METHODOLOGY
European Journal of Research Methods for the Behavioral and Social Sciences

PsychOpen GOLD

Original Article

Check for updates

Introducing an Efficient Alternative Technique to Optional Quantitative Randomized Response Models

Muhammad Azeem [1], Abdul Salam [1]

[1] *Department of Statistics, University of Malakand, Chakdara, KP, Pakistan.*

Methodology, 2023, Vol. 19(1), 24–42, https://doi.org/10.5964/meth.9921

Received: 2022-07-16 • **Accepted:** 2023-01-12 • **Published (VoR):** 2023-03-31

Handling Editor: Isabel Benítez, University of Granada, Granada, Spain

Corresponding Author: Muhammad Azeem, Department of Statistics, University of Malakand, Chakdara, KP, Pakistan. E-mail: azeemstats@uom.edu.pk

Abstract

In social surveys on sensitive characteristics, optional randomized response models give the respondents the option to either report the true response or report the scrambled response. If any respondent finds that the question being asked does not feel sensitive, he/she reports the true response. In the existing variants of optional randomized response models, the researcher doesn't know whether the respondent opted for the correct response or for a scrambled response. In practice, some of the respondents may have no problem in disclosing to the researcher that they are giving the true response and hence not opting for scrambling. This paper presents an alternative procedure to optional scrambling randomized response models, where each respondent has the choice whether or not to disclose to the researcher that he/she is giving the true response. Alternative modified versions of three existing scrambling randomized response models are presented. It is found that the efficiency of the quantitative randomized response models improves if the exact number of respondents who are opting for scrambling, is known to the researcher. Besides improvement in efficiency, the level of the respondent-privacy is the same as that of the existing models, thus resulting in an improvement in the overall quality of the existing models.

Keywords

optional randomized response, scrambling variable, sensitive surveys, privacy protection, efficiency, MSC 2020: 62D05, MSC 2020: 62F07

In sample surveys on sensitive characteristics, it is natural for the respondents to refuse to provide information. The sensitive characteristics under study may be illegal income, monthly expenditure, the number of cigarettes used per day, the marks obtained in an examination, and the amount of tax payable etc. Such refusals result in a high rate of non-response in the collected data which may badly affect the estimates of population parameters. In order to cope with refusals on sensitive variables, Warner (1965) proposed a strategy commonly called the randomized response technique. Warner's (1965) randomization technique was limited to binary variables. Warner (1971) introduced another technique for situations where the sensitive variable of interest is quantitative. Eichhorn and Hayre (1983) suggested a quantitative randomized response model where multiplicative scrambling is used as opposed to the additive scrambling model of Warner (1971).

The concept of optional randomized response techniques was first studied by Gupta et al. (2002). In all of the existing versions of optional randomized response models, the respondents are free to either report the true response or report a scrambled response. Another optional randomized response technique was introduced by Bar-Lev et al. (2004) where a multiplicative scrambling noise is utilized as opposed to the additive scrambling in the Gupta et al. (2002) technique. Yan et al. (2008) introduced a measure for the respondent-privacy level ensured by a quantitative randomized response model. Diana and Perri (2011) introduced a randomized response procedure which utilizes both additive and multiplicative scrambling. Hussain et al. (2016) introduced a randomized response strategy which uses additive and subtractive scrambling. Gupta et al. (2018) presented a joint measure of privacy protection and efficiency for assessing the overall quality of quantitative randomized response models. Narjis and Shabbir (2021) proposed a modified variant of the Gjestvang and Singh (2009) model. Khalil et al. (2021) analyzed the influence of measurement errors on the estimators of the mean in sensitive surveys. Gupta et al. (2022) introduced a scrambled randomized response procedure which improved the Diana and Perri (2011) technique in terms of efficiency and privacy protection. Further research studies on randomized response models can be found in Kalucha et al. (2016), Murtaza et al. (2021), Yan et al. (2008), Young et al. (2019), and Zhang et al. (2021).

Besides simple random sampling, the ranked set sampling scheme can also be combined with randomized response technique to obtain efficient estimates of the parameters of interest. For detailed literature, one may refer to the studies of Mahdizadeh and Zamanzade (2021a, 2021b) and Mahdizadeh and Zamanzade (in press, 2022a, 2022b).

The next section presents some of the existing quantitative randomized response models.

Some Existing Quantitative Models and Evaluation Metrics

Let the population under consideration consists of N units and a simple random sample of n units is obtained with replacement. Further, let Y denote the sensitive variable of interest and S denote an additive scrambling variable and let us assume that $E(Y_i) = \mu_Y$, $E(S) = 0$, $V(Y_i) = \sigma_Y^2$, $V(S) = \sigma_S^2$. Moreover, let T be a multiplicative scrambling variable such that $E(T) = 1$, and $V(T) = \sigma_T^2$, where σ_Y^2, σ_T^2, and σ_S^2 are population variances of variable Y, T, and S, respectively, and μ_Y is the mean of the sensitive variable Y. It is further assumed that all variables are independent of each other. In this section, some existing quantitative scrambling techniques are presented.

The Warner (1971) Additive Model

The reported responses under the Warner (1971) additive scrambling model are as follows:

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

An unbiased mean estimator of Y based on the Warner (1971) model is given as:

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

The variance of $\hat{\mu}_W$ is given as:

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

The Eichhorn and Hayre (1983) Model

The reported responses under the Eichhorn and Hayre (1983) technique are as follows:

$$Z = TY \tag{4}$$

An unbiased mean estimator of Y under the Eichhorn and Hayre (1983) technique is as follows:

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

The variance of $\hat{\mu}_{EH}$ is given as:

$$Var(\hat{\mu}_{EH}) = \frac{\sigma_Y^2}{n} + \frac{\sigma_T^2(\sigma_Y^2 + \mu_Y^2)}{n} \tag{6}$$

The Diana and Perri (2011) Quantitative Model

The reported responses under the Diana and Perri (2011) quantitative scrambling model are given as:

$$Z = TY + S \tag{7}$$

An unbiased mean estimator of the sensitive variable of interest on the basis of the Diana and Perri (2011) technique is given as:

$$\widehat{\mu}_{DP} = \frac{1}{n}\sum_{i=1}^{n} Z_i \tag{8}$$

The variance of $\widehat{\mu}_{DP}$ is given by:

$$Var(\widehat{\mu}_{DP}) = \frac{1}{n}\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2 + \sigma_S^2\right] \tag{9}$$

The measure of privacy level due to Yan et al. (2008) for comparison of randomized response models is as follows:

$$\nabla = E[Z - Y]^2 \tag{10}$$

The higher the value of ∇, the higher the level of privacy of the respondents provided by a particular randomized response model.

The joint measure of Gupta et al. (2018) for privacy and efficiency is as follows:

$$\delta = \frac{MSE}{\nabla} \tag{11}$$

From Equation 11, one can clearly observe that a lower value of δ is preferable.

For the Warner's (1971) model, the measure of respondent-privacy is as follows:

$$\nabla_W = E[Y + S - Y]^2 = E(S^2) = \sigma_S^2 \tag{12}$$

The joint measure of efficiency and privacy for the Warner's (1971) model is given as:

$$\delta_W = \frac{Var(\widehat{\mu}_W)}{\nabla_W} = \frac{1}{n}\left(\frac{\sigma_Y^2 + \sigma_S^2}{\sigma_S^2}\right) \tag{13}$$

For the Eichhorn and Hayre (1983) quantitative technique, the measure of privacy is given by:

$$\nabla_{EH} = E[TY - Y]^2 = E(T^2)E(Y^2) + E(Y^2) - 2E(T)E(Y^2)$$

or

$$\nabla_{EH} = \sigma_T^2(\sigma_Y^2 + \mu_Y^2) \tag{14}$$

The joint measure of model-efficiency and respondent-privacy for the Eichhorn and Hayre (1983) quantitative technique is given as:

$$\delta_{EH} = \frac{Var(\widehat{\mu}_{EH})}{\nabla_{EH}} = \frac{1}{n}\left[\frac{\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_Y^2}{\sigma_T^2(\sigma_Y^2 + \mu_Y^2)}\right] \tag{15}$$

The measure of privacy for the Diana and Perri (2011) model is given by:

$$\nabla_{DP} = E[TY + S - Y]^2 = \sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2 \tag{16}$$

The joint measure of privacy and efficiency for the Diana and Perri (2011) model is given as:

$$\delta_{DP} = \frac{Var(\widehat{\mu}_{DP})}{\nabla_{DP}} = \frac{1}{n}\left[\frac{\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2 + \sigma_Y^2}{\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2}\right] \tag{17}$$

In each of the proposed models, since the respondents in the first group give true response, so the measure of privacy is zero. In the second group, the responses provided by the respondents are the same as those of the corresponding existing models. The only difference is that the sample size n_2 is used in place of n. Since the mathematical expression for ∇ in the case of each model is independent of the sample size n, so the value of∇ for each proposed model is the same as that of the corresponding existing model. That is, for the proposed Model I, the measure of privacy is given by:

$$\nabla_{P1} = \sigma_S^2 \tag{18}$$

For the proposed Model II, the measure of privacy is given by:

$$\nabla_{P2} = \sigma_T^2(\sigma_Y^2 + \mu_Y^2) \tag{19}$$

For the proposed Model III, the measure of privacy is given by:

$$\nabla_{P3} = \sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2 \tag{20}$$

The joint measure of efficiency and privacy for the proposed Model I is given as:

$$\delta_{P1} = \frac{\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_S^2}{\sigma_S^2} \tag{21}$$

The joint measure of efficiency and privacy for the proposed Model II is given as:

$$\delta_{P2} = \frac{\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_T^2(\sigma_Y^2 + \mu_Y^2)}{\sigma_T^2(\sigma_Y^2 + \mu_Y^2)} \tag{22}$$

The joint measure of efficiency and privacy for the proposed Model III is given as:

$$\delta_{P3} = \frac{\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\left[\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2\right]}{\sigma_T^2\left(\sigma_Y^2 + \mu_Y^2\right) + \sigma_S^2} \tag{23}$$

Proposed Models

In the proposed technique, the researcher asks each respondent whether he/she wants to report the correct answer or prefers to report a scrambled response. The researcher not only collects response on the sensitive variable under study but also records whether it is true response or scrambled response. At the end of data collection process, the researcher knows how many of the collected responses are scrambled. This procedure enables the researcher to know the priority of the respondents about true or scrambled responses. Let n_1 out of n respondents disclose to the researcher that they are providing the true response without using the scrambling technique, and let the remaining $n_2 = n - n_1$ respondents prefer the scrambling technique for privacy protection. This section presents the modified versions of the models given in Section 2.

Proposed Model I

Motivated by Warner (1971) and Gupta et al. (2002), every participant is asked to either report the true response or use a scrambling procedure. Every respondent also has to tell the researcher whether his/her response is a true or scrambled response. This enables the researcher to know the exact number of respondents who opted for true response, and the number of respondents who opted for scrambled response. Under the proposed Model I, there are two groups of respondents:

i. The n_1 respondents who report the true response Y.
ii. The n_2 respondents who report the scrambled response $Z = Y + S$.

The mean of the first group is:

$$\bar{Y} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i \tag{24}$$

The mean of the second group is:

$$\bar{Z} = \frac{1}{n_2}\sum_{i=1}^{n_2} Z_i = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(Y_i + S_i\right) \tag{25}$$

The mean estimator of the sensitive variable under study is the weighted mean of the two groups. That is;

$$\hat{\mu}_{P1} = \frac{n_1\bar{Y} + n_2\bar{Z}}{n_1 + n_2} \tag{26}$$

where $n_1 + n_2 = n$.

Proposed Model II

Motivated by Eichhorn and Hayre (1983), every respondent is requested to either report the true response or use a multiplicative scrambling. Every respondent also has to tell the researcher whether his/her response is true or scrambled. This enables the researcher to know the exact number of respondents who opted for true response, and the number of respondents who opted for scrambled response. Under the proposed Model II, there are two groups of respondents:

i. The n_1 respondents who report the true response Y.
ii. The n_2 respondents who report the scrambled response $Z = TY$.

The mean of the first group is:

$$\bar{Y} = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i \tag{27}$$

The mean of the second group is:

$$\bar{Z} = \frac{1}{n_2}\sum_{i=1}^{n_2} Z_i \tag{28}$$

The mean estimator of the sensitive variable under study is the weighted mean of the two groups. That is;

$$\hat{\mu}_{P2} = \frac{n_1\bar{Y} + n_2\bar{Z}}{n_1 + n_2} \tag{29}$$

Proposed Model III

Motivated by Diana and Perri (2011), every respondent is requested to either report the true response or use a scrambling procedure. Every respondent also has to tell the researcher whether his/her response is true or scrambled. This enables the researcher to know the exact number of respondents who opted for true response, and the number of respondents who opted for scrambled response. Under the proposed Model III, there are two groups of respondents:

i. The n_1 respondents who report the true response Y.
ii. The n_2 respondents who report the scrambled response $Z = TY + S$.

The mean of the first group is:

$$\bar{Y} = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i \tag{30}$$

The mean of the second group is:

$$\bar{Z} = \frac{1}{n_2} \sum_{i=1}^{n_2} Z_i \tag{31}$$

The mean estimator of the sensitive variable under study is the weighted mean of the two groups. That is;

$$\widehat{\mu}_{P3} = \frac{n_1\bar{Y} + n_2\bar{Z}}{n_1 + n_2} \tag{32}$$

Mean and Variance

The section presents the proof of unbiasedness and derivation of variances of the mean estimators under the proposed models.

Theorem 1: The estimators $\widehat{\mu}_{P1}$, $\widehat{\mu}_{P2}$ and $\widehat{\mu}_{P3}$ are unbiased estimators of the population mean μ_Y.

Proof: Taking expectation on both sides of Equation 26 yields:

$$E\left(\widehat{\mu}_{P1}\right) = E\left(\frac{n_1\bar{Y} + n_2\bar{Z}}{n_1 + n_2}\right) = \frac{n_1E\left(\bar{Y}\right) + n_2E\left(\bar{Z}\right)}{n_1 + n_2} \tag{33}$$

Taking expectation of Equations 24 and 25 yields:

$$E\left(\bar{Y}\right) = E\left(\frac{1}{n_1} \sum_{i=1}^{n_1} Y_i\right) = \mu_Y \tag{34}$$

and

$$E\left(\bar{Z}\right) = \frac{1}{n_2} \sum_{i=1}^{n_2} E\left(Y_i + S_i\right) = \mu_Y \tag{35}$$

Using Equations 24 and 35 in 33 yields:

$$E\left(\widehat{\mu}_{P1}\right) = \frac{n_1\mu_Y + n_2\mu_Y}{n_1 + n_2} = \mu_Y \tag{36}$$

In a similar manner, the unbiasedness of $\widehat{\mu}_{P2}$ and $\widehat{\mu}_{P3}$ can be easily proved.

Theorem 2: The variances of the estimators $\widehat{\mu}_{P1}$, $\widehat{\mu}_{P2}$ and $\widehat{\mu}_{P3}$ are given by:

$$Var(\widehat{\mu}_{P1}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_S^2 \tag{37}$$

$$Var(\widehat{\mu}_{P2}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_T^2(\sigma_Y^2 + \mu_Y^2) \tag{38}$$

$$Var(\widehat{\mu}_{P3}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2] \tag{39}$$

Proof: Applying variance on both sides of Equation 26 yields:

$$Var(\widehat{\mu}_{P1}) = \frac{n_1^2 Var(\bar{Y}) + n_2^2 Var(\bar{Z})}{(n_1 + n_2)^2} \tag{40}$$

Applying variance on both sides of Equation 24 and 25 yields:

$$Var(\bar{Y}) = \frac{1}{n_1^2}\sum_{i=1}^{n_1} Var(Y_i) = \frac{\sigma_Y^2}{n_1} \tag{41}$$

and

$$Var(\bar{Z}) = \frac{1}{n_2^2}\sum_{i=1}^{n_2} Var(Y_i + S_i) = \frac{1}{n_2}(\sigma_Y^2 + \sigma_S^2) \tag{42}$$

Using Equation 41 and 42 in Equation 40 and simplification yields:

$$Var(\widehat{\mu}_{P1}) = \frac{1}{(n_1 + n_2)^2}[(n_1 + n_2)\sigma_Y^2 + n_2\sigma_S^2]$$

or

$$Var(\widehat{\mu}_{P1}) = \frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_S^2$$

Using the same procedure and assuming independence of variables, the variances of $\widehat{\mu}_{P2}$ and $\widehat{\mu}_{P3}$ can be easily obtained.

An Application of the Proposed Technique

The proposed Model III was applied to the problem of estimation of the true mean of the Grade Point Average (GPA) of the 175 students of the Department of Statistics, University of Malakand, Pakistan. A simple random sample of 40 students was obtained from the undergraduate students currently enrolled in the department. Currently, a total of 175 students are studying in the undergraduate program of the Department of Statistics in the University of Malakand, Pakistan. Each of the 40 selected students was asked whether he/she wants to report the true GPA. If the student's answer was 'yes', he/she reported his/her true GPA. If a respondent did not want to report his/her true GPA, he/she was given a deck of 100 cards along with a calculator. Each card had two random numbers printed on it—one for variable T and the other for variable S. The random numbers for both scrambling variables were generated using a normal distribution. The random numbers for the additive scrambling variable S were generated using a normal distribution having mean 0 and variance 0.5. The random numbers for the multiplicative scrambling variable T were generated using a normal distribution having mean 1 and variance 0.5. The respondents who opted for scrambled response were told not to disclose their true GPA to the interviewer, and hence their privacy protection was ensured. The respondents were also told not to show the selected card to the interviewer. Out of 40 students, 16 students wanted to report the true GPA, whereas the remaining 24 students opted for scrambled response. The responses reported by the 40 sampled students are presented in Table 1.

Table 1

Responses Reported by Students

True Responses				Scrambled Responses					
2.78	3.41	2.88	3.16	2.9677	4.3116	2.7810	3.3618	3.5319	2.4298
3.75	2.47	1.99	3.33	1.5986	2.9468	2.6090	3.7874	4.0074	1.9924
3.90	3.64	2.43	1.88	3.8477	1.8653	2.9668	4.4793	1.3270	4.6992
2.58	3.16	2.24	1.98	2.7437	3.3362	1.6973	3.4518	3.1946	2.6173

In Table 1, one may observe that some of the reported scrambled responses exceed 4.0 although the students' actual GPA was on the scale of 4.0. If the researcher generates random numbers from normal distribution having a large mean or variance, then the reported scrambled responses may result in large values which will look unnatural for students' GPA dataset. Moreover, it may also lead to overestimate the true mean GPA since the estimates are calculated from the observed responses. It is therefore advised that the researchers should keep in mind to always choose appropriate choices of the parameters of the distribution from which random numbers are to be generated. The parameters should be chosen in such a way that the reported scrambled responses do not deviate too much from the possible range of the quantitative variable of interest. In the given example, one may observe that most of the scrambled responses cover the possible range of the GPA which is from 0 to 4.

Efficiency Comparison

The suggested Model I is more efficient than Warner's (1971) model if:

$$Var(\hat{\mu}_{P1}) \leq Var(\hat{\mu}_{W})$$

or

$$\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_S^2 \leq \frac{\sigma_Y^2}{n} + \frac{n\sigma_S^2}{n^2}$$

or

$$n_2 \leq n \tag{43}$$

Condition 43 always holds.

The suggested Model II is more efficient than the Eichhorn and Hayre (1983) model if:

$$Var(\hat{\mu}_{P2}) \leq Var(\hat{\mu}_{EH})$$

or

$$\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\sigma_T^2(\sigma_Y^2 + \mu_Y^2) \leq \frac{\sigma_Y^2}{n} + \frac{n\sigma_T^2(\sigma_Y^2 + \mu_Y^2)}{n^2}$$

or

$$n_2 \leq n \tag{44}$$

Condition 44 always holds.

The suggested Model III is more efficient than the Diana and Perri (2011) model if:

$$Var(\hat{\mu}_{P3}) \leq Var(\hat{\mu}_{DP})$$

or

$$\frac{\sigma_Y^2}{n} + \frac{n_2}{n^2}\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2\right] \leq \frac{\sigma_Y^2}{n} + \frac{n}{n^2}\left[\sigma_T^2(\sigma_Y^2 + \mu_Y^2) + \sigma_S^2\right]$$

or

$$n_2 \leq n \tag{45}$$

Condition 45 always holds.

Table 2 displays the variances of the mean estimator under the Warner (1971) and the Eichhorn and Hayre (1983) scrambling model, the Diana and Perri (2011) quantitative model, and the three proposed models for various choices of n_1 and n_2. One may clearly observe the improvement in efficiency of the proposed models over the existing models.

Table 2

Variances of the Mean Under Different Models

Population Variance		Number of Respondents		Variance of the Mean Estimator					
σ_T^2	σ_S^2	n_1	n_2	$Var(\hat{\mu}_W)$	$Var(\hat{\mu}_{EH})$	$Var(\hat{\mu}_{DP})$	$Var(\hat{\mu}_{P1})$	$Var(\hat{\mu}_{P2})$	$Var(\hat{\mu}_{P3})$
4	3	10	40	0.16	18.50	18.56	0.15	14.82	14.87
		20	30	0.16	18.50	18.56	0.14	11.14	11.18
		30	20	0.16	18.50	18.56	0.12	7.46	7.48
		40	10	0.16	18.50	18.56	0.11	3.78	3.79
	6	10	40	0.22	18.50	18.62	0.20	14.82	14.92
		20	30	0.22	18.50	18.62	0.17	11.14	11.21
		30	20	0.22	18.50	18.62	0.15	7.46	7.51
		40	10	0.22	18.50	18.62	0.12	3.78	3.80
8	5	10	40	0.20	36.90	37.00	0.18	29.54	29.62
		20	30	0.20	36.90	37.00	0.16	22.18	22.24
		30	20	0.20	36.90	37.00	0.14	14.82	14.86
		40	10	0.20	36.90	37.00	0.12	7.46	7.48
	10	10	40	0.30	36.90	37.10	0.26	29.54	29.70
		20	30	0.30	36.90	37.10	0.22	22.18	22.30
		30	20	0.30	36.90	37.10	0.18	14.82	14.90
		40	10	0.30	36.90	37.10	0.14	7.46	7.50
12	8	10	40	0.26	55.30	55.46	0.23	44.26	44.39
		20	30	0.26	55.30	55.46	0.20	33.22	33.32
		30	20	0.26	55.30	55.46	0.16	22.18	22.24
		40	10	0.26	55.30	55.46	0.13	11.14	11.17
	15	10	40	0.40	55.30	55.60	0.34	44.26	44.50
		20	30	0.40	55.30	55.60	0.28	33.22	33.40
		30	20	0.40	55.30	55.60	0.22	22.18	22.30
		40	10	0.4	55.3	55.6	0.16	11.14	11.2

Note. $\mu_Y = 15$, $\sigma_Y^2 = 5$, $n = 50$. W, EH, DP, p1, p2, p3 = the Warner (1971), the Eichhorn and Hayre (1983), the Diana and Perri (2011), and the three proposed models, respectively.

Table 3 displays the improvement in terms of δ values over the existing models.

Table 3

δ Values for Different Models

σ_T^2	σ_S^2	n_1	n_2	δ_W	δ_{EH}	δ_{DP}	δ_{P1}	δ_{P2}	δ_{P3}
4	3	10	40	0.053333	0.020109	0.020108	0.049333	0.016109	0.016108
		20	30	0.053333	0.020109	0.020108	0.045333	0.012109	0.012108
		30	20	0.053333	0.020109	0.020108	0.041333	0.008109	0.008108
		40	10	0.053333	0.020109	0.020108	0.037333	0.004109	0.004108
	6	10	40	0.036667	0.020109	0.020108	0.032667	0.016109	0.016108
		20	30	0.036667	0.020109	0.020108	0.028667	0.012109	0.012108
		30	20	0.036667	0.020109	0.020108	0.024667	0.008109	0.008108
		40	10	0.036667	0.020109	0.020108	0.020667	0.004109	0.004108
8	5	10	40	0.04	0.020054	0.020054	0.036	0.016054	0.016054
		20	30	0.04	0.020054	0.020054	0.032	0.012054	0.012054
		30	20	0.04	0.020054	0.020054	0.028	0.008054	0.008054
		40	10	0.04	0.020054	0.020054	0.024	0.004054	0.004054
	10	10	40	0.03	0.020054	0.020054	0.026	0.016054	0.016054
		20	30	0.03	0.020054	0.020054	0.022	0.012054	0.012054
		30	20	0.03	0.020054	0.020054	0.018	0.008054	0.008054
		40	10	0.03	0.020054	0.020054	0.014	0.004054	0.004054
12	8	10	40	0.0325	0.020036	0.020036	0.0285	0.016036	0.016036
		20	30	0.0325	0.020036	0.020036	0.0245	0.012036	0.012036
		30	20	0.0325	0.020036	0.020036	0.0205	0.008036	0.008036
		40	10	0.0325	0.020036	0.020036	0.0165	0.004036	0.004036
	15	10	40	0.026667	0.020036	0.020036	0.022667	0.016036	0.016036
		20	30	0.026667	0.020036	0.020036	0.018667	0.012036	0.012036
		30	20	0.026667	0.020036	0.020036	0.014667	0.008036	0.008036
		40	10	0.026667	0.020036	0.020036	0.010667	0.004036	0.004036

Note. $\mu_Y = 15$, $\sigma_Y^2 = 5$, $n = 50$.

Simulation Study

In order to show improvement in efficiency and privacy protection, a simulation study was carried out by generating an artificial population of $N = 5000$ units from a normal distribution having mean 200 and variance 25. For the additive scrambling variable S, the random numbers were generated using a normal distribution with mean 0 and variance 1.5625. For the multiplicative scrambling variable T, the random numbers were generated using a normal distribution with mean 1 and variance 1.5625. A total of 1000 iterations of sample selection were run, using the sample size $n = 1000$ at each iteration. The results of the amount of bias in the mean estimator under each of the three proposed models are presented in Table 4. Likewise, the results of the simulated variances can be observed in Table 5 with δ values in Table 6. Observing Tables 4, 5, and 6, one may clearly see the improvement over the existing models. In Table 4, most of the simulated values of bias are close to zero for all of three proposed models, which is consistent with the unbiasedness proved in Equation 36.

Table 4

Simulated Bias in the Mean Estimator Under the Proposed Models

Variance	Population Value		Simulated Bias		
σ_S / σ_T	n_1	n_2	$Bias(\hat{\mu}_{P1})$	$Bias(\hat{\mu}_{P2})$	$Bias(\hat{\mu}_{P3})$
1.25	200	800	-0.04231722	-0.3359392	-0.3343549
	400	600	-0.05178339	0.06455729	0.06581804
	500	500	-0.04452886	0.04366002	0.04456227
	600	400	-0.04710829	-0.03758789	-0.03710422
	800	200	-0.03654629	-0.2208454	-0.2205013
1.5	200	800	-0.04200036	-0.3943467	-0.3924455
	400	600	-0.05153124	0.08807758	0.08959048
	500	500	-0.04434841	0.06147825	0.06256095
	600	400	-0.04701155	-0.03558708	-0.03500667
	800	200	-0.03647746	-0.2576364	-0.2572234
1.75	200	800	-0.04168349	-0.4527542	-0.4505362
	400	600	-0.05127909	0.1115979	0.1133629
	500	500	-0.04416796	0.07929647	0.08055963
	600	400	-0.04691482	-0.03358627	-0.03290912
	800	200	-0.03640864	-0.2944274	-0.2939456
2	200	800	-0.04136663	-0.5111618	-0.5086269
	400	600	-0.05102694	0.1351182	0.1371354
	500	500	-0.04398751	0.0971147	0.09855831
	600	400	-0.04681808	-0.03158545	-0.03081157
	800	200	-0.03633981	-0.3312184	-0.3306678

Table 5

Simulated Variances of the Mean Under the Proposed and Existing Models

Variance	Population Value		Population Mean					
σ_S / σ_T	n_1	n_2	$Var(\hat{\mu}_W)$	$Var(\hat{\mu}_{P1})$	$Var(\hat{\mu}_{EH})$	$Var(\hat{\mu}_{P2})$	$Var(\hat{\mu}_{DP})$	$Var(\hat{\mu}_{P3})$
1.25	200	800	0.02383096	0.02360185	61.79515	48.81937	61.81969	48.83813
	400	600	0.02319297	0.02303067	58.02994	35.89469	58.05022	35.90795
	500	500	0.02290962	0.02259029	64.59142	30.46517	64.57387	30.45292
	600	400	0.02425798	0.02365736	61.87235	24.96826	61.91008	24.97025
	800	200	0.02252473	0.02109507	63.99624	11.9735	64.02271	11.97824
1.5	200	800	0.02453836	0.02419761	88.99663	70.30539	89.03195	70.33238
	400	600	0.0238225	0.02350492	83.54286	51.69235	83.57212	51.71143
	500	500	0.02352331	0.02296447	93.01462	43.88139	92.98943	43.86372
	600	400	0.02486492	0.0239517	89.089	35.94049	89.14334	35.94329
	800	200	0.02324659	0.02123525	92.14935	17.22565	92.18741	17.23246
1.75	200	800	0.02537135	0.02489702	121.1471	95.70026	121.1951	95.73697
	400	600	0.02457703	0.02406322	113.6929	70.36438	113.7328	70.39033
	500	500	0.02426035	0.02340347	126.6076	59.74012	126.5734	59.71605
	600	400	0.02558184	0.02429185	121.2546	48.90709	121.3285	48.91086
	800	200	0.02409082	0.02139922	125.4218	23.43166	125.4736	23.44092
2	200	800	0.02632994	0.02570009	158.2465	125.004	158.3092	125.0519
	400	600	0.02545657	0.02470559	148.48	91.91078	148.5322	91.94466
	500	500	0.02512075	0.02390727	165.3703	78.04135	165.3257	78.0099
	600	400	0.02640872	0.02467781	158.3691	63.86808	158.4657	63.87295
	800	200	0.02505742	0.02158696	163.8136	30.59154	163.8811	30.60362

Table 6

Simulated δ Values of the Proposed and Existing Models

Variance	Population Value	Change Value					
σ_S / σ_T	n_1	δ_W	δ_{P1}	δ_{EH}	δ_{P2}	δ_{DP}	δ_{P3}
1.25	200	0.0151905	0.0150673	0.0009907802	0.0007811312	0.0009911831	0.0007814225
	400	0.01490502	0.0148246	0.0009260774	0.0005744241	0.0009263716	0.0005746333
	500	0.01472327	0.0145062	0.001029526	0.0004839192	0.001029244	0.0004837062
	600	0.01557841	0.0152525	0.0009908548	0.0004004777	0.0009914397	0.0004004859
	800	0.01445102	0.01365198	0.001022456	0.0001915471	0.001022894	0.0001916116
1.5	200	0.01086234	0.01072791	0.0009909185	0.0007811924	0.000991321	0.0007814833
	400	0.01062884	0.01050509	0.0009258519	0.0005744713	0.0009261468	0.0005746804
	500	0.01049757	0.01024015	0.001029553	0.0004840504	0.001029271	0.000483837
	600	0.01108908	0.01072322	0.000990769	0.00040031	0.0009913538	0.0004003176
	800	0.01035557	0.009544028	0.001022394	0.0001913655	0.001022831	0.0001914299

Variance	Population Value	Change Value					
σ_S / σ_T	n_1	δ_W	δ_{P1}	δ_{EH}	δ_{P2}	δ_{DP}	δ_{P3}
1.75	200	0.008251483	0.008109662	0.0009910293	0.0007812483	0.0009914316	0.0007815389
	400	0.008054057	0.007899927	0.0009257027	0.0005745169	0.0009259983	0.0005747259
	500	0.007953403	0.007666759	0.001029583	0.000484156	0.001029302	0.0004839422
	600	0.008381798	0.007989669	0.0009907202	0.0004002028	0.0009913048	0.0004002098
	800	0.007883378	0.007066392	0.001022361	0.0001912472	0.001022797	0.0001913115
2	200	0.006556218	0.006409192	0.000991119	0.0007812969	0.0009915211	0.0007815872
	400	0.006385273	0.006208672	0.0009255975	0.0005745577	0.0009258934	0.0005747667
	500	0.006304594	0.005995812	0.001029612	0.0004842417	0.001029332	0.0004840276
	600	0.006624443	0.006213837	0.0009906904	0.0004001292	0.000991275	0.0004001358
	800	0.006277043	0.005457874	0.001022342	0.0001911647	0.001022778	0.0001912289

Discussion and Conclusion

This paper presents an alternative procedure to the so-called optional quantitative randomized response models. Modified versions of the Warner (1971), the Eichhorn and Hayre (1983), and the Diana and Perri (2011) models were analyzed in previous sections. The efficiency conditions are strong and always hold, which shows that suggested modified variants are superior to the existing versions.

Observing Table 2 and Table 3, the improvement over the existing methods may be seen for various choices of n_1 and n_2. Table 3 shows the improvement in terms of δ values over the existing models. It is observed that the suggested Model I is superior to the Warner (1971) model, Model II is better than the Eichhorn and Hayre (1983) quantitative model, and the proposed Model III is better than the Diana and Perri (2011) model. Moreover, one may observe that among the proposed models, Model I is the best model in terms of efficiency. However, the proposed Model III is the best model if δ values are taken into account. It is also observed that as n_1 increases, the variance of the mean for each of the proposed models decreases. This means that as the number of respondents opting for true response increases, the efficiency of the models increases. Therefore, it is advised to the researchers to motivate the respondents to opt for true response as far as possible. This will minimize the number of those opting for scrambled responses, thus resulting in efficient estimates of the mean.

Table 4 shows that among the three proposed models, the proposed Model I produces less amount of simulated bias compared to the proposed Model II and Model III, which makes Model I the best of the three models, in situations where unbiasedness is the priority for model selection. Moreover, the proposed Model I utilizes only additive scrambling, which makes it simpler than the proposed Model III where the respondents have to scramble their response using both additive and multiplicative scrambling. Moreover, the proposed Model I is also much more efficient than the proposed Model II and Model

III. However, Table 6 shows that the simulated values of the joint measure of privacy and efficiency under Model I are the worst among the three proposed models. Further, one may also observe from Table 5 that the proposed Model II and Model III are nearly equally efficient but Model II is better in terms of simplicity as it only uses multiplicative scrambling. The proposed Model III, on the other hand, provides a higher level of privacy protection since the respondents use both additive and multiplicative scrambling to report their responses.

The current study analyzed the efficiency of the mean estimator under the suggested alternative to the optional randomized response models. It may be interesting if researchers study estimation of other parameters like population median, variance, population proportion etc. under the suggested randomized response models.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have no conflict of interest to declare.

Data Availability: Data is freely available at Supplementary Materials.

Supplementary Materials

For this article, the R code used to construct the data sets, and the efficiency comparison tables are available via PsychArchives (for access see Index of Supplementary Materials below):

Index of Supplementary Materials

Azeem, M., & Salam, A. (2023). *Supplementary materials to "Introducing an efficient alternative technique to optional quantitative randomized response models"* [R code, efficiency tables]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.12592

References

Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika, 60*(3), 255–260. https://doi.org/10.1007/s001840300308

Diana, G., & Perri, P. F. (2011). A class of estimators of quantitative sensitive data. *Statistische Hefte, 52*(3), 633–650. https://doi.org/10.1007/s00362-009-0273-1

Eichhorn, B. H., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference, 7*(4), 307–316. https://doi.org/10.1016/0378-3758(83)90002-2

Gjestvang, C. R., & Singh, S. (2009). An improved randomized response model: Estimation of mean. *Journal of Applied Statistics, 36*(12), 1361–1367. https://doi.org/10.1080/02664760802684151

Gupta, S., Gupta, B., & Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference, 100*(2), 239–247. https://doi.org/10.1016/S0378-3758(01)00137-9

Gupta, S., Mehta, S., Shabbir, J., & Khalil, S. (2018). A unified measure of respondent privacy and model efficiency in quantitative RRT models. *Journal of Statistical Theory and Practice, 12*(3), 506–511. https://doi.org/10.1080/15598608.2017.1415175

Gupta, S., Zhang, J., Khalil, S., & Sapra, P. (2022). Mitigating lack of trust in quantitative randomized response technique models. *Communications in Statistics. Simulation and Computation*. https://doi.org/10.1080/03610918.2022.2082477

Hussain, Z., Al-Sobhi, M. M., Al-Zahrani, B., Singh, H. P., & Tarray, T. A. (2016). Improved randomized response approaches for additive scrambling models. *Mathematical Population Studies, 23*(4), 205–221. https://doi.org/10.1080/08898480.2015.1087773

Kalucha, G., Gupta, S., & Shabbir, J. (2016). A two-step approach to ratio and regression estimation of finite population mean using optional randomized response models. *Hacettepe Journal of Mathematics and Statistics, 45*(6), 1819–1830.

Khalil, S., Zhang, Q., & 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. https://doi.org/10.1080/03610918.2019.1584298

Mahdizadeh, M., & Zamanzade, E. (2021a). Smooth estimation of the area under ROC curve in multistage ranked set sampling. *Statistische Hefte, 62*(4), 1753–1776. https://doi.org/10.1007/s00362-019-01151-6

Mahdizadeh, M., & Zamanzade, E. (2021b). New estimators of the variances under strata in ranked set sampling. *Soft Computing, 25*(13), 8007–8013. https://doi.org/10.1007/s00500-021-05787-1

Mahdizadeh, M., & Zamanzade, E. (in press). On estimating the area under ROC curve in ranked set sampling. *Statistical Methods in Medical Research*. https://doi.org/10.1177/09622802221097211

Mahdizadeh, M., & Zamanzade, E. (2022a). On interval estimation of the population mean in ranked set sampling. *Communications in Statistics. Simulation and Computation, 51*(5), 2747–2768. https://doi.org/10.1080/03610918.2019.1700276

Mahdizadeh, M., & Zamanzade, E. (2022b). Using a rank-based design in estimating prevalence of breast cancer. *Soft Computing, 26*(7), 3161–3170. https://doi.org/10.1007/s00500-022-06770-0

Murtaza, M., Singh, S., & Hussain, Z. (2021). Use of correlated scrambling variables in quantitative randomized response technique. *Biometrical Journal. Biometrische Zeitschrift, 63*(1), 134–147. https://doi.org/10.1002/bimj.201900137

Narjis, G., & Shabbir, J. (2021). An efficient new scrambled response model for estimating sensitive population mean in successive sampling. *Communications in Statistics. Simulation and Computation*. https://doi.org/10.1080/03610918.2021.1986528

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63–69. https://doi.org/10.1080/01621459.1965.10480775

Warner, S. L. (1971). The linear randomized response model. *Journal of the American Statistical Association, 66*(336), 884–888. https://doi.org/10.1080/01621459.1971.10482364

Yan, Z., Wang, J., & 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. https://doi.org/10.1080/03610920802220785

Young, A., Gupta, S., & Parks, R. (2019). A binary unrelated-question RRT model accounting for untruthful responding. *Involve: A Journal of Mathematics, 12*(7), 1163–1173. https://doi.org/10.2140/involve.2019.12.1163

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. *Journal of Statistical Computation and Simulation, 91*(17), 3492–3504. https://doi.org/10.1080/00949655.2021.1941018

EUROPEAN ASSOCIATION OF METHODOLOGY

Methodology is the official journal of the European Association of Methodology (EAM).

leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.