VFAST Transactions on Mathematics
http://vfast.org/journals/index.php/VTM@ 2024 ISSN(e): 2309-0022, ISSN(p): 2411-6343
Volume 12, Number 1, January-June, 2024 pp: 176-188

# A versatile optional randomized response technique for use with sensitive surveys

**Muhammad Azeem**[1*] **and Abdul Salam**[1]

[1]Department of Statistics, University of Malakand, Khyber Pakhtunkhwa, Pakistan

***Keywords***:*Efficiency, Mean Estimator, Privacy Protection, Scrambling Variable, Randomized Response* . ***Subject Classification***: *2020 Mathematics Subject Classification: 62D05, 62F10.*

***Journal Info***:
*Submitted: March 16, 2024*
*Accepted: April 22, 2024*
*Published: April 28, 2024*

**Abstract** In data collection from human participants, researchers in almost every survey get refusals and/or false responses from the respondents. Such refusals and false reporting are particularly common in sample surveys where the participants are asked to answer questions on sensitive topics such as cheating in examination, illegal income, marks obtained in last examination, students' satisfaction from the teaching method, and amount of money spent on luxury items, etc. A popular approach to deal with the problem of refusals and untruthful responses is the randomized response technique. This paper introduces a randomized response model which is more precise than the available models. The proposed model combines both additive and multiplicative scrambling techniques to achieve improved efficiency. We show the improvement using different values of constants. The proposed randomized scrambling procedure guarantees the privacy protection of the respondents for motivating them to participate in the survey.

***Correspondence Author Email Address:**
azeemstats@uom.edu.pk

## 1 Introduction

In sample surveys, the respondents often get confused and do not provide truthful information when they are asked questions related to sensitive issues such as illegal income, monthly salary, purchase of luxury items, smoking habits, and cheating in an examination, etc. Survey participants often refuse to report truthful information about sensitive characteristics and their refusals result in a high non-response

This work is licensed under a Creative Commons Attribution 3.0 License.

rate. Besides refusals, false responses are also common in sensitive surveys. Attempting to minimize the non-response rate in sample surveys, Warner [17] presented a procedure called the randomized response technique, which got a wide popularity among survey researchers. The Warner's [17] procedure was applicable only to binary-type data. Warner [18] presented a scrambling technique for collection of reliable data on quantitative sensitive variables. Gupta et al. [6] presented the concept of optional scrambling models, offering the choice of true response to the survey participants. Diana and Perri [4] developed a linear combination scrambling model to scramble the responses. Gjestvang and Singh [5] presented an additive optional randomization technique. Hussain et al. [10] presented a new randomized response method for use with sensitive surveys. Narjis and Shabbir [15] improved the randomization technique of Gjestvang and Singh [5] by adding the option of true response. Khalil et al. [11] studied the impact of observational errors on the mean estimators of a sensitive variable. A recent study of Azeem [2] presented an exponential scrambling technique for collection of data in sample surveys. Mahdizadeh and Zamanzade [12] discussed reliability estimation under ranked set sampling design. In another study, Mahdizadeh and Zamanzade [13] analyzed dynamic reliability estimation methods.

The two main objectives of using the randomized response technique are protection of privacy and model's efficiency. The usefulness of a given randomized response technique can be assessed by these two features. In order to jointly quantify the level of privacy and model's efficiency into a single measure, Gupta et al. [7] suggested a unified metric of privacy level and efficiency for evaluation of randomized response techniques.

Abbasi et al. [1] developed a partial randomized response model under ranked set sampling scheme. Mahdizadeh and Zamanzade [14**?** ] developed novel estimators of the area under the receiver operating characteristics curve. Zaman et al. [21] suggested that focused group discussions can be used for efficient estimation of population proportion. Using robust regression approach, Zaman et al. [22] developed a Hartley-Ross type estimator and analyzed its properties.

For further studies regarding randomized response models, one can refer to Yan et al. [19], Young et al. [20], Gupta et al. [8], Zhang et al. [23], and Azeem and Salam [3].

Gupta et al. [9] suggested an optional model which offers the survey participants the option of using an additive-type scrambling variable or using an additive as well as a multiplicative variable simultaneously. The Gupta et al. [9] method achieved improvement over the Diana and Perri [4] method in the respondents' privacy levels and in the efficiency of the sample mean. The model of Gupta et al. [9] gives the following three options to the respondents.

(i) Report the true response,
(ii) Use additive scrambling only,
(iii) Use additive and multiplicative scrambling simultaneously.

Motivated by Gupta et al. [9], the proposed scrambling model adds a fourth option for the respondents – the use of a multiplicative only scrambling. Thus, the respondents have four options:

(i) Report the true response,
(ii) Use additive scrambling only,
(iii) Use multiplicative scrambling only,
(iv) Use both additive and multiplicative scrambling.

The inclusion of multiplicative-only scrambling makes the randomized response technique more versatile as it gives an additional option to the respondents to motivate them to participate in the survey. The improvement of the proposed model over the Gupta et al. [9] model in terms of privacy protection and efficiency is shown in Table 2-3.

# 2 Selected Models from the Literature

Suppose that the population under study has a total of $N$ units and a suppose a random sample of n units is taken with replacement from the population under study. Let $Y$ denotes the sensitive variable under study and let S be an additive-type random / scrambling variable. It is further assumed that $E(Y_i) = \mu_Y$ ,$E(S) = 0$ ,$V(Y_i) = \sigma_Y^2$ , $V(S) = \sigma_S^2$ . Further, let $T$ denote another random variable of multiplicative-type such that , and . Here, and denote the population variance of variable $Y$, $T$, and $S$, respectively, and let denote the population mean of the variable Y. We also assume that, for the sake of respondents' privacy protection, all variables are uncorrelated with each other. This section presents some of the already available models for comparison with the proposed model.

## 2.1 2.1 Warner's [18] Scrambling Model

The scrambling model of Warner [18] may be mathematically written as:

$$Z = Y + S, \tag{1}$$

where $Z$ denotes the response recorded by the researcher. An unbiased estimator of the mean of Y by using the Warner's [18] model may be written as:

$$\hat{\mu}_W = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{2}$$

The sampling variance of $\hat{\mu}_W$ may be derived as:

$$Var(\hat{\mu}_W) = \frac{1}{n}\left(\sigma_Y^2 + \sigma_S^2\right). \tag{3}$$

## 2.2 Diana and Perri [4] Model

The model of Diana and Perri [4] may be expressed as:

$$Z = TY + S. \tag{4}$$

An unbiased mean estimator using the Diana and Perri [4] linear combination model may be written as:

$$\hat{\mu}_{DP} = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{5}$$

The variance of the mean estimator $\hat{\mu}_{DP}$ can be derived as:

$$Var(\hat{\mu}_{DP}) = \frac{1}{n}\left[\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_Y^2 + \sigma_S^2\right]. \tag{6}$$

## 2.3 Narjis and Shabbir [15] Randomization Technique

Narjis and Shabbir [15] suggested the following model:

$$Z = \begin{cases} Y - \beta S & \text{with probability } \frac{\alpha}{\alpha+\beta+\gamma} \\ Y + \alpha S & \text{with probability } \frac{\beta}{\alpha+\beta+\gamma} \\ Y & \text{with probability } \frac{\gamma}{\alpha+\beta+\gamma}, \end{cases} \tag{7}$$

where $\alpha$ ,$\beta$ , and $\gamma$ are constants. Utilizing the Narjis and Shabbir [15] method, an unbiased mean estimator may be written as:

$$\hat{\mu}_{NS} = \frac{1}{n}\sum_{i=1}^{n} Z_i. \tag{8}$$

The variance of the mean estimator $\hat{\mu}_{NS}$ can be obtained as:

$$Var(\hat{\mu}_{NS}) = \frac{1}{n}\left[\frac{\alpha\beta(\alpha+\beta)\sigma_S^2}{\alpha+\beta+\gamma} + \sigma_Y^2\right]. \tag{9}$$

## 2.4 2.4 Gupta et al. [9] Optional Scrambling Model

Gupta et al. [9] suggested the following model:

$$Z = \begin{cases} Y & \text{with probability } 1 - W \\ Y + S & \text{with probability } WA \\ TY + S & \text{with probability } W(1 - A), \end{cases} \tag{10}$$

where $W$ denotes the level of sensitivity, and A is a constant such that$0 < A < 1$. Using the Gupta et al. [9] optional model, an unbiased mean estimator may be expressed as:

$$\hat{\mu}_G = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{11}$$

The sampling variance of the estimator $\hat{\mu}_G$ may be obtained as:

$$Var(\hat{\mu}_G) = \frac{1}{n}\left[W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + W\sigma_S^2\right] \tag{12}$$

# 3 Proposed Model

Motivated by Gupta et al. [9], the following optional randomization model is proposed.

$$Z = \begin{cases} Y & \text{with probability } 1 - W \\ Y + S & \text{with probability } WA \\ TY & \text{with probability } WB \\ TY + S & \text{with probability } W(1 - A - B). \end{cases} \tag{13}$$

where $0 < A < 1$, $0 < B < 1$, and $A + B < 1$. It is to be noted that for $B = 0$, the new suggested quantitative randomized response model reduces to the model proposed by Gupta et al. [9].

**Theorem 1.** *Using the suggested model, the mean estimator of Y is unbiased for the population mean, that is,*

$$E(\hat{\mu}_P) = \mu_Y \tag{14}$$

*Proof.* Taking expectation of Eq. 14 yields:

$$E(\hat{\mu}_P) = E\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Z_i). \tag{15}$$

Now

$$E(Z_i) = (1-W)E(Y) + (WA)E(Y+S) + (WB)E(TY) + W(1-A-B)E(TY+S). \tag{16}$$

Since the variables $Y$, $T$, and $S$ are independent of each other, so using $E(TY) = E(T)E(Y)$ in the Eq. 13 yields:

$$E(Z_i) = (1-W)\mu_Y + (WA)E(\mu_Y) + (WB)\mu_Y + W(1-A-B)\mu_Y. \tag{17}$$

On simplification, Eq. 17 reduces to:

$$E(Z_i) = \mu_Y. \tag{18}$$

Hence

$$E(\mu_p) = \frac{1}{n}\sum_{i=1}^{n} \mu_Y = \mu_Y. \tag{19}$$

□

**Theorem 2.** *The variance of the mean estimator $\hat{\mu_p}$ can be derived as:*

$$Var(\hat{\mu}_P) = \frac{1}{n}\left[W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + W(1-B)\sigma_S^2\right] \tag{20}$$

*Proof.* The variance of $Z_i$ is obtained as:

$$Var(Z_i) = E(Z_i^2) - [E(Z_i)]^2. \tag{21}$$

Now

$$E(Z_i^2) = (1-W)E(Y_i)^2 + (WA)E(Y+S)^2 + (WB)E(TY)^2 + W(1-A-B)E(TY+S)^2 \tag{22}$$

Using $E(Y) = \mu_Y$,$E(T) = 1$, $E(Y^2) = \sigma_Y^2 + \mu_Y^2$, $E(T^2) = \sigma_T^2 + 1$, $E(S^2) = \sigma_S^2$ and assuming independence of variables, Eq. 22 simplifies to:

$$E(Z_i^2) = (1-W)(\sigma_Y^2 + \mu_Y^2) + (WA)(\sigma_Y^2 + \mu_Y^2 + \sigma_S^2) + (WB)(\sigma_T^2 + 1)(\sigma_Y^2 + \mu_Y^2) + W(1-A-B)\left[(\sigma_T^2 + 1)(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2\right] \tag{23}$$

Further algebraic simplification yields:

$$E(Z_i^2) = \sigma_Y^2 + \mu_Y^2 + W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + W(1-B)\sigma_S^2 \tag{24}$$

Also

$$E(Z_i) = \mu_Y \tag{25}$$

Using Eq. 24 and Eq. 25 in Eq. 21, the variance of $Z_i$ simplifies to:

$$Var(Z_i) = \sigma_Y^2 + W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + W(1-B)\sigma_S^2 \tag{26}$$

Taking variance of Eq. 14 gives:

$$Var(\hat{\mu}_P) = \frac{1}{n}\left[\sigma_Y^2 + W(1-A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + W(1-B)\sigma_S^2\right]. \tag{27}$$

This completes the proof. □

**Remark:** If $W$, $A$, and $B$ are unknown, we may estimate the variance by:

$$Var(\hat{\mu}_P) = \frac{S_Z^2}{n} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(Z_i - \bar{Z})^2. \tag{28}$$

where $\bar{Z}$ and $S_Z^2$ denote the mean and variance of Z for the sample data, respectively.

# 4 An Application of the Suggested Model

For the application of the suggested model to a real-life sample survey, a sample of 40 students of undergraduate level was drawn from the undergraduate level students currently studying in the University of Malakand, Pakistan. There are a total of 195 students currently enrolled in the university. The researcher was interested in estimating the population mean of the Grade Point Averages (GPA) of all students enrolled in the department. The 40 sampled students were provided with a deck of 100 cards and a calculator. The cards of the deck displayed two random values – one each for variable S and variable T. The random numbers were created using a normal distribution. For the scrambling variable $S$, the researcher created random numbers using normal distribution having mean zero and variance 0.5. For the random variable $T$, the research created random numbers using a normal distribution with mean one and variance equal to 0.5. Following Gjestvang and Singh [15], the numerical values for the constants $W$, $A$, and $B$ were decided by the researcher based on his prior knowledge about proportion of people who feel that the question is of sensitive nature. In the case of non-availabilit of prior information, a pilot survey can be planned to get estimates of these constants. In the given case, the researcher decided to take $W = 0.4$, $A = 0.5$, and $B = 0.3$, so that $1 - W = 0.6$ , $WA = 0.2$ , $WB = 0.12$ and $W(1 - A - B) = 0.08$ . Using these values, the reported response using the proposed model in Eq. 10 may be expressed as:

$$Z = \begin{cases} Y & \text{with probability } 0.60, \\ Y + S & \text{with probability } 0.20, \\ TY & \text{with probability } 0.12, \\ TY + S & \text{with probability } 0.08. \end{cases} \tag{29}$$

Using Eq. 29, each card of the deck displayed one of the following four questions:

(i) 60 out of 100 cards displayed the question: "Report your true GPA in last exam."
(ii) 20 out of 100 cards displayed the question: "Add the value of S to your true GPA and report the result of the addition."
(iii) 12 out of 100 cards displayed the question: "Multiply the value of T with your true GPA and report the result of the product."
(iV) 8 out of 100 cards displayed the question: "Multiply the value of T with your actual GPA and then add the value of $S$ and report the result."

These four questions on different cards correspond to the four components of the proposed technique defined in Eq. 13. The students were asked to draw one card at random from 100 cards and add or multiply the numbers as per instruction on the selected card. The respondents had neither to disclose their true grade point average nor to show the selected card to the researcher, hence assuring their privacy protection. The response reported by the respondents are presented in Table 1. Observing Table 1, some of the observed values are greater than 4.0 despite the fact that the students' actual grade point average was measured on the scale of 4.0. However, most of the reported responses lie well within the acceptable range of grade point average. The proportion of the responses lying inside the acceptable range depends on the researcher's choice of parameters for the distribution which is used to generate the random numbers. If the researcher chooses the mean equal to 50 or 100 for variable S or T to generate the random numbers, this will lead to reported responses having very large scores which will look too unnatural for data on students' grade point average which generally ranges from 0 to 4, and this will also result in getting an overestimate of the true mean grade point average. It is therefore advised to choose the parameters of the distribution in such a manner that the observed values do not disperse too much from the acceptable range of responses.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2.713 | 3.044 | 2.984 | 3.569 | 1.988 | 0.584 | 0.00202 | 0.00180 | 3.231 | 2.429 |
| 1.852 | 3.499 | 4.879 | 3.187 | 2.529 | 1.949 | 2.609 | 3.987 | 3.003 | 1.712 |
| 3.559 | 2.639 | 3.039 | 2.886 | 2.348 | 2.765 | 4.569 | 2.439 | 1.327 | 3.899 |
| 3.884 | 3.101 | 2.197 | 1.990 | 4.443 | 1.736 | 2.678 | 3.751 | 3.096 | 2.110 |

**Table 1.** Observed Responses

# 5 Evaluation of Models

Yan et al. [19] presented a quantification metric of the privacy level for evaluation of randomized response models which can be expressed as:

$$\Delta = E[Z - Y]^2. \tag{30}$$

A larger value of $\Delta$ translates to a higher privacy level provided by a particular model.
The evaluation of a given model is dependent upon two considerations: the privacy level provided by the model, and its efficiency. A unified metric of the efficiency and level of privacy suggested by Gupta et al. [**?** ] may be written as:

$$\delta = \frac{MSE}{\Delta}. \tag{31}$$

Observing Eq. 31, one may note that a smaller value of $\delta$ is preferable as it shows either a smaller variance, or a better privacy level, or both.
Under the Warner's [18] additive scrambling model, the privacy measure may be obtained as:

$$\Delta_W = E[Y + S - Y]^2 = E[S^2] = \sigma_S^2. \tag{32}$$

The unified metric of efficiency and level of privacy using the Warner's [18] scrambling technique may be expressed as:

$$\delta_W = \frac{Var(\hat{\mu}_W)}{\Delta_W} = \frac{1}{n}\left(\frac{\sigma_Y^2 + \sigma_S^2}{\sigma_S^2}\right). \tag{33}$$

The privacy measure using the technique of Diana and Perri [4] may be obtained as:

$$\Delta_{DP} = E[TY + S - Y]^2 = \sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2. \tag{34}$$

The unified metric of efficiency and level of privacy using the Diana and Perri [4] scrambling model can be obtained as:

$$\delta_{DP} = \frac{Var(\hat{\mu}_{DP})}{\Delta_{DP}} = \frac{1}{n}\left(\frac{\left[\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_Y^2 + \sigma_S^2\right]}{\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2}\right). \tag{35}$$

The privacy measure using the Gupta et al. [9] optional model may be obtained as:

$$\Delta_G = (1 - A)\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2)\right] + \sigma_S^2. \tag{36}$$

The unified metric of efficiency and level of privacy using the Gupta et al. [9] procedure may be obtained as:

$$\delta_G = \frac{Var(\hat{\mu}_G)}{\Delta_G} = \frac{1}{n}\left(\frac{\left[W(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + W\sigma_S^2\right]}{(1 - A)\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2)\right] + \sigma_S^2}\right). \tag{37}$$

Using the proposed model, the level of privacy may be obtained as:

$$\Delta_P = (1 - A)\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2)\right] + (1 - B)\sigma_S^2. \tag{38}$$

The unified metric of efficiency and privacy using the suggested model may be obtained as:

$$\delta_P = \frac{Var(\hat{\mu}_P)}{\Delta_P} = \frac{1}{n}\left(\frac{\left[\sigma_Y^2 + W(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + W(1 - B)\sigma_S^2\right]}{(1 - A)\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2)\right] + (1 - B)\sigma_S^2}\right). \tag{39}$$

# 6 Estimation of Sensitivity Level

We can estimate the sensitivity level $W$ and the mean sensitive trait $\mu_Y$ by using the split sample approach with equal split. For this, we assume that the constant $B$ is known. We also use the following assumptions about the expected values of variable $T$ and $S$.

$$E(S) = \theta,$$

and

$$E(T) = 1.$$

Consider two sub-samples having sizes $n_1$ and $n_2$ such that $n_1 = n_1 = n$. In the $i^{th}$ sub-sample, the respondents use variable $S_i$ and $T$ with mean $\theta_i$ and $\mu_T$ , and variance $\sigma_S^2$ and $\sigma_T^2$ , respectively for $i = 1, 2$.
Using the proposed model, the expected value of $Z_i$ may be expressed as:

$$E(Z_i) = \mu_Y + W\theta(1 - B), \tag{40}$$

where B is known.
Replacing $E(Z_i)$ by its estimator $\bar{Z}_i$ yields:

$$\bar{Z_1} = \hat{\mu}_Y + \hat{W}\theta_1(1 - B), \tag{41}$$

and

$$\bar{Z}_2 = \hat{\mu}_Y + \hat{W}\theta_2(1 - B), \tag{42}$$

Solving Eq. 39 and Eq. 42 simultaneously for $\hat{\mu_Y}$ and $\hat{W}$ yields the estimators:

$$\hat{\mu}_Y = \frac{\theta_1\bar{Z}_2 - \theta_2\bar{Z}_1}{\theta_1 - \theta_2}, \tag{43}$$

$$\hat{W} = \frac{\bar{Z}_1 - \bar{Z}_2}{(1 - B)(\theta_1 - \theta_2)}. \tag{44}$$

# 7 Comparison of Models

The suggested quantitative randomized response model is better than the Gupta et al. [9] model if:

$$Var(\hat{\mu}_P) < Var(\hat{\mu}_G), \tag{45}$$

or if

$$\frac{1}{n}\left[\sigma_Y^2 + W(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + W(1 - B)\sigma_S^2\right] < \frac{1}{n}\left[W(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + W\sigma_S^2\right], \tag{46}$$

or if

$$W(1 - B)\sigma_S^2 < W\sigma_S^2, \tag{47}$$

or if

$$(1 - B) < 1,$$

or if

$$B > 0, \tag{48}$$

which is always true. If $B = 0$, the proposed and the Gupta et al. [9] randomized response techniques are equally efficient. The suggested quantitative randomized response model is better than the Narjis and Shabbir [15] model if:

$$Var(\hat{\mu}_P) < Var(\hat{\mu}_{NS}), \tag{49}$$

$$\left[W(1 - A)\sigma_T^2(\sigma_Y^2 + \mu_Y^2)\right] < \left[\frac{\alpha\beta(\alpha + \beta)}{\alpha + \beta + \gamma} - W(1 - B)\right]\sigma_S^2, \tag{50}$$

or if

$$\sigma_T^2 < \left[\frac{\alpha\beta(\alpha + \beta)}{\alpha + \beta + \gamma} - W(1 - B)\right]\frac{\sigma_S^2}{W(1 - A)(\sigma_Y^2 + \mu_Y^2)}, \tag{51}$$

The results of the sampling variances under the suggested and the other models are provided in Table 2 for different values of constants and parameters. In addition, Table 3 displays the values of the unified metric of efficiency and privacy, $\delta$ , using various models for different values of constants.

| $\alpha$ | $\beta$ | $\gamma$ | $A$ | $B$ | $\hat{\mu}_{DP}$ | $\hat{\mu}_{NS}$ | $\hat{\mu}_{G}$ | $\hat{\mu}_{P}$ |
|---|---|---|---|---|---|---|---|---|
| 5 | 3 | 1 | 0.3 | 0.5 | 0.584 | 0.537 | 0.376 | 0.358 |
| | | | 0.5 | 0.3 | 0.584 | 0.537 | 0.280 | 0.269 |
| | | 2 | 0.3 | 0.5 | 0.584 | 0.484 | 0.338 | 0.322 |
| | | | 0.5 | 0.3 | 0.584 | 0.484 | 0.252 | 0.242 |
| | 6 | 1 | 0.3 | 0.5 | 0.584 | 1.104 | 0.387 | 0.369 |
| | | | 0.5 | 0.3 | 0.584 | 1.104 | 0.288 | 0.277 |
| | | 2 | 0.3 | 0.5 | 0.584 | 1.019 | 0.358 | 0.341 |
| | | | 0.5 | 0.3 | 0.584 | 1.019 | 0.266 | 0.256 |
| 10 | 3 | 1 | 0.3 | 0.5 | 0.584 | 1.118 | 0.392 | 0.374 |
| | | | 0.5 | 0.3 | 0.584 | 1.118 | 0.292 | 0.281 |
| | | 2 | 0.3 | 0.5 | 0.584 | 1.044 | 0.366 | 0.349 |
| | | | 0.5 | 0.3 | 0.584 | 1.044 | 0.273 | 0.262 |
| | 6 | 1 | 0.3 | 0.5 | 0.584 | 2.263 | 0.397 | 0.379 |
| | | | 0.5 | 0.3 | 0.584 | 2.263 | 0.296 | 0.284 |
| | | 2 | 0.3 | 0.5 | 0.584 | 2.137 | 0.376 | 0.358 |
| | | | 0.5 | 0.3 | 0.584 | 2.137 | 0.280 | 0.269 |
| 20 | 3 | 1 | 0.3 | 0.5 | 0.584 | 2.304 | 0.405 | 0.385 |
| | | | 0.5 | 0.3 | 0.584 | 2.304 | 0.301 | 0.290 |
| | | 2 | 0.3 | 0.5 | 0.584 | 2.212 | 0.389 | 0.370 |
| | | | 0.5 | 0.3 | 0.584 | 2.212 | 0.289 | 0.278 |
| | 6 | 1 | 0.3 | 0.5 | 0.584 | 4.626 | 0.407 | 0.387 |
| | | | 0.5 | 0.3 | 0.584 | 4.626 | 0.303 | 0.291 |
| | | 2 | 0.3 | 0.5 | 0.584 | 4.461 | 0.392 | 0.374 |
| | | | 0.5 | 0.3 | 0.584 | 4.461 | 0.292 | 0.281 |

**Table 2.** Variances of the mean for $\mu_Y = 5$ , $\sigma_Y^2 = 2$ , $\sigma_T^2 = 10$ , and $\sigma_S^2 = 20$

# 8 Discussion and Conclusion

This article presented an efficient quantitative randomized response model obtained by adding the multiplicative-only scrambling component to the already existing Gupta et al. [9] randomized response model. The inclusion of multiplicative-only component gives more choice to the respondents, hence motivating them more to participate in the survey. Besides achieving improvement in precision over the Gupta et al. [9] model, the proposed model has also been found more efficient than the Narjis and Shabbir [15] model and the Diana and Perri [4] model. Observing Table 2, it is also clearly observed that the sampling variance under the proposed model and under the Gupta et al. [9] model increases as $\alpha$ increases.

Table 3 shows the values of the unified metric of efficiency and privacy level, $\delta$ , under the suggested and other models. Different choices of the values of $\alpha$ , $\beta$ , and $\gamma$ have been considered to observe the values for the proposed technique and the already available randomization techniques. Since the Diana and Perri [4] model is independent of $\alpha$ , $\beta$ , and $\gamma$ Table 3 indicates that the value remains constant for all values of $\alpha$ , $\beta$ , and $\gamma$ . Table 3 also indicates that the proposed quantitative technique and the Gupta et al.

| $\alpha$ | $\beta$ | $\gamma$ | $A$ | $B$ | $\hat{\mu}_{DP}$ | $\hat{\mu}_{NS}$ | $\hat{\mu}_{G}$ | $\hat{\mu}_{P}$ |
|---|---|---|---|---|---|---|---|---|
| 5 | 3 | 1 | 0.3 | 0.5 | 0.00201 | 0.00202 | 0.00180 | 0.00180 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00202 | 0.00180 | 0.00180 |
| | | 2 | 0.3 | 0.5 | 0.00201 | 0.00202 | 0.00162 | 0.00162 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00202 | 0.00162 | 0.00162 |
| | 6 | 1 | 0.3 | 0.5 | 0.00201 | 0.00201 | 0.00185 | 0.00185 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00201 | 0.00186 | 0.00186 |
| | | 2 | 0.3 | 0.5 | 0.00201 | 0.00201 | 0.00171 | 0.00171 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00201 | 0.00172 | 0.00172 |
| 10 | 3 | 1 | 0.3 | 0.5 | 0.00201 | 0.00201 | 0.00188 | 0.00188 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00201 | 0.00188 | 0.00188 |
| | | 2 | 0.3 | 0.5 | 0.00201 | 0.00201 | 0.00175 | 0.00175 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00201 | 0.00176 | 0.00176 |
| | 6 | 1 | 0.3 | 0.5 | 0.00201 | 0.00200 | 0.00190 | 0.00190 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00200 | 0.00191 | 0.00191 |
| | | 2 | 0.3 | 0.5 | 0.00201 | 0.00200 | 0.00180 | .00180 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00200 | 0.00180 | 0.00180 |
| 20 | 3 | 1 | 0.3 | 0.5 | 0.00201 | 0.00200 | 0.00194 | 0.00194 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00200 | 0.00194 | 0.00194 |
| | | 2 | 0.3 | 0.5 | 0.00201 | 0.00200 | 0.00186 | 0.00186 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00200 | 0.00187 | 0.00187 |
| | 6 | 1 | 0.3 | 0.5 | 0.00201 | 0.00200 | 0.00195 | 0.00195 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00200 | 0.00195 | 0.00195 |
| | | 2 | 0.3 | 0.5 | 0.00201 | 0.00200 | 0.00188 | 0.00188 |
| | | | 0.5 | 0.3 | 0.00201 | 0.00200 | 0.00188 | 0.00188 |

**Table 3.** Variances of the mean for$\mu_Y$ = 5 ,$\sigma_Y^2$ = 2 ,$\sigma_T^2$ = 10 , and $\sigma_S^2$ = 20

[16] model produce equal values amongst, although the proposed model is more efficient than the Gupta et al. [9] model. Moreover, the $\delta$ values for the proposed model are smaller than the Diana and Perri [4] technique and the Narjis and Shabbir [15] technique. Keeping in view the improvements in $\delta$ values and efficiency over previous models, the proposed model is recommended for use in sensitive surveys.

It is recommended for future researchers to use some complex survey designs such as ranked set sampling and adaptive cluster sampling to achieve further improvement in respondents' privacy and/or efficiency.

# Author Contributions

**Muhammad Azeem**: Conceptualization, Methodology, Data curation, Supervision, Writing- Original draft preparation,
**Abdul Salam**: Software, Data curation, Validation, Writing- Reviewing and Editing.

# Compliance with Ethical Standards

It is declared that the author doesn't have any conflict of interest. It is also declared that informed consent was obtained from all individual participants included in the study.

# Funding Information

The authors have no funding to report.

# Author Information

**ORCID:**
Muhammad Azeem: 0000-0002-6475-6072
Abdul Salam: 0009-0007-3056-8712

# References

[1] Abbasi, A. M., Shad, M. Y. and Ahmed, A. [2022], 'On partial randomized response model using ranked set sampling', *Plos one* **17**(11), e0277497.

[2] Azeem, M. [2023], 'Using the exponential function of scrambling variable in quantitative randomized response models', *Mathematical Methods in the Applied Sciences* **46**(13), 13882–13893.

[3] Azeem, M. and Salam, A. [2023], 'Introducing an efficient alternative technique to optional quantitative randomized response models', *Methodology* **19**(1), 24–42.

[4] Diana, G. and Perri, P. F. [2011], 'A class of estimators for quantitative sensitive data', *Statistical Papers* **52**, 633–650.

[5] Gjestvang, C. R. and Singh, S. [2009], 'An improved randomized response model: Estimation of mean', *Journal of Applied Statistics* **36**(12), 1361–1367.

[6] 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.

[7] Gupta, S., Mehta, S., Shabbir, J. and Khalil, S. [2018], 'A unified measure of respondent privacy and model efficiency in quantitative rrt models', *Journal of Statistical Theory and Practice* **12**, 506–511.

[8] Gupta, S., Qureshi, M. N. and Khalil, S. [2020], 'Variance estimation using randomized response technique', *REVSTAT-Statistical Journal* **18**(2), 165–176.

[9] Gupta, S., Zhang, J., Khalil, S. and Sapra, P. [2022], 'Mitigating lack of trust in quantitative randomized response technique models', *Communications in Statistics-Simulation and Computation* pp. 1–9.

[10] Hussain, Z., Al-Sobhi, M. M., Al-Zahrani, B., Singh, H. P. and Tarray, T. A. [2016], 'Improved randomized response in additive scrambling models', *Mathematical Population Studies* **23**(4), 205–221.

[11] Khalil, S., Zhang, Q. and Gupta, S. [2021], 'Mean estimation of sensitive variables under measurement errors using optional rrt models', *Communications in Statistics-Simulation and Computation* **50**(5), 1417–1426.

[12] Mahdizadeh, M. and Zamanzade, E. [2017], 'Reliability estimation in multistage ranked set sampling', *REVSTAT-Statistical Journal* **15**(4), 565–581.

[13] Mahdizadeh, M. and Zamanzade, E. [2019], 'Dynamic reliability estimation in a rank-based design', *Probab Math Stat* **39**, 1–18.

[14] Mahdizadeh, M. and Zamanzade, E. [2022], 'On estimating the area under the roc curve in ranked set sampling', *Statistical Methods in Medical Research* **31**(8), 1500–1514.

[15] Narjis, G. and Shabbir, J. [2023], 'An efficient new scrambled response model for estimating sensitive population mean in successive sampling', *Communications in Statistics-Simulation and Computation* **52**(11), 5327–5344.

[16] Shahzad, U., Perri, P. F. and Hanif, M. [2019], 'A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information', *Communications in Statistics-Simulation and Computation* **48**(9), 2566–2585.

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

[18] Warner, S. L. [1971], 'The linear randomized response model', *Journal of the American Statistical Association* **66**(336), 884–888.

[19] Yan, Z., Wang, J. and Lai, J. [2008], 'An efficiency and protection degree-based comparison among the quantitative randomized response strategies', *Communications in Statistics-Theory and Methods* **38**(3), 400–408.

[20] Young, A., Gupta, S. and Parks, R. [2019], 'A binary unrelated-question rrt model accounting for untruthful responding', *Involve, A Journal of Mathematics* **12**(7), 1163–1173.

[21] Zaman, Q., Ijaz, M. and Zaman, T. [2023], 'A randomization tool for obtaining efficient estimators through focus group discussion in sensitive surveys', *Communications in Statistics-Theory and Methods* **52**(10), 3414–3428.

[22] Zaman, T., Shazad, U. and Yadav, V. K. [2024], 'An efficient hartley–ross type estimators of nonsensitive and sensitive variables using robust regression methods in sample surveys', *Journal of Computational and Applied Mathematics* **440**, 115645.

[23] Zhang, Q., Khalil, S. and Gupta, S. [2021], 'Mean estimation in the simultaneous presence of measurement errors and non-response using optional rrt models under stratified sampling', *Journal of statistical computation and Simulation* **91**(17), 3492–3504.