B-RSS samples (say M), and compute the mean estimate of the original RSS sample and its corresponding NPI-B-RSS samples. Let the mean estimate from the original sample be denoted by $\hat{\theta}_R$, and the mean estimates from the reproduced samples using the NPI-B approach are denoted by $\hat{\theta}r$. The following formulas are used to compute the Absolute Average Deviation (AAD) and Mean Square Deviation (MSD) between these estimates.

$$AAD = \frac{1}{M} \sum_{i=1}^{M} \left| \hat{\theta}_R - \hat{\theta}_{r_i} \right|, \tag{13}$$

and

$$MSD = \frac{1}{M} \sum_{i=1}^{M} \left(\hat{\theta}_R - \hat{\theta}_{r_i} \right)^2. \tag{14}$$

Let ε be any real-valued positive quantity, then $RP_1(\varepsilon)$ is the reproducibility probability that AAD is equal to or less than ε. Similarly, $RP_2(\varepsilon)$ is the reproducibility probability that MSD is equal to or less than ε. These probabilities are mathematically computed as

$$RP_1(\varepsilon) = \Pr(AAD \leqslant \varepsilon) \tag{15}$$

and

$$RP_2(\varepsilon) = \Pr(MSD \leqslant \varepsilon). \tag{16}$$

It is straightforward to compare the reproducibility of different RSS methods when $RP_1(\varepsilon)$ and $RP_2(\varepsilon)$ are plotted across different ε values. We define this as ε-reproducibility with regard to AAD and MSD, where $\varepsilon \in [0, +\infty]$.

4.1 Algorithm for NPI reproducibility of RSS methods

Algorithm 1 outlines the step-by-step procedure for computing the reproducibility probabilities for the different RSS methods. We employ NPI-B to reproduce the original samples, which sample future observations from the entire range of possible observations and extend beyond the bounds of the original samples. In Algorithm 1, the provided inputs are the values of the study variable and the concomitant variable. Original setups for basic RSS methods are produced, and samples of size n are drawn using their respective procedures. Each set of RSS methods is then reproduced using the NPI-B method, and the corresponding NPI-B-RSS sets are created and estimate their corresponding mean. The number of runs for bootstrapped samples is M. The estimated mean of the NPI-B-RSS samples is then used for calculating the AAD and MSD. In order to compute reproducibility probabilities $RP_1(\varepsilon)$ and $RP_2(\varepsilon)$, this entire process is repeated D times, where D denotes a population that has changed while maintaining the same parameters.

5 Simulations

In this section, we compare the ε-reproducibility of basic RSS methods through simulations. The above algorithm 1 is used for computation of ε-reproducibility using R software. The process begins with generating population values for the concomitant variable, denoted by X_i. The standardized normal population Z_i is also generated, which aids in establishing the desired correlation (ρ_{yx}) between the study variable Y and the concomitant variable X. Using the relationship $Y_i = \rho_{yx} X_i + Z_i \sqrt{1 - \rho_{yx}^2}$, we produced values for the study variable. The ε-reproducibility is evaluated using two hypothetical populations: (1) a normal distribution with parameters $\mu = 100$ and

Algorithm 1 Generating NPI-B-RSS samples and computing ε-reproducibility

Require: Original samples of size n

Ensure: AAD, MSD, $RP(\varepsilon_1)$ and $RP(\varepsilon_2)$

1: Draw original samples using procedures of RSS, MRSS, ERSS, and PRSS
2: Calculate original means of these samples
3: **for** each method of RSS **do**
4:     Apply NPI-B method to generate bootstrapped sets
5:     Replace bootstrapped set with original sets
6:     Draw NPI-B-RSS samples from bootstrapped data
7:     Compute sample means based on NPI-B-RSS samples
8: **end for**
9: **for** $i \leftarrow 1$ **to** M **do**
10:     Repeat Steps 3–8 and compute AAD and MSD using Equations (13) and (14)
11: **end for**
12: **for** $j \leftarrow 1$ **to** D **do**
13:     Repeat Steps 1–12 and compute $RP_1(\varepsilon)$ and $RP_2(\varepsilon)$ using Equations (15) and (16), respectively
14: **end for**

$\sigma^2 = 5$, and (2) an exponential distribution with rate parameter $\lambda = 0.3$. Both perfect and imperfect ranking criteria are considered, corresponding to weak ($\rho_{yx} = 0.50$) and strong ($\rho_{yx} = 0.90$) correlations between the study and concomitant variables. The ε-reproducibility is further examined for different sample sizes, with $n = 6$ and $n = 10$ considered. Using the above algorithm, we use $M = 1000$ to compute AAD and MSD given in (13) and (14), respectively. Finally, this entire process is repeated $D = 1000$ times to compute the respective reproducibility probabilities $RP_1(\varepsilon)$ and $RP_2(\varepsilon)$. We use different colors on plots to show the $\varepsilon-$reproducibility of different RSS estimates for different methods. The magnitude of ε is plotted along the x-axis, while the y-axis represents the likelihood of its occurrence. The resulting outcomes are illustrated below.

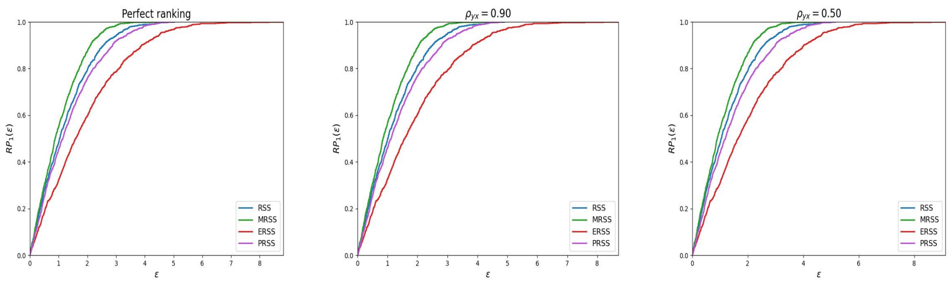

Fig. 1 $RP_1(\varepsilon)$ of estimates under different ranking criteria for normal distribution when $n = 6$

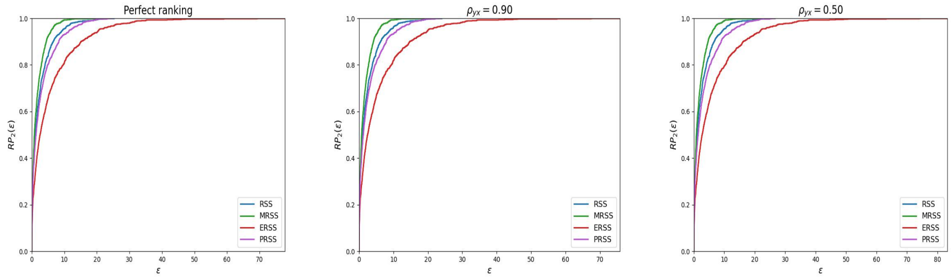

Fig. 2 $RP_2(\varepsilon)$ of estimates under different ranking criteria for normal distribution when $n = 6$

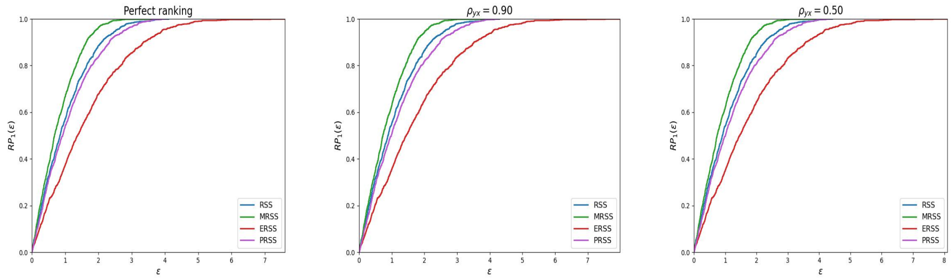

Fig. 3 $RP_1(\varepsilon)$ of estimates under different ranking criteria for normal distribution when $n = 10$

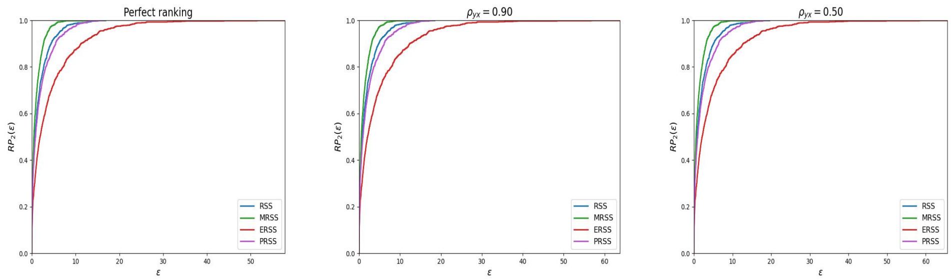

Fig. 4 $RP_2(\varepsilon)$ of estimates under different ranking criteria for normal distribution when $n = 10$

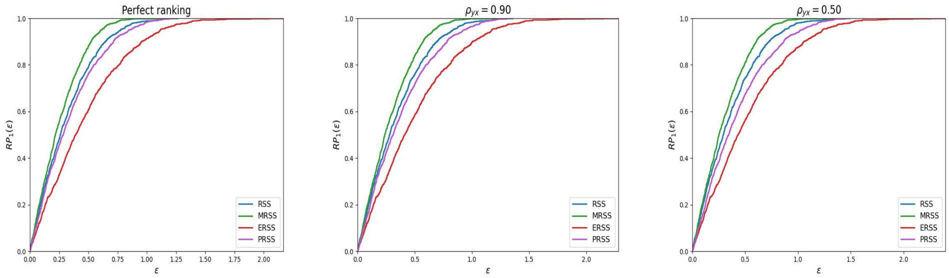

Fig. 5 $RP_1(\varepsilon)$ of estimates under different ranking criteria for exponential distribution when $n = 6$

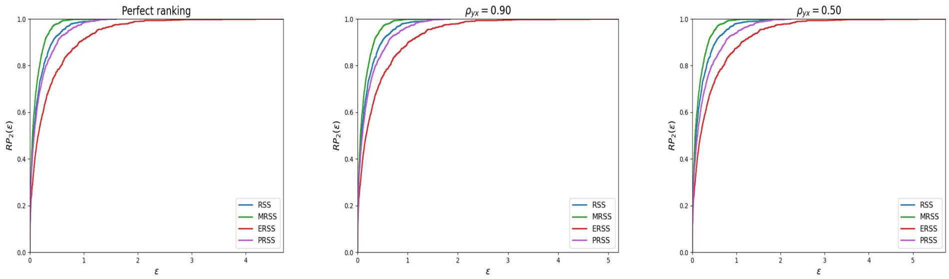

Fig. 6 $RP_2(\varepsilon)$ of estimates under different ranking criteria for exponential distribution when $n = 6$

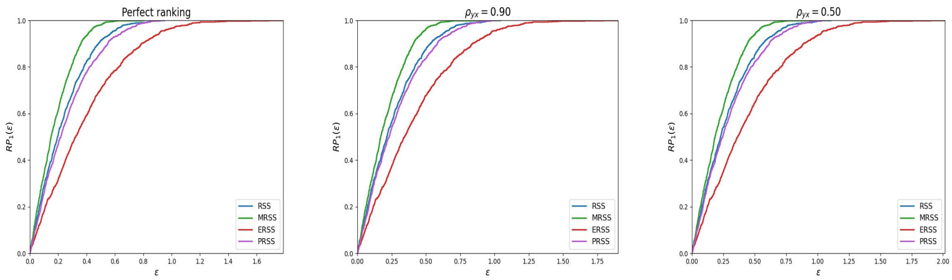

Fig. 7 $RP_1(\varepsilon)$ of estimates under different ranking criteria for exponential distribution when $n = 10$

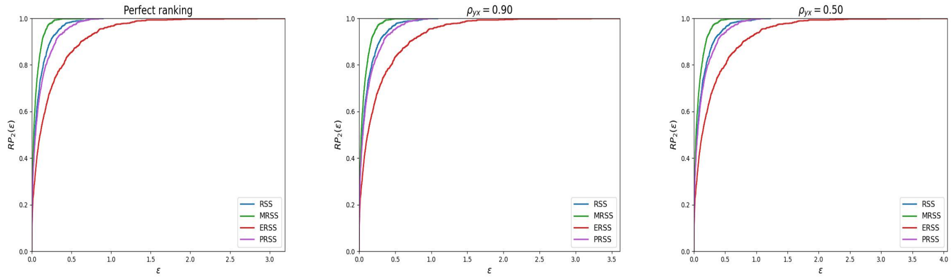

Fig. 8 $RP_2(\varepsilon)$ of estimates under different ranking criteria for exponential distribution when $n = 10$

Simulation results show the $\varepsilon-$reproducibility of mean estimates based on MRSS estimates increases earlier than $\varepsilon-$reproducibility of other methods, indicating that it is the highest reproducible method. Similarly, the $\varepsilon-$reproducibility of mean estimates based on ERSS estimates shows the fluctuation as compared to other methods, indicating that it is least reproducible method. On other hand, the RSS and PRSS methods display similar and intermediate levels of $\varepsilon-$reproducibility between the MRSS and ERSS methods. The plots also shows that $\varepsilon-$reproducibility is higher for larger samples as compared to small samples. Furthermore, $\varepsilon-$reproducibility for mean estimates of the MRSS method is higher in the case of the exponential distribution as compared to the normal distribution. The $\varepsilon-$reproducibility of MRSS estimates is also higher than other methods for both ranking criteria. In the case of imperfect ranking, $\varepsilon-$reproducibility for MRSS is higher when the correlation is high as compared to weak correlation. Generally, the $\varepsilon-$reproducibility of the MRSS method is higher than other RSS methods, while ERSS exhibits the lowest $\varepsilon-$reproducibility in all cases.

6 Application to real-life data

For real-life applications, we consider data on agriculture collected by Singh and Mangat (2013). The study variable is considered to be the total irrigated area in a village (Punjab, India), for which mean is estimated based on basic RSS methods. In the case of imperfect ranking, the concomitant variable is taken as the total number of tubewells in village. The simulation procedure is same as described in Algorithm 1, the graphs below show the results.

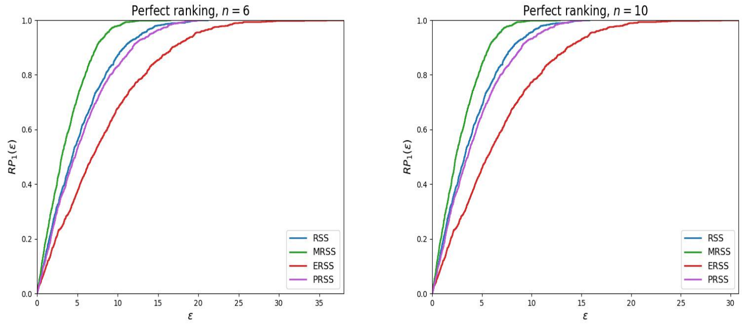

Fig. 9 $RP_1(\varepsilon)$ of estimates under perfect ranking

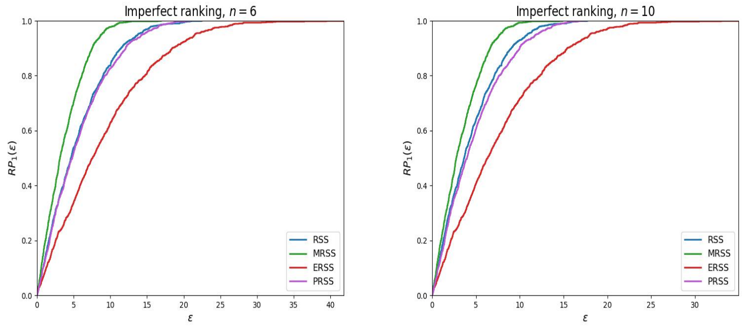

Fig. 10 $RP_1(\varepsilon)$ of estimates under imperfect ranking

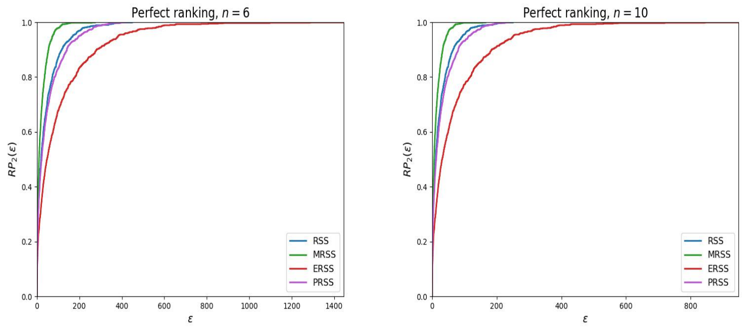

Fig. 11 $RP_2(\varepsilon)$ of estimates under perfect ranking

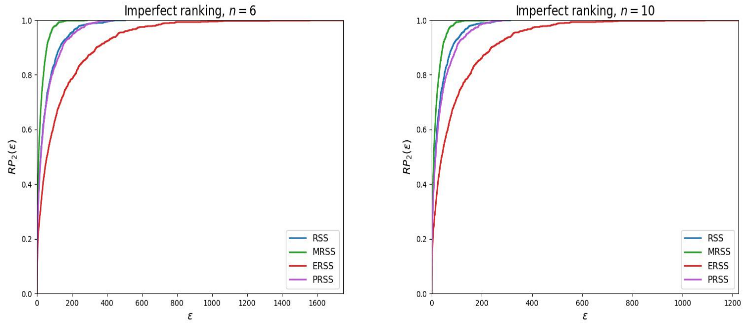

Fig. 12 $RP_2(\varepsilon)$ of estimates under imperfect ranking

7 Conclusions and future directions

This study investigated the reproducibility of four basic Ranked Set Sampling (RSS) methods: Classical RSS (RSS), Median RSS (MRSS), Extreme RSS (ERSS), and Paired RSS (PRSS). By employing Nonparametric Predictive Inference (NPI) bootstrapping, we explored the reproducibility of these methods for varying sample sizes and correlation coefficients between the study variable and the concomitant variable. Our analysis included both perfect and imperfect rankings, providing a comprehensive evaluation of the methods' performance. The findings of this study showed that MRSS exhibited the highest level of reproducibility among the four RSS methods considered. This highlights the importance of selecting the MRSS method for sample selection. In contrast, ERSS exhibited the lowest reproducibility, highlighting the need for caution when applying this method, particularly in situations where extreme observations may introduce variability.

Our work establishes the foundation for future studies in several ways. Firstly, researchers can build upon our findings to further investigate the factors influencing the reproducibility of RSS methods, such as sample size and ranking criteria. Additionally, exploring alternative approaches to improve the reproducibility of ERSS could lead to advancements in sampling methodology. Additionally, our study can serve as a foundation for assessing the reproducibility of different mean estimators that involve the auxiliary variable while designing stages. Furthermore, our study highlights the importance of considering reproducibility in survey methods, paving the way for the development of a more robust and reliable measure for comparing various survey methods and estimators of population parameters.

Acknowledgements. The author wishes to express sincere appreciation for the remarkable contributions of Professor Sat Gupta in the field of statistics and survey sampling. His pioneering research and dedication to advancing statistical methodology have inspired countless researchers, including the present study. This manuscript is submitted to the Special Issue in Honour of Professor Sat Gupta as a humble tribute to his lasting impact on the discipline of statistics.

Declarations

- **Funding**
  The authors did not receive any specific funding for this work.
- **Conflict of interest / Competing interests**
  The authors declare that they have no conflicts of interest or competing interests.
- **Ethics approval and consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Data availability**
  The data used in this study were hypothetically generated using R software and are available upon reasonable request.
- **Materials availability**
  Not applicable.
- **Code availability**
  The R code used for hypothetical data generation and analysis is available upon reasonable request.
- **Author contribution**
  All authors contributed equally to the conception, design, analysis, and manuscript preparation of this study.

References

Augustin, T., Coolen, F.P.A.: Nonparametric predictive inference and interval probability. Journal of Statistical Planning and Inference **124**(2), 251–272 (2004)

Augustin, T., Coolen, F.P.A., De Cooman, G., Troffaes, M.C.M.: Introduction to Imprecise Probabilities. John Wiley & Sons, Chichester, UK (2014)

Abu-Dayyeh, W.A., Ahmed, M.S., Ahmed, R., Muttlak, H.A.: Some estimators of a finite population mean using auxiliary information. Applied Mathematics and Computation **139**(2-3), 287–298 (2003)

Alghamdi, F.M.: Reproducibility of statistical inference based on randomised response data. PhD thesis, Durham University (2022)

Al-Saleh, M.F., Al-Kadiri, M.A.: Double-ranked set sampling. Statistics & Probability Letters **48**(2), 205–212 (2000)

Begley, C., Ellis, L.: Drug development: Raise standards for preclinical cancer research. Nature. **483**, 531–533 (2012)

Bhushan, S., Kumar, A., Hussam, E., Mustafa, M.S., Zakarya, M., Alharbi, W.R.: On stratified ranked set sampling for the quest of an optimal class of estimators. Alexandria Engineering Journal **86**, 79–97 (2024)

Coolen, F.P.A., Alqifari, H.N.: Nonparametric predictive inference for reproducibility of two basic tests based on order statistics. REVSTAT: Statistical Journal. **16**(2), 167–185 (2018)

Coolen, F.P.A., Bin Himd, S.: Nonparametric predictive inference for reproducibility of basic nonparametric tests. Journal of Statistical Theory and Practice **8**(4), 591–618 (2014) https://doi.org/10.1080/15598608.2013.819792 http://dx.doi.org/10.1080/15598608.2013.819792

Coolen, F.P., Himd, S.B.: Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample kolmogorov–smirnov test. Journal of Statistical Theory and Practice **14**, 1–13 (2020)

De Martini, D.: Reproducibility probability estimation for testing statistical hypotheses. Statistics & Probability Letters **78**(9), 1056–1061 (2008)

Goodman, S.N.: A comment on replication, p-values and evidence. Statistics in Medicine **11**(7), 875–879 (1992)

Hill, B.M.: Posterior distribution of percentiles: Bayes' theorem for sampling from a population. Journal of the American Statistical Association **63**(322), 677–691 (1968)

Killeen, P.R.: An alternative to null-hypothesis significance tests. Psychological Science **16**(5), 345–353 (2005)

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

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

Miller, J.: What is the probability of replicating a statistically significant effect? Psychonomic Bulletin & Review **16**, 617–640 (2009)

Muttlak, H.A.: Pair rank set sampling. Biometrical Journal **38**(7), 879–885 (1996)

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

Posavac, E.J.: Using p values to estimate the probability of a statistically significant replication. Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences **1**(2), 101–112 (2002)

Rehman, S.A., Coolen-Maturi, T., Coolen, F.P.A., Shabbir, J.: Reproducibility of mean estimators under ranked set sampling. Franklin Open (2024) https://doi.org/10.1016/j.franklinopen.2024.100690

Samawi, H.M., Ahmed, M.S., Abu-Dayyeh, W.: Estimating the population mean using extreme ranked set sampling. Biometrical Journal **38**(5), 577–586 (1996)

Shao, J., Chow, S.-C.: Reproducibility probability in clinical trials. Statistics in Medicine **21**(12), 1727–1742 (2002)

Simkus, A., Coolen, F.P.A., Coolen-Maturi, T., Karp, N.A., Bendtsen, C.: Statistical reproducibility for pairwise t-tests in pharmaceutical research. Statistical Methods in Medical Research **31**(4), 673–688 (2022)

Senn, S.: Comment on 'a comment on replication, p-values and evidence', by sn goodman (letter to the editor). Statistics in Medicine **21**, 2437–2444 (2002)

Singh, R., Mangat, N.S.: Elements of Survey Sampling. Springer, Dordrecht (2013)

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 **20**(1), 1–31 (1968)