
A Novel Randomized Response Survey Technique for Sensitive Surveys

Muhammad Azeem^{1,*}, Musarrat Ijaz², Najma Salahuddin³,
Soofia Iftikhar³ and Abdul Salam¹

¹*Department of Statistics, University of Malakand, Khyber Pakhtunkhwa, Pakistan*

²*Department of Statistics, Rawalpindi Women University, Rawalpindi, Pakistan*

³*Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan*

E-mail: azeemstats@uom.edu.pk

**Corresponding Author*

Received 28 November 2024; Accepted 18 June 2025

Abstract

Survey statisticians employ randomized response techniques (RRT) to gather data from the respondents. From time to time, researchers make modifications to the existing methods, with the aim to achieve some sort of improvement over the previous methods. The improvement may be in terms of the privacy levels or model-efficiency, or both. In this paper, we introduce an efficient quantitative randomized response technique which provides efficient estimates of the finite population mean. Moreover, the unified quantitative measure under the new suggested technique is also observed to be smaller than the competitor models. A practical data collection example using the new suggested technique is also provided to illustrate its real-world application. Our findings suggest the new technique performs better than the

Journal of Reliability and Statistical Studies, Vol. 18, Issue 2 (2025), 271–288.

doi: 10.13052/jrss0974-8024.1821

© 2025 River Publishers

competitor techniques in efficiency as well as in respondents' privacy level. Besides empirical results, we have also conducted a simulation study to show the improved performance. The comparative analysis reveals that our proposed technique is appropriate for implementation in real-world sample surveys.

Keywords: Privacy protection, randomized response sampling, relative efficiency, scrambling variable, sensitive characteristics.

1 Introduction

In this section, we revisit some of the key developments in the history of randomized response models.

The concept of randomized response models originated as early as 1965 when Warner [1] introduced the randomized response technique as a remedy to the high rates of refusals in sample surveys. A limitation of the original randomized response technique was that its applicability only to qualitative variables. A few years later, this limitation was addressed by Warner [2] who modified and extended the original technique to accommodate quantitative variables. Eichhorn and Hayre [3] developed a new scrambling technique which was based on a multiplicative random noise.

Gupta et al. [4] suggested the concept of optional versions of the quantitative randomized response techniques in sampling theory. Gupta et al. [4] used an additive variable as a scrambling option, in addition to offering the option of true response. Later on, Bar-Lev et al. [5] introduced an enhanced form of the Gupta et al. [4] randomization strategy. Gjostvang and Singh [6] presented a novel survey procedure by utilizing an additive-type scrambling noise. Diana and Perri [7] developed a new scrambling procedure using additive as well as multiplicative scrambling variables. Another quantitative technique utilizing additive and subtractive noise was introduced by Hussain et al. [8]. For quantifying the overall quality of a given model as a single value, Gupta et al. [9] suggested a new quantitative metric for comparison of various models. Murtaza et al. [10] used the correlation among scrambling variables to introduce a new randomized technique.

Narjis and Shabbir [11] introduced another additive scrambling version of the Gjostvang and Singh [6] randomization method which improved the previously developed methods. Khalil et al. [12] worked on analyzing the effects of measurement errors on the estimators of the population mean of a quantitative variable of interest. In a recent research, Gupta et al. [13]

developed a versatile randomized response strategy which enhanced the Diana and Perri [7] randomization method in overall quality. For further research studies on randomized response models, one may refer to the research findings of Chaudhuri [14], Yan et al. [15], Young et al. [16], Saleem et al. [17], Zhang et al. [18], Azeem and Ali [19], Azeem et al. [20], and Azeem [21], etc.

Over the decades, although randomized response techniques have achieved a significant boost in efficiency compared to the early models, there is still a room for further improvement not only in efficiency but also in the levels of privacy offered to the respondents. Keeping in view the need for improvement in the quality of the existing techniques, the objectives of this study are as follows:

1. To achieve further improvement in efficiency of randomized survey techniques.
2. To improve the levels of privacy of the survey participants.
3. To efficiently estimate the population's mean.

Keeping in view the above research objectives, we introduce a novel randomized response survey technique and show its improved efficiency. Besides efficiency, we also prove that our proposed technique also achieves improvement in the unified measure of efficiency and privacy levels. The improved privacy level compared to the competitor techniques means that the proposed technique is applicable in sample surveys where the respondents hesitate to participate due to privacy concerns.

Before proceeding further, we introduce some notations which have been used in the subsequent sections.

Consider a population containing N units and let a probability sample of n units be chosen from the population. Let Y denote the sensitive variable and let us consider two scrambling variables S and T . Moreover, we assume that $E(Y_i) = \mu_Y$, $E(S) = 0$, $E(T) = 1$, $V(Y_i) = \sigma_Y^2$, $V(T) = \sigma_T^2$, $V(S) = \sigma_S^2$, where σ_Y^2 , σ_T^2 , and σ_S^2 denote the population-based variance of the variable Y , T , and S , respectively, whereas μ_Y denotes the mean of Y . Likewise, we also assume that all three variables are unrelated, which adds to increased levels of respondent-privacy protection.

2 Some Available Models

This section gives the brief descriptions of some popular competitor models and their underlying estimators.

2.1 Murtaza et al. [10] Optional Scrambling Technique

This model uses optional response approach where the respondents report their responses using the following relation.

$$Z = \begin{cases} Y, & \text{with probability } 1 - W, \\ TY + \alpha S, & \text{with probability } W, \end{cases} \quad (1)$$

where α is a pre-assigned constant to be determined by the researcher.

An unbiased estimator is as follows:

$$\hat{\mu}_M = \frac{1}{n} \sum_{i=1}^n Z_i. \quad (2)$$

The variance of $\hat{\mu}_M$ can be obtained as:

$$Var(\hat{\mu}_M) = \frac{1}{n} [W \{ \sigma_T^2 (\sigma_Y^2 + \mu_Y^2) + \alpha^2 \sigma_S^2 \} + \sigma_Y^2]. \quad (3)$$

2.2 Salemian et al. [22] Technique

The Salemian et al. [16] is given by:

$$Z = \begin{cases} S, & \text{with probability } p_1, \\ Y, & \text{with probability } p_2, \\ \alpha - Y, & \text{with probability } p_3, \end{cases} \quad (4)$$

where α denotes a pre-defined constant. A simpler variant of the Salemian et al. [22] method is given as:

$$Z = \begin{cases} S, & \text{with probability } p_1, \\ \alpha - Y, & \text{with probability } p_2. \end{cases} \quad (5)$$

Based on Equation (5), the population mean can unbiasedly be estimated by the estimator:

$$\hat{\mu}_{Sal} = \alpha p_2 - \bar{Z}, \quad (6)$$

where \bar{Z} denotes the sample mean of the scrambled responses. The variance of $\hat{\mu}_{Sal}$ is:

$$Var(\hat{\mu}_{Sal}) = \frac{1}{n} [p_1 \sigma_S^2 + p_2 \sigma_Y^2 + (1 - p_2)(p_2 \alpha^2 + \mu_Y^2)]. \quad (7)$$

2.3 Azeem et al. [23] Technique

This model uses optional response approach where the respondents report their responses using the following relation.

$$Z = \alpha(Y + S) + (1 - \alpha)(Y + YS), \tag{8}$$

where α is a pre-assigned constant to be determined by the researcher. An unbiased estimator is as follows:

$$\hat{\mu}_A = \frac{1}{n} \sum_{i=1}^n Z_i, \tag{9}$$

The variance of $\hat{\mu}_A$ can be obtained as:

$$Var(\hat{\mu}_A) = \frac{1}{n} [\sigma_Y^2 + \{\alpha^2 + (1 - \alpha)^2(\sigma_Y^2 + \mu_Y^2) + 2\alpha(1 - \alpha)\mu_Y\}\sigma_S^2]. \tag{10}$$

3 Proposed Randomized Response Model

Taking motivation from the study of Murtaza et al. [10] model, we introduce a quantitative randomized response model. The proposed model is expressed as:

$$Z = \begin{cases} Y, & \text{with probability } 1 - W, \\ \alpha S + [\beta(T - 1) + 1]Y, & \text{with probability } W, \end{cases} \tag{11}$$

where β denotes a predefined constant.

An unbiased mean estimator using the suggested technique can be written as:

$$\hat{\mu}_P = \frac{1}{n} \sum_{i=1}^n Z_i. \tag{12}$$

Here we show unbiasedness of the mean estimator and also derive the sampling variance of the mean.

Theorem 1: The estimator $\hat{\mu}_P$ unbiasedly estimates the population mean μ_Y .

Proof: Applying expected value on Equation (12) yields:

$$E(\hat{\mu}_P) = E\left(\frac{1}{n} \sum_{i=1}^n Z_i\right) = \frac{1}{n} \sum_{i=1}^n E(Z_i). \tag{13}$$

Now,

$$E(Z) = (1 - W)E(Y) + WE[\{\beta(T - 1) + 1\}Y + \alpha S].$$

Simplification gives:

$$E(Z) = \mu_Y. \quad (14)$$

Using Equation (14) in Equation (13) yields:

$$E(\hat{\mu}_P) = \frac{1}{n} \sum_{i=1}^n \mu_Y = \mu_Y. \quad (15)$$

Theorem 2: The sampling variance may be derived in the form:

$$Var(\hat{\mu}_P) = \frac{1}{n} [W\{\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2\} + \sigma_Y^2]. \quad (16)$$

Proof: Applying variance on Equation (12) yields:

$$Var(\hat{\mu}_P) = \frac{1}{n^2} \sum_{i=1}^n Var(Z_i). \quad (17)$$

$$Var(Z_i) = E(Z_i^2) - [E(Z_i)]^2. \quad (18)$$

We can simplify $E(Z_i^2)$ as follows:

$$E(Z_i^2) = (1 - W)E(Y^2) + WE[\{\beta(T - 1) + 1\}Y + \alpha S]^2,$$

or

$$E(Z_i^2) = (1 - W)(\sigma_Y^2 + \mu_Y^2) + W[E\{\beta^2(T - 1)^2 + 1 + 2\beta(T - 1)\}(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2]$$

Further simplification yields:

$$E(Z_i^2) = W\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2W\sigma_S^2 + \sigma_Y^2 + \mu_Y^2. \quad (19)$$

Using Equations (14) and (19) in Equation (18) yields:

$$Var(Z_i) = W\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2W\sigma_S^2 + \sigma_Y^2. \quad (20)$$

Using Equation (20) in Equation (17) yields the required result as:

$$Var(\hat{\mu}_P) = \frac{1}{n} [W\{\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2\} + \sigma_Y^2].$$

Remark: the sampling variance may be unbiasedly estimated as:

$$var(\hat{\mu}_P) = \frac{s_z^2}{n} = \frac{1}{(n-1)n} \sum_{i=1}^n (Z_i - \bar{Z})^2, \quad (21)$$

where \bar{Z} and s_z^2 represent the mean and variance calculated from the sample data.

4 Privacy and Efficiency Measures

The Yan et al. [15] privacy protection metric is given as:

$$\nabla = E[Z - Y]^2. \quad (22)$$

The Gupta et al. [9] unified measure can be expressed as:

$$\delta = \frac{MSE}{\nabla}. \quad (23)$$

It is obvious that a smaller value of δ corresponds to better model quality. Under the Warner's [2] scrambling method, the metric of respondents' privacy level is given by:

$$\nabla_W = \sigma_S^2. \quad (24)$$

The unified metric of efficiency and privacy level can be obtained as:

$$\delta_W = \frac{Var(\hat{\mu}_W)}{\nabla_W} = \frac{1}{n} \left[\frac{\sigma_S^2 + \sigma_Y^2}{\sigma_S^2} \right]. \quad (25)$$

The privacy can be quantified as:

$$\nabla_M = W[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2]. \quad (26)$$

The unified metric can be expressed as:

$$\delta_M = \frac{Var(\hat{\mu}_M)}{\nabla_M} = \frac{1}{n} \left[\frac{W\{\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2\} + \sigma_Y^2}{W[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2]} \right]. \quad (27)$$

The privacy level using the new suggested model is given as:

$$\nabla_P = W[\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2]. \quad (28)$$

The unified measure under the suggested model is given by:

$$\delta_P = \frac{Var(\hat{\mu}_P)}{\nabla_P} = \frac{1}{n} \left[\frac{W\{\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2\} + \sigma_Y^2}{W[\beta^2\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \alpha^2\sigma_S^2]} \right]. \quad (29)$$

5 An Example of Data Collection

A study was conducted to estimate the average number of times the undergraduate level students cheated in an examination. The target population for this study consisted of all of the undergraduate students enrolled in the University of Malakand, Pakistan. For sample selection, the simple random sampling scheme was used to choose 50 respondents from the population. The researcher then generated a total of 100 random numbers for variable S using a normal distribution $N(0, 3)$. Likewise, the researcher obtained random numbers for variable T from a normal distribution $N(1, 3)$. The researcher chose the values of α, β , and W as $\alpha = 0.4$, $\beta = 0.3$, and $W = 0.6$. To reduce the cognitive burden on the respondents, the factor $[\beta(T - 1) + 1]$ in the proposed model was transformed to another scrambling variable, V , using the transformation $[\beta(T - 1) + 1] = V$. The 100 random numbers for variables S and V were printed on a deck consisting of 100 cards, where each card presented one random number each for variable S and V . The 50 respondents selected in the sample were presented with the deck of cards and a calculator. The respondents were instructed to randomly select a card from the deck and then follow the statements printed on the chosen card. The respondents were instructed not to show the card selected by him/her to the researcher.

Using the values of α, β , and W in Equation (11), our proposed model takes the form:

$$Z = \begin{cases} Y, & \text{with probability } 0.4, \\ VY + 0.4S, & \text{with probability } 0.6, \end{cases} \quad (30)$$

where $V = [\beta(T - 1) + 1]$. Corresponding to Equation (30), one of the following two statements were written on each card:

- (i) 40 out of 100 cards showed the statement: "How many times did you cheat in your last examination? Report your true response".
- (ii) The remaining 60 cards displayed: "Multiply V with your true response, add 0.4 times the value of variable S , and then report the result".

The reported responses are presented in Table 1.

One may clearly observe from the above table that some of the reported responses are negative, due to the scrambling process. We recommend the researchers to choose the values of α , and β , in such a way as to ensure maximum possible levels of respondents' privacy protection. It was interesting to observe that, thanks to the respondents' interest in the survey, we didn't

Table 1 Responses reported by students

4	3	6	2	3	5	2	0	4	-1
0	5	3	1	4	2	5	6	1	2
7	-2	4	5	1	0	3	2	2	5
1	3	2	1	5	5	0	3	4	6
3	4	1	6	-1	2	1	4	3	2

experience any case of non-response, and thus the model was successfully applied to the student survey.

6 Efficiency Comparison

The efficiency condition for the proposed vs. the Murtaza et al. [10] technique may be derived as follows:

$$Var(\hat{\mu}_P) \leq Var(\hat{\mu}_M),$$

or,

$$\beta^2 \leq 1.$$

The efficiency condition for the proposed vs. the Saleman et al. [22] technique may be derived as follows:

$$Var(\hat{\mu}_P) \leq Var(\hat{\mu}_{Sal}),$$

or,

$$\sigma_Y^2 \leq \frac{(1 - p_2)(\alpha^2 p_2 - \mu_Y^2) + (p_1 - W\alpha^2)\sigma_S^2 - W\beta^2\sigma_T^2\mu_Y^2}{W\beta^2\sigma_T^2 - p_2}.$$

The efficiency condition for the proposed vs. the Azeem et al. [23] technique may be derived as follows:

$$Var(\hat{\mu}_P) \leq Var(\hat{\mu}_A),$$

or,

$$\begin{aligned} \sigma_Y^2 \leq & \frac{1}{W\beta^2\sigma_T^2} [\{\alpha^2 + (1 - \alpha)^2(\sigma_Y^2 + \mu_Y^2) \\ & + 2\alpha(1 - \alpha)\mu_Y - W\alpha^2\}\sigma_S^2 - W\beta^2\sigma_T^2\mu_Y^2]. \end{aligned}$$

Table 2 PRE's for $\mu_Y = 20, \sigma_Y^2 = 5$

σ_T^2	σ_S^2	W	$\alpha = 1,$ $\beta = 0.2$	$\alpha = 1,$ $\beta = 0.5$	$\alpha = 1,$ $\beta = 0.8$	$\alpha = 5,$ $\beta = 0.2$	$\alpha = 5,$ $\beta = 0.5$	$\alpha = 5,$ $\beta = 0.8$	
3	2	0.1	1259.443	356.149	152.7242	884.926	325.6966	149.8405	
		0.3	1833.994	382.6312	154.9313	1111.914	346.0067	151.8083	
		0.5	2024.752	388.5986	155.3951	1174.033	350.5155	152.2206	
		0.7	2119.99	391.2339	155.5963	1203.053	352.4988	152.3994	
		0.9	2177.087	392.7188	155.7087	1219.863	353.6142	152.4992	
	4	0.1	1236.842	354.717	152.5974	687.3112	300.8264	147.1539	
		0.3	1783.927	380.8888	154.7937	805.7685	316.7493	148.9116	
		0.5	1963.259	386.7821	155.2552	835.4351	320.2417	149.279	
		0.7	2052.367	389.3841	155.4553	848.9268	321.7731	149.4381	
		0.9	2105.655	390.8503	155.5672	856.6383	322.6332	149.5269	
	6	2	0.1	1663.539	376.3457	154.4301	1282.961	357.5972	152.8516
			0.3	2113.349	391.0567	155.5829	1523.596	370.3337	153.9378
			0.5	2236.264	394.1889	155.8193	1583.969	373.0337	154.1605
			0.7	2293.659	395.552	155.9212	1611.44	374.2074	154.2565
			0.9	2326.899	396.3147	155.978	1627.146	374.8638	154.3099
4		0.1	1642.857	375.5102	154.3624	1043.689	340.5941	151.3019	
		0.3	2079.186	390.13	155.5123	1190.773	351.6686	152.3247	
		0.5	2197.842	393.2422	155.7482	1225.869	354.007	152.5242	
		0.7	2253.165	394.5965	155.8498	1241.611	355.0225	152.6245	
		0.9	2285.179	395.3543	155.9065	1250.548	355.5902	152.6748	

The Percentage Relative Efficiency (PRE) can be computed by using the formula:

$$PRE = \frac{Var(\hat{\mu}_M)}{Var(\hat{\mu}_P)} \times 100. \tag{31}$$

Table 2 presents the PRE for our suggested scrambling model with respect to the Murtaza et al. [10] technique for different values of $\alpha, \beta,$ and W . It may be clearly examined that the new technique is more precise than the optional scrambling model of Murtaza et al. [10].

7 Simulation Study

Simulation study was carried out to show the improvement in our suggested technique over the competitor techniques. Different choices of the values of $\alpha, \beta,$ and W were considered for comparison of the models. Tables 4 and 5 present the results of the simulation. By examining Table 4, one may clearly notice that the new proposed model achieves a huge gain in efficiency over

Table 3 δ values for $\mu_Y = 20, \sigma_Y^2 = 5, n = 100$

σ_T^2	σ_S^2	W	$\alpha = 1, \beta = 2$		$\alpha = 3, \beta = 5$		$\alpha = 5, \beta = 10$	
			δ_P	δ_M	δ_P	δ_M	δ_P	δ_M
3	2	0.1	0.010103	0.010411	0.010016	0.010406	0.010004	0.010395
		0.3	0.010034	0.010137	0.010005	0.010135	0.010001	0.010132
		0.5	0.010021	0.010082	0.010003	0.010081	0.010001	0.010079
		0.7	0.010015	0.010059	0.010002	0.010058	0.010001	0.010056
		0.9	0.010011	0.010046	0.010002	0.010045	0.010000	0.010044
	4	0.1	0.010103	0.010411	0.010016	0.010406	0.010004	0.010395
		0.3	0.010034	0.010137	0.010005	0.010135	0.010001	0.010132
		0.5	0.010021	0.010082	0.010003	0.010081	0.010001	0.010079
		0.7	0.010015	0.010059	0.010002	0.010058	0.010001	0.010056
		0.9	0.010011	0.010046	0.010002	0.010045	0.010000	0.010044
6	2	0.1	0.010051	0.010206	0.010008	0.010204	0.010002	0.010202
		0.3	0.010017	0.010069	0.010003	0.010068	0.010001	0.010067
		0.5	0.010010	0.010041	0.010002	0.010041	0.010000	0.010040
		0.7	0.010007	0.010029	0.010001	0.010029	0.010000	0.010029
		0.9	0.010006	0.010023	0.010001	0.010023	0.010000	0.010022
	4	0.1	0.010051	0.010206	0.010008	0.010204	0.010002	0.010202
		0.3	0.010017	0.010069	0.010003	0.010068	0.010001	0.010067
		0.5	0.010010	0.010041	0.010002	0.010041	0.010000	0.010040
		0.7	0.010007	0.010029	0.010001	0.010029	0.010000	0.010029
		0.9	0.010006	0.010023	0.010001	0.010023	0.010000	0.010022

the competitors, for all choices of α, β , and W . Likewise, Table 5 clearly indicates that the proposed model gives smaller values of δ for all choices of α, β , and W , which shows that the overall quality of the new suggested technique is better than the existing techniques.

8 Discussion and Conclusion

We introduced a new modification of the Murtaza et al. [10] optional randomized response technique. We presented the mean estimator under the proposed technique and proved its unbiasedness. The variance of the mean was also derived and was compared with that of the competitor models. Further, we also derived the conditions for efficiency comparison between the new technique and its competitors. The proposed technique was also applied to a real-life sample survey to illustrate its practical implementation. Finally, empirical and simulated variances under different techniques were also computed and the results were presented in tables.

Table 4 Simulated variance under the proposed and competitor models

Parameters	W	α	β	$Var(\hat{\mu}_M)$	$Var(\hat{\mu}_{Sat})$	$Var(\hat{\mu}_A)$	$Var(\hat{\mu}_P)$
$\sigma_S^2 = 2,$ $\sigma_T^2 = 0.25$	0.2	1	0.2	0.6707021	1286.00168	13.659742	0.2094616
		1.2	0.3	0.6732895	1283.13446	10.614648	0.2366506
		1.4	0.4	0.6762361	1280.27043	7.9589959	0.2736427
		1.6	0.5	0.6795421	1277.40960	5.6927858	0.3204379
		1.8	0.6	0.6832073	1274.55198	3.8160175	0.3770362
$\sigma_S^2 = 4,$ $\sigma_T^2 = 0.5$	0.5	2	0.7	0.6872318	1271.69755	2.3286912	0.4434377
		1	0.2	2.652366	1994.73135	24.801216	0.3038224
		1.2	0.3	2.66482	1985.80902	19.239497	0.4383895
		1.4	0.4	2.678898	1976.90669	14.388773	0.6235688
		1.6	0.5	2.694602	1968.02436	10.249044	0.8593605
$\sigma_S^2 = 6,$ $\sigma_T^2 = 0.75$	0.8	1.8	0.6	2.711931	1959.16203	6.8203100	1.145764
		2	0.7	2.730886	1950.31970	4.1025710	1.482781
		1	0.2	5.918617	2887.67109	40.842404	0.4834438
		1.2	0.3	5.943735	2870.28681	31.685364	0.7942259
		1.4	0.4	5.973122	2852.95502	23.695343	1.221772
		1.8	0.5	6.006778	2835.67571	16.872342	1.766083
		1.3	0.6	6.044704	2818.44889	11.216359	2.427158
		2	0.7	6.086900	2801.27456	6.7273960	3.204998

The comparative analysis suggests that the new suggested technique is more precise than its competitor techniques. We observe the improvement in efficiency over the competitors for different parameter values. We can also observe from Table 2 that as the sensitivity level W of the respondents increases, the percentage relative efficiency (PRE) also enhances. Likewise, it may also be observed from tables that as σ_T^2 increases, the PRE also improves.

For comparative analysis in terms of the *overall quality* of the new suggested and the competitor techniques, the values of δ under different techniques have been provided in Table 3 using different values of α , β , and W . Glancing at Table 3, one may also observe that the values of δ for the proposed model are smaller compared to those of the competitor techniques. This means that our new suggested technique performs better than the competitors in circumstances where respondent-privacy and efficiency are simultaneously considered. Table 3 also indicates that δ value decreases as W increases. Moreover, similar findings can also be observed from the simulation analysis, presented in Tables 4–5. Based on the findings of our study, we recommend survey researchers to use the suggested model in sample surveys on sensitive quantitative variables.

Table 5 Simulated δ values under the proposed and competitor models

Parameters	W	α	β	δ_M	δ_{Sat}	δ_A	δ_P
$\sigma_S^2 = 2,$ $\sigma_T^2 = 0.25$	0.2	1	0.2	0.01433974	0.87934929	0.01114381	0.01107801
		1.2	0.3	0.01434669	0.88025380	0.01114313	0.01099863
		1.4	0.4	0.01435142	0.88115673	0.01114234	0.01092955
		1.6	0.5	0.01435394	0.88205804	0.01114153	0.01086910
		1.8	0.6	0.01435427	0.88295769	0.01114071	0.01081590
$\sigma_S^2 = 4,$ $\sigma_T^2 = 0.5$	0.5	2	0.7	0.01435245	0.88385563	0.01113991	0.01076886
		1	0.2	0.01148184	0.87228605	0.01098035	0.01073001
		1.2	0.3	0.01145912	0.87310891	0.01098204	0.01070270
		1.4	0.4	0.01143552	0.87393202	0.01098336	0.01067860
		1.6	0.5	0.01141112	0.87475535	0.01098441	0.01065721
$\sigma_S^2 = 6,$ $\sigma_T^2 = 0.75$	0.8	1.8	0.6	0.01138596	0.87557885	0.01098526	0.01063812
		2	0.7	0.01136011	0.87640250	0.01098595	0.01062101
		1	0.2	0.01038722	0.93484520	0.01115781	0.01006330
		1.2	0.3	0.01039179	0.93505507	0.01115997	0.01005572
		1.4	0.4	0.01039670	0.93526493	0.01116174	0.01004914
		1.8	0.5	0.01040192	0.93547477	0.01116323	0.01004340
		1.3	0.6	0.01040744	0.93568457	0.01116449	0.01003836
		2	0.7	0.01041325	0.93589433	0.01116558	0.01003392

9 Limitations and Future Research Recommendations

Our proposed technique is based on simple random sampling design which is only applicable situations where the population units are homogeneous with regard to the variable under study. In the case of heterogeneous units, the proposed technique may not perform well. Further, we have only analyzed the simple estimator of the population mean under the proposed technique. We suggest the following recommendations for future researchers.

1. The proposed technique may be extended to stratified, ranked-set, and cluster sampling designs.
2. Future researchers can also extend the proposed technique to unequal probability sampling designs such as PPS (probability proportional to size) sampling.
3. The efficiency and/or privacy level may be further improved by making some modifications in the mathematical equation of the proposed technique.
4. Future researchers may explore simple and ratio estimators of population variance under the proposed technique.

References

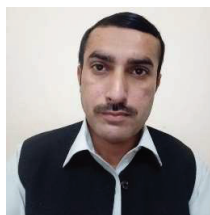
- [1] S.L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309): 63–69, 1965. <https://doi.org/10.1080/01621459.1965.10480775>.
- [2] S.L. Warner. The linear randomized response model. *Journal of the American Statistical Association*, 66(336): 884–888, 1971. <https://doi.org/10.1080/01621459.1971.10482364>.
- [3] B.H. Eichhorn, and L.S. Hayre. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, 7(4): 307–316, 1983.
- [4] S. Gupta, B. Gupta, and S. Singh. Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference*, 100(2): 239–247, (2002). [https://doi.org/10.1016/S0378-3758\(01\)00137-9](https://doi.org/10.1016/S0378-3758(01)00137-9).
- [5] S.K. Bar-Lev, E. Bobovitch, and B. Boukai. A note on randomized response models for quantitative data. *Metrika*, 60(3): 255–260, 2004.
- [6] C.R. Gjestvang, and S. Singh. An improved randomized response model: Estimation of mean. *Journal of Applied Statistics*, 36(12): 1361–1367, 2009. <https://doi.org/10.1080/02664760802684151>.
- [7] G. Diana, and P.F. Perri. A class of estimators of quantitative sensitive data. *Statistical Papers*, 52(3): 633–650, 2011. <https://doi.org/10.1007/s00362-009-0273-1>.
- [8] Z. Hussain, M.M. Al-Sobhi, B. Al-Zahrani, H.P. Singh, and T.A. Tarray. Improved randomized response approaches for additive scrambling models. *Mathematical Population Studies*, 23(4): 205–221, 2016. <https://doi.org/10.1080/08898480.2015.1087773>.
- [9] S. Gupta, S. Mehta, J. Shabbir, and S. Khalil. A unified measure of respondent privacy and model efficiency in quantitative rrt models. *Journal of Statistical Theory and Practice*, 12(3): 506–511, 2018. <https://doi.org/10.1080/15598608.2017.1415175>.
- [10] M. Murtaza, S. Singh, and Z. Hussain. An innovative optimal randomized response model using correlated scrambling variables. *Journal of Statistical Computation and Simulation*, 90(15): 2823–2839, 2020. <https://doi.org/10.1080/00949655.2020.1791118>.
- [11] G. Narjis, and J. Shabbir. An efficient new scrambled response model for estimating sensitive population mean in successive sampling. *Communications in Statistics – Simulation and Computation*, 52(11): 5327–5344, 2023. <https://doi.org/10.1080/03610918.2021.1986528>.

- [12] S. Khalil, Q. Zhang, and S. Gupta. Mean estimation of sensitive variables under measurement errors using optional rrt models. *Communications in Statistics – Simulation and Computation*, 50(5): 1417–1426, 2021. <https://doi.org/10.1080/03610918.2019.1584298>.
- [13] S. Gupta, J. Zhang, S. Khalil, and P. Sapra. Mitigating lack of trust in quantitative randomized response technique models. *Communications in Statistics – Simulation and Computation*, 53(6): 2624–2632, 2024. <https://doi.org/10.1080/03610918.2022.2082477>.
- [14] A. Chaudhuri. *Randomized Response and Indirect Questioning Techniques in Surveys*, Boca Raton, FL: Chapman & Hall/CRC, 2011.
- [15] Z. Yan, J. Wang, and J. Lai. An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics – Theory and Methods*, 38(3): 400–408, 2008. <https://doi.org/10.1080/03610920802220785>.
- [16] A. Young, S. Gupta, and R. Parks. A binary unrelated – question rrt model accounting for untruthful responding. *Involve, A Journal of Mathematics*, 12(7): 1163–1173, 2019. <https://doi.org/10.2140/involve.2019.12.1163>.
- [17] I. Saleem, A. Sanaullah, L.A. Al-Essa, and S. Bashir. Efficient estimation of population variance of a sensitive variable using a new scrambling response model. *Scientific Reports*, 13: Article ID 19913, 2023. <https://doi.org/10.1038/s41598-023-45427-2>.
- [18] Q. Zhang, S. Khalil, and S. Gupta. 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, 2021. <https://doi.org/10.1080/00949655.2021.1941018>.
- [19] M. Azeem, and S. Ali. A neutral comparative analysis of additive, multiplicative, and mixed quantitative randomized response models. *PLOS ONE*, 18(4): e0284995, 2023. <https://doi.org/10.1371/journal.pone.0284995>.
- [20] M. Azeem, A. Salam, O. Albalawi, and S. Hussain. A new unified measure for evaluation of randomized response techniques. *Heliyon*, 10(16): e35852, 2024. <https://doi.org/10.1016/j.heliyon.2024.e35852>.
- [21] M. Azeem. Introducing a weighted measure of privacy and efficiency for comparison of quantitative randomized response models. *Pakistan Journal of Statistics*, 39(3): 377–385, 2023.
- [22] H. Salemian, E. Mahmoudi, O.A. Alamri, and J. Shabbir. Improving a novel quantitative randomized response method using auxiliary variable

information. *Heliyon*, 10: e40367, 2024. <http://dx.doi.org/10.1016/j.heliyon.2024.e40367>.

- [23] M. Azeem, N. Salahuddin, S. Hussain, M. Ijaz, and A. Salam. An efficient estimator of population variance of a sensitive variable with a new randomized response technique. *Heliyon*, 10(5): e27488, 2024. <https://doi.org/10.1016/j.heliyon.2024.e27488>.

Biographies

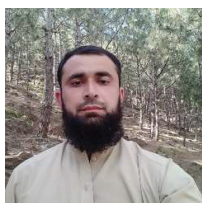


Muhammad Azeem holds PhD degree in Statistics with specialization in Survey Sampling. He is currently working as Assistant Professor in the Department of Statistics, University of Malakand, Pakistan. He has authored more than 30 peer-reviewed research publications in pure and applied Statistics. He has 10 years of post-PhD teaching experience at undergraduate and postgraduate level. He is also working as a referee for reputable impact factor journals.

Musarrat Ijaz holds PhD degree in Statistics with specialization in machine learning. She is currently working as an Assistant Professor in the Department of Statistics, Rawalpindi Women University, Rawalpindi, Pakistan. Before joining Rawalpindi Women University, she worked as a lecturer in the Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan. She has a vast experience of teaching at undergraduate and postgraduate level.

Dr. Najma Salahuddin received her PhD degree in Statistics from University of Peshawar, Pakistan. She is currently working as a lecturer in the Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan. She has over 15 years of teaching and research experience in the field of Statistics.

Soofia Iftikhar received her PhD degree in Statistics from University of Peshawar, Pakistan, with specialization in survey sampling. She is currently working as an Assistant Professor in the Department of Statistics, Shaheed Benazir Bhutto Women University, Peshawar, Pakistan. She has vast experience of teaching at undergraduate and postgraduate level.



Abdul Salam is working as an Assistant Professor in the Department of Statistics, University of Malakand, Pakistan. He is a young researcher and has completed his PhD from the University of Groningen, Netherlands. His research interests include Bayesian modelling, computational statistics, and dynamic Bayesian network models.

