Communications in Advanced Mathematical Sciences
Vol. 7, No. 2, 104-113, 2024
Research Article
e-ISSN: 2651-4001
https://doi.org/10.33434/cams.1435108

An Improved Quantitative Optional Randomised Response Technique with Additive Scrambling using Two Questions Approach

Neeraj Tiwari[1]*, Tanuj Kumar Pandey[2]

Abstract
In this paper, an improved two-stage and three-stage optional randomized response (ORR) models for quantitative variables that make the use of additive scrambling was proposed. These two-stage and three-stage models achieve efficient estimation of the mean and sensitivity level simultaneously in the single sample by using two questions. It is found that the proposed models perform better than the existing ORR models in terms of estimating sensitive attribute and sensitivity level simultaneously. It is found that the proposed three stage ORR model provides better estimates than the two-stage and one-stage ORR models and offers more privacy to the respondents with suitable choice of design parameters. The properties of the proposed models are demonstrated with the help of a numerical study.

Keywords: Optional randomised response, Sensitive surveys, Sensitivity level, Two-questions approach
2010 AMS: 62D05

[1] *Department of Statistics, S.S.J. University, Almora, neerajtiwari.amo@gmail.com, ORCID:0000-0003-2198-884X*
[2] *Department of Statistics, Kumaun University, Nainital, tanujkumar.613@gmail.com, ORCID: 0000-0002-2508-6221*
***Corresponding author**
Received: 11 February 2024, **Accepted:** 21 June 2024, **Available online:** 22 June 2024
How to cite this article: N. Tiwari, T. K. Pandey, *An improved quantitative optional randomised response technique with additive scrambling using two questions approach*, Commun. Adv. Math. Sci., 7(2) (2024), 104-113.

1. Introduction

In the realm of research, assessing the sensitive characteristics poses a unique challenge. Questions about illegal activities, stigmatized behaviours, or private beliefs often fall prey to biased responses, fuelled by factors like social desirability or fear of judgment. This is where the randomized response technique (RRT) emerges as a beacon of hope, offering a robust method to collect accurate data while preserving the privacy of the respondent. First proposed by Warner [1], RRT is a statistical method designed to introduce controlled randomness into the response of sensitive questions. Different modifications of Warner's [1] original method have been developed and empirically applied to different situations concerning the sensitive data. While RRT has revolutionized data collection in sensitive surveys, limitations arise when quantitative information is required. To handle quantitative data, Warner [2] proposed randomised response (RR) model under additive scrambling in which a random number is added to the response and the response so obtained is known as scrambled response. Pollock and Bek [3] investigated and compared additive and multiplicative RR models including Greenberg et al. [4] model. Multiplicative RR model was discussed in detail by Eichhorn and Hayre [5] which was earlier briefly discussed by Pollock and Bek [3]. This contributed to understanding the statistical implications of applying RRT to quantitative data.

Gupta et al. [6] introduced an optional randomized response technique (ORRT) where the respondents decide themselves whether they want to tell the truth (or scramble their true response) depending upon whether the question being asked is perceived by them as nonsensitive (or sensitive). The proportion of respondents who consider the question sensitive is called the sensitivity level (usually denoted by ω) of the question. The Gupta et al. [6] model used multiplicative scrambling as proposed by Eichhorn and Hayre [5] to estimate the mean of a sensitive variable. Gupta et al. [7] introduced additive scrambling in optional randomised response (ORR) model. Gupta et al. [8] and Mehta et al. [9] further improved the one stage additive scrambling ORR models by extending it to the two stage and three stage models respectively. Huang [10] used linear combination of scrambling variables to scramble the response under ORR model. Gupta et al. [11] observed that the ORR model under additive scrambling performs better than the ORR model under linear combination scrambling. In an ORR model, there are two parameters of interest: the sensitivity level of the question and prevalence of the sensitive characteristic in the population. In the above discussed ORR models, split sample approach is used to estimate the mean of sensitive variable and sensitivity level of the research question. However, the split sample approach requires a larger total sample size for estimation.

Gupta et al. [12] estimated the finite population mean and sensitivity level using ORR model in the presence of nonsensitive auxiliary information from a single sample. Tiwari and Mehta [13] proposed an improved methodology for ORR models in which the sensitivity level (ω) was considered to be known and the RRT was applied only for those respondents who considered the particular question a sensitive one. Tiwari and Mehta [14] also proposed an improved ORRT for quantitative variable.

In ORRT, the approach to estimate the unknown sensitivity level of the main research question by means of using RRT is called two-questions approach [15]. In this approach, all respondents are asked two separate questions. The question about sensitivity level of the sensitive question is asked first via randomization device. In this randomization process, the question is “Is the main research question sensitive?” This question can be denoted by Question no. 1. It can be asked along with an unrelated innocuous question. The underlying sensitivity level and its variance can be estimated from the sample by using any binary RRT. The main research question is denoted by Question no. 2, where respondent answer the question using second randomisation device. The two-questions approach eliminates the need of a split sample to estimate the mean and sensitivity level separately. The two-questions approach also increases the precision of the estimate of sensitivity level.

Sihm et al. [16] used two-questions approach to estimate the sensitivity level when using unrelated ORRT. Chhabra et al. [17] extended this method to the multi stage unrelated ORRT. In a similar way, Kalucha et al. [18] used two-questions approach to estimate sensitivity level and estimated the mean of the sensitive variable by using one stage additive scrambling ORRT propounded by Gupta et al. [7].

Narjis and Shabir [15] proposed three unrelated ORR models under two-questions approach to simultaneously estimate the proportion of sensitive attribute and sensitivity level. In addition, Narjis and Shabir [19] also proposed a multi-question approach to estimate the proportion of sensitive attribute and sensitivity level when an unrelated innocuous attribute is unknown. Recently, Gupta et al. [20] addressed lack of trust in RRTs by proposing an optional enhanced trust (OET) model for quantitative RRT. In OET model, respondents can choose between revealing their true answer or using a scrambling technique either proposed by Warner [2] or Diana and Perry [21] based on their trust in the respective RR model. The OET model introduces three unknowns: mean of sensitive variable, sensitivity level (ω), and trust level (A).

Azeem et al. [22] simplify the OET model by assuming sensitivity level and trust level to be known. However, it is imperative to acknowledge the inherent limitation of this assumption as the sensitivity level and trust level are rarely, if ever, truly known. For a comprehensive summary of RRT, one may refer to Fox and Tracy [23], Chaudhury et al. [24], and Le et al. [25].

There are several models to scramble the quantitative response in RRT such as additive scrambling, multiplicative scrambling, linear combination of scrambling variables etc. However, in this paper we restrict ourselves to additive scrambling in ORRT. In ORR models under additive scrambling in two-stage [8] and three-stage [9], variance of the estimate of sensitivity level inflates as the second stage and third stage probability increases. In addition, these two models used split-sample approach to estimate the prevalence of sensitive characteristics and the sensitivity level. This negatively impacts the estimation of sensitivity level and requires a larger total sample size.

To overcome these limitations in two-stage and three-stage ORR models, in this paper, we propose two improved two-stage and three-stage ORR models under additive scrambling. The proposed models estimate the prevalence of the sensitive characteristic and the sensitivity level of the main research question from two different sets of responses from the same sample. The proposed models are compared with the existing ORR models using additive scrambling. In Section 2, the ORR models using split sample approach discussed by Gupta et al. [8] and Mehta et al. [9] and ORR model using two-questions approach given by Kalucha et al. [18] are discussed in brief. Section 3 deals with the proposed improved two-stage and three stage ORR models using additive scrambling. In Section 4, the efficiency of the proposed ORR models is compared with the existing ORR models using a numerical study. Privacy protection of the proposed ORR models is discussed in Section 5, followed by the conclusion of the study in Section 6.

2. Brief Description of Quantitative ORRT

Let μ and σ^2 be the unknown mean and variance of the sensitive variable X and S is a scrambling variable (independent of X) with known mean θ and known variance σ_s^2. Let ω be the unknown sensitivity level of the survey question in the population. Under these assumptions, a brief discussion of Gupta et al. [8], Mehta et al. [9] and Kalucha et al. [18] models are as follows:

2.1 The split sample spproach - Gupta et al. and Mehta et al. models

In Gupta et al. [8] two stage ORR model, a known proportion (T) of the respondents provide truthful response to the sensitive question. From the remaining known proportion of respondents $(1-T)$, an unknown proportion (ω) provides scrambled responses and the rest unknown proportion $(1-\omega)$ provide truthful responses to the question. To estimate the mean of sensitive variable (X) and sensitivity level (ω), the sample size n is split into two sub-samples with sizes n_1 and n_2, respectively. Under this model, reported responses $(Z_i;\ i=1,2)$ in the two sub-samples are given by,

$$Z_i = \begin{cases} X \text{ with probability } T+(1-T)(1-\omega), \\ (X+S_i) \text{ with probability } (1-T)\omega, \end{cases} \qquad i=1,2$$

Here, $S_i,\ i=1,\ 2$, are independent scrambling variables, independent of X. The unbiased estimators of the mean of sensitive variable and sensitivity level respectively from the sub-samples are given by,

$$\widehat{\mu_G} = \frac{\theta_1 \bar{z}_2 - \theta_2 \bar{z}_1}{\theta_1 - \theta_2} \quad \text{and} \quad \widehat{\omega_G} = \frac{\bar{z}_1 - \bar{z}_2}{(1-T)(\theta_1-\theta_2)}, \quad \theta_1 \neq \theta_2.$$

Here, $\bar{z}_1$ and $\bar{z}_2$ respectively are the sample mean of reported responses in the two sub-samples. The variances of these estimators are given by,

$$Var\left(\widehat{\mu_G}\right) = \frac{1}{(\theta_1-\theta_2)^2}\left(\theta_2^2 \frac{\sigma_{Z_1}^2}{n_1} + \theta_1^2 \frac{\sigma_{Z_2}^2}{n_2}\right) and\ Var\ \left(\widehat{\omega_G}\right) = \frac{1}{(1-T)^2(\theta_1-\theta_2)^2}\left(\frac{\sigma_{Z_1}^2}{n_1} + \frac{\sigma_{Z_2}^2}{n_2}\right),$$

where $\sigma_{Z_i}^2 = \sigma^2 + (1-T)\,\omega\sigma_{S_i}^2 + (1-T)\,\omega\,\{1-(1-T)\,\omega\}\,\theta_i^2;\ i=1,\ 2.$

Mehta et al. [9] extended the two-stage ORR model of Gupta et al. [8] to three stages. In three-stage ORRT, in each sub sample a fixed predetermined proportion (T) of respondents is instructed to tell the truth and a fixed predetermined proportion (F) of respondents is instructed to scramble their response. The remaining proportion $(1-T-F)$ of respondents have an option of scrambling their responses additively if they consider the question to be sensitive, or else they can report their true response X. For $F=0$, the model is same as the Gupta et al. [8] model. For both T and F equal to zero, the model is same as the Gupta et al. [7] model. The reported response in the sub-samples is given by,

$$Z_i = \begin{cases} X \text{ with probability } T+(1-T-F)(1-\omega), \\ (X+S_i) \text{ with probability } F+(1-T-F)\,\omega, \end{cases} \qquad i=1,2.$$

The unbiased estimators of the mean of sensitive variable and sensitivity level from the sub-samples are given by,

$$\widehat{\mu_M} = \frac{\theta_1 \bar{z}_2 - \theta_2 \bar{z}_1}{\theta_1 - \theta_2}\ and\ \widehat{\omega_M} = \frac{1}{(1-T-F)}\left(\frac{\bar{z}_1 - \bar{z}_2}{(\theta_1-\theta_2)} - F\right),\ \theta_1 \neq \theta_2,\ T+F \neq 1.$$

Here, $\bar{z}_1$ and $\bar{z}_2$, respectively are the sample mean of reported responses in the two sub-samples. The variances of these estimators are given by,

$$Var\left(\widehat{\mu_M}\right) = \frac{1}{(\theta_1-\theta_2)^2}\left(\theta_2^2 \frac{\sigma_{Z_1}^2}{n_1} + \theta_1^2 \frac{\sigma_{Z_2}^2}{n_2}\right)$$

and

$$Var\ \left(\widehat{\omega_M}\right) = \frac{1}{(1-T-F)^2(\theta_1-\theta_2)^2}\left(\frac{\sigma_{Z_1}^2}{n_1} + \frac{\sigma_{Z_2}^2}{n_2}\right).$$

Here for $i=1,\ 2$, $\sigma_{Z_i}^2 = \sigma^2 + \{F+(1-T-F)\,\omega\}\,\sigma_{S_i}^2 + \{F+(1-T-F)\,\omega\}\,[1-\{F+(1-T-F)\,\omega\}]\,\theta_i^2.$

2.2 The two-questions approach- Kalucha et al. model

In this model, from sample of size n, a proportion $(1-\omega)$ of respondents truthfully answer the sensitive question (main research question) directly while the remaining proportion (ω) scramble their responses additively. Here, the underlying sensitivity level ω and its variance are estimated by using the Greenberg et al. [26] model. Hence, the reported quantitative response is given by,

$$Z = \begin{cases} X \text{ with probability } (1-\omega), \\ (X+S) \text{ with probability } \omega, \end{cases}$$

Here, S is scrambling variable independent of X. The unbiased estimators of the mean of sensitive variable and sensitivity level respectively from the sample are given by,

$$\widehat{\mu_K} = \bar{z} - \widehat{\omega_K}\theta \text{ and } \widehat{\omega_K} = \frac{\widehat{P_y} - (1-P)\pi}{P}.$$

Here, $\bar{z}$ is the sample mean of reported quantitative responses in the sample obtained by asking main research question, $\widehat{P_y}$ is the proportion of 'yes' responses in the sample from first question (viz., Is the main research question sensitive?), P and π respectively are design parameters of Greenberg et al. [26] model to estimate sensitivity level ω. The variances of estimators proposed by Kalucha et al. [18] are given by,

$$Var\left(\widehat{\mu_K}\right) = \frac{\sigma_Z^2}{n} + \theta^2 \frac{P_y(1-P_y)}{nP^2} \text{ and } Var\left(\widehat{\omega_K}\right) = \frac{P_y(1-P_y)}{nP^2},$$

where $\sigma_Z^2 = \sigma^2 + \omega\sigma_S^2 + \omega(1-\omega)\theta^2$ and $P_y = P\omega + (1-P)\pi$.

3. The Proposed Two-Stage and Three-Stage Improved ORR Models

In the proposed two-stage and three-stage models, all respondents are asked two separate questions. The question to estimate the sensitivity level is asked first via randomization device 1. In this randomization process, the question is "Is the main research question sensitive?" It is asked along with an unrelated innocuous question. The underlying sensitivity level ω and its variance are estimated by using the Greenberg et al. [26] model. Let π be the known probability of the binary innocuous unrelated question and P be the known probability of the respondent selecting Question no. 1. The probability of getting "yes" response to the Question no. 1 is $P_y = P\omega + (1-P)\pi$. Solving for ω, we get,

$$\omega = \frac{P_y - (1-P)\pi}{P}.$$

Thus, the unbiased estimate of ω, as per the Greenberg et al. [26] model is given by,

$$\hat{\omega} = \frac{\widehat{P_y} - (1-P)\pi}{P}, \tag{3.1}$$

where $\widehat{P_y}$ is the proportion of 'yes' responses in the sample. The variance of the estimator is given by,

$$Var\ (\hat{\omega}) = \frac{P_y(1-P_y)}{nP^2}.$$

3.1 Two stage improved ORR model

In the same sample, to answer Question no. 2 (main research question), a known proportion (T) of the respondents provide truthful response. From the remaining known proportion of respondents $(1-T)$, an unknown proportion (ω) provides scrambled responses and the rest unknown proportion $(1-\omega)$ provide truthful responses to the main research question. Therefore, the reported quantitative response Z to the main research question according to two-stage ORR model is given by,

$$Z = \begin{cases} X \text{ with probability } T + (1-T)(1-\omega), \\ (X+S) \text{ with probability } (1-T)\omega, \end{cases}$$

The mean and variance of Z respectively are given by,

$$E(Z) = \{T + (1-T)\ (1-\omega)\}E(X) + \{(1-T)\ \omega\}E(X+S)$$
$$E(Z) = E(X) + \{(1-T)\ \omega\}E(S)$$

$$E(Z) = \mu + \omega(1-T)\theta \tag{3.2}$$

and

$$Var\ (Z) = \sigma_Z^2 = E\left(Z^2\right) - \{E(Z)\}^2$$
$$= \{T + (1-T)\ (1-\omega)\}E\left(X^2\right) + \{(1-T)\omega\}E(X+S)^2 - \{E(Z)\}^2$$
$$Var\ (Z) = \sigma^2 + (1-T)\omega\sigma_S^2 + (1-T)\omega\{1-(1-T)\omega\}\theta^2.$$

From equation (3.2), the unbiased estimator of μ under this model (denoted by $\widehat{\mu_1}$) is given by,

$$\widehat{\mu_1} = \bar{z} - (1-T)\theta\hat{\omega},$$

here $\bar{z}$ is sample mean of reported responses and $\hat{\omega}$ is an unbiased estimator of ω given in equation (3.1). The variance of the estimator $\widehat{\mu_1}$ is given by,

$$Var\ \left(\widehat{\mu_1}\right) = Var\ (\bar{z}) + (1-T)^2\theta^2 Var(\hat{\omega})$$
$$Var\ \left(\widehat{\mu_1}\right) = \frac{\sigma_Z^2}{n} + \theta^2(1-T)^2\frac{P_y(1-P_y)}{nP^2}$$

3.2 Three stage improved ORR model

In three stage model, to answer Question no. 2 (main research question) in the same sample, a fixed predetermined proportion (T) of respondents is instructed to tell the truth and a fixed predetermined proportion (F) of respondents is instructed to scramble their response. The remaining proportion $(1-T-F)$ of respondents have an option of scrambling their responses additively if they consider the question to be sensitive, else they can report their true response X. Thus, the reported quantitative response Z to the main research question according to three-stage ORR model is given by,

$$Z = \begin{cases} X \text{ with probability } T + (1-T-F)(1-\omega), \\ (X+S) \text{ with probability } F + (1-T-F)\omega, \end{cases}$$

The mean and variance of Z are given by,

$$E(Z) = \{T + (1-T-F)(1-\omega)\}E(X) + \{F + (1-T-F)\omega\}E(X+S)$$
$$E(Z) = E(X) + \{F + (1-T-F)\omega\}E(S)$$

$$E(Z) = \mu + \{F + (1-T-F)\omega\}\theta \tag{3.3}$$

and

$$Var\ (Z) = \sigma_Z^2 = \{T + (1-T-F)(1-\omega)\}E\left(X^2\right) + \{F + (1-T-F)\omega\}E(X+S)^2 - \{E(Z)\}^2$$
$$\sigma_Z^2 = \sigma^2 + \{F + (1-T-F)\omega\}\sigma_s^2 + \{F + (1-T-F)\omega\}[1 - \{F + (1-T-F)\omega\}]\theta^2.$$

From equation (3.3), the unbiased estimator of μ under this model (denoted by $\widehat{\mu_2}$) is given by,

$$\widehat{\mu_2} = \bar{z} - (1-T-F)\theta\hat{\omega} - F\theta,$$

here $\bar{z}$ is sample mean of reported responses and $\hat{\omega}$ is an unbiased estimator of ω given in equation (3.1). The variance of the estimator $\widehat{\mu_2}$ is given by,

$$Var\ \left(\widehat{\mu_2}\right) = Var\ (\bar{z}) + (1-T-F)^2\theta^2 Var(\hat{\omega})$$
$$Var\ \left(\widehat{\mu_2}\right) = \frac{\sigma_Z^2}{n} + \theta^2(1-T-F)^2\frac{P_y(1-P_y)}{nP^2}$$

Table 1. PRE of proposed estimators of mean (μ) w.r.t Gupta et al. [8][G], Mehta et al. [9][M] and Kalucha et al. [18][K] estimators. ($n=1000,\ n_1=n_2=500,\ X\sim Poisson\ (4),\ S_1\sim Poisson\ (2),\ S_2\sim Poisson\ (5),\ P=0.70,\ \pi=0.25$).

ω	T	F	$PRE\,(\hat{\mu_1},\hat{\mu_G})$	$PRE\,(\hat{\mu_1},\hat{\mu_K})$	$PRE\,(\hat{\mu_2},\hat{\mu_G})$	$PRE\,(\hat{\mu_2},\hat{\mu_M})$	$PRE\,(\hat{\mu_2},\hat{\mu_K})$
0.70	0.55	0.30	682.58	139.78	692.00	735.31	141.71
0.70	0.45	0.40	668.72	130.40	698.43	738.94	136.19
0.70	0.35	0.50	650.07	122.18	706.43	737.56	132.77
0.70	0.25	0.60	627.47	114.91	716.20	731.47	131.16
0.70	0.15	0.70	601.61	108.45	728.04	720.62	131.25
0.70	0.05	0.80	573.01	102.67	742.43	704.57	133.03
0.80	0.55	0.30	690.78	134.98	710.57	736.52	138.84
0.80	0.45	0.40	676.82	126.31	716.55	739.36	133.73
0.80	0.35	0.50	657.52	118.93	722.25	737.25	130.64
0.80	0.25	0.60	633.67	112.57	727.91	730.45	129.31
0.80	0.15	0.70	605.82	107.04	733.73	718.85	129.64
0.80	0.05	0.80	574.41	102.20	739.95	701.96	131.65
0.90	0.55	0.30	699.41	128.44	726.29	737.83	133.38
0.90	0.45	0.40	685.60	120.82	730.55	739.87	128.75
0.90	0.35	0.50	665.87	114.57	732.41	737.04	126.02
0.90	0.25	0.60	640.83	109.40	732.08	729.53	124.98
0.90	0.15	0.70	610.88	105.11	729.60	717.17	125.54
0.90	0.05	0.80	576.21	101.55	724.85	699.42	127.74

4. Efficiency Comparison

The efficiency of the proposed estimators with respect to the estimators suggested by Gupta et al. [8] [G], Mehta et al. [9] [M] and Kalucha et al. [18] [K] is numerically established using the following formula of percent relative efficiency:

$$PRE\,(\widehat{\tau}_i,\widehat{\tau}_j)=\frac{Var\,(\widehat{\tau}_j)}{Var\,(\widehat{\tau}_i)}\times 100;\ \tau=\hat{\mu},\ \hat{\omega};\ i=1,2;\ j=G,M,K$$

The PRE of the proposed estimators has been computed at various values of model parameters. The results of the numerical study are illustrated in the Table 1 and Table 2. For the numerical analysis, we used distribution of sensitive variable and scrambling variable similar to what is used in Mehta et al. [9]. The distribution of S_1 is used for single sample in proposed ORR models. Table 1 illustrates the PRE of proposed estimator $\hat{\mu}_i$, $(i=1,2)$ of mean (μ) w.r.t the estimators suggested by Gupta et al. [8], Mehta et al. [9] and Kalucha et al. [18].

It is observed from Table 1 that all the PREs corresponding to the proposed estimators for the mean of sensitive variable, under two-stage and three-stage ORR models are greater than 100. The results indicate that the proposed two-stage and three stage ORR models under two-questions approach are more efficient than ORR models of Gupta et al. [8] and Mehta et al. [9] under split sample approach. Moreover, the proposed two-stage and three-stage ORR models under two questions approach performs better than the one stage ORR model under two-questions approach as suggested by Kalucha et al. [18] for all values of the model parameters. Moreover, it is seen from Table 1 that the proposed three-stage ORR model performs much better than the two-stage and one-stage ORR models under two-questions approach for estimating the mean of sensitive variable. The main focus of the present study is to check whether the estimation of sensitivity level improved under proposed models or not. In this regard, the results concerning the PREs of estimators under proposed models for sensitivity level are demonstrated in Table 2.

It is observed from Table 2 that while comparing with the Gupta et al. [8] and Mehta et al. [9] models, the PRE for the estimators of sensitivity level for the proposed two-stage and three-stage ORR models are significantly higher than 100. A large gain in PRE is observed for proposed ORR models under two-questions approach in comparison to the two-stage and three-stage ORR models under split-sample approach due to the reason that the variance of estimate of sensitivity level inflates as the second stage and third stage probability increases in split samples. However, the two-questions approach used in Kalucha et al. [18] ORR model and in proposed two-stage and three-stage ORR models yields same precision to estimate the sensitivity level but the proposed two-stage and three-stage ORR models also improved the precision of estimate of the mean of sensitive variable under various practical situations (see, Table 1). Hence, the proposed two-stage and three-stage ORR models under two questions approach outperforms the Gupta et al. [8] and Mehta et al. [9] ORR models under split sample approach and Kalucha et al. [18] ORR model under two-questions approach.

Table 2. PRE of proposed estimators of sensitivity level (ω) w.r.t Gupta et al. [8][G], Mehta et al. [9][M] and Kalucha et al. [18][K] estimators. ($n = 1000,\ n_1 = n_2 = 500,\ X \sim Poisson\ (4),\ S_1 \sim Poisson\ (2),\ S_2 \sim Poisson\ (5),\ P = 0.70,\ \pi = 0.25$).

ω	T	F	$PRE\,(\hat{\omega}, \hat{\omega_G})$	$PRE\,(\hat{\omega}, \hat{\omega_M})$	$PRE\,(\hat{\omega}, \hat{\omega_K})$
0.70	0.55	0.30	3601.77	32415.93	100.00
0.70	0.45	0.40	2572.06	34579.95	100.00
0.70	0.35	0.50	1926.98	36184.35	100.00
0.70	0.25	0.60	1489.17	37229.15	100.00
0.70	0.15	0.70	1174.49	37714.33	100.00
0.80	0.55	0.30	3990.81	35917.27	100.00
0.80	0.45	0.40	2830.57	38055.40	100.00
0.80	0.35	0.50	2099.21	39418.46	100.00
0.80	0.25	0.60	1600.26	40006.45	100.00
0.80	0.15	0.70	1240.05	39819.36	100.00
0.90	0.55	0.30	4608.24	41474.14	100.00
0.90	0.45	0.40	3239.07	43547.47	100.00
0.90	0.35	0.50	2371.29	44527.59	100.00
0.90	0.25	0.60	1776.58	44414.50	100.00
0.90	0.15	0.70	1345.58	43208.20	100.00

5. Privacy Protection

The aspect of privacy protection of respondents is an integral part of the RRT. We examine this aspect for the proposed ORR models. Lanke [27], Yan et al. [28] and Giordano and Perri [29] have discussed this issue in detail. Lanke [27] and Giordano and Perri [29] devised a privacy measure to assess the privacy protection of binary RRTs while Yan et al. [28] derived a privacy measure for the quantitative RRTs. Yan et al. [28] defined the measure of privacy protection as $\nabla = E\,(Z-X)^2$ where X is the true response of the sensitive variable and Z is the reported response. For a given model, the larger the value of ∇, the larger the privacy provided by the model.

The privacy measure for one-stage additive scrambling ORR model is given by $\nabla_1 = \left(\theta^2 + \sigma_S^2\right)\omega$ and for two-stage additive scrambling ORR model it is given by $\nabla_2 = \left(\theta^2 + \sigma_S^2\right)\omega\,(1-T)$. Thus, comparing ∇_1 and ∇_2, it is observed that for same precision Gupta et al. [7] model (which is used by Kalucha et al. [18] under two questions approach) is more protective than Gupta et al. [8] model (proposed two-stage improved ORR model). This shows that the proposed two stage ORR model with two-questions approach may be made more protective as compared to Kalucha et al. [18] model with two questions approach, but at the cost of precision. In fact, it is a trade-off between the efficiency and privacy protection. That is, we can have highly efficient estimator by compromising on privacy. Similarly, we can build a more protective model by compromising on the efficiency.

Hussain and Al-Zehrani [30] discussed that Gupta et al. [7] one-stage model is more protective compared to Gupta et al. [8] two-stage model and they argued that Gupta et al. [7] model remain more protective among all other existing ORR models under additive scrambling. This argument is not true in case of Mehta et al. [9] three-stage ORR model. In case of three-stage ORR model due to Mehta et al. [9], the reported response is given by,

$$Z = \begin{cases} X \text{ with probability } T + (1-T-F)(1-\omega), \\ (X+S) \text{ with probability } F + (1-T-F)\omega, \end{cases}$$

then

$$Z - X = \begin{cases} 0 \text{ with probability } T + (1-T-F)(1-\omega), \\ S \text{ with probability } F + (1-T-F)\omega, \end{cases}$$

Therefore, the privacy measure in case of the Mehta et al. [9] model is given by,

$$\nabla_3 = E\,(Z-X)^2 = \{F + (1-T-F)\,\omega\}E(S^2) = \{F + (1-T-F)\,\omega\}\left(\theta^2 + \sigma_S^2\right)$$

The privacy comparison of the proposed three-stage ORR model with one-stage and two-stage ORR models can be summarised in the following theorems.

Theorem 5.1. *Three stage ORRT with two questions approach offers more privacy than one stage ORRT with two-questions approach if,* $F\,(1-\omega) > T\omega$.

Proof. Considering the difference of the privacy measures of the models, we observe

$$\begin{aligned}\nabla_3-\nabla_1 &= \left(\theta^2+\sigma_S^2\right)\{F+(1-T-F)\,\omega\}-\left(\theta^2+\sigma_S^2\right)\omega \\ &= \left(\theta^2+\sigma_S^2\right)(F-F\omega-T\omega) \\ &= \left(\theta^2+\sigma_S^2\right)\{F(1-\omega)-T\omega\}\end{aligned}$$

For $\nabla_3-\nabla_1>0$, we get $F\,(1-\omega)>T\omega$. Hence the theorem. □

Theorem 5.2. *Three stage ORRT with two-questions approach always offers more privacy than two stage ORRT with two-questions approach.*

Proof. Considering the difference of the privacy measures of the models, we observe

$$\begin{aligned}\nabla_3-\nabla_2 &= \left(\theta^2+\sigma_S^2\right)\{F+(1-T-F)\,\omega\}-\left(\theta^2+\sigma_S^2\right)\omega(1-T) \\ &= \left(\theta^2+\sigma_S^2\right)(F-F\omega)\end{aligned}$$

For $\nabla_3-\nabla_2>0$, we get $F>F\omega$ which is always true. This proves the theorem. □

The above results establish the superiority of the proposed three-stage improved ORR model. Hence, our two-questions approach in three-stage ORR model is more protective in comparison to Gupta et al. [8] and Kalucha et al. [18] ORR models. Numerically, from Table 1 it can be observed that for the suitable choice of parameters, the proposed three stage ORR model with two-questions approach performs better in terms of efficiency and at the same time protect the privacy of respondents more than the existing models. For example, when $\omega=0.80$ taking $T=0.15$ and $F=0.70$, the proposed three-stage ORR model has greater precision and the parameter values also satisfy the conditions under Theorem 5.1 and Theorem 5.2. Thus, it may be concluded that if the parameters are chosen carefully, the three-stage quantitative ORR model with two questions approach offers better efficiency and more privacy than the Gupta et al. [8], Mehta et al. [9] and Kalucha et al. [18] ORR models.

6. Conclusion

Using two-questions approach, improved two-stage and three-stage ORR models for quantitative variables have been proposed and their properties are discussed. It is observed from the numerical comparisons that the proposed two-stage and three-stage ORR models using two-questions approach are found to be more efficient as compared to the two-stage and three-stage ORR models using the usual split-sample approach. It is also observed that proposed ORR models can be made more efficient than the existing ORR models by choosing appropriate design parameters. It is also found that the proposed three-stage ORR model under two-questions approach offers more privacy than one-stage and two-stage ORR models.

In addition, the proposed three-stage ORR model using two-questions approach with suitable choices of design parameters performs better than the two-stage and one-stage ORR models and provides more privacy as compared to one stage ORRT using two-questions approach. It is found that there is a significant gain in precision under the two-questions approach to estimate the sensitivity level. Moreover, the precision of estimate of sensitivity level can also be increased by using forced RRT or two-stage RRT.

On the basis of our study, we may conclude that, the proposed three-stage ORR model under the two-questions approach stands out as a particularly valuable tool for surveys that grapple with highly sensitive issues. This model may be highly useful for data collection in areas where respondents might be reluctant to answer truthfully due to fear of judgment or social stigma. This model may be quite relevant in research on illegal activities, stigmatized health conditions, or unpopular opinions. Thus, it may be recommended that to estimate the mean of sensitive variable along with the sensitivity level in a survey concerning sensitive information, the proposed three stage ORR model under two-questions approach is better choice among other existing ORR models.

Article Information

Acknowledgements: The authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

Author's contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflict of Interest Disclosure: No potential conflict of interest was declared by the authors.

Supporting/Supporting Organizations: No grants were received from any public, private or non-profit organizations for this research.

Ethical Approval and Participant Consent: It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography.

Plagiarism Statement: This article was scanned by the plagiarism program.

References

[1] S. L. Warner, *Randomized response: a survey technique for eliminating evasive answer bias*, J. Amer. Stat. Assoc., **60**(309) (1965), 63–69.

[2] S. L. Warner, *Linear randomized response models*, J. Amer. Stat. Assoc., **66** (1971), 884–888.

[3] K. H. Pollock, Y. Bek, *A comparison of three randomized response models for quantitative data*, J. Amer. Stat. Assoc., **71**(356) (1976), 884-886.

[4] B. G. Greenberg, R. R. Kuebler, J. R. Abernathy, D. G. Horvitz, *Application of the randomized response technique in obtaining quantitative data*, J. Amer. Statist. Assoc., **66**(334) (1971), 243-250.

[5] B. H. Eichhorn, L. S. Hayre, *Scrambled randomized response methods for obtaining sensitive quantitative data*, J. Stat. Plan. Infer., **7**(4) (1983), 307-316.

[6] S. Gupta, B. Gupta, S. Singh, *Estimation of sensitivity level of personal interview survey questions*, J. Stat. Plan. Infer., **100**(2) (2002), 239-247.

[7] S. N. Gupta, B. Thornton, J. Shabbir, S. Singhal, *A comparison of multiplicative and additive optional RRT models*, J. Stat. Theo. Appl., **5** (2006), 226–239.

[8] S. Gupta, J. Shabbir, S. Sehra, *Mean and sensitivity estimation in optional randomized response models*, J. Stat. Plan. Infer. **140** (2010), 2870-2874.

[9] S. Mehta, B. K. Dass, J. Shabbir, S. N. Gupta, *A three-stage optional randomised response model*, J. Stat. Theo. Pract., **6** (2012), 417-427.

[10] K. C. Huang, *Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling*, Metrika, **71** (2010), 341-352.

[11] S. N. Gupta, S. Mehta, J. Shabbir, B. Dass, *Generalized scrambling in quantitative optional randomized response models*, Comm. Stat.: Theo. Meth., **42**(22) (2013), 4034-4042.

[12] S. Gupta, G. Kalucha, J. Shabbir, B. K. Dass, *Estimation of finite population mean using optional RRT models in the presence of nonsensitive auxiliary information*, Amer. J. Math. Manag. Sci., **33**(2) (2014), 147–159.

[13] N. Tiwari, P. Mehta, *An improved two stage optional RRT model*, J. Ind. Soc. Agr. Stat., **70**(3) (2016), 197-203.

[14] N. Tiwari, P. Mehta, *Additive randomized response model with known sensitivity level*, Int. J. Comp. Theor. Stat. **4**(2) (2017).

[15] G. Narjis, J. Shabbir, *Estimation of population proportion and sensitivity level using optional unrelated question randomized response techniques*, Comm. Stat. Sim. Comp., **49**(12) (2020), 3212-3226.

[16] J. S. Sihm, A. Chhabra, S. N. Gupta, *An optional unrelated question RRT model*, Involve J. Math., **9**(2) (2016), 195-209.

[17] A. Chhabra, B. K. Dass, S. Mehta, *Multi-stage optional unrelated question RRT model*, J. Stat. Theo. Appl., **15**(1) (2016), 80-95.

[18] G. Kalucha, S. N. Gupta, J. Shabbir, *A two-step approach to ratio and regression estimation of finite population mean using optional randomized response models*, Hac. J. Math. Stat., **45**(6) (2016), 1819–1830.

[19] G. Narjis, J. Shabbir, *Estimating the prevalence of sensitive attribute with optional unrelated question randomized response models under simple and stratified random sampling*, Sci. Iran., **28**(5) (2021), 2851-2867.

[20] S. Gupta, J. Zhang, S. Khalil, P. Sapra, *Mitigating lack of trust in quantitative randomized response technique models*, Commun. Stat. Sim. Comp., (2022), 1-9.

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

[22] M. Azeem, I. M. Asadullah, S. Hussain, N. Salahuddin, A. Salam, *A novel randomized scrambling technique for mean estimation of a finite population*, Hel., **10**(11), e31690.

[23] J. A. Fox, P. E. Tracy, *Quantitative Applications In The Social Sciences: Randomized Response*, Newbury Park, CA: SAGE Publications, Inc., **1986**.

[24] A. Chaudhuri, T. C. Christofides, C. R. Rao, *Handbook of Statistics, Volume 34, Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits*, North-Holland. **2016**.

[25] T. N. Le, S. M. Lee, P. L. Tran, C. S. Li, *Randomised response techniques: A systematic review from the pioneering work of Warner (1965) to the present*, Mathematics, (2023), 1718.

[26] B. G. Greenberg, A. L. A. Abul-Ela, W. R. Simmons, D. G. Horvitz, *The unrelated question randomized response model: Theoretical framework*, J. Amer. Statist. Assoc., **64**(326) (1969), 520-539.

[27] J. Lanke, *On the degree of protection in randomized interviews*, Int. Stat. Rev., **44**(2), (2016), 197-203.

[28] Z. Yan, J. Wang, J. Lai, *An efficiency and protection degree-based comparison among the quantitative randomized response strategies*, Comm. Stat. Theo. Meth., **38** (2009), 400-408.

[29] S. Giordano, P. F. Perri, *Efficiency comparison of unrelated question models based on same privacy protection degree*, Stat. Papers, **53** (2012), 987-999.

[30] Z. Hussain, M. M. Al-Sobhi, B. Al-Zahrani, H. P. Singh, T. A. Tarray, *Improved randomized response approaches for additive scrambling models*, Math. Pop. Stud., **23**(4) (2016), 205–221.