Sri Lankan Journal of Applied Statistics, Vol (26-3)

Efficient Estimator of Population Mean of Sensitive Variable in Presence of Scrambled Response Using Auxiliary Information

P. V. Singh[1], M. K. Srivastava[1], and N. Srivastava[2,*]

[1]Department of Statistics (I.S.S), Dr. Bhimrao Ambedkar University, Agra, India

[2]Department of Statistics, St. John's College, Agra, U. P., India

*Corresponding author: drnamita.sjc@gmail.com

Received: 07th June, 2025/ Revised: 10th August, 2025/Accepted: 19th September, 2025/
Published: 31st December, 2025

ABSTRACT

In this paper, we propose a modified regression exponential estimator of population mean of the sensitive variable for additive model of Randomized Response Technique (RRT) given by Pollack and Bek (1976). The proposed estimator utilizes a sufficiently correlated non-sensitive auxiliary variable to improve estimation accuracy. We derive the expression for bias and mean square error (MSE) of the estimator up to the first order of approximation. Furthermore, an empirical and theoretical study is conducted using artificially generated populations for comparing efficiency of the proposed estimator with existing estimators. The results demonstrate that the present relative efficiency of the proposed estimator generally is comparison of the other existing estimators for different values of correlation coefficient.

Keywords: RRT, Sensitive variables, Bias, MSE, Auxiliary variable

1 Introduction

In social survey research, respondents sometimes encounter sensitive questions seeking the information regarding corruption, criminality, abortion, illegal income, tax evasion, drug addiction, gambling, political views, sexual and physical abuses, etc. People are not willing to provide information that might be considered incriminating and stigmatizing. As an outcome, respondent may refuse to reply such questions or deliberately provide incorrect information. To address this situation, one of the techniques is randomized response technique (RRT), given by Warner (1965). This technique attempts to maintain the respondent's privacy so that more honest answers can be extracted. The estimation of the finite population mean for a sensitive variable based on RRT under different sampling designs has been addressed by many researchers without using auxiliary information. Greenberg et al. [1971] and Pollock and Bek [1976] suggested quantitative randomized response model, Eichhorn and Hayre [1983] proposed multiplication RRT model. Gupta et al. [2002] suggested Optional Randomized Response Technique (ORRT) model. Gupta and Shabbir [2004], Saha [2008], Gupta et al. [2010] suggest additive (ORRT) model. In the recent decade, many authors have utilized auxiliary information for the mean estimation of sensitive variable in randomized response models. Sausa et al. [2010], Diana and Perri [2011], Gupta et al. [2012] introduced a regression and generalized regression-cum-ratio type estimators based on RRT models to deal with sensitive situation. Koyuncu et al. [2014], Ozgul and Cingi [2017] proposed an improved mean estimator using partial additive randomized response model. Waseem et al. [2022] generalized the exponential type estimator for estimating the mean of sensitive variable using auxiliary information. Murtaza et al. [2021] presented quantitative (RRT) model. In the present article, motivated by Shabbir and Gupta [2010] and Gupta et al. [2012], we propose an estimator for the finite population mean using a transformation proposed by Bedi (1996) for scrambled RRT model in presence of non-sensitive auxiliary information. Expressions for the bias and mean squared error [MSE] of the proposed estimator are derived up to the first order of approximation. The minimum MSE of the proposed estimator is compared with existing methods. It has been shown that the proposed estimator outperforms other existing estimators both theoretically and empirically.

2 Terminology

Let Y be the sensitive study variable which is not observed directly and X be a non-sensitive auxiliary variable correlated with Y, and S be a scrambled variable independent of Y and X. The respondent is asked to report a scrambled response for Y given as $Z = Y + S$, and a true response for X. A random sample of size n is drawn without replacement from a finite population $U = (U_1, U_2, \ldots, U_N)$.. Let y_i and x_i be the values of the sensitive study variable Y and auxiliary variable X respectively, and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i, \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$. be the sample means of $Y, X, S$and Z respectively. Let $\bar{Y} = E(Y), \bar{X} = E(X), \bar{Z} = E(Z)$, also S is defined as $\bar{S} = E(S) = 0$. Thus,$E(Z) = E(Y) + E(S) = E(Y)$, that is, one can take $\bar{Z} = \bar{Y}$. We define below:

$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ be the population mean of sensitive study variable Y,
$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ be the population mean of non-sensitive study variable X,
$\bar{S} = \frac{1}{N}\sum_{i=1}^{N} s_i$ be the population mean of the scrambling variable S,
$\bar{Z} = \frac{1}{N}\sum_{i=1}^{N} z_i$ be the population mean of the observed scramble response Z.
$S_y^2 = \frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N-1}$ be the population variance of the sensitive study variable Y.
$S_x^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{X})^2}{N-1}$, be the population variance of the non-sensitive study variable S.
$Z^2 = \frac{\sum_{i=1}^{N}(z_i-\bar{Z})^2}{N-1}$,be the population variance of the observed response variable Z.
$S_s^2 = \frac{\sum_{i=1}^{N}(s_i-\bar{S})^2}{N-1}$, be the population variance of the scrambled variable Z
$C_y^2 = \frac{S_y^2}{\bar{Y}^2}$, be the coefficient of variation of the sensitive study variable Y.
$C_x^2 = \frac{S_x^2}{\bar{X}^2}$,be the coefficient of variation of the non-sensitive study variable X.

$C_z^2 = C_y^2 + \frac{S_s^2}{\bar{Y}^2}$ be the coefficient of variation of the observed scrambled response Z, ρ_{zx} be the population correlation coefficient between sensitive study variable and auxiliary variable, and $\rho_{zx} = \frac{\rho_{yx}}{\sqrt{1+\frac{S_s^2}{S_y^2}}}$.

The bias and mean square error (MSE) expressions of the proposed estimator are obtained by using the following relative error terms:

$$e_0 = \frac{\bar{z}-\bar{Z}}{\bar{Z}}, \quad e_1 = \frac{\bar{x}-\bar{X}}{\bar{X}},$$

We have

$$E(e_0) = E(e_1) = 0,$$

$$E(e_0^2) = \theta C_z^2, \quad E(e_1^2) = \theta C_x^2, \quad E(e_0 e_1) = \theta \rho_{zx} C_z C_x,$$

where $\theta = \frac{N-n}{nN}$, N is population size, n is the sample size, and $R = \frac{\bar{Y}}{\bar{X}}$ is the population ratio.

3 Some existing RRT literature for population mean of a sensitive variable.

(i). The ordinary randomized response mean estimator for the sensitive variable, suggested by Sousa et al. (2010), is given as

$$\hat{\mu}_Y = \bar{z} \tag{1}$$

The mean square error (MSE) is

$$\text{MSE}(\hat{\mu}_Y) = \theta\left(S_y^2 + S_s^2\right) \tag{2}$$

(ii). The ratio estimator suggested by Sousa et al. (2010) for the mean of the sensitive variable Y using non-sensitive auxiliary variable X is given as

$$\hat{\mu}_R = \bar{z}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{3}$$

The Bias and MSE of $\hat{\mu}_R$, to first degree of approximation, are given by

$$\text{Bias}(\hat{\mu}_R) \cong \theta\bar{Y}\left(C_x^2 - \rho_{zx}C_zC_x\right) \tag{4}$$
$$\text{MSE}(\hat{\mu}_R) \cong \theta\bar{Y}^2\left(C_z^2 + C_x^2 - 2\rho_{zx}C_zC_x\right) \tag{5}$$

(iii). The regression estimator suggested by Gupta et al. (2012) for estimation of the mean of a sensitive variable is given as

$$\hat{\mu}_{Reg} = \bar{z} + \hat{\beta}_{zx}\left(\bar{X} - \bar{x}\right) \tag{6}$$

where $\hat{\beta}_{zx}$ is the regression coefficient between Z and X under the additive scrambling model $Z = Y + S$. The expression (6) can be re-written using Taylor's approximation and retaining terms up to second order as

$$(\hat{\mu}_{Reg} - \bar{Z}) \cong \bar{Z}e_0 - \hat{\beta}_{zx}\bar{X}\left(e_1 + e_1e_3 - e_1e_2\right) \tag{7}$$

By Mukhopadhyay (1998, p. 123), we have

$$E(e_1e_2) = \theta\frac{1}{\bar{X}}\left(\frac{\mu_{03}}{\mu_{02}}\right), \quad E(e_1e_3) = \theta\frac{1}{\bar{X}}\left(\frac{\mu_{12}}{\mu_{11}}\right).$$

Hence, the Bias and MSE of $\hat{\mu}_{Reg}$ to first-order approximation are given by

$$\text{Bias}(\hat{\mu}_{Reg}) \cong -\beta_{zx}\theta\left(\frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{03}}{\mu_{02}}\right) \tag{8}$$
$$\text{MSE}(\hat{\mu}_{Reg}) \cong \bar{Y}^2C_z^2\left(1 - \rho_{zx}^2\right) = \theta S_y^2\left\{\left(1 + \frac{S_s^2}{S_y^2}\right) - \rho_{zx}^2\right\} \tag{9}$$

(iv). The regression-cum-ratio estimator under the same RRT model for mean of a sensitive variable, suggested by Gupta et al. (2012), is given by

$$\hat{\mu}_{GRR} = \left[k_1\bar{z} + k_2(\bar{X} - \bar{x})\right]\left(\frac{\bar{X}}{\bar{x}}\right) \tag{10}$$

where k_1 and k_2 are suitably chosen constants. The Bias of $\hat{\mu}_{GRR}$ to first order of approximation is given by

$$\text{Bias}(\hat{\mu}_{GRR}) \cong (k_1 - 1)\bar{Y} + k_1^2\bar{Y}\theta\left(C_z^2 - \rho_{zx}C_zC_x\right) + k_2\bar{X}\theta C_x^2 \tag{11}$$

The optimum values of k_1 and k_2 are

$$k_{1(opt)} = \frac{(1 - \theta C_x^2)}{1 - \theta\{C_x^2 - C_z^2(1 - \rho_{zx}^2)\}}, \tag{12a}$$

$$k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[1 + k_{1(opt)}\left(\frac{\rho_{zx}C_z}{C_x} - 2\right)\right] \tag{12b}$$

The minimum MSE is

$$\text{MSE}(\hat{\mu}_{GRR})_{min} \cong \left[\frac{\bar{Y}^2C_z^2(1 - \rho_{zx}^2)\theta\{1 - \theta C_x^2\}}{C_z^2(1 - \rho_{zx}^2)\theta + \{1 - \theta C_x^2\}}\right] \tag{13}$$

4 Proposed Estimator

Motivated by Shabbir and Gupta (2010) and Gupta et al. (2012), we propose a randomized scrambled estimator to estimate the population mean of a sensitive study variable Y using a non-sensitive auxiliary variable X.

The proposed estimator is given as

$$\hat{\mu}_t = \left[w_1\bar{z} + w_2(\bar{X} - \bar{x})\right]\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x} + 2N\bar{X}}\right) \tag{14}$$

where w_1 and w_2 are suitably chosen constants obtained by minimizing the MSE of $\hat{\mu}_t$.

4.1 Bias and MSE of the Proposed Estimator($\hat{\mu}_t$)

Expanding $\hat{\mu}_t$ in terms of error terms e_i, we have

$$\hat{\mu}_t = \Big[w_1\bar{Z}(1+e_0) + w_2\big(\bar{X} - \bar{X}(1+e_1)\big)\Big]\exp\left(\frac{\bar{X} - \bar{X}(1+e_1)}{\bar{X} + \bar{X}(1+e_1) + 2N\bar{X}}\right) \tag{15}$$

Simplifying (14) by using Taylor's approximation up to second order and subtracting $\bar{Y}$, we obtain

$$\begin{aligned}(\hat{\mu}_t - \bar{Y}) &= \Big[\bar{Y}(w_1-1) + w_1\bar{Y}\left(e_0 - \tfrac{e_1}{2(1+N)} - \tfrac{e_0e_1}{2(1+N)} + \tfrac{3e_1^2}{8(1+N)^2}\right) \\ &\quad - w_2\bar{X}\left(e_1 - \tfrac{e_1^2}{2(1+N)}\right)\Big] \\ &= \bar{Y}\Big[(w_1-1) + w_1\left(e_0 - \tfrac{1}{2}\delta e_1 - \tfrac{1}{2}\delta e_0e_1 + \tfrac{3}{8}\delta^2e_1^2\right) \\ &\qquad - w_2\tfrac{1}{R}\left(e_1 - \tfrac{1}{2}\delta e_1^2\right)\Big]\end{aligned} \tag{16}$$

where $\delta = \frac{1}{1+N}$ and $R = \frac{\bar{Y}}{\bar{X}}$.

Taking expectations on both sides, the Bias of $\hat{\mu}_t$ to first-order approximation is

$$\text{Bias}(\hat{\mu}_t) \cong \bar{Y}\Big[(w_1-1) + w_1\left(\tfrac{3}{8}\delta^2\theta C_x^2 - \tfrac{1}{2}\delta\theta\rho_{zx}C_zC_x\right) + w_2\tfrac{\theta\delta}{2R}C_x^2\Big] \tag{17}$$

Squaring both sides of equation (16), neglecting higher-order terms,

$$\begin{aligned}(\mu_t - \bar{Y})^2 &= \bar{Y}\Big[(w_1-1) + w_1\left(e_0 - \tfrac{1}{2}\delta e_1 - \tfrac{1}{2}\delta e_0e_1 + \tfrac{3}{8}\delta^2e_1^2\right) \\ &\qquad - w_2\tfrac{1}{R}\left(e_1 - \tfrac{1}{2}\delta e_1^2\right)\Big]^2 \\ &= \bar{Y}^2\Big[(w_1-1)^2 + w_1^2\left(e_0^2 + \tfrac{1}{4}\delta^2e_1^2 - \delta e_0e_1 - \delta e_0e_1 + \tfrac{3}{4}\delta^2e_1^2\right) \\ &\qquad + w_2^2\tfrac{1}{R^2}e_1^2 - 2w_1\left(\tfrac{3}{8}\delta^2e_1^2 - \tfrac{1}{2}\delta e_0e_1\right) \\ &\qquad + 2w_1w_2\tfrac{1}{R}\left(\delta e_1^2 - e_0e_1\right)\Big]\end{aligned}$$

By taking expectations, we have

$$\begin{aligned}\mathrm{MSE}(\hat{\mu}_t) \cong \bar{Y}^2\Big[&(w_1-1)^2 + w_1^2\theta\Big(C_z^2 + \delta^2 C_x^2 - 2\delta\rho_{zx}C_zC_x\Big)\\ &+ w_2^2\tfrac{\theta}{R^2}C_x^2 + 2w_1w_2\tfrac{\theta}{R}\{\delta C_x^2 - \rho_{zx}C_zC_x\}\\ &- 2w_1\theta\Big(\tfrac{3}{8}\delta^2C_x^2 - \tfrac{1}{2}\delta\rho_{zx}C_zC_x\Big) - 2w_2\tfrac{\theta\delta}{2R}C_x^2\Big] \end{aligned} \tag{18}$$

Minimum Mean Square Error of the proposed estimator
To get the optimum values of w_1 and w_2, differentiate $\mathrm{MSE}(\hat{\mu}_t)$ in (18) w.r.t. w_i $(i = 1, 2)$ and set equal to zero:

$$w_{1(opt)} = \frac{1 - \frac{1}{8}\theta\delta^2C_x^2}{1 + \theta C_z^2(1-\rho_{zx}^2)}, \qquad w_{2(opt)} = R\left[\tfrac{\delta}{2} - w_{1(opt)}\left(\delta - \rho_{zx}\tfrac{C_z}{C_x}\right)\right] \tag{19}$$

Substituting the optimum values of $w_{1(opt)}$ and $w_{2(opt)}$ in (18) and putting the value of δ, we get

$$\mathrm{MSE}_{min}(\hat{\mu}_t) \cong \bar{Y}^2\left[1 - \tfrac{1}{4(1+N)^2}\theta C_x^2 - \frac{\left(1 - \frac{1}{8(1+N)^2}\theta C_x^2\right)^2}{1 + \theta C_z^2(1-\rho_{zx}^2)}\right] \tag{20}$$

5 Efficiency Comparisons

We compare the efficiency of $\hat{\mu}_t$ with the existing estimators $\hat{\mu}_Y, \hat{\mu}_R, \hat{\mu}_{Reg}, \hat{\mu}_{GRR}$.

(i) Comparison with $\hat{\mu}_Y$
From (20) and (2):

$$\mathrm{MSE}(\hat{\mu}_Y) - \mathrm{MSE}_{min}(\hat{\mu}_t) > 0 \quad \text{if}$$

$$\begin{aligned}\theta\bar{Y}^2C_z^2 - \bar{Y}^2\Big[1 - \tfrac{1}{4(1+N)^2}\theta C_x^2 - \tfrac{\left(1-\frac{1}{8(1+N)^2}\theta C_x^2\right)^2}{1+\theta C_z^2(1-\rho_{zx}^2)}\Big] &> 0\\ \Big[\theta C_z^2\rho_{zx}^2 + \tfrac{\theta C_x^2}{4(1+N)^2}\Big(\theta C_z^2(1-\rho_{zx}^2) + \tfrac{\theta C_x^2}{16(1+N)^2}\Big)\Big] &> 0 \end{aligned} \tag{21}$$

Since $C_z^2 = C_y^2 + \frac{S_s^2}{\bar{Y}^2} = (\frac{S_y^2}{\bar{Y}^2} + \frac{S_s^2}{\bar{Y}^2})$, then $(S_y^2 + S_s^2) = \bar{Y}^2C_z^2$ this is always true. Hence, $\hat{\mu}_t$ is more efficient than $\hat{\mu}_Y$.

(ii) Comparison with $\hat{\mu}_R$
From (20) and (5):

$$\theta\bar{Y}^2(C_z^2+C_x^2-2\rho_{zx}C_zC_x)-\bar{Y}^2\Big[1-\tfrac{1}{4(1+N)^2}\theta C_x^2-\frac{\left(1-\frac{1}{8(1+N)^2}\theta C_x^2\right)^2}{1+\theta C_z^2\left(1-\rho_{zx}^2\right)}\Big]>0$$

$$\Big[(1-\rho_{zx}C_z/C_x)^2+\theta\tfrac{1}{4(1+N)^2}\Big(\theta C_z^2(1-\rho_{zx}^2)+\tfrac{\theta C_x^2}{16(1+N)^2}\Big)\Big]>0 \qquad (22)$$

This inequality is always satisfied, so $\hat{\mu}_t$ is more efficient than $\hat{\mu}_R$.

(iii) Comparison with $\hat{\mu}_{Reg}$
From (20) and (9):

$$\theta\bar{Y}^2C_z^2(1-\rho_{zx}^2)-\bar{Y}^2\Big[1-\tfrac{1}{4(1+N)^2}\theta C_x^2-\frac{\left(1-\frac{1}{8(1+N)^2}\theta C_x^2\right)^2}{1+\theta C_z^2(1-\rho_{zx}^2)}\Big]>0$$

$$\Big[C_z^2(1-\rho_{zx}^2)+\tfrac{C_x^2}{4(1+N)^2}\Big(1+\tfrac{C_x}{16(1+N)^2C_z^2(1-\rho_{zx}^2)}\Big)\Big]>0 \qquad (23)$$

Thus, $\hat{\mu}_t$ is always more efficient than $\hat{\mu}_{Reg}$.

(iv) Comparison with $\hat{\mu}_{GRR}$
From (20) and (13):

$$\frac{\bar{Y}^2C_z^2(1-\rho_{zx}^2)\theta(1-\theta C_x^2)}{C_z^2(1-\rho_{zx}^2)\theta+(1-\theta C_x^2)}-\bar{Y}^2\Big[1-\tfrac{1}{4(1+N)^2}\theta C_x^2-\frac{\left(1-\frac{1}{8(1+N)^2}\theta C_x^2\right)^2}{1+\theta C_z^2(1-\rho_{zx}^2)}\Big]>0$$

$$\Big[\tfrac{1}{4(1+N)^2}\Big(1+\tfrac{1-\theta C_x^2}{\theta C_z^2(1-\rho_{zx}^2)}\Big)\Big\{\tfrac{\theta C_x^2}{16(1+N)^2}+\theta C_z^2(1-\rho_{zx}^2)-\tfrac{1-\theta C_x^2}{\theta C_z^2-\theta C_x^2+1}\Big\}\Big]>\theta C_z^2(1-\rho_{zx}^2) \qquad (24)$$

Hence, $\hat{\mu}_t$ is also more efficient than $\hat{\mu}_{GRR}$.

6 Empirical Study

For the purpose of comparing the efficiency of the proposed estimator, we conducted an empirical study and computed the percentage relative efficiency (PRE) of the estimator $\hat{\mu}_t$, as compared to the mean estimator $\hat{\mu}_Y$, the ratio estimator $\hat{\mu}_R$, the regression estimator $\hat{\mu}_{Reg}$, and the generalized regression-cum-ratio estimator $\hat{\mu}_{GRR}$. The comparisons are made theoretically, i.e., using the formulas derived for the MSE of the estimators, and also empirically by conducting a simulation study for the comparisons in terms of efficiency.

For empirical investigations, four artificial populations of size $N = 1000$ each from bivariate normal populations with different covariance matrices for (Y, X) are generated using the `mvrnorm` function from the `MASS` package in R software. All of the simulated populations have theoretical mean of (Y, X) as $\mu = [2, 2]$. The covariance matrices (Σ) are given in Table 1.

Table 1: Data Summary

Population	μ	Covariance Matrix (Σ)	ρ_{XY}
1	$[2, 2]$	$\begin{bmatrix} 9 & 1.9 \\ 1.9 & 4 \end{bmatrix}$	0.3209
2	$[2, 2]$	$\begin{bmatrix} 0.4 & 3.85 \\ 3.85 & 0.4 \end{bmatrix}$	0.9636
3	$[2, 2]$	$\begin{bmatrix} 0.6 & 3 \\ 3 & 0.2 \end{bmatrix}$	0.8684
4	$[2, 2]$	$\begin{bmatrix} 10 & 3 \\ 3 & 0.2 \end{bmatrix}$	0.6746

The scrambling variable S is taken to follow a normal distribution with mean equal to zero and standard deviation equal to 10% of the standard deviation of X. The reported response is given by $Z = Y + S$.

To calculate the percentage relative efficiency (PRE) for various estimators with respect to μ_Y we use the formula given by

$$PRE(\hat{\mu}_i) = \frac{MSE(\hat{\mu}_Y)}{MSE(\hat{\mu}_i)} \times 100, \quad i = Y, R, Reg, GRR, t \tag{25}$$

The simulation study has been carried out to illustrate and compare the performance of the proposed RRT estimator $\hat{\mu}_t$, compared to the mean estimator $\hat{\mu}_Y$, the ratio estimator $\hat{\mu}_R$, the regression estimator $\hat{\mu}_{Reg}$, and the generalized regression-cum-ratio estimator $\hat{\mu}_{GRR}$. For this, $50{,}000$ samples of sizes $n = 50, 100, 200,$ and 300 were selected from the populations $1, 2, 3,$ and 4 respectively, and the PRE of all the estimators are compared with the conventional RRT mean estimator $\hat{\mu}_Y$, using the formula

$$PRE(\hat{\mu}_i) = \frac{\frac{1}{50{,}000}\sum_{j=1}^{50{,}000}(\hat{\mu}_Y - \mu_Y)^2}{\frac{1}{50{,}000}\sum_{j=1}^{50{,}000}(\hat{\mu}_i - \mu_Y)^2} \times 100 \tag{26}$$

The empirical MSE and theoretical MSE, with the PRE outcomes of all estimators are reported in Tables 2–5 for populations 1–4 respectively.

Table 2: Empirical and Theoretical MSE and PRE population 1 $\rho_{xy} = 0.3209$

n	Estimator	MSE Theoretical	MSE Empirical	PRE Theoretical	PRE Empirical
50	$\hat{\mu}_Y$	0.16951454	0.16912941	100.0000	100.0000
	$\hat{\mu}_R$	0.17586399	0.18390183	96.3895	91.9673
	$\hat{\mu}_{Reg}$	0.15252218	0.14799680	111.1409	114.2791
	$\hat{\mu}_{GRR}$	0.14671045	0.24605369	115.5446	68.7679
	$\hat{\mu}_t$	0.14673853	0.14668346	115.5215	115.3023
100	$\hat{\mu}_Y$	0.08029636	0.07952336	100.0000	100.0000
	$\hat{\mu}_R$	0.08330400	0.08265997	96.3896	96.2055
	$\hat{\mu}_{Reg}$	0.07224735	0.06932824	111.1409	114.7056
	$\hat{\mu}_{GRR}$	0.07093005	0.01141653	113.2050	69.6564
	$\hat{\mu}_t$	0.07090035	0.06864207	113.2524	115.8257
200	$\hat{\mu}_Y$	0.03568727	0.03537699	100.0000	100.0000
	$\hat{\mu}_R$	0.03702400	0.03644440	96.3896	97.0712
	$\hat{\mu}_{Reg}$	0.03210993	0.03107996	111.1409	113.8257
	$\hat{\mu}_{GRR}$	0.03184811	0.04997621	112.0546	70.7877
	$\hat{\mu}_t$	0.03210993	0.03078599	112.1179	114.9126
300	$\hat{\mu}_Y$	0.02081758	0.02062086	100.0000	100.0000
	$\hat{\mu}_R$	0.02159733	0.02075667	96.3896	99.3457
	$\hat{\mu}_{Reg}$	0.01873079	0.01794152	111.1409	114.9338
	$\hat{\mu}_{GRR}$	0.01864140	0.02936566	111.6739	70.2210
	$\hat{\mu}_t$	0.01863042	0.01779003	111.7397	115.9124

Table 3: Empirical and Theoretical MSE and PRE for population 2 $\rho_{xy} = 0.9636$

n	Estimator	MSE Theoretical	MSE Empirical	PRE Theoretical	PRE Empirical
50	$\hat{\mu}_Y$	0.07563102	0.07569070	100.0000	100.0000
	$\hat{\mu}_R$	0.00581001	0.00600044	1301.7365	1261.4183
	$\hat{\mu}_{Reg}$	0.00561965	0.00549932	1345.8315	1376.3632
	$\hat{\mu}_{GRR}$	0.00558714	0.08405358	1353.6625	90.0504

	$\hat{\mu}_t$	0.00540286	0.00670650	1399.8330	1128.6176
100	$\hat{\mu}_Y$	0.03515265	0.03561302	100.0000	100.0000
	$\hat{\mu}_R$	0.00269996	0.00276024	1301.9693	1290.5372
	$\hat{\mu}_{Reg}$	0.00261325	0.00260919	1345.1699	1365.2506
	$\hat{\mu}_{GRR}$	0.00260001	0.03943915	1352.0198	90.3214
	$\hat{\mu}_t$	0.00251296	0.00285646	1398.8544	1246.9862
200	$\hat{\mu}_Y$	0.01580272	0.01582029	100.0000	100.0000
	$\hat{\mu}_R$	0.00121389	0.00121511	1301.8248	1301.9551
	$\hat{\mu}_{Reg}$	0.00113552	0.00117141	1391.6725	1350.5283
	$\hat{\mu}_{GRR}$	0.00113523	0.01747702	1392.0280	90.5205
	$\hat{\mu}_t$	0.00113013	0.00122353	1398.3099	1292.9982
300	$\hat{\mu}_Y$	0.00921825	0.00924589	100.0000	100.0000
	$\hat{\mu}_R$	0.00068280	0.00070196	1350.0659	1217.1533
	$\hat{\mu}_{Reg}$	0.00066238	0.00067739	1383.6271	1364.9215
	$\hat{\mu}_{GRR}$	0.00066228	0.01021143	1391.8772	90.5445
	$\hat{\mu}_t$	0.00065901	0.00069679	1398.8027	1342.0696

Table 4: Empirical and Theoretical MSE and PRE for population 3 ρ_{xy}=0.8684

n	Estimator	MSE Theoretical	MSE Empirical	PRE Theoretical	PRE Empirical
50	$\hat{\mu}_Y$	0.11398924	0.11364658	100.0000	100.0000
	$\hat{\mu}_R$	0.03700200	0.03770869	308.0623	301.3803
	$\hat{\mu}_{Reg}$	0.02850886	0.02740591	399.8379	414.6791
	$\hat{\mu}_{GRR}$	0.02829489	0.06670809	402.8616	170.3640
	$\hat{\mu}_t$	0.02815282	0.02799097	404.8936	406.0115
100	$\hat{\mu}_Y$	0.05399490	0.05350707	100.0000	100.0000
	$\hat{\mu}_R$	0.01752726	0.01714527	308.0623	312.0806
	$\hat{\mu}_{Reg}$	0.01350420	0.01290467	399.8379	414.6334
	$\hat{\mu}_{GRR}$	0.01345482	0.03154142	401.3053	169.6406
	$\hat{\mu}_t$	0.01338631	0.01304111	403.3591	410.2954
200	$\hat{\mu}_Y$	0.02399773	0.02375628	100.0000	100.0000
	$\hat{\mu}_R$	0.00778989	0.00762475	308.0623	311.5681
	$\hat{\mu}_{Reg}$	0.00600186	0.00581219	399.8379	408.7322
	$\hat{\mu}_{GRR}$	0.00599144	0.01410488	400.5334	168.4259
	$\hat{\mu}_t$	0.00596080	0.00582445	402.5919	407.8714
300	$\hat{\mu}_Y$	0.01399868	0.01388205	100.0000	100.0000
	$\hat{\mu}_R$	0.00454410	0.00437276	308.0623	317.4669
	$\hat{\mu}_{Reg}$	0.00350108	0.00333777	399.8379	415.9083
	$\hat{\mu}_{GRR}$	0.00349724	0.00820108	400.2771	169.2710
	$\hat{\mu}_t$	0.00347934	0.00333856	402.3362	415.8097

Table 5: Empirical and Theoretical MSE and PRE for population 4 ρ_{xy}=0.6746

n	Estimator	MSE Theoretical	MSE Empirical	PRE Theoretical	PRE Empirical
50	$\hat{\mu}_Y$	0.18028084	0.18087624	100.0000	100.0000
	$\hat{\mu}_R$	0.10760219	0.11421800	167.5438	158.3605
	$\hat{\mu}_{Reg}$	0.09916139	0.10412676	181.8055	173.7077
	$\hat{\mu}_{GRR}$	0.09677296	0.13635692	186.2926	132.6491
	$\hat{\mu}_t$	0.09666304	0.10407210	186.5044	173.7990
100	$\hat{\mu}_Y$	0.08539619	0.08480617	100.0000	100.0000
	$\hat{\mu}_R$	0.05096946	0.05270696	167.5438	160.9013
	$\hat{\mu}_{Reg}$	0.04697119	0.04900004	181.8055	173.0737
	$\hat{\mu}_{GRR}$	0.04643015	0.06408078	183.9240	132.3426
	$\hat{\mu}_t$	0.04636886	0.04887988	184.1671	173.4991
200	$\hat{\mu}_Y$	0.03795386	0.03776570	100.0000	100.0000
	$\hat{\mu}_R$	0.02265309	0.02334730	167.5438	161.7562
	$\hat{\mu}_{Reg}$	0.02087608	0.02181529	181.8055	173.1158
	$\hat{\mu}_{GRR}$	0.02076838	0.02861482	182.7483	131.9795
	$\hat{\mu}_t$	0.02073999	0.02173156	182.9984	173.7827
300	$\hat{\mu}_Y$	0.02213975	0.02220577	100.0000	100.0000
	$\hat{\mu}_R$	0.01321430	0.01365837	167.5438	162.5800
	$\hat{\mu}_{Reg}$	0.01217772	0.01277893	181.8055	173.7686
	$\hat{\mu}_{GRR}$	0.01214084	0.01681132	182.3577	132.0882
	$\hat{\mu}_t$	0.01212414	0.01272923	182.6089	174.4472

7 Interpretation of Computational Results

The following interpretations can be read from Tables 2–5, which exhibit the results of theoretical comparisons and the simulation study carried out drawing $50{,}000$ samples from all populations that are artificially generated and named Populations 1–4.

It is clear that the proposed estimator is theoretically better for all the populations considered for the sample sizes $n = 50, 100, 200, 300$ consistently. The empirical results have shown slightly better performance for $\hat{\mu}_{Reg}$ in some cases, which is a matter of further investigation. In general, our results show a significant improvement over the other estimators, especially over $\hat{\mu}_{GRR}$. Empirical investigation shows a good amount of improvement.

8 Conclusion

In this paper we have suggested an estimator for the finite population mean $\bar{Y}$ of the sensitive variable Y utilizing non-sensitive auxiliary varibale for the additive model in RRT. We have derived the bias and MSE of the suggested estimator upto the first oder of approximation. Optimum conditions are obtained at which the envisaged estimator attained the minimum MSE. The theoretical comparisions of the proposed RRT estimator with some existing competitive estimators have been presented. The expression of theoretical comparisons provide the conditions which ensures the least MSE of the proposed estimators. For empirical investigations to judge the merits of the proposed estimator, we have carried out an empirical study through simulation based on four artificially generated bivariate normal populations. The simulation study confirms that among the competiting estimatros the suggested RRT estimator has least MSE and largest PRE. Thus, we recommended the use of suggested RRT estimators for precisely estimating the finite population mean of sensutive variable. The present work can be easily extended under the stratified random sampling and other sampling designs to provide better efficient estimator. One can also attempt the problem of variance when the study variable is sensitive and information on non-sensitive auxiliary information is available.

References

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

Eichhorn, B. H., and Hayre, L. S. (1983) *Scrambled randomized response methods for obtaining sensitive quantitative data*. Journal of Statistical Planning and inference, 7(4), 307-316.

Gupta, S., Gupta, B., and Singh, S. (2002). *Estimation of sensitivity level of personal interview survey questions*. Journal of Statistical Planning and inference, 100(2), 239-247.

Gupta, S., Shabbir, J., and Sehra, S. (2010) *Mean and sensitivity estimation in optional randomized response models*. Journal of Statistical Planning and Inference, 140(10), 2870-2874.

Gupta, S., Shabbir, J., Sousa, R., and Corte-Real, P. (2012) *Estimation of the mean of a sensitive variable in the presence of auxiliary information*. Communication in Statistics-Theory and Methods, 41(13-14), 2394-2404.

Gupta, S., and Shabbir, J. (2004). *Sensitivity estimation for personal interview survey questions*. Statistica, 64(4), 643-653.

Koyuncu, N., Gupta, S., and Sousa, R. (2014). *Exponential-type estimators of the mean of a sensitive variable in the presence of non-sensitive auxiliary information*. Communications in statistics-Simulation and Computation, 43(7), 1583-1594.

Murtaza, M., Singh, S., and Hussain, Z. (2021). *Use of correlated scrambling variables in quantitative randomized response technique*. Biometrical Journal, 63(1), 134-147.

Pollock, K. H. and Bek, Y. (1976). *A comparison of three randomized response models for quantitative data*. Journal of the American statistical Association, 71(356), 884-886.

Saha, A. (2008). *A randomized response technique for quantitative data under unequal probability sampling*. Journal of Statistical Theory and Practice, 2, 589-596.

Shabbir, J., and Gupta, S. (2010). *On estimating finite population mean in simple and stratified random sampling*. Communications in Statistics-Theory and Methods, 40(2), 199-212.

Ozgul, N., and Cingi, H. (2017) *A new improved estimator of population means in partial additive randomized response models*. Hacettepe Journal of Mathematics and Statistics, 46(2), 325-338.

Waseem, Z., Khan, H., Shabbir, J., and Fatima, S. E. (2022) A generalized class of exponential type estimators for estimating the mean of the sensitive variable when using optional randomized response model. *Communications in Statistics-Simulation and Computation*, 51(12), 7602-7612.

Warner, S. L. (1965) Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63-69.