Journal of Statistical Theory and Practice (2022) 16:28
https://doi.org/10.1007/s42519-022-00251-1

ORIGINAL ARTICLE

Using Randomized Response to Estimate the Population Mean of a Sensitive Variable under the Influence of Measurement Error

Kuldeep Kumar Tiwari[1] · **Sandeep Bhougal**[1] · **Sunil Kumar**[2] · **Khalid Ul Islam Rather**[3]

Accepted: 25 February 2022 / Published online: 6 April 2022

Abstract

There are situations in survey sampling where the study characters are sensitive. Due to the sensitivity of characters, practitioners don't get the actual response. Randomized response technique (RRT) models are developed to reduce the bias raised by an evasive response on the sensitive variable. The measurement error (ME) is usually always present in the surveys so we need to study the RRT models with ME. We propose an estimator to predict the population mean of a sensitive variable in the influence of ME. The properties of the proposed estimator are studied and comparisons are made with the existing estimators. At last, a simulation study is executed to illustrate the results numerically.

Keywords Sensitive variable · Bias · Mean squared error · Measurement error · Efficiency

Mathematics Subject Classification 62D05

Kuldeep Kumar Tiwari, Sunil Kumar and Khalid Ul Islam Rather have contributed equally to this study.

✉ Sandeep Bhougal
sandeep.bhougal@smvdu.ac.in

Kuldeep Kumar Tiwari
kuldeep.smvd@gmail.com

Sunil Kumar
sunilbhougal06@gmail.com

Khalid Ul Islam Rather
khalidstat34@gmail.com

1 School of Mathematics, Shri Mata Vaishno Devi University, Katra, Jammu and Kashmir 182320, India

2 Department of Statistics, University of Jammu, Jammu, Jammu and Kashmir 180016, India

3 Division of Statistics and Computer Science, SKUAST, Jammu, Jammu and Kashmir 180009, India

1 Introduction

A variable having some sensitive information about a person or enterprise can be classified as a sensitive variable. Direct observation on the study variable is sometimes not possible in surveys because the information may be sensitive. That is, a respondent may be uncomfortable to provide the information that the interviewer required due to some personal or any reasons, e.g., the questions seeking the information regarding corruption, criminality, abortion, drug addiction, etc. To handle such situations, Warner [1] proposed a randomized response technique (RRT) to reduce the bias on evasive response. In RRT, a scramble variable which is independent of sensitive study variable and auxiliary variable is used to estimate the population mean of a sensitive variable. It is assumed that the distribution of the scramble variable is known. The interviewee is asked to provide a scrambled response to sensitive variable but give true response to the auxiliary variable. Pollock and Bek [2] proposed an additive RRT model for a quantitative sensitive variable, which is further discussed by Himmelfarb and Edgell [3]. Based on multiplicative scrambling, Eichhorn and Hayre [4] introduced an RRT model to get information on the sensitive variable. A question in the survey questionnaire may be sensitive for one respondent but not for another. That is, for the same question some respondents provide scrambled response while some of them may provide true response. Addressing this issue, Gupta, Gupta and Singh [5] proposed the concept of the optional randomized response technique (ORRT) model and explained that an ORRT model is generally more efficient than the corresponding RRT model. Gupta et al. [6] show that there is no extra loss of privacy in using ORRT models as compared to the corresponding RRT models. The problem of estimation of population mean of a sensitive variable considered by many authors as Wu, Tian and Tang [7], Gupta, Shabbir and Sehra [8], Sousa et al. [9], Gupta et al. [10], Koyuncu, Gupta and Sousa [11], Tarray and Singh [12], Shahzad et al. [13], Mushtaq and Noor-ul-Amin [14], Saleem and Sanaullah [15], Su et al. [16], etc. None of these studies have examined the impact of measurement errors (ME) that happen commonly in surveys. ME is the difference between observed and true value. ME is one of the very common contributors to non-sampling error. The problem of ME is inherent in survey sampling. It may increase in the case of a sensitive issue as the surveyor has to deal with an evasive response. And so, RRT in the presence of ME seeks attention for an extensive study. Very limited efforts have been made to estimate the finite population mean of a sensitive variable in the presence of ME. Recently some researchers focuses on this issue. Blattman et al. [17] developed a survey validation technique for qualitative variables to check for ME when dealing with sensitive attributes. Khalil, Gupta and Hanif [18] propose a study in stratified sampling in the presence of scrambled response and ME. Khalil, Zhang and Gupta [19] use the ORRT model under ME to study some estimators of population mean. Zahid and Shabbir [20] use dual auxiliary variable to estimate population mean of a sensitive variable in the presence of ME. Onyango, Oduor and Odundo [21] propose an estimator using RRT and ME in double stratified sampling. Zhang, Khalil and Gupta [22] propose a study on mean estimation comprising sensitive variable, ME and non-response. Some more recent work on the sensitive issue in the presence of ME are Khalil, Noor-ul-Amin and Hanif [23], Zhang, Khalil and Gupta [24], Zahid, Shabbir and Alamri [25].

We propose an estimator of the population mean of a sensitive variable using a non-sensitive auxiliary variable when there is a presence of ME in the study.

2 Notations and Existing Estimators

Let $\Omega = \{\Omega_1, \Omega_{2,..,\Omega_N}\}$ be a finite population of size N. A sample of size n is taken using simple random sampling without replacement from Ω. Let Y be the sensitive study variable, which cannot be observed directly and X be a non-sensitive auxiliary variable correlated with Y. Let S be a scrambling variable independent of Y and X. We assume that S has a known distribution with mean zero and variance σ_s^2. Here we use the additive RRT model of Pollock and Bek [2]. In the survey, the scrambled response variable Z is observed as $Z = Y + S$. It is assumed that $E(S) = 0$ and $E(Z) = E(Y) + E(S)$, so $E(Z) = E(Y)$. That is, $\bar{Z} = \bar{Y}$. So we have to estimate the population mean of the scrambled response variable Z and that will be the population mean of sensitive variable Y. The degree of protection for the additive RRT model is $\Delta = E(Z - Y)^2 = \sigma_s^2$. For detail, see Yan, Wang and Lai [26] and Saleem, Sanaullah and Hanif [27].

Let (x_i, z_i) be the observed values and (X_i, Z_i) be the true values of the variables X, Z. Let u be the ME on Z and v be the ME on X. The measurement errors on i^{th} observed unit are $u_i = z_i - Z_i$ and $v_i = x_i - X_i$. Since the measurement errors are independent of each other and there is both under and over reporting so it is assumed that u and v are uncorrelated with mean zero and variances σ_u^2 and σ_v^2, respectively.

Some other notations used in the article are: $\bar{X}$, $\bar{Y}$, $\bar{Z}$ and $\bar{x}$, $\bar{y}$, $\bar{z}$ population and sample means of X, Y and Z, respectively, coefficient of skewness for auxiliary variable $\beta_1(x)$, coefficient of kurtosis for auxiliary variable $\beta_2(x)$, σ_z^2 population variance for Z, σ_x^2 population variance for X, ρ_{zx} correlation coefficient between Z and X, $C_x = \frac{\sigma_x}{\bar{X}}$, $C_z = \frac{\sigma_z}{\bar{Z}}$, $\lambda = \frac{1}{n} - \frac{1}{N}$.

The need of efficiency motivate the researchers to work on development of estimators. Sample mean per unit estimator $\hat{\bar{Y}}_1 = \bar{y}$ is the usual estimator of population mean. To get better efficiency over $\hat{\bar{Y}}_1$, Cochran [28] uses ratio method of estimation as $\hat{\bar{Y}}_2 = \bar{y}\frac{\bar{X}}{\bar{x}}$. It is found that $\hat{\bar{Y}}_2$ works better than $\hat{\bar{Y}}_1$ when there is high positive correlation between Y and X. Using coefficient of variation of auxiliary variable, Sisodia and Dwivedi [29] propose $\hat{\bar{Y}}_2 = \bar{y}\left[\frac{\bar{X}+C_x}{\bar{x}+C_x}\right]$ and show its worth over $\hat{\bar{Y}}_2$. Following Sisodia and Dwivedi [29], Upadhaya and Singh [30], Singh [31], Singh and Tailor [32] and Singh et al. [33] propose various estimators using coefficient of kurtosis, coefficient of skewness, correlation coefficient, standard deviation, etc. Some other works in the process are Kadilar and Cingi [34], Yan and Tian [35], Abid et al. [36], etc.

The RRT sample mean per unit estimator is $\hat{\mu}_1 = \bar{z}$. The MSE of $\hat{\mu}_1$ is

$$MSE(\hat{\mu}_1) = \lambda\sigma_z^2.$$

When there is sensitivity on the study variable, Sousa et al. [9] defined the ratio estimator as $\hat{\mu}_2 = \bar{z}\left(\frac{\bar{X}}{\bar{x}}\right)$ and obtained the bias and MSE of $\hat{\mu}_2$ up to first order of approximation as

$$Bias(\hat{\mu}_2) = \lambda\bar{Z}\left(C_x^2 - \rho_{zx}C_zC_x\right) \tag{1}$$

$$MSE(\hat{\mu}_2) = \lambda\bar{Z}^2\left(C_z^2 + C_x^2 - 2\rho_{zx}C_zC_x\right). \tag{2}$$

Further, Sousa et al. [9] propose four more estimators to estimate the population mean of the sensitive variable using auxiliary information as

$$\hat{\mu}_3 = \bar{z}\left[\frac{\bar{X}+\beta_1(x)}{\bar{x}+\beta_1(x)}\right]$$

$$\hat{\mu}_4 = \bar{z}\left[\frac{\bar{X}+\beta_2(x)}{\bar{x}+\beta_2(x)}\right]$$

$$\hat{\mu}_5 = \bar{z}\left[\frac{\beta_1(x)\bar{X}+\beta_2(x)}{\beta_1(x)\bar{x}+\beta_2(x)}\right]$$

$$\hat{\mu}_6 = \bar{z}\left[\frac{\beta_2(x)\bar{X}+\beta_1(x)}{\beta_2(x)\bar{x}+\beta_1(x)}\right].$$

The bias and MSE of $\hat{\mu}_3$, $\hat{\mu}_4$, $\hat{\mu}_5$ and $\hat{\mu}_6$ are

$$Bias(\hat{\mu}_i) = \lambda\bar{Y}\left(q_i^2C_x^2 - q_i\rho_{zx}C_zC_x\right) \tag{3}$$

$$MSE(\hat{\mu}_i) = \lambda\bar{Y}^2\left(C_z^2 + q_i^2C_x^2 - 2q_i\rho_{zx}C_zC_x\right) \tag{4}$$

where $i = 3, 4, 5, 6$ and $q_3 = \frac{\bar{X}}{\bar{X}+\beta_1(x)}$, $q_4 = \frac{\bar{X}}{\bar{X}+\beta_2(x)}$, $q_5 = \frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+\beta_2(x)}$, $q_6 = \frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+\beta_1(x)}$.

Using a simulation study, Sousa et al. [9] show that $\hat{\mu}_3$ achieve modest gain over other member estimators. The performance of these estimators in the presence of ME will be discussed hereafter.

The MSEs of the estimators $\hat{\mu}_i$, $i = 1, 2, \ldots, 6$ in the presence of measurement error can be derived as

$$MSE(\hat{\mu}_i) = \lambda(\sigma_z^2 + \sigma_u^2) + q_i^2\lambda(\sigma_x^2 + \sigma_v^2) - (2q_i\lambda\rho_{zx}\sigma_z\sigma_x) \tag{5}$$

where $i = 1, 2, 3, 4, 5, 6$ and $q_1 = 0$, $q_2 = \frac{\bar{Z}}{\bar{X}}$, $q_3 = \frac{\bar{X}}{\bar{X}+\beta_1(x)}$, $q_4 = \frac{\bar{X}}{\bar{X}+\beta_2(x)}$, $q_5 = \frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+\beta_2(x)}$, $q_6 = \frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+\beta_1(x)}$.

3 Proposed Estimator

As it is known that ratio estimators usually do not perform better than regression type estimators and so to get a better estimate than Sousa et al. [9], we propose a class of difference type randomized response estimator to estimate the population mean of a sensitive variable Y using a non-sensitive auxiliary variable X in the presence of measurement error. The proposed estimator is

$$t_{(\delta)} = \left[\eta_1 \bar{z} + \eta_2 (\bar{X} - \bar{x})\right] \exp\left[\frac{\delta(\bar{X} - \bar{x})}{\bar{X} + \bar{x}}\right] \tag{6}$$

where η_1, η_2 are constants to be optimize for minimum MSE of $t_{(\delta)}$ and δ is suitable constant which takes real or parametric values. By giving different values to δ, we can generate new members of the class of estimators $t_{(\delta)}$, e.g., for $\delta = c(constant)$, the member estimator will be denoted by $t_{(c)}$.

4 Bias and MSE

To derive the bias and mean squared error (MSE) of the proposed class of estimator, let

$\omega_z = \sum_{i=1}^{N} \left(Z_i - \bar{Z}\right)$, $\omega_u = \sum_{i=1}^{N} u_i$, $\omega_x = \sum_{i=1}^{N} \left(X_i - \bar{X}\right)$ and $\omega_v = \sum_{i=1}^{N} v_i$.

Adding ω_z and ω_u, we get $\omega_z + \omega_u = \sum_{i=1}^{N} \left(Z_i - \bar{Z}\right) + \sum_{i=1}^{N} u_i$. Now, divide both sides by n and simplify, we get

$$\bar{z} = \bar{Z} + \xi_z \text{ where } \xi_z = \frac{1}{n}(\omega_z + \omega_u).$$

Similarly, using ω_x and ω_v we get

$$\bar{x} = \bar{X} + \xi_x \text{ where } \xi_x = \frac{1}{n}(\omega_x + \omega_v).$$

The error terms to get bias and MSE of estimators are $e_z = \frac{\bar{z}-\bar{Z}}{\bar{Z}} = \frac{\xi_z}{\bar{Z}}$ and $e_x = \frac{\bar{x}-\bar{X}}{\bar{X}} = \frac{\xi_z}{\bar{X}}$.

The expected values are $E(e_z^2) = \frac{\lambda(\sigma_z^2+\sigma_u^2)}{\bar{Z}^2}$, $E(e_x^2) = \frac{\lambda(\sigma_x^2+\sigma_v^2)}{\bar{X}^2}$ and $E(e_z e_x) = \frac{\lambda \rho_{zx}\sigma_z\sigma_x}{\bar{Z}\bar{X}}$.

Now, express proposed estimator $t_{(\delta)}$ in terms of errors e_z and e_x, we have

$$t_{(\delta)} = \left[\eta_1 \bar{Z}(1 + e_z) + \eta_2 \left\{\bar{X} - \bar{X}(1 + e_x)\right\}\right] \exp\left[\frac{\delta\{\bar{X} - \bar{X}(1 + e_x)\}}{\bar{X} + \bar{X}(1 + e_x)}\right]$$

$$t_{(\delta)} = \left[\eta_1 \bar{Z} + \eta_1 \bar{Z} e_z - \eta_2 \bar{X} e_x\right] \exp\left[\frac{-\delta e_x}{2}\left(1 + \frac{e_x}{2}\right)^{-1}\right].$$

Assuming $\mid e_x \mid < 1$ expand above equation and terminate the terms having e's degree greater than two and simplify, using $\bar{Z} = \bar{Y}$ we get

$$t_{(\delta)} - \bar{Y} = (\eta_1 - 1)\bar{Y} + \eta_1 \bar{Y} e_z - \left(\eta_2 \bar{X} + \frac{\eta_1 \bar{Y}\delta}{2}\right) e_x + \left[\frac{\eta_2 \bar{X}\delta}{2} + \left(1 + \frac{\delta}{2}\right)\frac{\eta_1 \bar{Y}\delta}{4}\right] e_x^2 - \frac{\eta_1 \bar{Y}\delta}{2} e_z e_x. \quad (7)$$

Taking expectation on both sides of Eq. (7), we get the bias.

$$Bias(t_{(\delta)}) = (\eta_1 - 1)\bar{Y} + \frac{1}{2\bar{X}}\left[\left\{\eta_2 \delta + \left(1 + \frac{\delta}{2}\right)\frac{\eta_1 \bar{Y}\delta}{2\bar{X}}\right\} \times \lambda(\sigma_x^2 + \sigma_v^2) - \eta_1 \delta \lambda \rho_{zx} \sigma_z \sigma_x\right]. \quad (8)$$

Squaring Eq. (7) and terminate terms having e's degree greater than two and simplify, we get

$$(t_{(\delta)} - \bar{Y})^2 = (\eta_1 - 1)^2 \bar{Y}^2 + \eta_1^2 \bar{Y}^2 e_z^2 + \left[\eta_2^2 + \eta_1^2\left(\frac{\delta R^2}{2} + \frac{\delta^2 R^2}{4}\right) + 2\eta_1 \eta_2 \delta R\right]\bar{X}^2 e_x^2 - [2\eta_1 \eta_2 + 2\delta \eta_1^2 R]\bar{X}\bar{Y} e_z e_x \quad (9)$$

where $R = \frac{\bar{Y}}{\bar{X}}$.

Taking expectation on both sides of Eq. (9), we get

$$MSE(t_{(\delta)}) = (\eta_1 - 1)^2 \bar{Y}^2 + \eta_1^2 \nu_1 + \left[\eta_2^2 + \eta_1^2\left(\frac{\delta R^2}{2} + \frac{\delta^2 R^2}{4}\right) + 2\eta_1 \eta_2 \delta R\right]\nu_2 - [2\eta_1 \eta_2 + 2\delta \eta_1^2 R]\nu_3 \quad (10)$$

where $\nu_1 = \lambda(\sigma_z^2 + \sigma_u^2)$, $\nu_2 = \lambda(\sigma_x^2 + \sigma_v^2)$ and $\nu_3 = \lambda \rho_{zx} \sigma_z \sigma_x$.

$$MSE(t_{(\delta)}) = \bar{Y}^2 - 2\bar{Y}^2 \eta_1 + \left[\bar{Y}^2 + \nu_1 + \left(\frac{\delta R^2}{2} + \frac{\delta^2 R^2}{4}\right)\nu_2 - 2\delta R \nu_3\right]\eta_1^2 + \nu_2 \eta_2^2 + (2\delta R \nu_2 - 2\nu_3)\eta_1 \eta_2. \quad (11)$$

The optimum values of η_1 and η_2 to minimize $MSE(t_{(\delta)})$ are
$\eta_{1_O} = \frac{4\bar{Y}^2 \nu_2}{4\bar{Y}^2 \nu_2 + 4\nu_1 \nu_2 - 4\nu_3^2 + 2\delta R^2 \nu_2^2 - 3\delta^2 R^2 \nu_2^2}$ and
$\eta_{2_O} = \frac{4\bar{Y}^2 \nu_3 - 4\bar{Y}^2 \delta R \nu_2}{4\bar{Y}^2 \nu_2 + 4\nu_1 \nu_2 - 4\nu_3^2 + 2\delta R^2 \nu_2^2 - 3\delta^2 R^2 \nu_2^2}$.

For optimum values of η_1 and η_2, the minimum MSE of $t_{(\delta)}$ is

$$MSE_{min}(t_{(\delta)}) = \bar{Y}^2 - \frac{4\bar{Y}^4\nu_2}{4\bar{Y}^2\nu_2 + 4\nu_1\nu_2 - 4\nu_3^2 + 2\delta R^2\nu_2^2 - 3\delta^2 R^2\nu_2^2}. \quad (12)$$

As $MSE_{min}(t_{(\delta)})$ depends on δ so the optimum MSE of any member $t_{(c)}$ of the class of estimators $t_{(\delta)}$ can be obtained by putting $\delta = c$ in Eq. (12). So, for $\delta = c$ the member estimator is $t_{(c)}$ and optimum MSE is $MSE_{min}(t_{(c)})$.

5 Efficiency Comparison

The proposed class of estimators $t_{(\delta)}$ will be more efficient than the estimators $\hat{\mu}_i$, $i = 1, 2, \ldots, 6$ whenever the condition $MSE_{min}(t_{(\delta)}) < MSE(\hat{\mu}_i)$ satisfied.

From Eqs. (5) and (12), we found that $MSE_{min}(t_{(\delta)}) < MSE(\hat{\mu}_i)$ if

$$\bar{Y}^2 + 2q_i\nu_3 < \nu_1 + q_i^2\nu_2 + \frac{4\bar{Y}^4\nu_2}{4\bar{Y}^2\nu_2 + 4\nu_1\nu_2 - 4\nu_3^2 + 2\delta R^2\nu_2^2 - 3\delta^2 R^2\nu_2^2}$$

where $i = 1, 2, \ldots, 6$ and $\nu_1 = \lambda(\sigma_z^2 + \sigma_u^2)$, $\nu_2 = \lambda(\sigma_x^2 + \sigma_v^2)$, $\nu_3 = \lambda\rho_{zx}\sigma_z\sigma_x$.

6 Monte Carlo Simulation

In this section, we do a simulation to see the performance of the proposed estimator against the existing estimator. For that, we have generated three different populations using R software. The scrambling variable S is employed to get the observation from the bivariate normal population as $Z = Y + S$. Also, there is measurement error u and v on z and x, respectively. Here, μ is the population mean vector as $\mu = \begin{bmatrix} \bar{X} \\ \bar{Y} \end{bmatrix}$ and $\sum$ is the covariance matrix. We have made 15000 replication to get a reliable result. The input data used to derive the three different populations and their descriptive statistics are given below.

6.1 Population 1

$N = 250$, $n = 80$, $\mu = \begin{bmatrix} 40 \\ 20 \end{bmatrix}$, $\sum = \begin{bmatrix} 75 & 60 \\ 60 & 55 \end{bmatrix}$, $S = rnorm(N, 0, 10)$, $u = rnorm(N, 0, 7)$ and $v = rnorm(N, 0, 9)$. Derived parametric values are $\bar{Z} = 20.0074$, $\bar{X} = 40.0029$, $\sigma_z = 12.4384$, $\sigma_x = 8.6517$, $\sigma_u = 6.9926$, $\sigma_v = 8.9946$, $\rho_{zx} = 0.5557$, $\beta_1(x) = 0.00085$, $\beta_2(x) = 2.97348$.

6.2 Population 2

$N = 1000$, $n = 150$, $\mu = \begin{bmatrix} 2300 \\ 200 \end{bmatrix}$, $\sum = \begin{bmatrix} 12000 & 2900 \\ 2900 & 1500 \end{bmatrix}$, $S = rnorm(N, 0, 25)$, $u = rnorm(N, 0, 10)$ and $v = rnorm(N, 0, 13)$. Derived parametric values are

Table 1 PRE of the estimators with respect to $\hat{\mu}_1$

Estimator	PRE		
	Population 1	Population 2	Population 3
$\hat{\mu}_1$	100	100	100
$\hat{\mu}_2$	111.4157	129.1675	112.4433
$\hat{\mu}_3$	111.4160	129.1675	112.4421
$\hat{\mu}_4$	112.0693	129.1295	111.3223
$\hat{\mu}_5$	100.3316	113.5626	100.2688
$\hat{\mu}_6$	111.4158	129.1675	112.4429
$t_{(-2)}$	821.4720	238.1295	171.6113
$t_{(-1)}$	154.7481	179.0240	134.1503
$t_{(0)}$	113.1446	160.8748	122.2855
$t_{(1)}$	119.5684	164.2041	124.4751
$t_{(2)}$	198.6460	192.0222	142.5413

$\bar{Z} = 200.0065$, $\bar{X} = 2300.0016$, $\sigma_z = 41.5169$, $\sigma_x = 109.5119$, $\sigma_u = 9.9999$, $\sigma_v = 12.9915$, $\rho_{zx} = 0.6370$, $\beta_1(x) = 0.00116$, $\beta_2(x) = 2.99341$.

6.3 Population 3

$N = 100$, $n = 25$, $\mu = \begin{bmatrix} 15 \\ 3 \end{bmatrix}$, $\sum = \begin{bmatrix} 20 & 12 \\ 12 & 14 \end{bmatrix}$, $S = rnorm(N, 0, 2)$, $u = rnorm(N, 0, 3)$ and $v = rnorm(N, 0, 5)$. Derived parametric values are $\bar{Z} = 3.0047$, $\bar{X} = 15.0018$, $\sigma_z = 4.2342$, $\sigma_x = 4.4610$, $\sigma_u = 2.9902$, $\sigma_v = 4.9886$, $\rho_{zx} = 0.6304$, $\beta_1(x) = 0.00301$, $\beta_2(x) = 2.93465$.

The percent relative efficiency (PRE) of an estimator with respect to RRT sample mean per unit estimator $\hat{\mu}_1 = \bar{z}$ is defined as

$$PRE(\hat{\mu}_1,\ .) = \frac{MSE(\hat{\mu}_1)}{MSE(.)} \times 100. \tag{13}$$

The steps of the simulation process are:

Step 1 Using R software, generate a random population by giving inputs to N, μ, $\sum$, S, u and v.
Step 2 Use scrambling variable S as $Z = Y + S$ to make study variable sensitive.
Step 3 Derive the required parametric values from the generated population.
Step 4 To get a stable value to the parameters, replicate Step 1 to Step 3 up to 15000 times and record it.
Step 5 Use the average of 15000 values to calculate the MSEs of the estimators.
Step 6 Calculate PREs of the estimators by using MSEs from Step 5 and Eq. (13).

From Table 1, we can see that the PRE values of the proposed estimator $t_{(\delta)}$ for $\delta = -2, -1, 0, 1, 2$ are higher than the all considered existing estimators $\hat{\mu}_i$; i=1,2,…,6. This conclude that the proposed class of estimators $t_{(\delta)}$ is more efficient than the existing estimators.

7 Conclusion

The article presents a study on the estimation of population mean of a sensitive variable in the influence of measurement error. An estimator of the population mean is proposed under randomized response technique with measurement error. The expressions for bias and mean squared error derived up to the first order of approximation. The theoretical comparison is made with the existing estimators. A simulation study is performed to see the results numerically. The theoretical and simulation results show that the proposed class of estimators are more efficient than the estimators of Sousa et al. [9].

Acknowledgements The authors are thankful to Professor Sat N. Gupta, Editor-in-Chief and the learned referees for their suggestions, which improved the quality of the paper.

Declarations

Conflict of interest There is no conflict of interest to disclose.

References

1. Warner SL (1965) Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc 60(309):63–69. https://doi.org/10.1080/01621459.1965
2. Pollock KH, Bek Y (1976) A comparison of three randomized response models for quantitative data. J Am Stat Assoc 71(356):884–886. https://doi.org/10.2307/2286855
3. Himmelfarb S, Edgell SE (1980) Additive constants model: a randomized response technique for eliminating evasiveness to quantitative response questions. Psychol Bull 87(3):525–530. https://doi.org/10.1037/0033-2909.87.3.525
4. Eichhorn BH, Hayre LS (1983) Scrambled randomized response methods for obtaining sensitive quantitative data. J Statist Plan Inference 7(4):307–316. https://doi.org/10.1016/0378-3758(83)90002-2
5. Gupta SN, Gupta BC, Singh S (2002) Estimation of sensitivity level of personal interview survey questions. J Statist Plan Inference 100(2):239–247. https://doi.org/10.1016/S0378-3758(01)00137-9 J Statist Plan Inference 100(2):239–247. https://doi.org/10.1016/S0378-3758(01)00137-9
6. Gupta S, Samridhi M, Shabbir J, Khalil S (2018) A unified measure of respondent privacy and model efficiency in quantitative rrt models. J Statist Theory Practice 12(3):506–511. https://doi.org/10.1080/15598608.2017.1415175
7. Wu JW, Tian GL, Tang ML (2008) Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika 67:251–263. https://doi.org/10.1007/s00184-007-0131-x
8. Gupta S, Shabbir J, Sehra S (2010) Mean and sensitivity estimation in optional randomized response models. J Statist Plan Inference 140(10):2870–2874. https://doi.org/10.1016/j.jspi.2010.03.010
9. Sousa R, Shabbir J, Real PC, Gupta S (2010) Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. J Statist Theory Practice 4(3):495–507. https://doi.org/10.1080/15598608.2010.10411999
10. Gupta S, Shabbir J, Sousa R, Real PC (2012) Estimation of the mean of a sensitive variable in the presence of auxiliary information. Commun Statist-Theory Methods 41(13–14):2394–2404. https://doi.org/10.1080/03610926.2011.641654
11. Koyuncu N, Gupta S, Sousa R (2014) Exponential type estimators of the mean of a sensitive variable in the presence of non-sensitive auxiliary information. Commun Statist-Simul Comput 43(7):1583–1594. https://doi.org/10.1080/03610918.2012.737492
12. Tarray TA, Singh HP (2017) An improved estimation procedure of the mean of a sensitive variable using auxiliary information. Biostat Biometr Open Acc J 3(2):26–33. https://doi.org/10.19080/BBOAJ.2017.03.555607

13. Shahzad U, Hanif M, Koyuncu N, Luengo AVG (2019) A regression type estimators for mean estimation under ranked set sampling alongside the sensitivity issue. Commun Fac Sci Univ Ank Ser A1 Math Stat 68(2):2037–2049. https://doi.org/10.31801/cfsuasmas.586057
14. Mushtaq N, Noor-ul-Amin M (2020) Joint influence of double sampling and randomized response technique on estimation method of mean. Appl Math 10(1):12–19. https://doi.org/10.5923/j.am.20201001.03
15. Saleem I, Sanaullah A (2021) Estimation of mean of a sensitive variable using efficient exponential-type estimators in stratified sampling. J Statist Comput Simul 92:232–248. https://doi.org/10.1080/00949655.2021.1940182
16. Su S, Salinas VI, Zamora ML, Sedory SA, Singh S (2021) Randomized response sampling with applications to tracking drugs for better life. Stat Sin 31:1–20. https://doi.org/10.5705/ss.202016.0177
17. Blattman C, Jamison J, Koroknay-Palicz T, Rodrigues K, Sheridan M (2016) Measuring the measurement error: a method to qualitatively validate survey data. J Dev Econ 120:99–112. https://doi.org/10.1016/j.jdeveco.2016.01.005
18. Khalil S, Gupta S, Hanif M (2018) Estimation of finite population mean in stratified sampling using scrambled responses in the presence of measurement errors. Commun Statist - Theory Methods 48(6):1553–1561. https://doi.org/10.1080/03610926.2018.1435817
19. Khalil S, Zhang Q, Gupta S (2019) Mean estimation of sensitive variables under measurement errors using optional rrt models. Commun Statist - Simul Comput 50:1417–1426. https://doi.org/10.1080/03610918.2019.1584298
20. Zahid E, Shabbir J (2019) Estimation of finite population mean for a sensitive variable using dual auxiliary information in the presence of measurement errors. PLoS One 14(2):e0212111. https://doi.org/10.1371/journal.pone.0212111
21. Onyango R, Oduor B, Odundo F (2021) Joint influence of measurement errors and randomized response technique on mean estimation under stratified double sampling. Open J Math Sci 5:192–199. https://doi.org/10.30538/oms2021.0156
22. Zhang Q, Khalil S, Gupta S (2021) Mean estimation of sensitive variables under non-response and measurement errors using optional rrt models. J Statist Theory Practice 15:1–15. https://doi.org/10.1007/s42519-020-00135-2
23. Khalil S, Noor-ul-Amin M, Hanif M (2018) Estimation of population mean for a sensitive variable in the presence of measurement error. Statist Manag Syst 21(1):81–91. https://doi.org/10.1080/09720510.2017.1367478
24. Zhang Q, Khalil S, Gupta S (2021) Mean estimation in the simultaneous presence of measurement errors and non-response using optional rrt models under stratified sampling. J Statist Comput Simul 91:3492–3504. https://doi.org/10.1080/00949655.2021.1941018
25. Zahid E, Shabbir J, Alamri OA (2022) A generalized class of estimators for sensitive variable in the presence of measurement error and non-response under stratified random sampling. J King Saud Univ - Sci 34:1–10. https://doi.org/10.1016/j.jksus.2021.101741
26. Yan Z, Wang J, Lai J (2008) An efficiency and protection degree-based comparison among the quantitative randomized response strategies. Commun Statist - Theory Methods 38(3):400–408. https://doi.org/10.1080/03610920802220785
27. Saleem I, Sanaullah A, Hanif M (2019) Double-sampling regression-cum-exponential estimator of the mean of a sensitive variable. Math Popul Stud 26(3):163–182. https://doi.org/10.1080/08898480.2019.1565273
28. Cochran WG (1940) The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. J Agric Sci 30(2):262–275. https://doi.org/10.1017/S0021859600048012
29. Sisodia BVS, Dwivedi BK (1981) A modified ratio estimator using coefficient of variation of auxiliary variable. J Indian Soc Agric Statist 33:13–18
30. Upadhyaya LN, Singh HP (1999) Use of transformed auxiliary variable in estimating the finite population mean. Biom J 41(5):627–636
31. Singh GN (2003) On the improvement of product method of estimation in sample surveys. J Indian Soc Agric Statist 56(3):267–275
32. Singh HP, Tailor R (1999) Use of known correlation coefficient in estimating the finite population mean. Statist Trans 6:555–560
33. Singh HP, Tailor R, Tailor R, Kakran MS (2004) An improved estimator of population mean using power transformation. J Indian Soc Agric Statist 58:223–230

34. Kadilar C, Cingi H (2004) Ratio estimators in simple random sampling. Appl Math Comput 151(5):893–902. https://doi.org/10.1016/S0096-3003(03)00803-8
35. Yan Z, Tian B (2010) Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. In: Zhu R, Zhang Y, Liu B, Liu C (eds) Inform Comput Appl. Springer, Berlin, Heidelberg, pp 103–110
36. Abid M, Abbas N, Nazir HZ, Lin Z (2016) Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. Revista Colombiana Estadist 39(1):63–79. https://doi.org/10.15446/rce.v39n1.55139