AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 11(6): 18692–18714.
DOI: 10.3934/math.2026760
Received: 26 March 2026
Revised: 11 June 2026
Accepted: 15 June 2026
Published: 25 June 2026

Research article

An improved randomized response model: Theory, efficiency analysis, and applications

Shehzad Ahmad Khan[1], **Moiz Qureshi**[1,2], **Hasnain Iftikhar**[1,3,*], **Fatimah E. Almuhayfith**[4,*] and **Paulo Canas Rodrigues**[5,6]

[1] Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan
[2] Department of Statistics, University of Sindh, Jamshoro, Pakistan
[3] Department of Statistics, University of Peshawar, Peshawar 25120, Pakistan
[4] Department of Mathematics and Statistics, College of Science, King Faisal University, Alahsa 31982, Saudi Arabia
[5] Department of Statistics, Federal University of Bahia, Salvador 40170-110, Brazil
[6] Department of Business Management, University of Pretoria, Pretoria 0002, South Africa

* **Correspondence:** Email: hasnainchill3@gmail.com, falmuhaifeez@kfu.edu.sa.

Abstract: Randomized response (RR) techniques are commonly used to estimate the proportion of a sensitive attribute in a finite population while the preserving respondents' privacy. In this paper, we develop a novel two-stage randomized response model to jointly estimate the population proportion η of a sensitive characteristic and the probability T that a respondent provides a true answer under direct questioning. First the proposed design first implements direct questioning and only applies a randomization device to non-respondents, thus reducing unnecessary randomization and improving data quality. The randomization mechanism incorporates three additional questions within a two-stage probabilistic framework, thus offering enhanced privacy protection and improved statistical efficiency. Closed-form expressions for the proposed estimators are derived, and their statistical properties, including unbiasedness and variance, are obtained under simple random sampling. Analytical efficiency comparisons with existing models are established through mean-squared error criteria. The methodology is further extended to stratified random sampling, and the corresponding estimators and variance expressions are derived. Monte Carlo simulations are conducted to assess the finite-sample performance and to support the theoretical findings. The results demonstrate that the proposed estimators outperform competing approaches in both efficiency and privacy protection. The proposed framework provides a mathematically rigorous and practically effective tool for inference on sensitive characteristics in survey sampling.

Keywords: randomized response technique; sensitive survey; two-stage randomization; stratified

sampling; unbiased estimation; efficiency comparison
Mathematics Subject Classification: 62D05, 62F10, 62P25

1. Introduction

In socioeconomic and behavioral surveys, two primary sources of non-sampling error are the respondents' refusal to participate and the deliberate provision of inaccurate answers. The bias induced by such errors can be substantial, often rendering sample estimates unreliable [1]. This issue becomes particularly severe when surveys address sensitive topics such as drug addiction, homosexuality, or substance abuse, where respondents may feel discomfort or fear social stigma, thus avoiding truthful disclosure.

To mitigate non-response and response bias in sensitive surveys, [2] introduced the pioneering randomized response (RR) technique to estimate the proportion of individuals who possess a confidential characteristic. The RR technique replaces direct questioning with a probabilistic randomization mechanism, thereby allowing respondents to answer based on the outcome of a random device whose realization is unknown to the investigator. This mechanism ensures respondent privacy while enabling unbiased estimation.

In [2] model, two mutually exclusive and exhaustive statements were considered: (i) "I belong to the sensitive group" and (ii) "I do not belong to the sensitive group", selected with probabilities p_r and $(1 - p_r)$, respectively. Respondents answer *yes* or *no* based on the selected statement. Assuming truthful compliance with the randomization device, the probability of a *yes* response, denoted by θ_{wa}, is given by the following:

$$\theta_{wa} = p_r\eta + (1 - p_r)(1 - \eta), \tag{1.1}$$

where η represents the population proportion of the sensitive attribute.

The maximum likelihood estimator

$$(MLE)$$

of η is

$$\hat{\eta}_{wa} = \frac{\hat{\theta}_{wa} - (1 - p_r)}{2p_r - 1}, \tag{1.2}$$

with variance

$$V(\hat{\eta}_{wa}) = \frac{\eta(1 - \eta)}{n} + \frac{p_r(1 - p_r)}{n(2p_r - 1)^2}. \tag{1.3}$$

Following the seminal contribution of [2], the RR technique has undergone extensive development, with numerous modifications and extensions proposed to enhance both the efficiency and the respondent's privacy. Early and foundational advances include the works of [3,4], which introduced improved designs and refined probability mechanisms to reduce bias while maintaining confidentiality.

Subsequent research further expanded the applicability and robustness of RR methodologies. Notable contributions include [5–7], who explored alternative scrambling and randomized mechanisms to improve the estimator performance and practical implementation. Later developments focused on efficiency gains and structural improvements, as seen in [8–10], along with extensions that addressed more complex survey settings in [11, 12].

An important advancement was made by [13, 14], who developed alternative *RR* procedures that allowed for a simultaneous estimation of the sensitive proportion η and the parameter T, defined as the probability that a respondent truthfully reports possessing the sensitive attribute under direct questioning. Subsequently, [15, 16] proposed improved RR models that enhanced the efficiency relative to the [14] procedure.

Motivated by recent advancements in RR techniques, this study proposes a novel two-stage randomized response procedure for the simultaneous estimation of η and T. In the proposed framework, respondents are first given a direct question, and only those who refuse to respond are subsequently subjected to a randomization device. This selective use of randomization minimizes the unnecessary perturbation of responses, enhances data quality, and improves the respondent cooperation by reducing the perceived survey burden. In addition, the proposed mechanism integrates three auxiliary statements within a two-stage probabilistic structure, thereby providing stronger privacy protection and improving the estimation efficiency compared to [14–16], while retaining the original modeling structure.

Under a simple random sampling scheme, closed-form estimators are derived along with their statistical properties [9]. Theoretical efficiency comparisons, supported by analytical results, demonstrate the superiority of the proposed estimators over those of [14–16]. A key feature of the proposed methodology is its ability to jointly estimate the sensitivity level T and the sensitive proportion η, whereas existing competing models do not simultaneously address these two parameters. The approach is further extended to stratified random sampling [17, 18], and efficiency comparisons are conducted against [16, 19, 20]. Extensive numerical studies confirm that the proposed estimators consistently achieve lower mean squared errors while providing stronger privacy protection than existing approaches [21, 22].

The remainder of the paper is organized as follows: Section 2 presents the proposed model under simple random sampling and derives the corresponding estimators and their properties; Section 3 consists of the literature on existing estimators; Section 4 extends the methodology to stratified random sampling and provides analytical efficiency comparisons; and concluding remarks are given in Section 5.

2. Literature review of existing estimators

[14, 15] developed RR models to estimate the population proportion of sensitive attributes and the associated sensitivity parameters. These models are briefly described below.

2.1. Huang2004 [14] model

[14] proposed a model in which a sample of size n is selected using simple random sampling with replacement ($SRSWR$) from a finite population of size N. First, the sampled individuals are asked a direct question regarding whether they belong to the sensitive group A. If the respondent answers “no”, then They are provided with a randomization device (RD) that consists of two statements: (a) “I belong to group A” and (b) “I do not belong to group A”, with associated probabilities p_r and $(1 - p_r)$, respectively. It is assumed that respondents who possess the non-sensitive attribute respond truthfully under both direct questioning and the RR procedure. For respondents in group A, it is assumed that they respond truthfully under the RR procedure, whereas under direct questioning, they may respond honestly with a probability T. The entire response process remains unobserved by the interviewer, thus ensuring respondent privacy and reducing the response bias.

Under the above model the probability of a *yes* response θ_{h_1} in the direct response is defined as:

$$pr\,(yes) = \theta_{h_1} = \eta T, \tag{2.1}$$

and the probability of *yes* response θ_{h2} in the *RRT* under the above model is given by

$$pr\,(yes) = \theta_{h_2} = p_r\eta\,(1-T) + (1-p_r)\,(1-\eta)\,, \tag{2.2}$$

where T is the probability of truthful reporting in direct questioning about a sensitive characteristic.

The estimator of η and T are given by

$$\hat{\eta}_h = \frac{\hat{\theta}_{h_2} + p\hat{\theta}_{h_1} - (1-p)}{(2p-1)} \tag{2.3}$$

$$\hat{T}_h = \frac{p\hat{\theta}_{h_1}}{\hat{\theta}_{h_2} + p_r\hat{\theta}_{h_1} - (1-p_r)} respectively, \tag{2.4}$$

where, $\hat{\theta}_{h_i}(i = 1, 2)$ is the proportion of a,*yes* response.

The variance of $\hat{\eta}_h$ is written as:

$$V(\hat{\pi}_h) = \frac{\eta\,(1-\eta)}{n} + \frac{p_r\,(1-p_r)\,(1-\eta T)}{n\,(2p_r-1)^2}, \tag{2.5}$$

and the MSE of $\hat{T}_h$ is given by

$$MSE(\hat{T}_h) = \frac{T\,(1-T)}{n\eta} + \frac{p_r\,(1-p_r)\,T^2\,(1-\eta T)}{n\,(2p_r-1)^2\,\eta^2}. \tag{2.6}$$

2.2. Singh2013 [15] model

In this technique, a sample of size n is selected by $SRSWR$ from a finite population of size N. The sampled individual is asked to answer a direct query about whether he/she belongs to A. When answering *no*, the respondent is provided with an *RD* consisting of three statements: (a) 'I am a member of a sensitive group', (b) say *yes*, and (c) say *no* with known probabilities p_r, $(1-p_r)/2$, $(1-p_r)/2$, respectively. It is assumed that the respondents who belong to the sensitive group have no reason to tell a lie; thus, they are completely truthful in their response, no matter whether a direct response or an *RR* procedure is adopted.

Under the [15] model, the probability of *yes* response in the direct response is defined as:

$$pr\,(yes) = \theta_{s_1} = \eta T, \tag{2.7}$$

and probability of *yes* response under [15] model in the indirect response is defined as:

$$pr\,(yes) = \theta_{s_2} = p_r\eta\,(1-T) + \frac{(1-p_r)}{2}. \tag{2.8}$$

The estimator of η and T are respectively given by

$$\hat{\eta}_s = \frac{\hat{\theta}_{s_2} + p_r\hat{\theta}_{s_1} - \dfrac{(1-p_r)}{2}}{p_r} \tag{2.9}$$

and

$$\hat{T}_s = \frac{p_r\hat{\theta}_{s_1}}{p_r\hat{\theta}_{s_1} + \hat{\theta}_{s_2} - \frac{(1-p_r)}{2}}, respectively. \tag{2.10}$$

The variance of $\hat{\eta}_s$ is:

$$V(\hat{\eta}_s) = \frac{\eta(1-\eta)}{n} + \frac{(1-p_r)(1+p_r-4p_r\eta T)}{4np_r^2}, \tag{2.11}$$

and the MSE is written as:

$$MSE(\hat{T}_s) = \frac{T(1-T)}{n\eta} + \frac{(1-p_r)T^2\{1+p_r+4p_r\eta(1-T)\}}{4np_r^2\eta^2}. \tag{2.12}$$

3. Estimation of sensitive trait and sensitivity level in simple random sampling

3.1. Proposed model under simple random sampling

The proposed procedure differs from the models of [14, 15] in its RR mechanism. In this framework, a simple random sample of size n is drawn with replacement from a finite population of size N. First, the sampled respondent is required to answer a direct question regarding whether he/she possesses the sensitive attribute A.

If the response is "no", then the respondent is provided with an RD. In this RD, a respondent is instructed to answer "yes" if he/she belongs to the sensitive group A. If the respondent does not belong to the sensitive attribute A, then he/she is instructed to respond using the RD proposed by [23], which consists of the following statements: (a) "I am in group A", (b) "saying yes", and (c) "saying no", with corresponding probabilities p_r, $(1-p_r)/2$, and $(1-p_r)/2$, respectively.

It is assumed that respondents who belong to the non-sensitive group have no incentive to misreport and therefore respond truthfully, whether they answer directly or through the randomization procedure. Respondents who belong to group A are assumed to provide honest responses under the RR mechanism, while a proportion T of them follow the conventional direct-response strategy.

The respondents are instructed not to disclose their answers to the interviewer, thus ensuring privacy protection. In the proposed model, the simultaneous estimation of the sensitive proportion η and the truthful response probability T does not pose an identification problem, as the model is constructed from two independent observational equations, each corresponding to one of the two unknown parameters. These equations yield a unique solution under the specified design assumptions, thus ensuring parameter identifiability.

Let T be the probability of an honest response in the first direct procedure. The probability of a *yes* response to the direct response procedure proposed by the model is:

$$\theta_1 = \eta T, \tag{3.1}$$

and the likelihood of a *yes* answer in the *RR* procedure according to the suggested model, is represented by:

$$\theta_2 = \eta(1-T) + (1-\eta)\frac{(1-p_r)}{2} \Rightarrow \pi = \frac{\theta_1 + \theta_2 - \frac{(1-p_r)}{2}}{\left(\frac{1+p_r}{2}\right)}. \tag{3.2}$$

The method of moments requires the use of an estimator of the population proportion, η, which can be defined as follows:

$$\hat{\eta} = \frac{\hat{\theta}_1 + \hat{\theta}_2 - \frac{(1-p_r)}{2}}{(\frac{1+p_r}{2})}. \tag{3.3}$$

Additionally, the estimator of T is given below, where $\hat{\pi}$ is shown in Eq (2.3):

$$\hat{T} = \frac{\hat{\theta}_1}{\hat{\pi}} = \frac{\hat{\theta}_1(\frac{1+p_r}{2})}{\hat{\theta}_1 + \hat{\theta}_2 - \frac{(1-p_r)}{2}}, \tag{3.4}$$

where the observed proportion of the *yes* response reported by the respondents in case j is denoted by $\hat{\theta}_j (j = 1, 2)$.

The following theorem presents the principal properties of the estimator.

Theorem 3.1. *The estimator $\hat{\eta}$ is an unbiased estimator with the following variance:*

$$V(\hat{\eta}) = \frac{\eta(1-\eta)}{n} + \frac{(1-p_r)(1-\eta)}{n(1+p_r)}.$$

Proof. In view of the obvious result that $E(\hat{\theta}_1) = \theta_1$ and $E(\hat{\theta}_2) = \theta_2$, which are unbiased estimators, we have

$$V(\hat{\eta}) = E\left[\frac{\hat{\theta}_1 + \hat{\theta}_2 - \frac{(1-p_r)}{2}}{(\frac{1+p_r}{2})}\right] = \frac{\theta_1 + \theta_2 - \frac{(1-p_r)}{2}}{(\frac{1+p_r}{2})} = \eta. \tag{3.5}$$

Hence, we prove that the estimator $\hat{\eta}$ is unbiased an estimator of η.

The variance of the estimator $\hat{\eta}$ is given by the following:

$$V(\hat{\eta}) = \frac{\left[V(\hat{\theta}_1) + V(\hat{\theta}_2) + 2cov(\hat{\theta}_1, \hat{\theta}_2)\right]}{(\frac{1+p_r}{2})^2}.$$

We note that $V(\hat{\theta}_1) = \frac{\theta_1(1-\theta_1)}{n}$, $V(\hat{\theta}_2) = \frac{\theta_2(1-\theta_2)}{n}$, and $cov(\hat{\theta}_1, \hat{\theta}_2) = -\frac{\theta_1\theta_2}{n}$; thus, we have

$$\begin{aligned}
V(\hat{\eta}) &= \frac{1}{(\frac{1+p}{2})^2}\left[\frac{\theta_1(1-\theta_1)}{n} + \frac{\theta_2(1-\theta_2)}{n} - \frac{2\theta_1\theta_2}{n}\right] \\
&= \frac{1}{n(\frac{1+p_r}{2})^2}\left[\eta(\frac{1+p_r}{2})^2 - \eta^2(\frac{1+p_r}{2})^2 + \frac{(1-p_r)}{2}(\frac{1+p_r}{2}) - \eta(\frac{1+p_r}{2})\right] \\
&= \frac{\eta(1-\eta)}{n} + \frac{\frac{(1-p_r)}{2}(1-\eta)}{n(\frac{1+p_r}{2})} \\
&= \frac{\eta(1-\eta)}{n} + \frac{(1-p_r)(1-\eta)}{n(1+p_r)}.
\end{aligned} \tag{3.6}$$

This complete the proof.

Corollary 3.1. *The unbiased estimator of $V(\hat{\eta})$ is given by the following:*

$$\hat{V}(\hat{\eta}) = \frac{\hat{\eta}(1-\hat{\eta})}{n-1} + \frac{(1-p_r)(1-\hat{\eta})}{(n-1)(1+p_r)}.$$

To derive the MSE of the estimator $\hat{T}$, we write $d_1 = \hat{\theta}_1(\frac{1+p_r}{2})$ and $d_2 = \hat{\theta}_1 + \hat{\theta}_2 - \frac{(1-p_r)}{2}$; it follows that $E(d_1) = \frac{(1-p_r)}{2}E(\hat{\theta}_1)$ and $E(d_2) = \theta$. The estimator $\hat{T}$ can be written as $\hat{T} = \dfrac{d_1}{d_2}$ and we have $T = \dfrac{E(d_1)}{E(d_2)}$. Furthermore, we define the following terms:

$$e_1 = \frac{(d_1 - E(d_1))}{E(d_1)},$$

and

$$e_2 = \frac{(d_2 - E(d_2))}{E(d_2)}.$$

Assuming that $| e_1 |< 1$ so that the function $(1+e_2)^{-1}$ can be validly expanded as a power series. It can be easily provided that

$$E(e_1)^2 = \frac{\theta_1(1-\theta_1)}{n\eta^2 T^2},$$

$$E(e_2)^2 = \frac{\theta_1(1-\theta_1) + \theta_2(1-\theta_2) - 2\theta_1\theta_2}{n\eta^2(\frac{1+p_r}{2})^2},$$

$$E(e_1 e_2) = \frac{\theta_1(1-\theta_1) - \theta_1\theta_2}{n\eta^2 T(\frac{1+p_r}{2})}.$$

Additionally, the estimation error of the estimator $\hat{T}$ can be expressed as:

$$\hat{T} - T = \frac{T\eta(1+e_1)}{\eta(1+e_2)} - T, \tag{3.7}$$

$$\hat{T} - T = T(1+e_1)(1+e_2) - T,$$

while the high-order term error component

$$\hat{T} - T = T(e_1 - e_2) - T. \tag{3.8}$$

Theorem 3.2. *The MSE of the estimator $\hat{T}$, up to terms of order $0(n^{-1})$, is given by the following:*

$$MSE(\hat{T}) = \frac{T(1-T)}{n\eta} + \frac{(1-p_r)T^2(1-\eta)}{n\eta^2(1+p_r)}.$$

Proof. We consider the following:

$$\begin{aligned} MSE(\hat{T}) &= E\left(T - \hat{T}\right)^2 \\ &= T^2\left[E(e_1)^2 - 2E(e_1e_2) + E(e_2)^2\right] \\ &= T^2\left[\frac{\theta_1(1-\theta_1)}{n\eta^2T^2} - \frac{2\{\theta_1(1-\theta_1) - \theta_1\theta_2\}}{n\eta^2 T(\frac{1+p_r}{2})^2} + \frac{\theta_1(1-\theta_1) + \theta_2(1-\theta_2) - 2\theta_1\theta_2}{n\eta^2(\frac{1+p_r}{2})^2}\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n\eta^2(\frac{1+p_r}{2})^2}\left[\left\{\theta_1\left((\frac{1+p_r}{2})-T\right)^2+T^2\theta_2\right\}+\left\{\theta_1\left((\frac{1+p_r}{2})-T\right)+T\theta_2\right\}^2\right]\\
&= \frac{1}{n\eta^2(\frac{1+p_r}{2})^2}\left[\eta T(\frac{1+p_r}{2})^2+\eta T^2(\frac{1+p_r}{2})+\frac{(1-p_r)}{2}T^2(\frac{1+p_r}{2})\right]\\
&= \frac{T(1-T)}{n\eta}-\frac{\frac{(1-p_r)}{2}T^2}{n\eta(\frac{1+p_r}{2})}+\frac{\frac{(1-p_r)}{2}T^2}{n\eta^2(\frac{1+p_r}{2})}\\
&= \frac{T(1-T)}{n\eta}+\frac{\frac{(1-p_r)}{2}T^2(1-\eta)}{n\eta^2(\frac{1+p_r}{2})}\\
&= \frac{T(1-T)}{n\eta}+\frac{(1-p_r)\,T^2\,(1-\eta)}{n\eta^2\,(1+p_r)}. \qquad (3.9)
\end{aligned}
$$

This complete the proof.

3.2. Numerical comparison based on theoretical MSE expressions

In this section, we obtained the conditions under which our proposed estimators $\hat{\eta}$ and $\hat{T}$ are preferable over existing estimators. Here, we present the numerical results using Eqs (3.14)–(3.18) by considering fixed values of the parameters n, T,η, and p_r to demonstrate the superiority of the proposed estimators over the existing estimators. We employ these theoretical expressions because previous studies, such as Huang et al. [14–16], also used theoretical comparisons to demonstrate the superiority of their proposed estimators over the existing estimators. Therefore, the performance of the proposed estimators in this study is evaluated through theoretical *PRE* comparisons under different parameter settings.

The proposed estimator $\hat{\eta}$ is more precise than [2] $\hat{\eta}_{wa}$ if

$$\eta > 1-\frac{p_r\,(1+p_r)}{(2p_r-1)^2}. \qquad (3.10)$$

The value of η should be lie between 0 and 1; thus the inequality given in (11) must be lie between 0 and 1, that is, we must have

$$0<\frac{p_r\,(1+p_r)}{(2p_r-1)^2}<1, \quad p_r \neq \frac{1}{2}.$$

The proposed estimator $\hat{\eta}$ will be more efficient than the [14] estimator $\hat{\eta}_h$ if

$$\pi < \frac{p_r\,(1+p_r)-(2p_r-1)^2}{p_rT\,(1+p_r)-(2p_r-1)^2}. \qquad (3.11)$$

The proposed estimator $\hat{T}$ is not less efficient than the [14] estimator $\hat{T}_h$ if

$$\pi < \frac{p_r\,(1+p_r)-(2p_r-1)^2}{p_rT\,(1+p_r)-(2p_r-1)^2}. \qquad (3.12)$$

To determine the performances of the suggested estimators $\hat{\eta}$ and $\hat{T}$, the percent relative efficiencies (*PRE*) of the modified estimators with respect to the [14–16] estimators are calculated. The (*PRE*) of the proposed estimators $(\hat{\eta},\hat{T})$ with respect to the [14] estimators $(\hat{\eta}_h,\hat{T}_h)$, the [15] estimators $(\hat{\eta}_s,\hat{T}_s)$, and the [16] estimators $(\hat{\eta}_k, \hat{T}_k)$ are computed using the following formulae for different cases of p_r, η, and T, respectively. The findings are presented in Tables 1–6.

$$PRE\left(\hat{\eta}, \hat{\eta}_h\right) = \frac{(1+p_r)\left[(2p_r-1)^2 \eta (1-\eta) + p_r (1-p_r)(1-\eta T)\right]}{(2p_r-1)^2 (1-\eta)\left[\eta(1+p_r) + (1-p_r)\right]} \times 100, \tag{3.13}$$

$$PRE\left(\hat{T}, \hat{T}_h\right) = \frac{(1+p_r)\left[(2p_r-1)^2 \eta T (1-T) + p_r (1-p_r) T^2 (1-\eta T)\right]}{(2p_r-1)^2 \left[(1+p_r) T (1-T)\eta + (1-p_r) T^2 (1-\eta)\right]} \times 100, \tag{3.14}$$

$$PRE\left(\hat{\eta}, \hat{\eta}_s\right) = \frac{(1+p_r)\left[4p_r^2 \eta (1-\eta) + (1-p_r)(1+p_r-4p_r\eta T)\right]}{4\eta^2 (1-\eta)\left[\eta(1+p_r) + (1-p_r)\right]} \times 100, \tag{3.15}$$

$$PRE\left(\hat{T}, \hat{T}_s\right) = \frac{(1+p_r)\left[4p_r^2 \eta T (1-T) + (1-p_r)\left\{1+p_r+4p_r\eta(1-T)\right\}\right]}{(2p_r-1)^2 \left[(1+p_r) T (1-T)\eta + (1-p_r) T^2 (1-\eta)\right]} \times 100, \tag{3.16}$$

$$PRE\left(\hat{\eta}, \hat{\eta}_k\right) = \frac{(1+p_r)\left[w^2 \eta (1-\eta) + (1-p_{r0})(1-p_r)\left\{1-(1-p_{r0})(1-p_r)\right\} - p_r \eta T (1-p_r)\right]}{(1-p_r) w^2 (1-\eta)\left[\eta(1+p_r) + (1-p_r)\right]} \times 100, \tag{3.17}$$

$$PRE\left(\hat{T}, \hat{T}_k\right) = \frac{(1+p_r)\left[w^2 \eta T (1-T) + (1-p_{r0})(1-p_r) T^2 \left\{1-(1-p_{r0})(1-p_r)\right\} + (1-p_r)\eta T^2 (2p_{r0}w - p_r T)\right]}{(2p_r-1)^2 \left[(1+p_r) T (1-T)\eta + (1-p_r) T^2 (1-\eta)\right]} \times 100. \tag{3.18}$$

From Tables 1–6, it is evident that the *PREs* of the proposed estimators are greater than 100, indicating that the proposed estimators $\hat{\eta}$ and $\hat{T}$ are more efficient than the estimators $\hat{\pi}_h$ and $\hat{T}_h$ proposed by [14], the estimators $\hat{\pi}_s$ and $\hat{T}_s$ proposed by [15], and the estimators $\hat{\eta}_k$ and $\hat{T}_k$ proposed by [16].

From Table 1, it can be further observed that the *PRE* of the proposed estimator $\hat{\eta}_k$ decreases as T increases for a fixed combination of p_r and η. Moreover, the *PRE* of the proposed estimator $\hat{\eta}$ is relatively higher when the values of p_r and η are closer to 0.5 for a fixed combination of T and η.

Table 1. Percent relative efficiency of the proposed estimator $\hat{\eta}$ with respect to the [14] estimator $\hat{\eta}_h$.

		η								
p_r	T	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	1914.29	1677.78	1566.23	1538.46	1586.67	1729.41	2031.58	2704.76	4826.09
	0.3	1876.19	1611.11	1472.73	1415.39	1426.67	1517.65	1736.84	2247.62	3886.96
0.6	0.5	1838.10	1544.44	1379.22	1292.31	1266.67	1305.88	1442.11	1790.48	2947.83
	0.7	1800.00	1477.78	1285.71	1169.23	1106.67	1094.12	1147.37	1333.33	2008.70
	0.9	1761.90	1411.11	1192.21	1046.15	946.67	882.35	852.63	876.19	1069.57
	0.1	558.38	480.20	444.68	433.67	442.55	474.50	544.09	700.23	1193.14
	0.3	547.83	462.77	421.06	403.32	403.75	423.79	474.20	592.70	973.67
0.7	0.5	537.28	445.34	397.45	372.96	364.95	373.08	404.32	485.17	754.20
	0.7	526.73	427.91	373.84	342.60	326.14	322.37	334.44	377.64	534.73
	0.9	516.18	410.47	350.23	312.24	287.34	271.66	264.56	270.11	315.27
	0.1	278.95	239.29	222.78	217.39	220.00	231.25	256.16	312.20	489.01
	0.3	274.27	232.14	213.51	205.80	205.45	212.50	230.59	273.17	409.89
0.8	0.5	269.59	225.00	204.25	194.20	190.91	193.75	205.02	234.15	330.77
	0.7	264.91	217.86	194.98	182.61	176.36	175.00	179.45	195.12	251.65
	0.9	260.23	210.71	185.71	171.01	161.82	156.25	153.88	156.10	172.53

From Table 2, it is observed that the *PRE* of the proposed estimator $\hat{T}$ increases as the value of T increases for fixed values of p_r and η. On the other hand, the *PRE* of the proposed estimator $\hat{T}$ decreases with an increase in η for fixed values of p_r and T. Furthermore, the *PRE* of the proposed estimator $\hat{T}$ is relatively higher when the value of $p_r$is close to 0.5 for fixed values of η and T.

Table 2. Percent relative efficiency of the proposed estimator $\hat{\eta}$ with respect to the [15] estimator $\hat{\eta}_h$.

		η								
p_r	T	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	608.00	384.00	296.35	249.60	220.54	200.73	186.35	175.45	166.89
	0.3	1320.73	916.00	704.00	573.54	485.16	421.33	373.07	335.30	304.94
0.6	0.5	1784.62	1400.00	1136.84	945.45	800.00	685.71	593.55	517.65	454.05
	0.7	2099.20	1836.00	1603.76	1397.33	1212.63	1046.40	896.00	759.27	634.43
	0.9	2317.18	2224.00	2118.40	1997.71	1858.46	1696.00	1504.00	1273.60	992.00
	0.1	207.72	158.99	140.71	131.14	125.25	121.26	118.38	116.20	114.50
	0.3	384.15	279.75	230.03	200.95	181.87	168.39	158.35	150.59	144.41
0.7	0.5	520.38	404.85	334.24	286.63	252.34	226.48	206.27	190.05	176.74
	0.7	626.48	535.26	462.29	402.59	352.84	310.74	274.66	243.38	216.02
	0.9	709.38	672.27	632.05	588.35	540.67	488.45	431.00	367.51	296.97
	0.1	134.00	118.35	112.72	109.82	108.05	106.86	106.00	105.35	104.85
	0.3	199.33	159.20	142.00	132.44	126.36	122.15	119.07	116.71	114.84
0.8	0.5	261.11	207.69	179.41	161.90	150.00	141.38	134.85	129.73	125.61
	0.7	319.33	268.00	232.46	206.40	186.47	170.74	158.00	147.48	138.64
	0.9	374.00	348.00	322.00	296.00	270.00	244.00	218.00	192.00	166.00

Table 3 shows that the *PRE* of the proposed estimator $\hat{\eta}$ increases as η becomes larger, but decreases as T increases. Hence, a high gain in efficiency using the proposed estimators $\hat{\eta}$ and $\hat{T}$ is observed when p_r is close to 0.5. From Table 4, it is observed that the *PRE* of the proposed estimator $\hat{\eta}$ decreases as T increases for a specific combination of p_r and η. Furthermore, the *PRE* of the proposed estimator $\hat{\eta}$ is relatively higher when p_r is closer to 0.5 for a fixed combination of T and η.

The higher *PRE* observed when p_r is close to 0.5 is mainly attributed to the variance behavior of the estimator. At moderate values of p_r, the randomization mechanism achieves an optimal balance between privacy protection and information efficiency, thus reducing the variance and improving the estimator's efficiency. As a result, the *PRE* attains higher values in the neighborhood of $p_r = 0.5$.

Table 3. Percent relative efficiency of the proposed estimator $\hat{\eta}$ with respect to the [16] estimator $\hat{\eta}_k$.

		$p_{r0} = .2$				η	
p_r	T	0.1	0.2	0.3	0.4	0.5	0.6
0.2	0.2	433.04	435.16	447.24	471.17	511.66	579.05
	0.4	427.12	423.39	429.14	445.66	476.68	530.72
	0.6	421.21	411.62	411.05	420.15	441.69	482.38
	0.8	415.29	399.84	392.95	394.64	406.71	434.05
0.3	0.2	2460.71	2426.92	2461.25	2572.22	2785.71	3161.4
	0.4	2428.5	2362.82	2362.73	2433.33	2595.24	2898.25
	0.6	2396.3	2298.72	2264.2	2294.44	2404.76	2635.09
	0.8	2364.09	2234.61	2165.68	2155.56	2214.29	2371.93
0.4	0.2	22331.88	21946.15	22198.52	23162.5	25071.43	28468.42
	0.4	22042.03	21369.23	21311.82	21912.5	23357.14	26100
	0.6	21752.17	20792.31	20425.12	20662.5	21642.86	23731.58
	0.8	21462.32	20215.38	19538.42	19412.5	19928.57	21363.16
		$p_0 = .4$					
0.2	0.2	32181.08	29861.36	28864.71	29013.79	30423.08	33600
	0.4	31550.45	28668.18	27100	26600	27192.31	29225
	0.6	30919.82	27475	25335.29	24186.21	23961.54	24850
	0.8	30289.19	26281.82	23570.59	21772.41	20730.77	20475
0.3	0.2	1979.79	1849.56	1796.69	1811.48	1901.44	2097.72
	0.4	1940.37	1774.98	1686.4	1660.61	1699.52	1824.28
	0.6	1900.96	1700.41	1576.1	1509.75	1497.6	1550.85
	0.8	1861.54	1625.84	1465.81	1358.89	1295.67	1277.41
0.4	0.2	623.81	591.88	581.39	590.15	620.88	683.33
	0.4	610.94	567.53	545.38	540.89	554.95	594.05
	0.6	598.07	543.18	509.36	491.63	489.01	504.76
	0.8	585.2	518.83	473.35	442.36	423.08	415.48

Table 4. Percent relative efficiency of the proposed estimator $\hat{T}$ with respect to the [14] estimator $\hat{T}_h$.

		η								
p_r	T	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	168.25	165.43	166.52	170.94	179.26	193.46	218.71	270.90	429.95
	0.3	165.43	160.49	159.60	161.82	167.41	177.78	196.88	237.04	360.39
0.2	0.5	162.61	155.56	152.67	152.71	155.56	162.09	175.05	203.17	290.82
	0.7	159.79	150.62	145.74	143.59	143.70	146.41	153.22	169.31	221.26
	0.9	156.97	145.68	138.82	134.47	131.85	130.72	131.38	135.45	151.69
	0.1	292.06	287.61	290.81	301.72	322.40	358.20	422.55	556.54	966.73
	0.3	286.54	277.17	275.37	280.60	294.13	319.92	368.28	471.08	788.91
0.3	0.5	281.02	266.73	259.93	259.48	265.87	281.64	314.00	385.61	611.09
	0.7	275.51	256.29	244.49	238.36	237.60	243.36	259.73	300.15	433.27
	0.9	269.99	245.85	229.04	217.24	209.33	205.08	205.46	214.68	255.44
	0.1	873.91	871.15	891.13	937.50	1020.00	1160.53	1412.20	1936.36	3542.55
	0.3	856.52	836.54	837.93	862.50	917.14	1018.42	1207.32	1609.09	2853.19
0.4	0.5	839.13	801.92	784.73	787.50	814.29	876.32	1002.44	1281.82	2163.83
	0.7	821.74	767.31	731.53	712.50	711.43	734.21	797.56	954.55	1474.47
	0.9	804.35	732.69	678.33	637.50	608.57	592.11	592.68	627.27	785.11

From Table 5, it is observed that the PRE of the proposed estimator $\hat{T}$ increases with an increase in T for fixed values of p_r and η. However, the PRE decreases as η increases for fixed values of p_r and T. Moreover, the PRE of the proposed estimator $\hat{T}$ is relatively higher when the value of p_r is close to 0.5 for fixed values of η and T.

Table 5. Percent relative efficiency of the proposed estimator $\hat{T}$ with respect to the [15] estimator $\hat{T}_h$.

		η								
p_r	T	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.1	119.11	111.78	108.91	107.38	106.43	105.78	105.31	104.95	104.67
	0.3	144.97	132.67	126.22	122.26	119.57	117.63	116.16	115.01	114.09
0.2	0.5	160.68	150.00	142.69	137.37	133.33	130.16	127.60	125.49	123.72
	0.7	170.31	163.78	158.01	152.89	148.30	144.18	140.44	137.05	133.95
	0.9	176.00	174.00	171.73	169.14	166.15	162.67	158.55	153.60	147.56
	0.1	171.06	144.03	132.44	126.00	121.90	119.07	116.99	115.40	114.14
	0.3	243.35	210.05	189.44	175.42	165.27	157.58	151.55	146.70	142.71
0.3	0.5	277.30	254.44	235.94	220.65	207.81	196.88	187.45	179.23	172.02
	0.7	294.81	283.38	271.94	260.50	249.06	237.63	226.19	214.75	203.31
	0.9	303.78	300.89	297.44	293.26	288.11	281.58	273.03	261.39	244.56
	0.1	456.00	329.14	269.05	234.00	211.03	194.82	182.77	173.45	166.04
	0.3	726.40	610.67	528.00	466.00	417.78	379.20	347.64	321.33	299.08
0.4	0.5	828.57	763.64	704.35	650.00	600.00	553.85	511.11	471.43	434.48
	0.7	874.67	847.38	817.92	786.00	751.30	713.45	672.00	626.40	576.00
	0.9	895.27	889.60	882.67	874.00	862.86	848.00	827.20	796.00	744.00

Furthermore, Table 6 indicates that the *PRE* of the proposed estimator $\hat{\eta}$ increases as the value of T increases, whereas it decreases with an increase in η. Hence, a substantial efficiency gain is observed when using the proposed estimators $\hat{\pi}$ and $\hat{T}$ when p is close to 0.5.

Table 6. Percent relative efficiency of the proposed estimator $\hat{T}$ with respect to the [16] estimator $\hat{T}_k$.

		$p_{r0} = .2$		η			
p_r	T	0.1	0.2	0.3	0.4	0.5	0.6
0.2	0.2	320.95	261.02	225.06	201.09	183.97	171.12
	0.4	390.07	349.47	316.25	288.57	265.15	245.07
	0.6	421.22	401.63	382.04	362.45	342.86	323.27
	0.8	437.2	433.06	428.29	422.72	416.14	408.24
0.3	0.2	1699.26	1265.56	1005.33	831.85	707.94	615
	0.4	2207.65	1920.44	1685.45	1489.63	1323.93	1181.9
	0.6	2443.33	2320	2196.67	2073.33	1950	1826.67
	0.8	2569.93	2573.65	2577.95	2582.96	2588.89	2596
0.4	0.2	15606.67	11710	9372	7813.33	6700	5865
	0.4	20553.33	18276	16412.73	14860	13546.15	12420
	0.6	22860	22320	21780	21240	20700	20160
	0.8	24110.67	24922.86	25860	26953.33	28245.45	29796
		$p_{r0} = .4$					
0.2	0.2	19687.27	13420	10137.14	8116.92	6748.39	5760
	0.4	28428	23113.33	19317.14	16470	14255.56	12484
	0.6	33172.63	30268	27640	25250.91	23069.57	21070
	0.8	36006.09	35576.36	35105.71	34588	34015.79	33380
0.5	0.2	1265.4	907.3	719.73	604.3	526.11	469.64
	0.4	1784.25	1493.54	1285.89	1130.16	1009.03	912.12
	0.6	2064.9	1925.03	1798.48	1683.44	1578.4	1482.11
	0.8	2231.63	2243.98	2257.5	2272.37	2288.82	2307.08
0.8	0.2	434.95	340.92	291.66	261.35	240.82	225.99
	0.4	583.1	514.69	465.83	429.18	400.68	377.88
	0.6	662.43	641.47	622.51	605.27	589.53	575.1
	0.8	708.84	734.06	761.69	792.08	825.67	862.99

4. Estimation of sensitive trait and sensitivity level in stratified random sampling

In this section, we discuss the application of the proposed RR model under stratified random sampling. Consider a population that is divided into L non-overlapping strata, where the size of the i^{th} stratum is N_i, known in advance. Let N denote the total population size, and define $w_i = Ni/Ni$ as the proportion of units that belong to the i^{th} stratum. From each stratum, a simple random sample with replacement ($SRSWR$) of size n_i is selected such that $\sum_i^k n_i = n$. Then, the sampled individuals are then asked to respond with *yes* or *no* according to the specified randomization mechanism.

[24] were the first to propose a stratified (RR) technique under proportional allocation. Later, [19] proposed an optimal allocation strategy within stratified random sampling, thereby demonstrating that it is more efficient and cost-effective than the proportional allocation approach of [24]. Building on this

advancement, [25] applied the [19] stratified *RRT* to the two-stage *RR* procedure of [4] and investigated the properties of their proposed model under optimal allocation. The mixed *RR* model suggested by [25] was later extended to stratified random sampling. [26] modified the stratified *RR* model for the unrelated question *RR* technique and developed its properties based on the *Redberg 1969* framework. Subsequently, the [14] model in stratified random sampling using optimal allocation—which was more efficient than the [25] stratified sampling model—was further extended by [20]. [16] modified the Huang model and also extended it to stratified random sampling.

To estimate the proportion of sensitive issues such as *HIV*/*AIDS* status across different levels of urban and rural areas, age groups, or income groups, the extension of *RR* models to stratified random sampling can be highly useful. Motivated by these real-world applications and prior studies, the proposed *RR* model can be extended to stratified random sampling. The description and properties of the proposed *RR* model are discussed below.

4.1. Proposed model under stratified random sampling

In this proposed model, the sample is drawn using SRSWR from each stratum after the finite population has been partitioned into strata. This study considers the assumption that the number of units in each stratum is known as an advantage of stratification. Initially, the from stratum i, each respondent is directly asked whether they belong to group A. When the answer is *no*, the individual is provided with an RD. Under this R_{1i}, each interviewee is advised to report *yes* if he/she is a participant of the sensitive trait A. If the respondent is not in the sensitive group A, they are instructed to provide their answer using the RD R_{2i} proposed by [23]. This device contains three possible instructions: (a) 'I belong a sensitive group A', (b) to say *yes*, and (c) to say *no*, with preassigned probability p_{ri}, $(1 - p_{ri})/2$ and $(1 - p_{ri})/2$. Different RDs with preassigned probabilities are used to assign respondents to strata within the sample. It is required that respondents in the non-sensitive group always tell the truth, whether they use the direct response method or a randomization technique. It is assumed that interviewees in group A answer truthfully when using the randomization mechanism; however, following standard practice, the probability T_i denotes the direct questioning procedure in the i^{th} stratum. The respondents are conditioned to conceal their answers from the interviewer. It must be noted that in the suggested model, since the optimal allocation depends on unknown parameters η_i and T, a two-stage adaptive scheme can be used. First, a pilot sample is selected from each stratum to obtain preliminary estimates of η_i and T_i. Then, these estimates are used in the Neyman allocation formulas to determine the second-stage sample sizes. This approach improves the allocation efficiency and reduces the variance of the estimators.

In stratum i, T_i denotes the probability of an honest response to the direct question. Under the proposed technique, the probability of a "yes" response in the direct personal investigation can be defined as:

$$\theta_{1i} = \eta_{si} T_i, \tag{4.1}$$

and the probability of a "yes" reply in the *RR* method in a stratum i under the suggested model is given by:

$$\theta_{2i} = \eta_{si}(1 - T_i) + (1 - \eta_{si})\frac{(1 - p_{ri})}{2} \Rightarrow \eta_{si} = \frac{\theta_{1i} + \theta_{2i} - \frac{(1-p_{ri})}{2}}{\left(\frac{1+p_{ri}}{2}\right)}, \tag{4.2}$$

where η_{si} represents, in stratum i, the proportion of respondents who possess the sensitive attribute, and

p_{ri} denotes the probability that an interviewee in sample stratum i receives a stigmatizing question (s) card in the first question.

η_{si} and T_i are the estimators of the proposed model and are defined as follows:

$$\hat{\eta}_{si} = \frac{\hat{\theta}_{1i} + \hat{\theta}_{2i} - \frac{\left(1-p_{ri}\right)}{2i}}{\left(\frac{1+p_{ri}}{2}\right)}, \tag{4.3}$$

and

$$\hat{T}_i = \frac{\hat{\theta}_{1i}\left(\frac{1+p_i}{2}\right)}{\hat{\theta}_{1i} + \hat{\theta}_{2i} - \frac{\left(1-p_i\right)}{2}}, respectively, \tag{4.4}$$

where $\hat{\theta}_{ji}$ $(j = 1, 2)$ represents the perceived proportion of "yes" responses reported by the interviewee. The unbiased estimator $\hat{\eta}_{si}$ of the variance can be defined as follows:

$$V(\hat{\eta}_{si}) = \frac{\left[V(\hat{\theta}_{1i}) + V(\hat{\theta}_{2i}) + 2cov(\hat{\theta}_{1i}, \hat{\theta}_{2i})\right]}{\left(\frac{1+p_{ri}}{2}\right)^2}. \tag{4.5}$$

Since the selections are independently made across strata, an estimator for the whole population is obtained by aggregating the estimators from the individual strata. The estimator of $\eta_s = \sum_i^k w_i\eta_{si}$, which is the proportion of respondents who belongs, to the sensitive group, is given by the following:

$$\hat{\eta}_s = \sum_i^k w_i\left[\frac{\hat{\theta}_1 + \hat{\theta}_2 - \frac{\left(1-p_{ri}\right)}{2}}{\left(\frac{1+p_{ri}}{2}\right)}\right], \tag{4.6}$$

where N represents the total number of units in the population, and N_i denotes the total number of units in stratum i. The weight for stratum i is defined as $w_i = (Ni/N)$ for $(i = 1, 2, ...k)$, such that $w = \sum_i^k w_i = 1$.

Theorem 4.1. *The estimator $\hat{\eta}_s$ provides an unbiased estimate of the population proportion η_s.*

Proof. The unbiasedness of $\hat{\eta}_{si}$ follows from $E(\hat{\theta_{ji}}) = \theta_{ji}$, $(i = 1, 2, 3...k)$. Thus, the unbiasedness of $\hat{\eta}_s$ follows from taking the expected value of (2.5).

Theorem 4.2. *The variance of the estimator $\hat{\eta}_s$ is:*

$$V(\hat{\eta}_s) = \sum_i^k w_i^2\left[\frac{\eta_{si}(1-\eta_{si})}{n_i} + \frac{(1-p_{ri})(1-\eta_{si})}{n_i(1+p_{ri})}\right]. \tag{4.7}$$

We know that, under stratified random sampling, the variance of the [20] estimator $\hat{\eta}_{ts}$ is written as as follows:

$$V(\hat{\eta}_{ts}) = \sum_i^k w_i^2\left[\frac{\eta_{si}\,(1-\eta_{si})}{n_i} + \frac{p_{ri}\,(1-p_{ri})\,(1-\eta_{si}T_i)}{n_i\,(2p_{ri}-1)^2}\right]. \tag{4.8}$$

Using Eqs (26) and (27), we obtain the following:

$$V(\hat{\eta}_{ts}) - V(\hat{\eta}_s) = \left[\frac{p_{ri}\,(1-p_{ri})\,(1-\eta_{si}T_i)}{n_i\,(2p_{ri}-1)^2} - \frac{(1-p_{ri})(1-\eta_{si})}{n_i(1+p_{ri})}\right] > 0. \tag{4.9}$$

This demonstrates the superiority of the proposed technique over that of [20]. To obtain an unbiased estimator of V($\hat{\eta}_s$), we proceed as follows.
Corollary 4.1. *An unbiased estimator of the $V(\hat{\eta}_s)$ is defined as follows:*

$$\hat{V}(\hat{\eta}_s) = \hat{V}\Big(\sum_i^k w_i^2 \hat{\eta}_{si}\Big) = \sum_i^k w_i^2 \hat{V}(\hat{\eta}_{si}), \tag{4.10}$$

where

$$\hat{V}(\hat{\eta}_{si}) = \frac{\eta_{si}(1-\eta_{si})}{n_i} + \frac{(1-p_{ri})(1-\eta_{si})}{n_i(1+p_{ri})}. \tag{4.11}$$

Theorem 4.3. *To derive the minimum variance of estimator $\hat{T}_s$, the optimal allocation of $n_1, n_2 ... n_{k-1}$, and n_k subject to a fixed $n = \sum_i^k n_i$ is given by the following:*

$$n_i = \frac{n w_i \Big[\frac{\eta_{si}(1-\eta_{si})}{n_i} + \frac{(1-p_{ri})(1-\eta_{si})}{n_i(1+p_{ri})}\Big]^{\frac{1}{2}}}{\sum_i^k w_i \Big[\frac{\eta_{si}(1-\eta_{si})}{n_i} + \frac{(1-p_{ri})(1-\eta_{si})}{n_i(1+p_{ri})}\Big]^{\frac{1}{2}}}. \tag{4.12}$$

Proof. Follows from Section 5.5 *of Cochran* (1977).

The minimal variance of the estimator $\hat{\eta}_s$ is given by:

$$V(\hat{\eta}_s) = \frac{1}{n}\Big\{\sum_i^k w_i \Big[\frac{\eta_{si}(1-\eta_{si})}{n_i} + \frac{(1-p_{ri})(1-\eta_{si})}{n_i(1+p_{ri})}\Big]^{\frac{1}{2}}\Big\}^2. \tag{4.13}$$

The unbiased minimal variance estimator of $\hat{\eta}$ is obtained from (3.11) using (3.13).

To obtain the MSE of the estimator $\hat{T}_i$, we express it as $d_{1i} = \hat{\theta}_{1i}(1 - \frac{(1-p_{ri})}{2})$ and $d_{2i} = \hat{\theta}_1 + \hat{\theta}_2 - \frac{(1-p_{ri})}{2}$; it follows that $E(d_{1i}) = (1 - \frac{(1-p_{ri})}{2})\eta_i T_i$ and $E(d_{1i}) = \theta$. The estimator $\hat{T}_i$ can be obtain as $\hat{T}_i = \dfrac{d_{1i}}{d_{2i}}$ and we have $T_s = \dfrac{E(d_{1i})}{E(d_{2i})}$. We further introduce the following terms:

$$e_{1i} = \frac{(d_{1i} - E(d_{2i})}{E(d_{1i})},$$

and

$$e_{2i} = \frac{(d_{1i} - E(d_{2i}))}{E(d_{2i})}.$$

The function $(1 + e_{2i})^{-1}$ can be validly expanded as a power series by assuming that $| e_{1i} |< 1$. After some straightforward simplification,

$$E(e_{1i})^2 = \frac{\theta_{1i}\,(1-\theta_{1i})}{n_i \eta_i^2 T_i^2},$$

$$E(e_{2i})^2 = \frac{\theta_{1i}(1-\theta_{1i}) + \theta_{2i}(1-\theta_{2i}) - 2\theta_{1i}\theta_{2i}}{n_i\eta_i^2(\frac{1+p_{ri}}{2})^2},$$

$$E(e_{1i}e_{2i}) = \frac{\theta_{1i}(1-\theta_{1i}) - \theta_{1i}\theta_{2i}}{n_i\eta_i^2 T(\frac{1+p_{ri}}{2})}.$$

Additionally, the estimation error of the estimator $\hat{T}$ can be written as follows:

$$\hat{T}_i - T_i = \frac{T_i\eta_{si}(1+e_{1i})}{\eta_{si}(1+e_{2i})} - T_i, \tag{4.14}$$

$$\hat{T}_i - T_i = T_i(1+e_{1i})(1+e_{2i}) - T_i.$$

After ignoring the high-order term error component

$$\hat{T}_i - T_i = T_i(e_{1i} - e_{2i}) - T_i, \tag{4.15}$$

using (2.8), we obtain the following theorem.

Theorem 4.4. *The MSE of the estimator $\hat{T}_i$, up to terms of order $0(n^{-1})$ is given by the following:*

$$MSE(\hat{T}_i) = \frac{T_i(1-T_i)}{n_i\eta_{si}} + \frac{(1-p_{ri})T_i^2(1-\eta_{si})}{n_i\eta_{si}^2(1+p_{ri})}. \tag{4.16}$$

Suppose the weighted probability is denoted by T_s and is defined as $T_s = \sum_i^k T_i wi$, where T_i is the probability of truthful reporting of the sensitive trait in stratum i under direct questioning. We obtain an estimate of T_s using Eq (3.16) in $T_s = \sum_i^k T_i wi$:

$$\hat{T}_s = \sum_i^k w_i\hat{T}_i = \sum_i^k w_i\Big[\frac{\hat{\theta}_1(1-\frac{(1-p_{ri})}{2})}{\hat{\theta}_{2i} + \hat{\theta}_{1i} - \frac{(1-p_{ri})}{2})}\Big]. \tag{4.17}$$

We derived the MSE of the estimator $\hat{T}_S$ up to order $(n-1)$ as follows:

$$MSE(\hat{T}_s) = \frac{1}{n_i\eta_i}\Big\{\sum_i^k w_i^2\Big[\frac{T_i(1-T_i)}{n_i\eta_i} + \frac{(1-p_{ri})T_i^2(1-\eta_i)}{n_i\eta_i^2(1+p_{ri})}\Big]\Big\}. \tag{4.18}$$

Theorem 4.5. *To derive the minimum MSE M of the estimator $\hat{T}_s$, subject to $n = \sum_i^k n_i$, the optimal allocation of n to $n_1, n_2 ... n_{k-1}$, and n_k is given by the following:*

$$n_i = \frac{nw_i\Big[\frac{T_i(1-T_i)}{n_i\eta_{si}} + \frac{(1-p_{ri})T_i^2(1-\eta_{si})}{n_i\eta_{si}^2(1+p_{ri})}\Big]^{\frac{1}{2}}}{\sum_i^k w_i\Big[\frac{T_i(1-T_i)}{n_i\eta_{si}} + \frac{(1-p_{ri})T_i^2(1-\eta_{si})}{n_i\eta_{si}^2(1+p_{ri})}\Big]^{\frac{1}{2}}}. \tag{4.19}$$

The MSE of the estimator $\hat{T}_s$, following Section 5.5 of [27], is given by the following:

$$MSE(\hat{T}_s) = \frac{1}{n}\Big\{\sum_i^k w_i\Big[\frac{T_i(1-T_i)}{n_i\eta_{si}} + \frac{(1-p_{ri})T_i^2(1-\eta_{si})}{n_i\eta_{si}^2(1+p_{ri})}\Big]^{\frac{1}{2}}\Big\}^2. \tag{4.20}$$

4.2. Efficiency comparison

In this subsection a numerical comparison of the proposed estimators $(\hat{\eta}_s, \hat{T}_s)$ with the [19] randomized response estimator $\hat{\eta}_k$, [20] estimators $(\hat{\eta}_{ts}, \hat{T}_{ts})$, and the estimators of [16] $(\hat{\eta}_k, \hat{T}_k)$ is performed using optimal allocation under stratified random sampling.The empirical study is preformed using $V(\hat{\eta}_s)/V(\hat{\eta}_{km})$, $V(\hat{\eta}_s)/V(\hat{\eta}_{ts})$, $V(\hat{T}_s)/V(\hat{T}_{ts})$, and $V(\hat{\eta}_s)/V(\hat{\eta}_k)$, $V(\hat{T}_s)/V(\hat{T}_k)$ due to difficulty of derivation mathematically $V(\hat{\eta}_s)/V(\hat{\eta}_{km})$, $V(\hat{\eta}_s)/V(\hat{\eta}_{ts})$, $V(\hat{T}_s)/V(\hat{T}_{ts})$,$V(\hat{\eta}_s)/V(\hat{\eta}_k)$, and $V(\hat{T}_s)/V(\hat{T}_k)$. Using Eqs (4.21)–(4.25) for different values of $p_{ro}, p_r, , p_{r1}, p_{r2}, T, T_1, T_2, w_1, w_2, \eta_1, \eta_{r2}$, and assuming two strata in the population $(k = 2)$, $T = T_1 = T_2$, and $p_{ro1} = p_{ro2} = p_{ro}$, the $PREs$ of the proposed $\hat{\eta}_s$ and $\hat{T}_s$ are computed with respect to the [19] estimator $\hat{\eta}_{km}$ and [20] estimators $\hat{\eta}_{ts}$ and $\hat{T}_{ts}$ through variance comparison:

$$PRE(\hat{\eta}_s, \hat{\eta}_{kw}) = \frac{\left(w_1\sqrt{B_1} + w_2\sqrt{B_2}\right)^2}{\left(w_1\sqrt{A_1} + w_2\sqrt{A_2}\right)^2}, \tag{4.21}$$

where $B_1 = \eta_1(1-\eta_1) + \frac{p_{r1}(1-p_r)}{(2p_{r1}-1)^2}$, $B_1 = \eta_2(1-\eta_2) + \frac{p_{r2}(1-p_{r2})}{(2p_{r2}-1)^2}$, $A_1 = \eta_1(1-\eta_1) + \dfrac{(1-p_{r1})(1-\eta_1)}{(1+p_{r1})^2}$, $A_2 = \eta_1(1-\eta_2) + \dfrac{(1-p_{r2})(1-\eta_2)}{(1+p_{r2})^2}$,

$$PRE(\hat{\eta}_s, \hat{\eta}_{ts}) = \frac{\left(w_1\sqrt{C_1} + w_2\sqrt{C_2}\right)^2}{\left(w_1\sqrt{A_1} + w_2\sqrt{A_2}\right)^2}, \tag{4.22}$$

where $C_1 = \eta_1(1-\eta_1) + \frac{p_{r1}\left(1-p_{r1}\right)\left(1-\eta_1 T\right)}{\left(2p_{r1}-1\right)^2}$, $C_2 = \eta_2(1-\eta_2) + \frac{p_{r2}\left(1-p_{r2}\right)\left(1-\eta_1 T\right)}{\left(2p_{r2}-1\right)^2}$,

$$PRE(\hat{T}_s, \hat{T}_{ts}) = \frac{\left(w_1\sqrt{C_1^*} + w_2\sqrt{C_2^*}\right)^2}{\left(w_1\sqrt{A_1^*} + w_2\sqrt{A_2^*}\right)^2}, \tag{4.23}$$

where $C_1^* = T(1-T) + \frac{p_{r1}\left(1-p_{r1}\right)T^2\left(1-\eta_1 T\right)}{\left(2p_{r1}-1\right)^2}$, $C_2^* = T(1-T) + \frac{p_{r2}\left(1-p_{r2}\right)T^2\left(1-\eta_1 T\right)}{\left(2p_{r2}-1\right)^2}$, $A_1^* = T(1-T) + \dfrac{(1-p_{r1})T^2(1-\eta_1)}{(1-p_{r1})^2\eta_1^2}$, $A_2^* = T(1-T) + \dfrac{(1-p_{r2})T^2(1-\eta_2)}{(1-p_{r2})^2\eta_2^2}$,

$$PRE(\hat{\eta}_s, \hat{\eta}_k) = \frac{\left(w_1\sqrt{D_1} + w_2\sqrt{D_2}\right)^2}{\left(w_1\sqrt{A_1} + w_2\sqrt{A_2}\right)^2}, \tag{4.24}$$

where $D_1 = \eta_1(1-\eta_1) + \dfrac{(1-pr1)(1-p_{ro})\{1-(1-p_{r1})(1-p_{ro})\}}{W^2} - \frac{p_{r1}\eta_1 T\left(1-p_{r1}\right)}{W^2}$, $D_2 = \eta_1(1-\eta_1) + \dfrac{(1-p_{r2})(1-p_{ro})\{1-(1-p)(1-p_{ro})\}}{W^2} - \frac{p_{r2}\eta_2 T\left(1-p_{r2}\right)}{W^2}$, and $\eta_s = w_1\eta_1 + w_2\eta_2$.

$$PRE(\hat{T}_s, \hat{T}_k) = \frac{\left(w_1\sqrt{D_1^*} + w_2\sqrt{D_2^*}\right)^2}{\left(w_1\sqrt{A_1^*} + w_2\sqrt{A_2^*}\right)^2}, \tag{4.25}$$

where $D_1^* = T(1-T) + \dfrac{(1-p_{r1})(1-p_{ro})T^2\{1-(1-p_{r1})(1-p_{ro})\}}{\pi_{r1}W^2} + \frac{\left(1-p_{r1}\right)T^2\left(2p_{r0}W - p_{r1}T\right)}{W^2}$, $D_2^* = T(1-T) + \dfrac{(1-p_{r2})(1-p_{ro})T^2\{1-(1-p_{r2})(1-p_{ro})\}}{\eta_2 W^2} + \frac{\left(1-p_{r2}\right)T^2\left(2p_{r2}W - p_{r2}T\right)}{W^2}$, and $\pi_s = w_1\eta_1 + w_2\eta_2$.

Tables 7–11 suggest that the $PREs$ of the proposed estimators $\hat{\eta}_s$ and $\hat{T}_s$ are greater than 100, thus indicating that the suggested estimators $\hat{\eta}_s$ and $\hat{T}_s$ are more efficient than the [19] estimator $\hat{\eta}_{kw}$, the [20] estimators $\hat{\eta}_{ts}$ and $\hat{T}_{ts}$, and the [16] estimators $\hat{\eta}_k$ and $\hat{T}_k$.

Table 7. *PRE* of the proposed estimator $\hat{\eta}_s$ with the [19] estimator $\hat{\eta}_{kw}$ in stratified random sampling using optimal allocation.

					p_{r1}	0.6	0.7	0.8
η_s	η_1	η_2	w_1	w_2	p_{r2}	0.65	0.75	0.85
0.24	0.3	0.2	0.4	0.6		1087.8	375.52	201.57
0.32	0.5	0.2	0.4	0.6		1101.12	377.86	202.08
0.4	0.7	0.2	0.4	0.6		1193.74	405.64	213.53
0.39	0.3	0.45	0.4	0.6		1034.09	352.97	190.14
0.47	0.5	0.45	0.4	0.6		1046.10	355.15	190.78
0.55	0.7	0.45	0.4	0.6		1127.73	377.97	199.52
0.6	0.3	0.8	0.4	0.6		1458.44	471.09	233.81
0.68	0.5	0.8	0.4	0.6		1480.82	473.43	233.59
0.76	0.7	0.8	0.4	0.6		1655.40	519.48	250.06

Table 8. *PRE* of the proposed estimator $\hat{\eta}_s$ with the [20] estimator $\hat{\eta}_{ts}$ in stratified random sampling using optimal allocation.

					p_{r1}	0.6	0.7	0.8
					p_{r2}	0.65	0.75	0.85
η_s	η_1	η_2	w_1	w_2	T			
0.24	0.3	0.2	0.4	0.6	0.3	1009.59	352.45	192.10
0.32	0.5	0.2	0.4	0.6		989.98	346.66	189.54
0.4	0.7	0.2	0.4	0.6		1037.79	362.82	196.38
0.39	0.3	0.45	0.4	0.6	0.5	853.11	298.95	169.14
0.47	0.5	0.45	0.4	0.6		813.37	289.12	165.56
0.55	0.7	0.45	0.4	0.6		821.70	293.78	167.62
0.6	0.3	0.8	0.4	0.6	0.7	929.56	310.02	172.26
0.68	0.5	0.8	0.4	0.6		844.42	289.25	164.73
0.76	0.7	0.8	0.4	0.6		828.21	288.39	164.78

Table 9. *PRE* of the proposed estimator $\hat{T}_s$ with the [19] estimator $\hat{T}_{ts}$ in stratified random sampling using optimal allocation.

					p_{r1}	0.6	0.7	0.8
					p_{r2}	0.65	0.75	0.85
η_s	η_1	η_2	w_1	w_2	T			
0.24	0.3	0.2	0.4	0.6	0.3	554	218.52	138.66
0.32	0.5	0.2	0.4	0.6		530.2	217.56	139.6
0.4	0.7	0.2	0.4	0.6		518.01	217.08	140.04
0.39	0.3	0.45	0.4	0.6	0.5	583.87	215.22	134.82
0.47	0.5	0.45	0.4	0.6		542.65	212.5	136.38
0.55	0.7	0.45	0.4	0.6		517.38	210.97	137.19
0.6	0.3	0.8	0.4	0.6	0.7	617.9	211.54	129.07
0.68	0.5	0.8	0.4	0.6		558.68	204.8	130.71
0.76	0.7	0.8	0.4	0.6		510.74	200	131.73

Table 10. *PRE* of the proposed estimator $\hat{\eta}_s$ with the [16] estimator $\hat{\eta}_k$ in stratified random sampling using optimal allocation.

					p1	0.6	0.7	0.8	0.6	0.7	0.8
					p2	0.65	0.75	0.85	0.65	0.75	0.85
π_{rs}	π_{r1}	π_{r2}	w_1	w_2	T		$p_{r0} = 0.4$			$p_{r0} = 0.5$	
0.24	0.3	0.2	0.4	0.6	0.2	193.73	149.16	123.01	146.48	124.56	110.91
0.32	0.5	0.2	0.4	0.6		194.28	149.39	123.20	147.80	125.44	111.56
0.4	0.7	0.2	0.4	0.6		202.23	154.13	125.87	152.96	128.70	113.49
0.39	0.3	0.45	0.4	0.6	0.5	172.33	135.13	114.94	133.48	115.4	105.47
0.47	0.5	0.45	0.4	0.6		169.69	133.48	114.10	132.36	114.63	105.08
0.55	0.7	0.45	0.4	0.6		171.86	134.41	114.46	133.49	114.97	105.13

Table 11. *PRE* of the proposed estimator $\hat{T}_s$ with the [16] estimator $\hat{T}_k$ in stratified random sampling using optimal allocation.

					p1	0.6	0.7	0.8	0.6	0.7	0.8
					p_{r1}	0.65	0.75	0.85	0.65	0.75	0.85
η_s	η_1	η_2	w_1	w_2	T		$p_{ro} = 0.4$			$p_{ro} = 0.5$	
0.24	0.3	0.2	0.4	0.6	0.2	172.4	134.19	114	147.92	124.55	110.51
0.32	0.5	0.2	0.4	0.6		162.01	129.43	112.48	145.62	123.77	110.75
0.4	0.7	0.2	0.4	0.6		156.69	127.1	111.78	144.44	123.38	110.87
0.39	0.3	0.45	0.4	0.6	0.5	222.56	159.75	125.04	180.15	142.12	118.22
0.47	0.5	0.45	0.4	0.6		214.47	154.71	123.18	183.99	143.96	119.79
0.55	0.7	0.45	0.4	0.6		209.51	151.87	122.22	186.35	144.99	120.59

It is observed from Tables 7, 8, and 10 that the *PRE* of the suggested estimator $\hat{\eta}_S$ is higher when p_{r1}, p_{r2}, η_{r1}, and η_2 approach 0.5. Furthermore, Tables 9 and 11 show that a higher gain in efficiency is achieved using the suggested estimator $\hat{\eta}_S$ when the values of p_{r1}, p_{r2}, η_1, and η_2 are close to 0.5.

5. Conclusions

The RR technique is preferred when survey questions are sensitive, as respondents may conceal their true status or provide deliberately false answers. The proposed model differs from existing *RR* procedures in that respondents are first asked to directly answer the sensitive question directly; only those who do not provide a direct response are subsequently guided to use an *RR* device. Moreover, the proposed *RR* model employs a two-stage mechanism that involves three questions, which offers enhanced protection of respondent privacy compared to conventional *RR* models. The proposed procedure enables us to unbaisedly estimate the sensitive proportion η and obtain an effective estimator of T, a feature lacking in most competing procedures. The results of the numerical simulations and theoretical comparisons demonstrate that, under certain conditions, the proposed estimators η_s and T_s uniformly dominate the estimators developed under the models of [2, 14–16]. Furthermore, when extended to stratified random sampling, the suggested estimators are more efficient than the models of [16, 19, 20]. Thus, the proposed randomized response techniques are recommended for use in practical surveys. In the future, the methodological framework proposed in this work could be further extended and refined to effectively address the emerging challenges of modern and complex data environments. Moreover, the proposed approach may be adapted, generalized, and applied to a wide range of datasets in a similar manner across different application domains [28–30].

Author contributions

Shehzad Ahmad Khan: Conceptualization, methodology, investigation, formal analysis, writing–original draft, writing–review & editing; Moiz Qureshi: Conceptualization, data curation, software, formal analysis, validation, visualization, writing–original draft, writing–review & editing; Hasnain Iftikhar: Methodology, resources, project administration, supervision, validation, writing–original draft, writing–review & editing; Fatimah E. Almuhayfith: Investigation, resources, project administration, validation, funding acquisition, writing–review & editing; Paulo Canas Rodrigues: Conceptualization, methodology, supervision, funding acquisition, validation, writing–review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU263259].

Conflict of interest

The authors declare no conflicts of interest.

References

1. B. G. Greenberg, A. L. A. Abul-Ela, W. R. Simmons, D. G. Horvitz, The unrelated question randomized response model: Theoretical framework, *J. Amer. Stat. Assoc.*, **64** (1969), 520–539. https://doi.org/10.1080/01621459.1969.10500991
2. S. L. Warner, Randomized response: A survey technique for eliminating evasive answer bias, *J. Amer. Stat. Assoc.*, **60** (1965), 63–69. https://doi.org/10.1080/01621459.1965.10480775
3. A. Chaudhuri, R. Mukherjee, Randomized response techniques: A review, *Stat. Neerl.*, **41** (1987), 27–44. https://doi.org/10.1111/j.1467-9574.1987.tb01169.x
4. N. S. Mangat, R. Singh, An alternative randomized response procedure, *Biometrika*, **77** (1990), 439–442. https://doi.org/10.2307/2336829
5. S. Singh, A. H. Joarder, Unknown repeated trials in randomized response sampling, *J. Ind. Soc. Ag. Statist.*, **50** (1997), 103–105.
6. M. Mahmood, S. Singh, S. Horn, On the confidentiality guaranteed under randomized response sampling: A comparison with several new techniques, *Biom. J.*, **40** (1998), 237–242. https://doi.org/10.1002/(SICI)1521-4036(199806)40:2¡237::AID-BIMJ237¿3.0.CO;2-N
7. S. Singh, R. Singh, N. S. Mangat, Some alternative strategies to Moors' model in randomized response sampling, *J. Stat. Plann. Inference*, **83** (2000), 243–255. https://doi.org/10.1016/S0378-3758(99)00092-0
8. T. C. Chua, A. K. Tsui, Procuring honest responses indirectly, *J. Stat. Plann. Inference*, **90** (2000), 107–116. https://doi.org/10.1016/S0378-3758(00)00109-9
9. J. M. Kim, W. D. Warde, A mixed randomized response model, *J. Stat. Plann. Inference*, **133** (2005), 211–221. https://doi.org/10.1016/j.jspi.2004.03.011
10. H. P. Singh, T. A. Tarray, An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation, *Model Assisted Stat. Appl.*, **9** (2014), 37–62. https://doi.org/10.3233/MAS-130277
11. T. A. Tarray, H. P. Singh, An optional randomized response model for estimating a rare sensitive attribute using Poisson distribution, *Commun. Stat. Theor. Methods*, **46** (2017), 2638–2654. https://doi.org/10.1080/03610926.2015.1040506
12. Z. Hussain, S. A. Cheema, I. Hussain, A stratified randomized response model for sensitive characteristics using non identical trials, *Commun. Stat. Theor. Methods*, **49** (2020), 99–115. https://doi.org/10.1080/03610926.2018.1530791
13. H. J. Chang, K. C. Huang, Estimation of proportion and sensitivity of a qualitative character, *Metrika*, **53** (2001), 269–280. https://doi.org/10.1007/s001840100109
14. K. C. Huang, A survey technique for estimating the proportion and sensitivity in a dichotomous finite population, *Stat. Neerl.*, **58** (2004), 75–82. https://doi.org/10.1046/j.0039-0402.2003.00113.x
15. H. P. Singh, T. A. Tarray, A modified survey technique for estimating the proportion and sensitivity in a dichotomous finite population, *Int. J. Adv. Sci. Technol. Res.*, **6** (2013), 459–472.

16. S. A. Khan, M. A. Alomair, M. Qureshi, A. M. Alomair, A two-stage randomized response technique for simultaneous estimation of sensitivity and truthfulness, *Sci. Rep.*, **15** (2025), 35664. https://doi.org/10.1038/s41598-025-19658-4
17. S. Ahmad, H. Iftikhar, M. Qureshi, I. Khan, A. S. A. Omer, E. A. T. Armas, et al., A new auxiliary variables-based estimator for population distribution function under stratified random sampling and non-response, *Sci. Rep.*, **15** (2025), 13580. https://doi.org/10.1038/s41598-025-98246-y
18. S. Ahmad, M. Qureshi, H. Iftikhar, P. C. Rodrigues, M. Z. Rehman, An improved family of unbiased ratio estimators for a population distribution function, *AIMS Math.*, **10** (2025), 1061–1084. https://doi.org/10.3934/math.2025051
19. J. M. Kim, W. D. Warde, A stratified Warner's randomized response model, *J. Stat. Plann. Inference*, **120** (2004), 155–165. https://doi.org/10.1016/S0378-3758(02)00500-1
20. T. A. Tarray, H. P. Singh, A survey technique for estimating the proportion and sensitivity in a stratified dichotomous finite population, *Stat. Appl.*, **15** (2017), 173–191.
21. B. G. Greenberg, A. L. A. Abul-Ela, W. R. Simmons, D. G. Horvitz, The unrelated question randomized response model: Theoretical framework, *J. Amer. Stat. Assoc.*, **64** (1969), 520–539. https://doi.org/10.1080/01621459.1969.10500991
22. S. Shah, E. Mahmoudi, M. Qureshi, H. Iftikhar, P. C. Rodrigues, R. I. Gonzales Medina, et al., A new class of median estimators using auxiliary information under PPS sampling: Theoretical properties and empirical evaluation, *Comput. Stat.*, **41** (2026), 59. https://doi.org/10.1007/s00180-026-01724-8
23. R. Singh, S. Singh, N. S. Mangat, D. S. Tracy, An improved two stage randomized response strategy, *Stat. Papers*, **36** (1995), 265–271. https://doi.org/10.1007/BF02926038
24. K. H. Hong, J. K. Yum, H. Y. Lee, A stratified randomized response technique, *Korean J. Appl. Stat.*, **7** (1994), 141–147.
25. J. M. Kim, M. E. Elam, A two-stage stratified Warner's randomized response model using optimal allocation, *Metrika*, **61** (2005), 1–7. https://doi.org/10.1007/s001840400319
26. J. M. Kim, M. E. Elam, A stratified unrelated question randomized response model, *Stat. Papers*, **48** (2007), 215–233. https://doi.org/10.1007/s00362-006-0327-6
27. W. G. Cochran, *Sampling techniques*, Wiley, 1977.
28. J. Nasir, H. Iftikhar, M. Aamir, H. Iftikhar, P. C. Rodrigues, M. Z. Rehman, A hybrid LMD–ARIMA–machine learning framework for enhanced forecasting of financial time series: Evidence from the NASDAQ composite index, *Mathematics*, **13** (2025), 2389. https://doi.org/10.3390/math13152389
29. W. M. Cuba, J. C. Huaman Alfaro, H. Iftikhar, J. López-Gonzales, Modeling and analysis of monkeypox outbreak using a new time series ensemble technique, *Axioms*, **13** (2024), 554. https://doi.org/10.3390/axioms13080554
30. F. Khan, H. Iftikhar, I. Khan, P. C. Rodrigues, A. A. Alharbi, J. Allohibi, A hybrid vector autoregressive model for accurate macroeconomic forecasting: An application to the US economy, *Mathematics*, **13** (2025), 1706. https://doi.org/10.3390/math13111706