Reproducibility of estimates based on randomised response methods

Fatimah M. Alghamdi[1], Frank P. A. Coolen[2] and Tahani Coolen-Maturi[2*]

[1]Department of Mathematical Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, 84428, Saudi Arabia.
[2]Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, United Kingdom.

*Corresponding author. E-mail: tahani.maturi@durham.ac.uk;
Contributing authors: fmalghamdi@pnu.edu.sa;
frank.coolen@durham.ac.uk;

Abstract

A key aspect of statistical inference is estimation of population characteristic. This paper investigates the reproducibility of estimates of population characteristics. It focuses on estimates based on data collected by survey-based randomised response methods (RRT) to obtain the truth in cases when the characteristic is sensitive. This work introduces a new approach called ϵ-reproducibility of estimates based on data collected from quantitative RRT methods. This approach defines the $\epsilon-$reproducibility probability as the probability that, in the event that an experiment is repeated under similar circumstances, the estimate based on the data from the repeated experiment will not differ more ϵ from the estimated based on the original data. To address prediction issues, the quantification approach makes use of Nonparametric Predictive Inference (NPI). The findings demonstrate that lower reported response variability for RRT approaches increases reproducibility of estimates using bootstrap and representative sample while maintaining an equivalent degree of privacy for survey respondents. A number of RRT methods are compared, including the Greenberg method, the Eichhorn and Hayre method, and the optional multiplicative method.

Keywords: Reproducibility Probability, Nonparametric Predictive Inference, Bootstrap Method, Representative Sample, Randomised Response Data,

Greenberg Method, Eichhorn and Hayre Method, Optional Multiplication Method.

1 Introduction

Statistical inference is crucial for estimating population characteristics and drawing accurate conclusions about experiments. Goodman [1] emphasized the importance of statistical reproducibility in research, arguing that p-values are often misunderstood and inaccurately used. Senn [2] agreed with Goodman but disagreed, stating that p-values overstate the strength of evidence against the null hypothesis. Both argued for a more detailed approach that includes effect sizes, confidence intervals, and other measures.

Recent studies emphasize reproducibility of experiment conclusions, particularly when methods, circumstances, researchers, or sample size change. This literature review explores various experiments on reproducibility.

Additionally, Billheimer [7] highlights the significance of predictive inference and scientific reproducibility. According to his viewpoint, parametric modelling is a helpful approximation of the prior distribution of either parameters or future observations; the distribution of future observations is the only thing that is affected by parameter selections. Furthermore, conclusions or decisions based on predictions must to be evaluated in light of the overall research problem. In addtion, he highlights the importance of the predictive inference approach, which motivates statisticians to use observable quantities to define interesting findings and predict the probability of them in future investigations.

Alghamdi et al. [8, 9] studied the statistical reproducibility of hypothesis test results using RRT techniques. They used nonparametric predictive inference to quantify reproducibility, which is a prediction issue. The study found that reproducibility of RRT data with an equivalent level of privacy is higher when there is less variation in reported answers of RRT techniques.

This paper investigates the reproducibility of estimates, highlighting that for real-valued random quantities, an estimate of a parameter or population characteristic cannot be precisely reproduced. It introduces a new concept called ϵ-reproducibility, which is the probability that if an experiment is repeated under the same conditions, the future sample estimate will not differ more than ϵ from the original sample estimate. The objective of this paper is to introduce NPI for ϵ-reproducibility of estimates and compare ϵ-reproducibility of mean

when data are collected using RRT methods.

This paper is organised as follows. Section 2 introduces the RRT methods for the study in this paper. Nonparametric Predictive Inference (NPI) is introduced in Section 3. Section 4 introduces NPI for reproducibility. Then Section 5 shows reproducibility of estimates and uses both NPI-Bootstrap and a representative sample. Section 6 introduces numerical examples of $\epsilon-$Reproducibility of estimates based on RRT methods using NPI-Bootstrap and the representative sample. Section 7 presents a discussion of related topics for further research.

2 Randomised Response Techniques (RRT)

Warner [10] introduced randomised response techniques (RRT) to collect reliable data on sensitive variables. These techniques use a device like a coin or spinner to hide responses from the interviewer. However, Warner method cannot collect sensitive quantitative variables. Therefore, Greenberg et al. [11] suggested an unrelated question RRT technique to determine the number of people who have a sensitive characteristic in a population. This paper uses some quantitative RRT methods for a survey in which sensitive questions are answered using real numbers such as the Greenberg method (GM), the optional multiplicative method (MM) and where the Eichorn and Hayre method (EH) and their efficiency and privacy degree. Various RRT techniques have been suggested [11–13], For more information, those quantitative RRT are introduced as follows.

The Greenberg technique (GM) [14] uses the Warner technique, replacing nonsensitive questions with unrelated ones and using quantitative responses. Both answers of these question are real-valued quantities. Let the probability that the sensitive question is asked be γ, where $0 \leq \gamma \leq 1$, and let the answer be the random quantity X_i with expected value $E(X_i) = \mu_x$ and variance σ_x^2. The probability that the unrelated question is asked be $1 - \gamma$, The answer to this question is the random quantity Y_i with expected value $E(Y_i) = \mu_y$ and variance σ_y^2. Both μ_y and σ_y^2 are assumed to be known.

Assume that random quantity Z_i denotes response of the i^{th} respondent $(i = 1, 2, ..., n)$, so

$$Z_i = \begin{cases} X_i & \text{with probability} \quad \gamma \\ Y_i & \text{with probability} \quad 1 - \gamma \end{cases} \tag{1}$$

The expected value of Z_i is

$$E(Z_i) = \gamma E(X_i) + (1-\gamma)E(Y_i) = \gamma\mu_x + (1-\gamma)\mu_y \tag{2}$$

With $\bar{Z}$ denoting to the sample mean is μ_x by

$$\hat{\mu}_x = \frac{\bar{Z} - (1-\gamma)\mu_y}{\gamma} \tag{3}$$

The variance of Z_i [15] is

$$\text{Var}(Z_i) = \frac{1}{\gamma^2}\left[\sigma_y^2 + \gamma(\sigma_x^2 - \sigma_y^2) + \gamma(1-\gamma)(\mu_x - \mu_y)^2\right] \tag{4}$$

Another quantitative method is Eichhorn and Hayre method [12]. Assume that there is random quantity X_i as the true response with expected value $E(X_i) = \mu_x$ and variance $\sigma_x^2 = V(X_i)$, where $i = 1, ..., n$, and μ_x and σ_x^2 are unknown. The randomisation device provides a numerical value S_i that follows a predetermined probability distribution with a known mean $E(S_i) = \theta$ and variance r^2, where the random quantities S_i and X_i are assumed to be independent variables. In this method, respondents choose a number and report the product of the real responses X_i and S_i, as follows:

$$Z_i = X_i S_i \tag{5}$$

Because S_i and X_i are assumed to be independent, the expected value of Z_i is:

$$E(Z_i) = E(X_i S_i) = E(X_i)E(S_i) = \mu_x\theta \tag{6}$$

The variance of Z_i is

$$\text{Var}(Z_i) = \left[\sigma_x^2 + \frac{r^2}{\theta^2}(\sigma_x^2 + \mu_x^2)\right] \tag{7}$$

Gupta et. al [16] developed the optional multiplicative method (MM), a quantitative method where an unknown proportion of respondents scramble their responses if the question is sensitive, while others give their true responses. Assume that there is random quantity X_i as a sensitive characteristic for individual i with an unknown mean μ_x, and random quantity S_i as a scrambling variable with a known mean $E(S_i)$, where X_i and S_i are independent, and S_i can be produced from any distribution. Assume that there is a random quantity Z_i denoting the response of a person i where $i = 1, ..., n$. Giving the randomisation

device which gives a random quantity S_i that follows a known probability distribution with the known mean $E(S_i) = 1$ and known variance γ^2, where S_i and X_i are independent, and both random variables with positive values, therefore, $E(S_i)$ is assumed to be 1.

The respondent answers $Z_i = X_i$ if the question is not sensitive; if the question is sensitive, the answer is scrambling $Z_i = S_i X_i$. All respondents have an equal probability of scrambling ψ, which is a known quantity. Therefore, the reported responses Z_i are

$$Z_i = \begin{cases} X_i & \text{with probability} \quad \psi \\ X_i S_i & \text{with probability} \quad 1-\psi \end{cases} \tag{8}$$

The expected value of Z_i is

$$\mu_z = E(Z_i) = \psi E(X_i) + (1-\psi)E(X_i)E(S_i) = \mu_x$$

where S_i and X_i are independent. So, it can be noted that the estimator of scrambled responses $\hat{\mu}_z$ equals to the estimator of the true responses $\hat{\mu_x}$. The variance of Z_i [16] is

$$\mathrm{Var}(Z_i) = \sigma_x^2 + \psi\gamma^2(\sigma_x^2 + \mu_x^2)$$

Importantly, RRT techniques must consider efficiency and privacy as key considerations. Efficiency refers to the accuracy of RRT in estimating sensitive characteristics in a population. Privacy is crucial for ensuring precise estimates of sensitive characteristics while protecting respondents' privacy. Various privacy measures have been proposed for quantitative RRT techniques, aiming to increase respondents' accuracy and reduce bias. When efficiency decreases, privacy degree increases [11, 17].

Degree of Privacy of quantitative RRT is defined as the expectation of the square of the difference between the reported response Z_i and true response X_i of the sensitive question [18]. The privacy of the quantitative RRT method is

$$\Delta = E(Z_i - X_i)^2 \tag{9}$$

The RRT approach provides better privacy protection with a larger Δ value, resulting in reduced variance of reported answers Z_i and increased efficiency. The true responses almost match respondents' responses Z_i when respondents believe in the RRT technique and fully protect their privacy. Therefore, privacy degree of the Greenberg method Δ_{GM} [18] is

$$\Delta_{GM} = (1-\gamma)E(Y_i - X_i)^2 = (1-\gamma)[\sigma_y^2 + \sigma_x^2 + (\mu_x - \mu_y)^2] \tag{10}$$

The privacy degree of the EH [12] and the MM [16] method are

$$\Delta_{EH} = (\frac{r}{\theta})^2(\sigma^2 + \mu^2) \tag{11}$$

$$\Delta_{MM} = \psi(\gamma)^2(\sigma_x^2 + \mu_x^2) \tag{12}$$

Overall, the efficiency of a quantitative RRT approach is higher when it provides less privacy. So, we offer methods based on reproducibility probability, efficiency and degree of privacy to investigate the relationship between them.

3 Nonparametric Predictive Inference (NPI)

Nonparametric predictive inference (NPI) [19] provides direct conditional probabilities for a future observable random quantity based on observed values of related random quantities [4, 20]. It is based on Hill's assumption $A_{(n)}$. Assume that there are $n+1$ exchangeable real-valued random quantities, denoted as $Y_1, ..., Y_n, Y_{n+1}$. Set $y_0 = -\infty$ and $y_{n+1} = \infty$. Assume that the ordered observed values of the random quantities $Y_1, ..., Y_n$ are represented by $y_1 < y_2 < ... < y_n$. With $i = 1, ..., n+1$, the real-line is divided into $n+1$ intervals $I_i = (y_{i-1}, y_i)$ by the n observations. For a single future observation Y_{n+1}, the assumption $A_{(n)}$ [19] is

$$P(Y_{n+1} \in I_i) = \frac{1}{n+1} \quad \text{for} \quad i = 1, ..., n+1 \tag{13}$$

$A_{(n)}$ is a post-data assumption related to exchangeability [21]. Given the probabilities in Equation (13), NPI uses De Finetti's Fundamental Theorem of Probability [21] to find optimal bounds for the probability of an event of interest involving Y_{n+1} [4]. In the theory of imprecise probability [22] and interval probability [23], this theory has strong consistency properties and yields reliable predictive results [4].

Based on Equation (13) lower and upper probabilities for an event $Y_{n+1} \in \mathfrak{B}$, for any $\mathfrak{B} \subset \mathbb{R}$, are [4, 20]:

$$\underline{P}(Y_{n+1} \in \mathfrak{B}) = \sum_{i=1}^{n+1} \mathbf{1}\{I_i \subseteq \mathfrak{B}\} P(Y_{n+1} \in I_i) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{1}\{I_i \subseteq \mathfrak{B}\} \tag{14}$$

$$\overline{P}(Y_{n+1} \in \mathfrak{B}) = \sum_{i=1}^{n+1} \mathbf{1}\{I_i \cap \mathfrak{B} \neq \emptyset\} P(Y_{n+1} \in I_i) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{1}\{I_i \cap \mathfrak{B} \neq \emptyset\} \tag{15}$$

where $\mathbf{1}\{E\}$ is the indicator function which is equal to 1 if event E is true and 0 otherwise. NPI has been introduced for a number of applications,

including finance [24, 25], trading [26], and statistical process control [27, 28].

3.1 NPI for multiple future observations

NPI has been also developed for multiple future real-valued observations m, where we are interested in $m > 1$. Assume that the ordered observed values of the random quantities $Y_1, ..., Y_n$ are denoted by $y_1 < y_2 < ... < y_n$, with the lower bound denoted by y_0 and the upper bound by y_{n+1}. It should be noted that y_{n+m} is not an observed value for Y_{n+m}. The n observations split the real-line into $n + 1$ intervals $I_i = (y_{i-1}, y_i)$, where $i = 1, ..., n + 1$.

According to Hill's assumption $A_{(n)}$, all the orderings O_j of the future observations m among the original observations n have equal probability based on Equation (13), where $j = 1, 2, ..., \binom{n+m}{n}$. For the future observations Y_{n+i}, each ordering can be derived from $S_i^j = \#\{Y_{n+i} \in I_j, i = 1, ..., n, j = 1, 2, ..., \binom{n+m}{n}\}$. We link the data and future observations via Hill's assumption $A_{(n)}$ [29], or more precisely, via consecutive application of $A_{(n)}, A_{(n+1)}, ..., A_{((n+m)-1)}$ which can be considered as a post-data version of a finite exchangeability assumption for $n + m$ random quantities $Y_{n+1}, ..., Y_{n+m}$. A practical interpretation of the $A_{(n)}$ assumptions is that all possible orderings of data observations n and future observations m are equally likely.

Based on the $A_{(.)}$ assumptions, assume that the random quantity S_i^j is the number of m future observations in $I_j = (y_{j-1}, y_j)$ given a specific ordering O_i, the probability of each ordering [30] is:

$$P\left(\bigcap_{i=1}^{n+1}\{S_i^j = s_i^j\}\right) = P(O_j) = \binom{n+m}{n}^{-1} \tag{16}$$

where the s_i^j are non-negative integers with $\sum_{i=1}^{n+1} s_i^j = m$, and $A_{(.)}$ assumptions indicate that one does not know whether particular values of close known observations to increase or decrease the probability that a future observation will fall between them. We can count the number of such orderings for which an event occurs for any event involving the m future observations, according to Equation (16).

When there are several orderings among large data observations, the Sampling of Orderings Method (SOM) is used to generate large ordering of future observation data s_i^j. The method is based on NPI, ensuring each order has the same probability of selection and is independent of other selections. If the sample size or the number of orderings sampled is large, the total number of orderings becomes

large enough to ignore potential differences between sampling with or without replacement of these orderings [31].

In this method, we need to choose such vectors at random of the orderings $r_1, ..., r_n$ with $r_1 \geq 1$ and $r_{l-1} < r_l$ where $r_n \leq 2n$ for all $l = 2, ..., n$. Take the rank of the l-th ordered data observation among the $2n$ combined data and future observations to be r_l. Then, the future observation data S_l^j is specified as $S_l^j = (r_l - r_{l-1}) - 1$ where $l = 1, ..., n+1$, such that $r_0 = 0$ and $r_{n+1} = 2n + 1$. This method is used in this paper in reproducibility for an estimate using the representative sample to generate unlimited orderings for the future observations among the original data as explained in Section 5.2.

3.2 NPI-bootstrap

Statistical inference depends on quantifying the variability of sample estimates, which can be challenging in complicated situations. Efron [32] developed a bootstrap method that makes fewer assumptions but requires more computations, and it is easy use and accurate approximations. Nonparametric predictive inference bootstrap (NPI-B) is a computational implementation of NPI used to quantify uncertainty in statistical inference. The NPI-Bootstrap is used to sample observations from data sets and intervals, adding them to the data set [3, 33].

Assume that the ordered observed values of the random quantities $Y_1, ..., Y_n$ are denoted by $y_1 < y_2 < ... < y_n$, with the lower bound denoted by $y_{(0)}$ and the upper bound by y_{n+1}. The n observations split the real-line into $n + 1$ intervals $I_i = (y_{i-1}, y_i)$, where $i = 1, ..., n + 1$.

The assumption NPI-bootstrap is based on constructing $n+1$ intervals from n observations. As in $A_{(n)}$, we create intervals I_i between the observations n where $i = 1, ..., n + 1$, then draw one value from these intervals and add it to the dataset, and then sampling $m-1$ more values to produce a new sample called an NPI-B sample, which can be applied on a finite interval is as follows:

1- Assume there is a data set of n real-valued.
2- The partitions $n + 1$ created by n observations.
3- Choose one of the $n + 1$ intervals at random, with equal probability for each interval. From this chosen interval, choose one future value uniformly.
4- Increase n to $n + 1$ and add that future value to the data. Steps 2-4 must be repeated with $n + 1$ data to obtain a further future value.
5- Repeat this to produce m NPI-B samples $b_1, b_2, ..., b_m$.
6- Repeat all of these steps n_B times, to obtain a total of n_B NPI-Bootstrap samples of size m.

4 NPI for reproducibility

An important characteristic of the practical application of experiment results is a test's reproducibility of this experiment. An important characteristic of the practical application of experiment results is a test's reproducibility for this experiment. The reproducibility probability (RP), which its definition and interpretation as well as its estimate are studied in the traditional frequentist statistical framework, has attracted a lot of interest recently. The NPI method of frequentist statistics explicitly focuses on future observations while making few assumptions and using lower and upper probabilities to quantify uncertainty. This makes it possible to reach inferences about RP logically given the explicitly predictive nature of NPI.

NPI for reproducibility of statistical tests was introduced by Coolen and Bin Himd [5]. It is also known as NPI-RP and is defined as the probability that the test outcome- that is, whether or not the null hypothesis is rejected- would have the same if the test were repeated based on an experiment performed in the same way as the original experiment. A few basic nonparametric tests were considered for this, such as the two sample rank sum test, Wilcoxon's signed rank test, and the sign test [34]. NPI for reproducibility of statistical tests is used by NPI for Bernoulli quantities [33] and real-valued data [**?**]. This led to NPI lower and upper reproducibility probabilities, denoted by $\underline{RP}$ and $\overline{RP}$, respectively, rather than precisely determined reproducibility probabilities.

For more complicated test situations, NPI can be obtained using the NPI-Bootstrap approach, which Coolen and Bin Himd [6] developed and proved for the Kolmogorov-Smirnov test. The NPI-RP method is applied for two basic order statistical tests: quantile and precedence test to compare data of two populations and a test for a given population quantile value. For lifetime data studies, the precedence test tend to be used when one has to reach a conclusion before all observations are available. The lower and upper reproducibility probabilities for quantile and basic precedence test are provided for these inferences using the NPI for future order statistics [34].

Simkus et al. [35] uses an NPI algorithm to assess the reproducibility of the pairwise t-test and then use simulations to investigate the reproducibility both under the null and alternative hypotheses. This procedure is used to apply NPI reproducibility to real-life applications of a clinical experiment that involves numerous pairwise comparisons of test groups and varying drug concentrations for each group [35]. In

addition, the nonparametric predictive inference approach for reproducibility of likelihood ratio statistic is investigated [31]. The idea of this research is to investigate tests between two simple hypotheses on the mean value. The developed approximations' precision is shown by the numerical studies. So, in order to analyse the reproducibility probability of the tests and power, simulations are developed.

Coolen and Marques [36] introduce the sampling of orderings method. The reproducibility of likelihood ratio tests using the test criterion defined in terms of the sample mean is investigated using NPI to derive lower and upper probabilities of this test. This is done by taking into account all orderings of n future observations among the n data observations, all of which are equally likely based on an exchangeability assumption. However, computing limitations restrict the ability to determine accurate lower and upper probability to very small values of n. In order to approximate the lower and upper repeatability probability, the main goal of this study is to investigate sampling of the orderings of the future data among the seen data. The method is used for both normal and exponential distributions, and the effectiveness of ordered sampling is examined in order to approximate the NPI lower and upper reproducibility probability.

5 Reproducibility of estimates

Estimation of population parameters is an essential part of statistical inference. In this section, we investigate the reproducibility for such estimates. However, it is clear that for real-valued random quantities, an estimate of a parameter will not be reproduced precisely. Therefore, we define reproducibility of an estimate as the probability for the event that, if the experiment under the same circumstances is repeated, the estimate based on the future sample will be close to the estimate based on the original sample.

The objective of this section is to introduce NPI for reproducibility of estimates using two procedures. The first procedure is reproducibility of estimates using NPI-Bootstrap method. The second procedure is reproducibility of estimates using a representative sample of a population, which introduce different samples without using more assumptions. In this paper, we investigate the reproducibility of estimates of population parameters such as the mean. In addition, the reproducibility of estimates of population median, variance, quartiles and interquartile ranges can be investigated using the representative sample [9].

This section is structured as follows. Section 5.1 introduces reproducibility of estimates, in general, using NPI-Bootstrap. In Section 5.2, the concept of a representative sample of the underlying population distribution is introduced, together with its use to asses reproducibility of estimates and a comparison of both technique of NPI-Bootstrap and a a representative sample.

5.1 Reproducibility of estimates using NPI-Bootstrap

In this section, a new theory of reproducibility of estimates is proposed. Suppose that there are n real-valued random quantities $Y_1, Y_2, ..., Y_n$, which are assumed to be independent and identically distributed. Let the ordered observed values of these random quantities be denoted by $y_1 < ... < y_n$. For simplicity of implementing this theory, we set the lower and upper limits of these random quantities denoted by y_0 and y_{n+1} to avoid possible values of $y_0 = -\infty$ or $y_n = \infty$ that could impact the mean of the future m observations where the mean is the most tendency measure which does not influence the outliers. Therefore, we assume the assumption that Y_i has finite support (y_0, y_{n+1}) using

$$y_0 = \min_{1 \le i \le n} (y_i) - d, \quad y_{n+1} = \max_{1 \le i \le n} (y_i) + d \tag{17}$$

where d is the maximum distance between two consecutive observations. We use those limits in order to simplify the NPI-Bootstrap method.

As can be seen, for each $i = 1, ..., n+1$, the n observations partition the real line into $n+1$ intervals, denoted as $I_i = (y_{i-1}, y_i)$. Now, using the NPI-Bootstrap as described in Section 3.2, generate m real-valued random quantities $Y_1^f, Y_2^f, ..., Y_n^f$, where $m = n$. Then, calculate the estimate based on the original sample that is $\hat{\theta}$ which is of interest, and calculate the estimate based on the future samples which are $\hat{\theta}^f$ where $\hat{\theta}^f = \hat{\theta} \pm \epsilon$ and ϵ is the distance between the two estimates and takes values $\epsilon \ge 0$. Now, we can derive theory $\epsilon-$reproducibility of estimates as follows.

The $\epsilon-$reproducibility of an estimate can be quantified as the probability of the absolute value of the difference between the estimate based on the original sample and the estimate based on the future sample which is equal or less than a small real positive value ϵ, which can be written as follows:

$$RP(\epsilon) = P(|\hat{\theta} - \hat{\theta}^f| \le \epsilon) \tag{18}$$

where $\hat{\theta}$ is an estimate of a population parameter based on the original sample, and $\hat{\theta}^f$ is an estimate of a population parameter based on the future sample.

For more clarification, the NPI-Bootstrap method, as described in Section 3.2, is used to generate a future sample $b_1, ..., b_n$ and denote the estimate based on this bootstrap sample by $\hat{\theta}_{B_i}$. We perform this procedure n_B times. Based on these n_B bootstrap samples, we can estimate ϵ-reproducibility of $\hat{\theta}$ by:

$$\widehat{RP}(\epsilon) = \sum_{i=1}^{n_B} \frac{1}{n_B} \mathbf{1}\left\{|\hat{\theta} - \hat{\theta}_{B_i}| \leq \epsilon\right\} \tag{19}$$

with $\epsilon \geq 0$, and $\mathbf{1}\{A\}$ is an indicator function that is equal to 1 if event A is true and 0 otherwise. Note that, by using NPI-Bootstrap, we get a point estimate of $RP(\epsilon)$.

5.2 Reproducibility of estimates using a representative sample

This section introduces a new method for assessing $\epsilon-$reproducibility of estimates of a characteristic population using a representative sample procedure instead of NPI-Bootstrap. This method helps to avoid randomness of sampling from the distribution to eliminate sampling bias. For a population distribution with cumulative distribution function F for real-valued random quantities, we define $y_1, y_2, ..., y_n$ as a representative sample as follows:

$$y_i = F^{-1}\left(\frac{i}{n+1}\right) \tag{20}$$

So, y_i is the $100(\frac{i}{n+1})$-th percentile of F, for $i = 1, ..., n$. We call $y_1, y_2, ..., y_n$ as a representative sample of distribution F with ordering $y_{(1)} < y_{(2)} < ... < y_{(n)}$. The main idea of this method is to estimate the variability from random sampling, and focus on different RP for different methods to estimate the parameter $\hat{\theta}$.

As in Section 5.1, we assume finite support in order to simplify the NPI method, so we define the lower and upper bounds of the original sample y_0 and y_{n+1} are derived as follows:

$$y_0 = \min_{1 \leq i \leq n}(y_i) - d, \quad y_{n+1} = \max_{1 \leq i \leq n}(y_i) + d \tag{21}$$

with d again the maximal distance between two consecutive y_i values. We now have $n + 1$ intervals $I_i = (y_{i-1}, y_i)$, where $i = 1, ..., n + 1$. We assume that all the orderings O_j of the future observations among the original observations are equally likely as explained in Section 3.1, and each ordering includes the future observations $S_i^j = \#\{Y_{n+i}, i = 1, ..., n\}$

where $j = 1, 2, ..., \binom{2n}{n}$. We link the data and future observations via Hill's assumption $A_{(n)}$ [29], or more precisely, via consecutive application of $A_{(n)}, A_{(n+1)}, ..., A_{(2n-1)}$ which can be considered as a post-data version of a finite exchangeability assumption for $2n$ random quantities that are $Y_{n+1}, ..., Y_{2n}$. The $A_{(n)}$ assumptions imply that all possible orderings of n data observations and n future observations are equally likely, where the n data observations and n future observations cannot be separated from one another.

For a larger sample size, we use simple random sampling (SOM) [36] to generate the future observations as explained in Section 3.1. Based on the $A_{(n)}$ assumptions, Equation (22) derive the probability of each ordering [30] as follows.

$$P\left(\bigcap_{i=1}^{n+1}\{S_i^j = s_i^j\}\right) = P(O_j) = \binom{2n}{n}^{-1} \tag{22}$$

where the s_i^j are non-negative integers with $\sum_{i=1}^{n} s_i^j = n$. For ordering O_j, the lower and upper estimates denoted by $\hat{\theta}_{j,L}^f$ and $\hat{\theta}_{j,U}^f$, respectively, can be calculated by using the minimum and maximum possible values the future estimates given these orderings. For example, if interested in the mean, then

$$\hat{\theta}_{j,L}^f = \frac{1}{n}\sum_{i=1}^{n+1} S_i^j y_{i-1}, \qquad \hat{\theta}_{j,U}^f = \frac{1}{n}\sum_{i=1}^{n+1} S_i^j y_i \tag{23}$$

We now use these lower and upper estimates corresponding to ordering O_j to derive the lower and upper probabilities for $\epsilon-$reproducibility of the estimates based on a representative sample. This provides a tool to compare RRT as will be explained in examples. The estimate of $\hat{\theta}^f$ based on the original representative sample is $\hat{\theta}$. To obtain the NPI lower $\epsilon-$reproducibility probability for the event that $|\hat{\theta}^f - \hat{\theta}| \leq \epsilon$, we need to find all estimates $\hat{\theta}^f$ with $[\hat{\theta}_{j,L}^f,\ \hat{\theta}_{j,U}^f] \subset [\hat{\theta} - \epsilon, \hat{\theta} + \epsilon]$. To obtain the NPI upper $\epsilon-$reproducibility probability for the event that $|\hat{\theta}^f - \hat{\theta}| \leq \epsilon$ with the condition $[\hat{\theta}_{j,L}^f, \hat{\theta}_{j,U}^f] \cap [\hat{\theta} - \epsilon, \hat{\theta} + \epsilon] \neq \emptyset$ where $j = 1, ..., \binom{2n}{n}$.

This leads to the NPI lower $\epsilon-$reproducibility probability:

$$\underline{RP}(\epsilon) = \underline{P}(|\hat{\theta}^f - \hat{\theta}| \leq \epsilon) = \sum_{j=1}^{\binom{2n}{n}} P(O_j)\ \ \mathbf{1}\left\{\max(\hat{\theta} - \hat{\theta}_{j,L}^f, \hat{\theta}_{j,U}^f - \hat{\theta}) \leq \epsilon\right\} \tag{24}$$

and the NPI upper $\epsilon-$reproducibility probability:

$$\overline{RP}(\epsilon) = |\overline{P}(\hat{\theta}^f - \hat{\theta}| \leq \epsilon) = \sum_{j=1}^{\binom{2n}{n}} P(O_j) \;\; \mathbf{1}\left\{ \max(\hat{\theta}^f_{j,L} - \hat{\theta}, \hat{\theta} - \hat{\theta}^f_{j,U}) \leq \epsilon \right\} \tag{25}$$

6 Numerical examples of ϵ-Reproducibility of estimates based on RRT methods using NPI-Bootstrap and the representative sample

In this section, we derive the $\epsilon-$reproducibility of estimates using data generated by RRT using NPI-Bootstrap and the representative sample; the approach is detailed in Sections 5.1 and 5.2. We use the simulation to generate the original sample of responses of the respondent of different RRT data such as the true response X, the scrambling response S, the response of the unrelated question Y or the reported response Z. These random quantities are generated using a simulation of RRT method.

To investigate $\epsilon-$reproducibility of estimates using NPI-Bootstrap, we generate all possible responses which are generated from the original sample. Then, calculate the original mean $\hat{\mu}_x$ and the future mean $\hat{\mu}_B$ and then we use the $\epsilon-$reproducibility of estimates is explained in Section 5.1. The next example explains more details about this method.

Example 6.1. This example introduces $\epsilon-$reproducibility of the estimate based on real-valued random quantities generated from the Greenberg method [14] and using NBI-B method. Respondents use the randomisation device to answer one of two questions. One of these questions is sensitive while the other is nonsensitive. Both answers are real-valued quantities. Assume the probability of the sensitive question is $\gamma = 0.70$. We simulate the responses to the sensitive question $X_i \sim N(\mu_x = 1, \sigma_x^2 = 10)$ and the responses to the unrelated question $Y_i \sim N(\mu_y = 4, \sigma_y^2 = 20)$.

Suppose that for $n = 5$, we order the X_i values, which are $2.8516, 3.2435, 0.6544, -0.4341, 2.9160$ and order the Y_i values, which are $-4.1301, 6.8179, 2.7649, 2.7292, -0.1113$. These values are simulated from the given distributions. Assume that $\gamma = 0.70$ is the probability of the question of interest being X_i for each person in which can they give Y_i as an answer. Assume that the randomisation device generated the values $\{1, 1, 0, 0, 1\}$. If the value 1, the response is $Z_i = Y_i$. If the value is 0, the answer is $Z_i = X_i$.

The reported Z_i responses are $-4.1301, 6.8179, 0.6544, -0.4341, -0.1113$. The estimate of the reported responses Z_i based on the original

sample to the sensitive question $\hat{\mu}_x^z$ is

$$\hat{\mu}_x^z = \frac{\hat{\mu}_z - (1-\gamma)\mu_y}{\gamma} = -0.9152 \tag{26}$$

where $\hat{\mu}_z = \frac{\sum_{i=1}^{n} Z_i}{n} = 0.5593$.

To apply NPI-Bootstrap for determining the lower and upper bounds for the support Z_i:

$$z_0 = \min_{1\le i\le n}(z_i) - d = -10.2937 \tag{27}$$

$$z_{n+1} = \max_{1\le i\le n}(z_i) + d = 12.9814 \tag{28}$$

where $d = 6.1635$ is the maximal distance between two consecutive z_i values. We generate $n_B = 1000$ NPI-Bootstrap samples $b_1, ...b_n$ size n based on the z_i values.

We calculate the expected value of bootstrap samples b_i based on the sample Z_i as $\hat{\mu}_z^B$, and we use the mean of normal distribution of the unrelated responses μ_y to derive the estimate of each bootstrap sample $\hat{\mu}_x^B$ as follows:

$$\hat{\mu}_x^B = \frac{\hat{\mu}_z^B - (1-\gamma)\mu_y}{\gamma} \tag{29}$$

where the mean of each bootstrap sample is $\hat{\mu}_z^B = \frac{\sum_{i=1}^{n} b_i}{n}$ where n is the number of bootstrap observations. Then, we calculate the difference between $\hat{\mu}_x^z$ and bootstrap samples means $\hat{\mu}_x^B$ to derive $\epsilon-$reproducibility of the mean for n_B times. Then, find the number of the event that $|\hat{\mu}_x^z - \hat{\mu}_x^B| \le \epsilon$ divided by n_B, as follows:

$$|\hat{\mu}_x^z - \hat{\mu}_x^B| \le \epsilon = \left|\frac{\hat{\mu}_z - (1-\gamma)\mu_y}{\gamma} - \frac{\hat{\mu}_z^B - (1-\gamma)\mu_y}{\gamma}\right| \le \epsilon \tag{30}$$

This leads to derive $\epsilon -$ reproducibility of the mean as follows:

$$\widehat{RP}(\epsilon) = P\left(|\hat{\mu}_x^z - \hat{\mu}_x^B| \le \epsilon\right) = \sum_{i=1}^{n_B} \frac{1}{n_B}\, \mathbf{1}\left\{\left|\frac{\hat{\mu}_z}{\gamma} - \frac{\hat{\mu}_x^{B_i}}{\gamma}\right| \le \epsilon\right\} \tag{31}$$

Perform this procedure n^* times to get n^* original sample to derive n^* of $\epsilon-$reproducibility of estimates. Figure 1 shows the $\widehat{RP}(\epsilon)$ as function of ϵ. It can be seen that for larger values of ϵ, reproducibility increases,

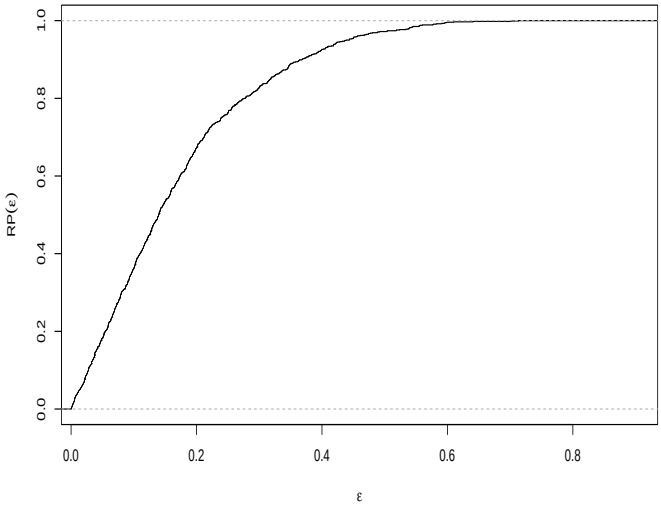

Fig. 1 The average of $\widehat{RP}(\epsilon)$ of the Greenberg method, $n = 5,\ n_B = 1000,\ n^* = 1000,$ $\mu_x = 1,\ \sigma_x^2 = 10,\ \mu_y = 4,\ \sigma_y^2 = 20, \gamma = 0.70$

because the difference between the estimate based on the original sample and the estimate based on the future sample is small which leads to increase the number of the event that $|\hat{\mu}_x^z - \hat{\mu}_x^B| \leq \epsilon$, and then that leads to increase the $\widehat{RP}(\epsilon)$. The results illustrate clearly that the lowest value of $\widehat{RP}(\epsilon)$ is for $\epsilon = 0$ whereas the highest value of $\widehat{RP}(\epsilon)$ for $\epsilon = 0.8$. Therefore, for any two values $\epsilon_2 > \epsilon_1$, the $\widehat{RP}(\epsilon_2) > \widehat{RP}(\epsilon_1)$.

For different original samples, the reproducibility of estimates based on the GM for each sample is calculated using different $\gamma \in [0, 1]$, where the variances of the normal distributions are $\sigma_x^2 > \sigma_y^2$ as shown in Table 1. It is shown that reproducibility gets higher as γ becomes larger because many people answered the unrelated question. The $\widehat{RP}(1)$ of the mean takes values between 0.04 and 0.28 and the $\widehat{RP}(1)$ of the median takes value between 0.03 and 0.27. The $\widehat{RP}(1)$ of $q(0.25)$ takes values between 0.02 and 0.19 whereas the $\widehat{RP}(1)$ of $q(0.75)$ takes values between 0.05 and 0.33. The $\widehat{RP}(1)$ of the IQR takes values between 0.02 and 0.14. The $\widehat{RP}(1)$ of the standard deviation sd takes small values between 0.02 and 0.12. The lower whisker takes the smallest value which is 0 whereas the highest whisker takes values between 0.08 and 0.55.

Table 1 $\widehat{RP}(\epsilon)$ of the GM method of $n = 5, n_B = 100,\ n^* = 100, \mu_x = 1,\ \sigma_x^2 = 10,$ $\mu_y = 4,\ \sigma_y^2 = 20,\ \epsilon = 1$

Summary	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$
$q(0.25)$	0.0530	0.0760	0.0978	0.1160	0.1360	0.1560	0.1740	0.1950
$q(0.75)$	0.0940	0.1340	0.1690	0.2040	0.2283	0.2613	0.2965	0.337
median	0.0690	0.1010	0.1280	0.1500	0.1710	0.1950	0.2215	0.2470
mean	0.0797	0.1155	0.1435	0.1711	0.1957	0.2240	0.2539	0.2820
sd	0.0419	0.0657	0.0698	0.0815	0.0970	0.1071	0.1192	0.1271
IQR	0.0410	0.0580	0.0713	0.0880	0.0923	0.1053	0.1225	0.1420

Table 2 $\widehat{RP}(\epsilon)$ of the GM method of $n = 500, n_B = 100,\ n^* = 100,\ \mu_x = 1, \sigma_x^2 = 10,$ $\mu_y = 4,\ \sigma_y^2 = 20,\ \epsilon = 0.5$

Summary	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$
$q(0.25)$	0.3000	0.4275	0.5300	0.6300	0.7000	0.7700	0.8300	0.8700
$q(0.75)$	0.3600	0.4825	0.6000	0.6900	0.7600	0.8200	0.8800	0.9100
median	0.3300	0.4600	0.5600	0.6500	0.7300	0.8000	0.8500	0.9000
mean	0.3301	0.4578	0.5645	0.6551	0.7309	0.7967	0.8496	0.8902
sd	0.0438	0.0440	0.0531	0.0506	0.0446	0.0462	0.04144	0.0335
IQR	0.0600	0.0550	0.0700	0.0600	0.0600	0.0500	0.0500	0.0400

Table 3 $\widehat{RP}(\epsilon)$ of the GM method of $n = 500,\ n_B = 100,\ n^* = 100,\ \mu_x = 1, \sigma_x^2 = 20,$ $\mu_y = 4,\ \sigma_y^2 = 10,\ \epsilon = 0.5$

Summary	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$
$q(0.25)$	0.2400	0.3800	0.500	0.6200	0.7300	0.8200	0.900	0.9500
$q(0.75)$	0.3000	0.4300	0.5700	0.6900	0.7900	0.8800	0.9400	0.9800
median	0.2750	0.4100	0.5400	0.6600	0.7600	0.8600	0.9200	0.9700
mean	0.2736	0.4077	0.5385	0.6553	0.7595	0.8498	0.9177	0.9602
sd	0.0408	0.0425	0.0526	0.0539	0.0449	0.04192	0.0307	0.0231
IQR	0.0600	0.0500	0.0700	0.0700	0.0600	0.0600	0.0400	0.0300

Table 4 $\widehat{RP}(\epsilon)$ of the GM method of $n = 500,\ n_B = 100,\ n^* = 100,\ \mu_x = 4, \sigma_x^2 = 20,$ $\mu_y = 1,\ \sigma_y^2 = 10,\ \epsilon = 0.5$

Summary	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$
$q(0.25)$	0.3000	0.4400	0.5700	0.6900	0.8000	0.8800	0.9325	0.9725
$q(0.75)$	0.3000	0.4400	0.5700	0.6900	0.8000	0.8800	0.9325	0.9725
median	0.2750	0.4100	0.5400	0.6600	0.7600	0.8600	0.9200	0.9600
mean	0.2742	0.4102	0.5403	0.6569	0.7616	0.8516	0.9163	0.9593
sd	0.0405	0.0445	0.0502	0.0506	0.0493	0.0449	0.0305	0.0219
IQR	0.0500	0.0600	0.0700	0.0700	0.0800	0.0525	0.0325	0.0250

A large sample size leads to higher reproducibility. However, a large number of replications n^* does not lead to more changes in the reproducibility because larger replication leads to more accurate for reproducibility of estimates as shown in Tables 1 and 2. For the assumptions, $\mu_y > \mu_x$ and $\sigma_y^2 > \sigma_x^2$, an increasing the variance of the distribution of the non-sensitive answers leads to an increase of the $\epsilon-$reproducibility for the estimates.

Table 2 shows that the $\widehat{RP}(0.5)$ of the mean takes values between 0.33 and 0.89 and the $\widehat{RP}(0.5)$ of the median takes value between 0.33 and 0.90. The $\widehat{RP}(0.5)$ of $q(0.25)$ takes values between 0.30 and 0.87 whereas the $\widehat{RP}(0.5)$ of $q(0.75)$ takes values between 0.36 and 0.91. The $\widehat{RP}(0.5)$ of the IQR takes values between 0.04 and 0.07. The $\widehat{RP}(0.5)$ of the standard deviation sd takes small values between 0.03 and 0.05.

Tables 2 shows the $\epsilon-$reproducibility of estimates are lower than the $\epsilon-$reproducibility of estimates as shown in Table 3, and that happens when the probability of the sensitive question is $\gamma \leq 0.5$.

Now, we increase the mean of the distribution of the sensitive responses and investigate the $\epsilon-$reproducibility of estimates where $\mu_x > \mu_y$ and $\sigma_y^2 < \sigma_x^2$. We find that the $\epsilon-$reproducibility of estimates of all characteristics increases than the $\epsilon-$reproducibility of estimates at $\mu_x < \mu_y$ and $\sigma_y^2 < \sigma_x^2$ as shown in Table 4. So, the $\epsilon-$reproducibility of estimates of the GM gets higher values if the mean and other measures of the distribution of the sensitive responses of the unrelated response increase for all the value of γ.

Now, an example of the $\epsilon-$reproducibility of estimates using the representative sample derived from a distribution will be given as discussed in Section 5.2. Data generated by the RRT method, including the GM technique. The next example provides additional details for more clarity.

Example 6.2. In this example, we derive the lower and upper reproducibility of estimates based on the GM method [14] using a representative sample with a size of $n = 3$. Assume the probability of the sensitive question is $\gamma = 0.70$. We simulate the random quantity of the responses to the sensitive question $X_i \sim N(\mu_x = 4, \sigma_x^2 = 3)$ and the responses to the unrelated question $Y_i \sim N(\mu_y = 1, \sigma_y^2 = 0.04)$.

We generate random quantities of the original sample Z_i from the normal distribution which has the mean and the variance of Z_i of the GM, where $Z_i \sim N(\gamma\mu_y + (1-\gamma)\mu_x,\ \frac{1}{\gamma^2}(\sigma_y^2 + \gamma(\sigma_x^2 - \sigma_y^2) + \gamma(1-\gamma)(\mu_x - \mu_y)^2)) = N(3.1, 8.1673)$. The first response is $z_1 = 1.1724$, the second response is $z_2 = 3.1$, and the third one is $z_3 = 5.0276$, where the mean of the original sample is $\bar{z} = 3.1$. The lower and upper bounds are $z_0 = -0.7552$ and $z_{n+1} = 6.9552$, where $d = 1.9276$ is the maximal distance between two consecutive of z_i values. Then, find all possible orderings of the future observations to calculate the lower and upper future averages $\underline{z}_l^f$ and $\overline{z}_u^f$. Then, calculate the maximum values of the difference between the lower and upper future averages and the original mean as $\epsilon_{j,l}$ and $\epsilon_{j,u}$, respectively. Then, derive the lower and upper

Table 5 The $\underline{z}^f_{j,l}$, $\overline{z}^f_{j,u}$, $\epsilon_{j,l}$, $\epsilon_{j,u}$, $\underline{RP}(\epsilon_{j,l})$ and $\underline{RP}(\epsilon_{j,u})$ of the GM method of $n = 3$, $n_o = 20$, $\mu_x = 4$, $\mu_y = 1$, $\sigma^2_x = 3$, $\gamma = 0.7$, $\sigma^2_y = 0.04$, $\bar{z} = 3.1$

$\underline{z}^f_{j,l}$	$\epsilon_{j,L}$	$\underline{RP}(\epsilon_{j,l})$	$\overline{z}^f_{j,u}$	$\epsilon_{j,u}$	$\overline{RP}(\epsilon_{j,u})$
-0.7552	1.2851	0.35	1.1724	0.0000	0.55
2.4575	1.2851	0.35	4.3851	0.0000	0.55
2.4575	1.2851	0.35	4.3851	0.0000	0.55
1.1724	1.2851	0.35	3.1000	0.0000	0.55
3.7425	1.2851	0.35	5.6701	0.6425	0.70
2.4575	1.2851	0.35	4.3851	0.6425	0.70
0.5299	1.2851	0.35	2.4575	0.6425	0.70
2.4575	1.9276	0.55	4.3851	0.6425	0.80
3.1000	1.9276	0.55	5.0276	0.6425	0.80
1.8149	1.9276	0.55	3.7425	1.2851	0.95
1.1724	1.9276	0.55	3.1000	1.2851	0.95
0.5299	2.5701	0.80	2.4575	1.2851	0.95
-0.1127	2.5701	0.80	1.8149	1.9276	1.00
4.3851	2.5701	0.80	6.3127	0.0000	0.55
1.1724	2.5701	0.80	3.1000	0.0000	0.55
3.7425	2.5701	0.80	5.6701	0.0000	0.55
2.4575	3.2127	0.95	4.3851	0.0000	0.55
3.7425	3.2127	0.95	5.6701	0.6425	0.70
-0.1127	3.2127	0.95	1.8149	0.6425	0.70
2.4575	3.8552	1.00	4.3851	0.6425	0.70

$\epsilon-$reproducibility of the difference between the average of the original sample and the future sample as Table 5 is shown.

It is noted that a larger distance between the original sample means and lower means $\underline{z}^f_l$ and upper means $\overline{z}^f_u$ leads to larger reproducibility of the mean. Therefore, largest value of $\epsilon_{j,l} = 3.8552$ leads to largest value of $\underline{RP}(3.8552) = 1$. Similarly, the largest value of $\epsilon_{j,u} = 1.9276$ leads to the largest value of $\overline{RP}(1.9276) = 1$ and the lowest value of $\epsilon_{j,u} = 0$ leads to the largest value of $\overline{RP}(0) = 0.55$ as Table 5 is shown.

Table 6 shows the lower and upper reproducibility probabilities for lower quartile $q(0.25)$, median and $q(0.25)$, the upper quartile $q(0.75)$, the interquartile range IQR, the mean and the lowest and highest whiskers of different sample size $n^* = 100, 500, 1000$. The lower and upper reproducibility probabilities for median does not change considerably with different n^*. The lower and upper reproducibility probabilities for mean takes a value between 0.74 and 0.91 respectively. The lowest values of the upper reproducibility probabilities is for standard deviation, and for IQR.

In addition, the results show that the increasing of the replication numbers n^* does not affect considerably on the lower and upper $\epsilon-$reproducibily of the mean. That means the increasing in n^* leads to

Table 6 Estimates of $\underline{RP}(1)$ and $\overline{RP}(1)$ using the GM method of $n = m = 30$, $n_o = 20$, $\mu_x = 4$, $\mu_y = 1$, $\gamma = 0.7$, $\sigma_x^2 = 3$, $\sigma_y^2 = 0.04$

	$\underline{RP}(1)$			$\overline{RP}(1)$		
n^*	100	500	1000	100	500	1000
$q(0.25)$	0.7000	0.7000	0.7000	0.8500	0.8500	0.900
$q(0.75)$	0.8000	0.8000	0.8000	0.9500	0.9500	0.9500
Median	0.7500	0.7500	0.7500	0.9000	0.9000	0.9000
Mean	0.7415	0.7506	0.7480	0.9155	0.9181	0.9152
sd	0.0935	0.0871	0.0932	0.0610	0.0595	0.0619
IQR	0.1000	0.1000	0.1000	0.1000	0.1000	0.0500

Table 7 Average of $\underline{RP}(1)$ and $\overline{RP}(1)$ using the GM method of $n^* = 100$, $\mu_x = 4, \mu_y = 1, \gamma = 0.7$, $\sigma_x^2 = 3$, $\sigma_y^2 = 0.04$

n	n_o	$\underline{RP}(1)$	$\overline{RP}(1)$	n	n_o	$\underline{RP}(1)$	$\overline{RP}(1)$
	100	0.1729	0.7920		100	0.9807	0.9948
5	500	0.1682	0.7865	100	500	0.9799	0.9936
	1000	0.1684	0.7853		1000	0.9795	0.9935
	100	0.7506	0.9181		100	1.0000	1.0000
30	500	0.7504	0.9176	1000	500	1.0000	1.0000
	1000	0.7512	0.9186		1000	1.0000	1.0000

a slight increase in $\underline{RP}(1)$ and $\overline{RP}(1)$.

For different sample sizes and orderings numbers, Table 7 presents the $\underline{RP}(1)$ and $\overline{RP}(1)$ for mean. It shows that increasing sample size leads to increasing the average of the lower and upper $\epsilon-$reproducibility. As a result, it is noted that the sample size and choosing parameters of the RRT methods have a basic role to obtain high the lower and upper $\epsilon-$reproducibility of RRT method.

Now, we computed the exact lower and upper $\epsilon-$reproducibility probabilities for estiamtes using a sample with a size of $n = m$, considering large orderings n_o and appling the SOM methodology to compute the lower and upper $\epsilon-$reproducibility probabilities in order to assess the precision of the SOM method of the computation of the lower and upper $\epsilon-$reproducibility based on GM method. The Normal distribution is assumed to be the underlying distributions. Then, the number of orderings n_o are generated equal to $1000, 2000, 5000, 10000, 20000, 50000, 100000$ using the SOM approach, and then the 95% confidence intervals are calculated for both the lower and upper bounds in each replication as shown on Table 8.

When the number of sampled orderings is larger than or equal 1000, we calculate the lower confidence interval of exact $\epsilon-$reproducibility of estimates, the interval is calculated using the normal approximation for the ordering number n_o. Therefore, the $(1-\alpha)\%$ confidence intervals of

Table 8 The lower and upper of CI(95%) of $\underline{RP}(1)$ and $\overline{RP}(1)$ using the GM method of $n = m = 30,\ n^* = 100,\ \mu_x = 4,\ \mu_y = 1,\ \gamma = 0.7,\ \sigma_x^2 = 3,\ \sigma_y^2 = 0.04$

n_o	Average $\underline{RP}(1)$	$\underline{CI}(0.95)$	Average $\overline{RP}(1)$	$\overline{CI}(0.95)$
1000	0.7512	(0.7244, 0.7780)	0.9186	(0.9017, 0.9355)
2000	0.7511	(0.7322, 0.7700)	0.9192	(0.9073, 0.9311)
5000	0.7501	(0.7311, 0.7691)	0.9186	(0.9066, 0.9306)
10000	0.7497	(0.7307, 0.7687)	0.9184	(0.9064, 0.9304)
20000	0.7496	(0.7436, 0.7556)	0.9184	(0.9146, 0.9222)
50000	0.7494	(0.7456, 0.7532)	0.9183	(0.9159, 0.9207)
100000	0.7497	(0.7470, 0.7524)	0.9184	(0.9167, 0.9201)

the $\underline{RP}(\epsilon)$ and $\overline{RP}(\epsilon)$ are derived by

$$\underline{RP}(\epsilon) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\underline{RP}(\epsilon)\left(1-\underline{RP}(\epsilon)\right)}{n_o}}, \qquad \overline{RP}(\epsilon) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{RP}(\epsilon)\left(1-\overline{RP}(\epsilon)\right)}{n_o}} \tag{32}$$

where $z_{\frac{\alpha}{2}}$ is $1-\frac{\alpha}{2}$ quantile of the standard Normal distribution. Table 8 show that increasing n_o leads to a slight decreases in the $\underline{RP}(\epsilon)$ and $\overline{RP}(\epsilon)$ of the mean of GM method and the confidence intervals $\underline{CI}(95\%)$ and $\overline{CI}(95\%)$.

In general, it is noted that increasing the number of orderings of future observations leads to an increase in the approximate lower and upper reproducibility of estimates based on RRT method and a decrease in the range of lower and upper confidence intervals for the lower and upper reproducibility probabilities. For further applications, several RRT methods are used to applying reproducibility of estimates supported with several examples [9].

Now, it is useful to compare the $\epsilon-$reproducibility of estimates derived from RRT techniques to find a link with it and the degree of privacy and variance change. The following example compare between those properties.

Example 6.3. This example compares RRT methods for real-valued quantities based on three properties; variance, privacy degree, and $\epsilon-$reproducibility of estimates based on GM, EH, and MM data using simulation with NPI-B and the representative sample, we set parameters values for each method to achieve the same privacy degree, then assess the variance and reproducibility of each method. To provide more clarification, the previous RRT techniques are used to $\epsilon-$reproducibility of estimates [9].

We first fix some parameters values of RRT methods such as the sample size $n = 100,\ 300,\ 500$, the number of ordering $n_o = 1000$, the number of NPI-Bootstrap samples $n_B = 1000$ and $\epsilon = 0.1,\ 0.7$.

Table 9 Reproducibility of the mean of GM of $n_B = 1000$, $n_o = 1000$, $\mu_x = 4$, $\mu_y = 4$, $\sigma_x^2 = 2.5$, $\sigma_y^2 = 1.5780$, $\epsilon = 0.1$

	$n = m =$	100	300	500	Δ_{GM}	$\text{Var}(Z_i)$
$\gamma = 0.7$	$\widehat{RP}(\epsilon)$	0.3330	0.4730	0.5690	1.2234	3.7849
	$\underline{RP}(\epsilon)$	0.1330	0.3680	0.5020		
	$\overline{RP}(\epsilon)$	0.4170	0.5260	0.6170		
$\gamma = 0.53$	$\widehat{RP}(\epsilon)$	0.2460	0.3680	0.4450	1.9167	7.3573
	$\underline{RP}(\epsilon)$	0.0680	0.2770	0.3840		
	$\overline{RP}(\epsilon)$	0.3740	0.4570	0.5280		

Table 10 Reproducibility of the mean of MM of $n_B = 1000$, $n_o = 1000$, $\mu_x = 2$, $\theta = 1$, $\sigma_x^2 = 4.8$, $\gamma^2 = 0.2958$, $\epsilon = 0.1$

	$n = m =$	100	300	500	Δ_{MM}	$\text{Var}(Z_i)$
$\text{MM}(\psi = 0.70)$	$\widehat{RP}(\epsilon)$	0.2850	0.4220	0.5010	0.7809	6.6221
	$\underline{RP}(\epsilon)$	0.0830	0.2890	0.4170		
	$\overline{RP}(\epsilon)$	0.3830	0.4680	0.5410		
$\text{MM}(\psi = 0.53)$	$\widehat{RP}(\epsilon)$	0.2850	0.4220	0.5010	1.2234	6.1796
	$\underline{RP}(\epsilon)$	0.0900	0.3100	0.4260		
	$\overline{RP}(\epsilon)$	0.3900	0.4780	0.5520		

Table 11 Reproducibility of the mean of EH of $n_B = 1000$, $n_o = 1000$, $\mu = 2$, $\theta = 2$, $\sigma_x^2 = 12.543$, $r^2 = 0.2958$, $\epsilon = 0.1$

EH	$n = m =$	100	300	500	Δ_{EH}	$\text{Var}(Z_i)$
	$\widehat{RP}(\epsilon)$	0.1160	0.1900	0.2340	1.2234	13.7664
	$\underline{RP}(\epsilon)$	0.0160	0.1770	0.2750		
	$\overline{RP}(\epsilon)$	0.3270	0.3580	0.4140		

We set the other parameter values to obtain the same privacy degree 1.2234 of the RRT methods as follows. The parameter values of the GM methods are $\mu_x = 4$, $\mu_y = 4$, $\sigma_x^2 = 2.5$, $\sigma_y^2 = 1.5780$ and $\gamma = 0.70$. The parameter values of the MM method are $\mu_x = 2$, $\theta = 1$, $\sigma_x^2 = 4.8$, $\gamma^2 = 0.2958$, and $\psi = 0.70$. The parameter values of the EH methods are $\mu_x = 4$, $\theta = 2$, $\sigma_x^2 = 12.543$, $r^2 = 0.2958$. Then, we change γ, ψ and ϵ to investigate the changes in the RRT method in terms of privacy degree, the variance and the $\epsilon-$reproducibility of the mean.

Tables 9, 10 and 11 show that the comparison between GM, MM and EH methods using the NPI-B and the representative sample. The results show that the lower and upper $\epsilon-$reproducibility of the mean of RRT methods increases if the variance decreases (the efficiency of the method increases) while the privacy degree decreases. It is also shown that the $\epsilon-$reproducibility of the mean based on the GM, MM and EH method increases if γ increases or the sample size increases. The $\epsilon-$reproducibility for an estimate using NPI-B gets values within the range of the lower and upper reproducibility of RRT using the representative sample except for the cases in which the difference

Table 12 Reproducibility of the mean of GM of $n_B = 1000$, $n_o = 1000$, $\mu_x = 4$, $\mu_y = 4$, $\sigma_x^2 = 2.5$, $\sigma_y^2 = 1.5780$, $\epsilon = 0.7$

	$n = m =$	100	300	500	Δ_{GM}	$\text{Var}(Z_i)$
$\gamma = 0.7$	$\widehat{RP}(\epsilon)$	0.9890	1.0000	1.0000	1.2234	4.5376
	$\underline{RP}(\epsilon)$	0.9750	1.0000	1.0000		
	$\overline{RP}(\epsilon)$	0.9900	1.0000	1.0000		
$\gamma = 0.53$	$\widehat{RP}(\epsilon)$	0.9270	1.0000	1.0000	1.9167	7.35728
	$\underline{RP}(\epsilon)$	0.9150	0.9990	1.0000		
	$\overline{RP}(\epsilon)$	0.9670	1.0000	1.0000		

Table 13 Reproducibility of the mean of MM of $n_B = 1000$, $n_o = 1000$, $\mu_x = 2$, $\theta = 1$, $\sigma_x^2 = 4.8$, $\gamma^2 = 0.2958$, $\epsilon = 0.7$

	$n = m =$	100	300	500	Δ_{MM}	$\text{Var}(Z_i)$
MM($\psi = 0.70$)	$\widehat{RP}(\epsilon)$	0.9850	1.0000	1.0000	0.7809	6.6221
	$\underline{RP}(\epsilon)$	0.9270	1.0000	1.0000		
	$\overline{RP}(\epsilon)$	0.9740	1.0000	1.0000		
MM($\psi = 0.53$)	$\widehat{RP}(\epsilon)$	0.9850	1.0000	1.0000	1.2234	6.1796
	$\underline{RP}(\epsilon)$	0.9370	1.0000	1.0000		
	$\overline{RP}(\epsilon)$	0.9780	1.000	1.0000		

Table 14 Reproducibility of the mean of EH of $n_B = 1000$, $n_o = 1000$, $\mu = 2$, $\theta = 2$, $\sigma_x^2 = 12.543$, $r^2 = 0.2958$, $\epsilon = 0.7$

EH	$n = m =$	100	300	500	Δ_{EH}	$\text{Var}(Z_i)$
	$\widehat{RP}(\epsilon)$	0.6990	0.8970	0.9640	1.2234	13.7664
	$\underline{RP}(\epsilon)$	0.7750	0.9740	0.9940		
	$\overline{RP}(\epsilon)$	0.8910	0.9890	0.9970		

between the mean $\hat{\mu}_x^z$ of original samples of the reproducibility using the NPI-Bootstrap are large than the mean $\hat{\mu}_x^z$ of original samples of the reproducibility using the representative sample more than 0.2.

In addition, An increasing the sample size leads to higher $\epsilon-$reproducibility and obtains higher values of lower and upper $\epsilon-$reproducibility using the representative sample and $\epsilon-$reproducibility using NPI-Bootstrap. Tables 10 and 11 show that $\epsilon-$reproducibility of the mean based on MM is higher than the RP for estimates based on EH. Increasing ϵ leads to an increase of $\epsilon-$reproducibility as shown in Tables 12, 13 and 14.

Based on the comparisons of the quantitative RRT methods, it is observed that the GM method has less variability of the reported responses than the EH and the MM method at the same privacy degree. Furthermore, the $\epsilon-$reproducibility of estimates is affected by the variance of the original sample (the variability in the reported responses). Therefore, the variance increases (the variability of the

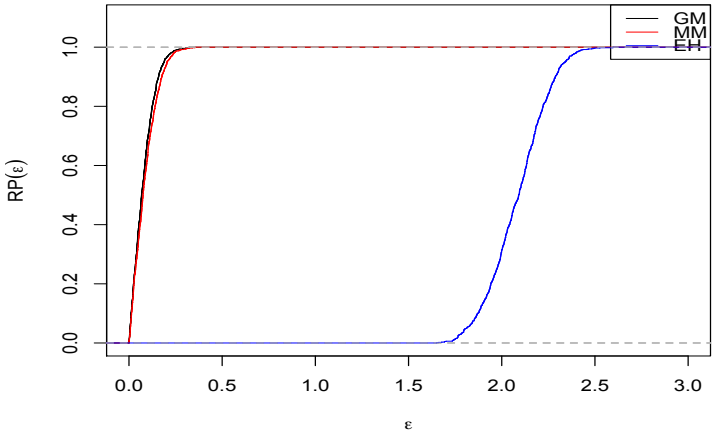

Fig. 2 $\widehat{RP}(\epsilon)$ values for the mean of GM, MM, and EH with the same $n = 100$, $n_B = 1000$, and $\Delta = 1.223$

reported responses becomes large), then the $\epsilon-$reproducibility of estimates decreases. In addition, higher $\epsilon-$reproducibility of estimates leads to lower privacy degree of the RRT methods. Figure 2 illustrates that, for varying values of ϵ, the reproducibility for the mean of the GM method is higher than reproducibility for the mean of the MM and EH methods. While reproducibility for the mean of the GM and MM methods at $\epsilon \geq 0.5$ has become the same, reproducibility the mean of all RRT methods has become the same as well at $\epsilon \geq 3$.

7 Concluding remarks

This paper studies $\epsilon-$reproducibility of estimates as introduced in Section 5 based on quantitative RRT methods in two ways; the first method uses NPI-Bootstrap and the other method uses the representative sample. $\epsilon-$reproducibility of estimates using NPI-Bootstrap investigates the $\epsilon-$reproducibility of estimates based on the randomised response method by using the simulation. The $\epsilon-$reproducibility of estimates has different sampled ordering depending on the design of RRT methods. This design could be the estimate of the reported response and then the mean of bootstrap samples of the RRT method.

Using NPI-Bootstrap method is an excellent procedure to generate large possible future observations of the original sample while SOM method is a helpful technique to generate large number of possible orderings of future observations. Therefore, for a large sample

size, if it cannot consider all the orderings of the future observations, we use sampling of ordering method (SOM) to obtain a large number of orderings to derive approximation of the lower and upper $\epsilon-$reproducibility. Using a larger sample size n leads to a decrease in the difference between the lower and upper $\epsilon-$reproducibility and gives accurate $\epsilon-$reproducibility. Additionally, a lower variance of the reported responses leads to higher $\epsilon-$reproducibility with the same privacy degree.

It is noted that $\epsilon-$reproducibility of an estimate of the GM method has less variability of the reported responses than the MM and EH methods. Therefore, there is a strong relationship between this variability and higher $\epsilon-$reproducibility of estimates of RRT method. That means less variability of the reported responses leads to high $\epsilon-$reproducibility of an estimate. In addition,increasing ϵ and the sample size n leads to higher $\epsilon-$reproducibility of an estimate. For further research, this work can be applied to different RRT methods that have different procedures or multiple samples. In addition, $\epsilon-$reproducibility of estimates can be improved to investigate a unified measure to connect the variability of the reported responses, respondents' privacy and $\epsilon-$reproducibility of estimates.

Finally, one may also compare the proposed method with other additive models aimed at enhancing privacy, such as the Warner additive RRT model ($Z = X + S$) [37] or the Diana and Perri linear combination model ($Z = TX + S$) [38]. This is left for future research.

Acknowledgements

The research described in this paper was conducted during Fatimah Alghamdi's PhD studies at the Department of Mathematical Sciences, Durham University. Thank you for Princess Nourah bint Abdulrahman University, and the Saudi Ministry of Education in Saudi Arabia for all contributed financial support for this research. We would like to thank Professor Sat Gupta for his insightful discussions and his important suggestions to our research.

References

[1] Goodman, S.N.: A comment on replication, p-values and evidence. *Statistics in Medicine*, **10**, 875-879 (1992).

[2] Senn, S.: A comment on replication, p-values and evidence. *Statistics in Medicine*, **21**, 2437-2444 (2002).

[3] Coolen, F.P.A., BinHimd, S.: Nonparametric predictive inference for reproducibility of basic nonparametric tests. *Journal of*

Statistical Theory and Practice, **8**, 591-618 (2014).

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

[5] Coolen, F.P.A.: Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters*, **36**, 349-357 (1998).

[6] Coolen, F.P.A., Bin Himd. S.: Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov-Smirnov test. *Journal of Statistical Theory and Practice*, **14**, (2020).

[7] Billheimer, D.: Predictive inference and scientific reproducibility. *The American Statistician*, **73**, 291-295 (2019).

[8] Alghamdi, F.M., Coolen, F.P.A., Coolen-Maturi, T.: Reproducibility of statistical tests based on randomised response data. *Journal of Statistical Theory and Practice*, **18**, 13 (2024).

[9] Alghamdi, F.M.: *Reproducibility of Statistical Inference Based on Randomised Response Data*. PhD thesis. Durham University (2022). Available at: http://etheses.dur.ac.uk/14783/.

[10] Warner, S.L.: Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **62**, 63-69 (1965).

[11] Greenberg, B.G. Abul-Ela, A.L.A. Simmons, W. R. , Horvitz, D. G.: The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, **65**, 520-539 (1969).

[12] Eichhorn, B.H. , Hayre, L. S. : Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, **7**, 307-316 (1983).

[13] Gupta, S.N. Thornton, Shabbir , B. J. , Singhal, S.: A comparison of multiplicative and additive optional RRT models, *Journal of Statistical Theory and Applications*, **5**, 226-239 (2006).

[14] Greenberg, B.G., Kuebler, R.R., Abernathy, J.R., Horvitz, D.G.: Application of the randomized response technique in obtaining quantitative data, *Journal of American Statistical*

Association, **66**, 243-250 (1971).

[15] Yan, Z., Wang, J., Lai, J.: An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics-Theory and Methods*, **38**, 400-408 (2008).

[16] Gupta, S.N., Gupta, B.C., Singh, S.: Estimation of sensitivity level of personal interview survey questions, *Journal of Statistical Planning and Inference*, **100**, 239-247 (2002).

[17] Young, A., Gupta, S., Parks, R.:A binary unrelated-question RRT model accounting for untruthful responding. *Involve - Journal of Mathematics*, **12**, 1163-1173 (2019).

[18] Zhimin, H., Zaizai, Y.: Measure of privacy in randomized response model. *Quality and Quantity*, **46**, 1167-1180 (2012).

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

[20] Coolen, F.P.A.: On nonparametric predictive inference and objective Bayesiansm. *Journal of Logic, Language and Information*, **15**, 21-47 (2006).

[21] De Finetti, B.: Theory of Probability. London: Wiley (1974).

[22] Walley, P.: *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall (1991).

[23] Weichselberger, K.: Elementare Grundbegriffe einer Allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als Umfassendes Konzept (In German). Physika, Heidelberg (2001).

[24] Baker, R.M., Coolen-Maturi, T., Coolen, F.P.A.: Nonparametric predictive inference for stock returns. *Journal of Applied Statistics*, **44**, 1333-1349 (2017).

[25] T. He, Coolen, F.P.A., Coolen-Maturi, T., Nonparametric predictive inference for European option pricing based on the Binomial Tree Model. *Journal of the Operational Research Society*, **70**, 1692-1708 (2019).

[26] Chen, J., Coolen, F.P.A., Coolen-Maturi, T.: On nonparametric predictive inference for asset and European option trading in the binomial tree model. *Journal of the Operational Research Society*, **70**, 1678-1691 (2019).

[27] Arts, G.R.J., Coolen, F.P.A.: Two nonparametric predictive control charts, *Journal of Statistical Theory and Practice*, **2**, 499-512 (2008).

[28] Arts, G.R.J., Coolen, F.P.A., Van der Laan, P.: Nonparametric predictive inference in statistical process control, *Quality Technology and Quantitative Management*, **1**, 201-216 (2004).

[29] B.M. Hill, De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference (with discussion). In J.M. Bernardo, et al. (Eds.), *Bayesian Statistics*. **3**, 211-241. Oxford University Press, (1988).

[30] Coolen, F.P.A., T. Maturi.: Nonparametric predictive inference for order statistics of future observations. In: *Combining Soft Computing and Statistical Methods in Data Analysis*. C. Borgelt, *et al.* (eds). Springer, Berlin (Advances in Intelligent and Soft Computing, 77, 97-104 (2010).

[31] Marques, F.J., Coolen, F.P.A., Coolen-Maturi, T.: Approximations for the likelihood ratio statistic for hypothesis testing between two Beta distributions, *Journal of Statistical Theory and Practice*, **13**, 17 (2019).

[32] Efron,B.: Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**, 1-26 (1979).

[33] BinHimd, S.: *Nonparametric predictive methods for bootstrap and test reproducibility*. PhD thesis. Durham University (2014). Available at: http://npi-statistics.com.

[34] Alqifari, H.N.:*Nonparametric Predictive Inference for Future Order Statistics*. PhD thesis. Durham University (2017). Available at : http://npi-statistics.com.

[35] 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**, 673-688 (2022).

[36] Coolen, F.P.A., Marques, F.J.: Nonparametric predictive inference for test reproducibility by sampling future data orderings. *Journal of Statistical Theory and Practice*, **14**, pp.1-22 (2020).

[37] Warner, S.L.: The linear randomized response model. *Journal of the American Statistical Association*, **66**, 884–888 (1971).

[38] Diana, G. and Perri, P.F.: A class of estimators for quantitative sensitive data. *Statistical Papers*, **52**, 633-650 (2011).