International Journal of Advanced Statistics and Probability, 13 (1) (2026) 17-26

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
https://doi.org/10.14419/03yt6271
Research paper

Efficient Optional Randomized Response Model for Estimating The Mean of a Sensitive Attribute with Known Sensitivity Level

I. B. Okafor [1] *, A. C. Onyeka [2], C. J. Ogbonna [2], C. H. Izunobi [2], F. C. Okafor [3]

[1] *Department of Statistics, Covenant Polytechnic, Aba, Abia State, Nigeria*
[2] *Department of Statistics, Federal University of Technology Owerri, Imo State, Nigeria*
[3] *Department of Statistics, University of Nigeria, Nsukka, Enugu State, Nigeria*
**Corresponding author E-mail: okafor_iyke@yahoo.com*

Received: November 22, 2025, Accepted: January 19, 2026, Published: February 8, 2026

Abstract

The paper proposed an efficient optional randomized response model for estimating the mean of a sensitive variable with known sensitivity level. The model introduced a multiplicative stochastic disturbance to further perturb respondents' responses and offers the respondents the option to either report a scrambled response or report a truthful response based on their perception of question sensitivity. Theoretical analysis confirms that the estimator is unbiased with minimum variance compared to some existing models. Efficiency conditions and privacy metrics were derived. The results of the empirical study show that the proposed model for different design parameter configurations recorded higher relative efficiency and privacy protection level. The log weighted privacy-efficiency measure further validates its robustness and practicality. By and large, the model offers a reliable and flexible framework for collecting sensitive survey data while minimizing the response bias.

Keywords: *Mean; Privacy Protection; Scrambling Response; Sensitivity Level; Stochastic Disturbance.*

1. Introduction

Estimating population parameters of a sensitive or delicate behaviour involving the human population is one of the problems encountered in survey sampling. A survey question (such as those on illicit drug use, sexual behaviour, tax evasion, cheating in examination etc.) is said to be sensitive if disclosure of information on it appears to have potential stigmatization, discrimination, or be severely intrusive. Due to the potential threats of social stigma and fear of reprisals, the conventional method of data collection on sensitive survey questions often results in respondents' refusal to respond or falsification of his/her response, thereby introducing response error bias which, if large enough, may estimate from such a survey unreliable.

To eliminate or minimize response error bias resulting from this situation, the randomized response technique model was introduced by Warner (1965). The model is meant to estimate the population proportion π of those possessing the sensitive attribute under study. Going further, Warner (1971), Greenberg et al. (1971) introduced randomized response technique models for quantitative sensitive attribute. Later on, Pollock & Bek (1976),

Eichhorn & Hayre (1983) suggested the scrambled randomized response technique for the estimation of the mean of quantitative sensitive attribute. The premise behind the scrambled randomized response technique model is to introduce a stochastic disturbance to the respondent's response so as to mask his/her response. In this way, their privacy and confidentiality protection is guaranteed, and their willingness to cooperate and participate in the survey increased.

The scrambled randomized response technique can either be a forced randomized response model or an optional randomized response model. In the forced randomized response technique, all respondents are instructed to report a scrambled response irrespective of their perception about the nature of the survey question, whether sensitive or nonsensitive. Using design parameters that are controlled by the interviewer, Bar-Lev et al. (2004) worked on the forced randomized response technique and proposed a randomized response technique that generalized Eichhorn & Hayre's (1983) model. In the model, certain proportion (P) of the randomization device (say a well shuffled card) contains an instruction instructing the respondent to report true value of the sensitive survey question irrespective of their perception of the nature of the survey question while remaining proportion (1−P) contains an instruction instructing respondents to report scrambled response to the sensitive survey question irrespective of their perception of the nature of the survey question. Other researchers like Saleem et al. (2019), Gjestvang & Singh (2009), Bouza-Herrera et al. (2022), Narjis

Copyright © B. Okafor et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

& Shabbir (2023), Azeem et al. (2024a) have contributed to the literature on forced randomized response technique models. In the optional randomized response technique models, respondents are instructed to either report a scrambled response if they consider the survey question sensitive or report a truthful response if they consider the survey question nonsensitive.

The idea behind the concept of the optional randomized response technique is that a survey question may be sensitive to one person but may not be sensitive to another person. The concept of the optional randomized response technique was introduced by Gupta et al. (2002). In recent times, the following researchers have contributed to the theory of optional randomized response technique models: Azeem et al. (2024a), Gupta et al. (2022), Parker et al. (2024).

The motivation for this article is based on the fact that the trade-off between privacy protection and data accuracy is disrupted with the use of the forced randomized response technique, most especially if the false response rate is high. Therefore, the paper proposed an optional randomized response technique model that will best balance the trade-off between privacy protection and data accuracy. The proposed model is an extension of Gjestvang & Singh's (2009) model, modified by Narjis & Shabbir's (2023) model under the forced randomized response technique to the optional randomized response technique.

By introducing multiplicative stochastic disturbance and allowing the respondents the choice of reporting a scrambled response or a truthful response based on their perception of the sensitivity of the survey question, the proposed model will strengthen confidentiality guarantees while preserving and improving the precision of statistical estimation. Below, Gjestvang & Singh (2009) and Narjis & Shabbir (2023) models were described in brief.

1.1. Gjestvang & Singh (2009) model

Using predetermined design parameters α and β that are controlled by the researcher, Gjestvang & Singh (2009) proposed an improved randomized response model for estimating the population mean of a sensitive variable. The model works as follows: Let Y be a sensitive quantitative variable with an unknown mean μ_Y and variance σ_Y^2. Suppose the mean θ and variance σ_s^2 of scrambling variable S are known. S follows a known probability distribution and is independent of Y. The respondent is asked to draw a card from a well-shuffled deck of cards and to report the scrambled response according to statements contained in the card selected. Each of the cards contains one of the following statements:

1) Multiply the scrambling variable S by α and add the result to the truthful value of sensitive variable Y with probability $p = \frac{\beta}{\alpha+\beta}$
2) Multiply the scrambling variable S with β and subtract the result from the truthful value of the sensitive variable Y with probability $1 - p = \frac{\alpha}{\alpha+\beta}$

Therefore, the i^{th} reported response h_i The sampled respondent is given as

$$h_i = \begin{cases} y_i + \alpha s_i \text{ with probability } p = \frac{\beta}{\alpha+\beta} \\ y_i - \beta s_i \text{ with probability } 1 - p = \frac{\alpha}{\alpha+\beta} \end{cases} \quad (1)$$

The unbiased estimator of the population mean μ_Y is given as

$$\hat{\mu}_{gs} = \frac{1}{n}\sum_{i=1}^{n} h_i \quad (2)$$

With variance given as

$$V(\hat{\mu}_{gs}) = \frac{1}{n}\left[\alpha\beta(\sigma_s^2 + \theta^2) + S_y^2\right] \quad (3)$$

1.2. Narjis & Shabbir (2023) model

Using design parameters α, β, and γ similar to those used by Gjestvang & Singh (2009), Narjis & Shabbir (2023) modified the Gjestvang & Singh (2009) model by introducing a third statement that allows some respondents to report the true value of the sensitive variable. Like in the case of Gjestvang & Singh's (2009) model, all the respondents are asked to scramble their response irrespective of their perception of the question being sensitive or not, and report their responses according to the three statements on the deck of cards provided.

1) Multiply the scrambling variable S with α and add the result to the truthful value of the sensitive variable Y with probability $p = \frac{\beta}{\alpha+\beta}$
2) Multiply the scrambling variable S with β and subtract the result from the truthful value of the sensitive variable Y with probability $1 - p = \frac{\alpha}{\alpha+\beta}$
3) Report the true value of the sensitive variable, say Y.

The response distribution of the new model is given by

$$z_i = \begin{cases} y_i + \alpha s_i \text{ with probability } p_1 = \frac{\beta}{\alpha+\beta+\gamma} \\ y_i - \alpha s_i \text{ with probability } p_2 = \frac{\alpha}{\alpha+\beta+\gamma} \\ y_i \text{ with probability } p_3 = \frac{\gamma}{\alpha+\beta+\gamma} \end{cases} \quad (4)$$

The unbiased estimator of the population mean is given by

$$\bar{y}_{ns} = \frac{1}{n}\sum_{i=1}^{n} z_i \quad (5)$$

And the variance is

$$V(\bar{y})_{ns} = \frac{1}{n}\left[\frac{\alpha\beta(\sigma_s^2+\theta^2)(\alpha+\beta)}{\alpha+\beta+\gamma} + S_y^2\right] \quad (6)$$

2. Methodology

2.1. Proposed optional randomized response model

Motivated by Gjestvang & Singh (2009) and Narjis & Shabbir (2023) models, the paper proposed an efficient optional randomized response technique model by introducing multiplicative stochastic disturbance and by giving the respondents the option to either use a randomization device to report a scrambled response if they perceive the survey question sensitive or report direct truthful response if they perceive the survey question nonsensitive.

In the proposed model, those respondents who consider the survey question nonsensitive will report a truthful response to the survey question, while those respondents who consider the survey question sensitive are provided with a randomization device, say a transparent box containing a set of well-shuffled cards. Each of the cards has one of the following instructions written on it.

1) If sensitive, report the scramble response as: Multiply scrambling variable S with α and add to the true value of the sensitive variable Y That is a report $Y + \alpha S$
2) If sensitive, report the scramble response as: Multiply scrambling variable S with $\beta$and subtract it from the true value of the sensitive variable Y That is a report $Y - \beta S$
3) If sensitive, report the scramble response as: Multiply scrambling variable S with the true value of the sensitive variable Y That is a report Y S

Let a sample of size. n be drawn from a finite population $\Omega = \Omega_1, \Omega_2, \ldots, \Omega_N$ of size N with simple random sampling without replacement (SRSWOR). Let Y be the sensitive study attribute with unknown mean μ_Y and unknown variance σ_Y^2

Let S be a scrambled variable independent of Y with known mean θ and variance σ_s^2. Let W, $(0 \leq W \leq 1)$ The probability that a respondent will scramble his/her response, W, is called the sensitivity level. It is assumed that the sensitivity level is known and can be obtained from an existing survey or can be estimated from a pilot survey, as $W = \frac{n_0}{n}$, where n_0 is the number of respondents who considered the survey question sensitive and n is the pilot sample size.

Let Y_i be the true response from i^{th} respondent and S_i be scramble variable selected by the i^{th} respondent in the population, let Z_i be the reported response by the i^{th} respondent in the population. Then, if the survey question is

1) sensitive, report scrambled response with probability W
2) non-sensitive, report non-scrambled response with probability $1 - W$

Let's define,

$$T_i = \begin{cases} \text{Yes if Sensitive for } i^{th} \text{ respondent} \\ \text{No if Non} - \text{Sensitive for } i^{th} \text{ respondent} \end{cases} \quad (7)$$

Then the reported response Z_i for the i^{th} respondent satisfies

$$Z_i = \begin{cases} T_i = \text{Yes} : \begin{cases} Y_i + \alpha S_i \text{ with probability } p_1 = \frac{W\beta}{\alpha+\beta+\gamma} \\ Y_i - \alpha S_i \text{ with probability } p_2 = \frac{W\alpha}{\alpha+\beta+\gamma} \\ Y_i S_i \text{ with probability } p_3 = \frac{W\gamma}{\alpha+\beta+\gamma} \end{cases} \\ T_i = \text{No}: Y_i \text{ with probability } p_4 = 1 - W \end{cases} \quad (8)$$

Thus,

$$\left.\begin{aligned} [Z_i|T_i = \text{Yes}] &= p_1(Y_i + \alpha S_i) + p_2(Y_i - \beta S_i) + p_3(Y_i S_i) \\ [Z_i|T_i = \text{No}] &= p_4 Y_i \end{aligned}\right\} \quad (9)$$

From (9), the reported response distribution model is then obtained as

$$Z_i = [Z_i|T_i = \text{Yes}] + [Z_i|T_i = \text{No}]$$

$$= p_1(Y_i + \alpha S_i) + p_2(Y_i - \beta S_i) + p_3(Y_i S_i) + p_4 Y_i$$

$$= \left(\frac{\beta W}{\alpha+\beta+\gamma}\right)(Y_i + \alpha S_i) + \left(\frac{\alpha W}{\alpha+\beta+\gamma}\right)(Y_i - \beta S_i)$$

$$+ \left(\frac{\gamma W}{\alpha+\beta+\gamma}\right)(Y_i S_i) + (1 - W)Y_i \quad (10)$$

where α, β and γ are predetermined real-valued positive constants such that $p_1 + p_2 + p_3 + p_4 = 1$ and p_1, p_2, p_3 are unconditional probabilities for the scramble rules. It should be noted that if $W = 1$ and $\gamma = 0$ The new model in (10) will reduce to the Gjestvang & Singh (2009) model. Also, if $W = 0$ The new model in (10) will reduce to a traditional direct interview; with this, the respondent's response to a sensitive question is no longer masked, hence, respondents' privacy is not guaranteed.

2.2. Estimation of the mean of the sensitive variable

For a sample of size n Respondents are drawn from a finite population. Ω With simple random sampling without replacement, let y_i be the true response from i^{th} respondent and s_i be scramble variable selected by the i^{th} respondent. Thus, from (10), the reported sample response distribution to the survey question for selected i^{th} The respondent is given as

$$z_i = \left(\frac{1}{\alpha+\beta+\gamma}\right)[W\beta(y_i + \alpha s_i) + W\alpha(y_i - \beta s_i) + W\gamma(y_i s_i) + (\alpha + \beta + \gamma)(1 - W)y_i] \quad (11)$$

Now, the sample mean of the response distribution from (11) is obtained as

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \begin{bmatrix} \left(\frac{W\beta}{\alpha+\beta+\gamma}\right)\frac{1}{n}\sum_{i=1}^{n}(y_i + \alpha s_i) + \left(\frac{W\alpha}{\alpha+\beta+\gamma}\right)\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta s_i) \\ + \left(\frac{W\gamma}{\alpha+\beta+\gamma}\right)\frac{1}{n}\sum_{i=1}^{n}(y_i s_i) + (1 - W)\frac{1}{n}\sum_{i=1}^{n} y_i \end{bmatrix}$$

$$= \frac{\mu_y}{(\alpha+\beta+\gamma)}[(\alpha + \beta + \gamma\theta)W + (1 - W)(\alpha + \beta + \gamma)] \quad (12)$$

This suggests that $\hat{\mu}_y$ is unbiased for $\mu_Y$if $\theta = 1$ThusThusThusThusThusThusWe have the following theorem.

Theorem 1: *$\hat{\mu}_y$ is an unbiased estimator of μ_Y*

Proof: Let the expectation over the sampling design d be E_d and expectation over the randomization device R be E_R, then

$$E(\hat{\mu}_y) = E_d E_R\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{W\beta}{\alpha+\beta+\gamma}\right)(y_i + \alpha s_i) + \left(\frac{W\alpha}{\alpha+\beta+\gamma}\right)(y_i - \beta s_i) + \left(\frac{W\gamma}{\alpha+\beta+\gamma}\right)(y_i s_i) + (1 - W)y_i\right\}\right]$$

$$= \frac{1}{n}E_d\left[\sum_{i=1}^{n} E_R\left\{\left(\frac{W\beta}{\alpha+\beta+\gamma}\right)(y_i + \alpha s_i) + \left(\frac{W\alpha}{\alpha+\beta+\gamma}\right)(y_i - \beta s_i) + \left(\frac{W\gamma}{\alpha+\beta+\gamma}\right)(y_i s_i) + (1 - W)y_i\right\}\right]$$

$$= E_d\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{W\beta}{\alpha+\beta+\gamma}\right)(Y + \alpha\theta) + \left(\frac{W\alpha}{\alpha+\beta+\gamma}\right)(Y - \beta\theta) + \left(\frac{W\gamma}{\alpha+\beta+\gamma}\right)(Y\theta) + (1 - W)Y\right\}\right]$$

$$= E_d\left[\frac{1}{n(\alpha+\beta+\gamma)}\sum_{i=1}^{n}\{W\beta Y + W\alpha Y + W\gamma Y\theta + (1 - W)Y\}\right]$$

$$= \left(\frac{\mu_Y}{\alpha+\beta+\gamma}\right)[W\gamma(\theta - 1) + (\alpha + \beta + \gamma)] \quad (13)$$

With the assumption that the scramble variable S comes from a normal distribution with mean $(\theta = 1)$, (13) reduced to

$$\hat{\mu}_y = \frac{(\alpha+\beta+\gamma)\mu_Y}{(\alpha+\beta+\gamma)} = \mu_Y \quad (14)$$

Hence the proof

Theorem 2: *The variance of $\hat{\mu}_y$ is given as*

$$\mathrm{Var}(\hat{\mu}_y) = \frac{1}{n}\left[\left(\frac{W}{\alpha+\beta+\gamma}\right)\left\{\gamma(\sigma_y^2 + \mu_y^2)\left((\sigma_s^2 + \theta^2) - 1\right) + \alpha\beta(\alpha + \beta)(\sigma_y^2 + \theta^2)\right\} + \sigma_y^2\right] \quad (15)$$

Proof: By definition,

$$\mathrm{Var}(\hat{\mu}_y) = \mathrm{Var}(\bar{\bar{z}}) = \frac{1}{n}E(z - \bar{z})^2$$

$$= \frac{1}{n}E(z^2 + \bar{z}^2 - 2z\bar{z}) = \frac{1}{n}\left(E(z^2) - \mu_y^2\right) \quad (16)$$

Now to compute the $E(z^2)$, recall that z is the response from two distributions – those who provide responses and those who provide scrambled responses. Thus,

$$E(z^2) = E_S(z^2) + E_T(z^2) \quad (17)$$

where, E_S and E_T There are expectations with respect to scrambled response and truthful response, respectively.

$$\mathrm{Var}(\hat{\mu}_y) = \frac{1}{n}[E_S(z^2) + E_T(z^2) - \mu_y^2] \quad (18)$$

$$= \frac{1}{n}\left[\left(\frac{W}{\alpha+\beta+\gamma}\right)E_S[\beta(y + \alpha S)^2 + \alpha(y - \beta S)^2 + \gamma y^2 S^2] + E_T\left((1 - W)y^2\right) - \mu_Y^2\right]$$

$$= \frac{1}{n}\left[\left(\frac{W}{\alpha+\beta+\gamma}\right)[(\alpha + \beta)\{(\sigma_Y^2 + \mu_Y^2) + \alpha\beta(\sigma_s^2 + \theta^2)\} + \gamma(\sigma_Y^2 + \mu_Y^2)(\sigma_s^2 + \theta^2)] + (1 - W)(\sigma_Y^2 + \mu_Y^2) - \mu_Y^2\right]$$

$$= \frac{1}{n}\left[\left(\frac{W}{\alpha+\beta+\gamma}\right)\left\{\gamma\left(\sigma_y^2 + \mu_y^2\right)\left((\sigma_s^2 + \theta^2) - 1\right) + \alpha\beta(\alpha + \beta)\left(\sigma_y^2 + \theta^2\right)\right\} + \sigma_y^2\right] \quad (19)$$

Hence the proof.

3. Results

3.1. Efficiency comparison

The theoretical conditions for the proposed optional randomized response model to be more efficient than the Gjestvang & Singh (2009) model, and Narjis & Shabbir (2023) model were:

i) Proposed Model Versus Gjestvang & Singh (2009) Estimator

The proposed model will be more efficient than Gjestvang & Singh's (2009) model if

$$Var_{pro}\left(\hat{\mu}_y\right) < Var_{gs}\left(\hat{\mu}_y\right) \quad (20)$$

Or

$$Var_{pro}\left(\hat{\mu}_y\right) - Var_{gs}\left(\hat{\mu}_y\right) < 0 \quad (21)$$

or

$$\frac{\left((\sigma_s^2+\theta^2)-1\right)}{\alpha\beta(\sigma_s^2+\theta^2)} < \frac{(\alpha+\beta+\gamma)-W(\alpha+\beta)}{W\gamma\left(\sigma_y^2+\mu_y^2\right)} \quad (22)$$

ii) Proposed Model Versus Narjis & Shabbir (2023)

The proposed model will be more efficient than Narjis & Shabbir's (2023) model if

$$Var_{pro}\left(\hat{\mu}_y\right) < Var_{ns}\left(\hat{\mu}_y\right) \quad (23)$$

or

$$Var_{pro}\left(\hat{\mu}_y\right) - Var_{ns}\left(\hat{\mu}_y\right) < 0 \quad (24)$$

or

$$\frac{\left((\sigma_s^2+\theta^2)-1\right)}{\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)} < \frac{1-W}{W\gamma\left(\sigma_y^2+\mu_y^2\right)} \quad (25)$$

3.2. Relative efficiency of the proposed model

Relative efficiency has been the most commonly used metric for evaluating the performance of a proposed randomized response technique model relative to some other existing models. The efficiency of any proposed randomized response technique model M_1 compared to another existing randomized response technique model M_2 for a parameter ϕ or the relative efficiency of M_1 over M_2 is defined as

$$\text{Relative Efficiency} = \theta_{re} = \frac{Var(M_2)}{Var(M_1)} \quad (26)$$

Since variance is always a positive quantity, $\theta_{re} \geq 0$. If $\theta_{re} > 1$ The proposed model is more efficient than the existing model under consideration. The larger the value of θ_{re}The more efficient the model will be. In this subsection, the relative efficiency of the optional randomized response model relative to the Gjestvang & Singh (2009) model and Narjis & Shabbir (2023) model would be evaluated. Thus, the proposed model will perform better than the Gjestvang & Singh (2009) model if the relative efficiency of the proposed model over the Gjestvang & Singh (2009) model is defined as

$$\theta_{re(1)} = \frac{Var_{gs}(\hat{\mu}_y)}{Var_{pro}(\hat{\mu}_y)} = \frac{(\alpha+\beta+\gamma)\left[\alpha\beta(\sigma_s^2+\theta^2)+\sigma_y^2\right]}{W\left\{\gamma\left(\sigma_y^2+\mu_y^2\right)\left((\sigma_s^2+\theta^2)-1\right)+\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)\right\}+\sigma_y^2(\alpha+\beta+\gamma)} > 1 \quad (27)$$

In the same vein, the proposed model will perform better than Narjis & Shabbir's (2023) model if the relative efficiency of the proposed model over Narjis & Shabbir's (2023) model is defined as

$$\theta_{re(2)} = \frac{Var_{ns}(\hat{\mu}_y)}{Var_{pro}(\hat{\mu}_y)} = \frac{\left[\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)+\sigma_y^2(\alpha+\beta+\gamma)\right]}{W\left\{\gamma\left(\sigma_y^2+\mu_y^2\right)\left((\sigma_s^2+\theta^2)-1\right)+\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)\right\}+\sigma_y^2(\alpha+\beta+\gamma)} > 1 \quad (28)$$

3.3. Privacy metric

The respondents' privacy protection is of utmost importance to survey statisticians. The reason why they continue seeking a randomized response model that offers high privacy protection to respondents. A measure of privacy proposed by Yan et al. (2009), often used in quantitative models, is given by

$$\nabla = E[Z - Y]^2 \quad (29)$$

Where Y is the true response by the respondents, and Z is the scrambled response by the respondents. It should be noted that when respondents' scrambled responses lie far from their true responses, this metric will be high. Larger values of ∇ are most appropriate. In this section, the privacy protection level of the proposed optional randomized response technique model was compared with those of the Gjestvang & Singh (2009) model and Narjis & Shabbir (2023) model.

Gjestvang & Singh (2009) Privacy Protection Level

The response distribution for the Gjestvang & Singh (2009) model is given as

$$Z = p(Y + \alpha S) + (1 - p)(Y - \beta S) \tag{30}$$

Where,

$$p = \frac{\beta}{\alpha+\beta}, 1 - p = \frac{\alpha}{\alpha+\beta}$$

Using (30),

$$\nabla_{gs} = E[p(Y + \alpha S) + (1 - p)(Y - \beta S) - Y]^2$$

$$= (\sigma_s^2 + \theta^2)\alpha\beta \tag{31}$$

Narjis & Shabbir (2023) Privacy Protection Level

The response distribution for Narjis & Shabbir's (2023) model is given as

$$Z = p_1(Y + \alpha S) + p_2(Y - \beta S) + p_3 Y \tag{32}$$

Where,

$$p_1 = \frac{\beta}{\alpha+\beta+\gamma}, p_2 = \frac{\alpha}{\alpha+\beta+\gamma}, p_3 = \frac{\gamma}{\alpha+\beta+\gamma}$$

Using (32),

$$\nabla_{ns} = E[p_1(Y + \alpha S) + p_2(Y - \beta S) + p_3 Y - Y]^2$$

$$= E(S^2)[p_1\alpha^2 + p_2\beta^2] = (\sigma_s^2 + \theta^2)\left[\frac{\alpha\beta(\alpha+\beta)}{\alpha+\beta+\gamma}\right] \tag{33}$$

Proposed Model Privacy Protection Level

In optional randomized response technique models, the privacy of those respondents who do not find the question sensitive and opted to provide a truthful response is not compromised (see Gupta et al., 2018), so they are not considered in computing the privacy protection level of the respondents.

Suppose that among n respondents who responded using the optional randomized response technique model, m' find the survey question sensitive, and opted for a scrambled response.

Yan et al. (2009) privacy protection measure for the optional randomized response technique models will be computed only from the group that provides scrambled responses. Thus, the Yan et al. (2009) privacy protection measure for the optional randomized response technique models is given as

$$\nabla = E[Z - Y]^2 = E'_m [E(Z - Y)^2 \mid m'] \tag{34}$$

Thus, the response distribution for the proposed model is given as

$$Z = p_1(Y + \alpha S) + p_2(Y - \beta S) + p_3 Y S \tag{35}$$

Where,

$$p_1 = \frac{\beta}{\alpha+\beta+\gamma}, p_2 = \frac{\alpha}{\alpha+\beta+\gamma}, p_3 = \frac{\gamma}{\alpha+\beta+\gamma}$$

Proposed Model Privacy Protection Level

In optional randomized response technique models, the privacy of those respondents who do not find the question sensitive and opted to provide a truthful response is not compromised (see Gupta et al., 2018), so they are not considered in computing the privacy protection level of the respondents.

Suppose that among n respondents who responded using the optional randomized response technique model, m' find the survey question sensitive, and opted for a scrambled response.

Yan et al. (2009) privacy protection measure for the optional randomized response technique models will be computed only from the group that provides scrambled responses. Thus, the Yan et al. (2009) privacy protection measure for the optional randomized response technique models is given as

$$\nabla = E[Z - Y]^2 = E'_m [E(Z - Y)^2 \mid m'] \tag{34}$$

Thus, the response distribution for the proposed model is given as

$$Z = p_1(Y + \alpha S) + p_2(Y - \beta S) + p_3 Y S \tag{35}$$

where,

$$p_1 = \frac{\beta}{\alpha+\beta+\gamma}, p_2 = \frac{\alpha}{\alpha+\beta+\gamma}, p_3 = \frac{\gamma}{\alpha+\beta+\gamma}$$

Using (34) and (35),

$$\nabla_{pro} = E_{m'}\left[E(p_1(Y + \alpha S)^2 + p_2(Y - \beta S)^2 + p_3 Y^2 S^2 \,|m') + E(Y^2|\, m') - 2E(ZY|m')\right]$$

$$= \left[\frac{\alpha^2\beta}{\alpha+\beta+\gamma}E(S^2\,|m') + \frac{\alpha\beta^2}{\alpha+\beta+\gamma}E(S^2\,|m') + \frac{\gamma}{\alpha+\beta+\gamma}E(Y^2S^2\,|m')\right]$$

$$= \left[\frac{\alpha^2\beta}{\alpha+\beta+\gamma}E(S^2) + \frac{\alpha\beta^2}{\alpha+\beta+\gamma}E(S^2) + \frac{\gamma}{\alpha+\beta+\gamma}E(Y^2S^2)\right]$$

$$= E(S^2)\left[\frac{\alpha^2\beta}{\alpha+\beta+\gamma} + \frac{\alpha\beta^2}{\alpha+\beta+\gamma}E(S^2\,)\right] + \frac{\gamma}{\alpha+\beta+\gamma}E(Y^2S^2)$$

$$= (\sigma_s^2 + \theta^2)\left(\frac{\alpha\beta(\alpha+\beta)}{\alpha+\beta+\gamma}\right) + \left(\frac{\gamma}{\alpha+\beta+\gamma}\right)(\sigma_s^2 + \theta^2)(\sigma_y^2 + \mu_y^2)$$

$$= (\sigma_s^2 + \theta^2)\left(\frac{\alpha\beta(\alpha+\beta)+\gamma(\sigma_y^2+\mu_y^2)}{\alpha+\beta+\gamma}\right) \tag{36}$$

3.4. Weighted privacy-efficiency measure

Further, the performance of the proposed model was evaluated by assigning relative weight to privacy and efficiency. Thus, using the weighted privacy-efficiency measure proposed by Azeem (2023), given as

$$\phi = \frac{\omega_1\theta_{re}+\omega_2 P}{\omega_1+\omega_2} \tag{37}$$

Where,

$$P = \frac{\nabla_p}{\nabla_o} \tag{38}$$

Is the ratio of privacy of the proposed model to privacy of the existing model and θ_{re} is the relative efficiency, while ω_1 is the weight assigned to efficiency and ω_2 Is the weight assigned to privacy? If $\phi = 1$The two models are of equal performance, if $\phi > 1$The proposed model has higher performance than the old model, and if $0 < \phi < 1$This indicates that the old model is of higher performance than the proposed model. For symmetric, Azeem (2023) suggested using log ϕ, whose value ranges from $-\infty < \phi < \infty$ and is symmetric around zero. Thus,

$$\log\phi = \log\left[\frac{\omega_1\theta_{re}+\omega_2 P}{\omega_1+\omega_2}\right] \tag{39}$$

log $\phi > 0$ indicates that the proposed model performs better than the old model while log $\phi < 0$ indicates that the proposed model performs poorly compared to the old model. For log $\phi = 0$ indicates that both models are of equal performance.
Thus, we define the measure log for the proposed model versus the Gjestvang & Singh (2009) model as

$$\log\phi_1 = \log\left[\frac{\omega_1\theta_{re(1)}+\omega_2 P_{(1)}}{\omega_1+\omega_2}\right] \tag{40}$$

And the measure log for the proposed model and Narjis & Shabbir's (2023) model is given as

$$\log\phi_1 = \log\left[\frac{\omega_1\theta_{re(2)}+\omega_2 P_{(2)}}{\omega_1+\omega_2}\right] \tag{41}$$

Thus, using (31) and (36), the ratio of privacy of the proposed model versus the Gjestvang & Singh (2009) model was obtained as

$$P_{(1)} = \frac{\nabla_{pro}}{\nabla_{gs}} = \frac{\alpha\beta(\alpha+\beta)+\gamma(\sigma_y^2+\mu_y^2)}{\alpha\beta(\alpha+\beta+\gamma)} \tag{42}$$

Also using (33) and (36), the ratio of privacy of the proposed model versus Narjis & Shabbir's (2023) model was obtained as

$$P_{(2)} = \frac{\nabla_{pro}}{\nabla_{ns}} = \frac{\alpha\beta(\alpha+\beta)+\gamma(\sigma_y^2+\mu_y^2)}{\alpha\beta(\alpha+\beta)} \tag{43}$$

Using (27) and (42), (40) becomes

$$\log\phi_1 = \log\left[\left(\frac{1}{\omega_1+\omega_2}\right)\left(\omega_1\frac{\alpha\beta(\sigma_s^2+\theta^2)+\sigma_y^2(\alpha+\beta+\gamma)}{W\{\gamma(\sigma_y^2+\mu_y^2)((\sigma_s^2+\theta^2)-1)+\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)\}+\sigma_y^2(\alpha+\beta+\gamma)}\right) + \omega_2\frac{\alpha\beta(\alpha+\beta)+\gamma(\sigma_y^2+\mu_y^2)}{\alpha\beta(\alpha+\beta+\gamma)}\right] \tag{44}$$

Using (28) and (43), (41) becomes

$$\log \phi_2 = \log\left[\left(\frac{1}{\omega_1+\omega_2}\right)\left(\omega_1 \frac{\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)+\sigma_y^2(\alpha+\beta+\gamma)}{W\left\{\gamma(\sigma_y^2+\mu_y^2)\left((\sigma_s^2+\theta^2)-1\right)+\alpha\beta(\alpha+\beta)(\sigma_s^2+\theta^2)\right\}+\sigma_y^2(\alpha+\beta+\gamma)}\right) + \omega_2 \frac{\alpha\beta(\alpha+\beta)+\gamma(\sigma_y^2+\mu_y^2)}{\alpha\beta(\alpha+\beta)}\right] \tag{45}$$

3.5. Empirical analysis

To evaluate the performance of the proposed model, an empirical study was carried out. The tables below show the results obtained. For different values of design parameters (α, β, γ)Table 1 presents the sample variances of the proposed model, Gjestvang & Singh (2009) (GS) model, and Narjis & Shabbir (2023) (NS) model, as well as the relative efficiency of the proposed model over the Gjestvang & Singh (2009) (GS) model and Narjis & Shabbir (2023) (NS) model.

Table 1: Variances and Relative Efficiency$\sigma_y^2 = 2, \sigma_s^2 = 8, \mu_y = 5, n = 40$

w	α	β	γ	variance $\left(\mathrm{Var}(\hat{\mu}_y)\right)$			Relative Efficiency (θ_{re})	
				Proposed Model	GS Model	NS Model	GS Model	NS Model
0.8	5	3	1	2.6033	3.0500	2.7167	1.1716	1.0435
			2	2.7260	3.0500	2.4500	1.1189	0.8988
		6	1	4.7650	6.0500	2.0500	1.2697	0.4302
			2	4.6931	6.0500	5.1269	1.2891	1.0924
	10	3	1	4.7771	6.0500	5.6214	1.2665	1.1767
			2	4.7140	6.0500	5.2500	1.2834	1.1137
		6	1	9.3076	12.0500	11.3441	1.2946	1.2188
			2	9.0033	12.0500	10.7167	1.3384	1.1903
0.5	8	6	2	4.5453	9.6500	8.4500	2.1231	1.8591
			1	4.6875	9.6500	9.0100	2.0587	1.9221
		10	2	7.4863	16.0500	14.4500	2.1439	1.9302
			15	5.4875	16.0500	8.7773	2.9248	1.5995
	10	15	25	5.7313	30.0500	15.0500	5.2432	2.6260
			20	9.4333	30.0500	16.7167	3.1855	1.7721
		10	30	5.4675	20.0500	8.0500	3.6671	1.4723
			2	9.3557	20.0500	18.2318	2.1431	1.9487
0.2	20	8	3	5.9221	32.0500	28.9532	5.4119	4.8890
			5	5.6235	32.0500	27.2015	5.6993	4.8371
		10	6	6.8742	40.0500	33.3833	5.8262	4.8563
			4	7.2200	40.0500	35.3441	5.5471	4.8953
	14	15	8	6.8381	42.0500	32.9689	6.1494	4.8214
			9	6.6843	42.0500	32.1026	6.2908	4.8027
		2	12	1.0950	5.6500	3.2500	5.1598	2.9680
			6	1.1223	5.6500	4.1227	5.0344	3.6736

Table 2 presents the privacy protection of the proposed model, Gjestvang & Singh (2009) model, and Narjis & Shabbir (2023) model for different values of design parameters.

Table 2: Privacy Protection Level $\sigma_y^2 = 2, \sigma_s^2 = 8, \mu_y = 5, n = 40$

w	α	β	Γ	Privacy Protection Level (∇)		
				Proposed Model	GS Model	NS Model
0.8	5	3	1	130.6667	120.0000	106.6667
			2	139.2000	120.0000	96.0000
		6	1	238.0000	240.0000	220.0000
			2	236.3077	240.0000	203.0769
	10	3	1	238.2857	240.0000	222.8571
			2	236.8000	240.0000	208.0000
		6	1	464.4706	480.0000	451.7647
			2	450.6667	480.0000	426.6667
0.5	8	6	2	363.0000	384.0000	336.0000
			1	372.8000	384.0000	358.4000
		10	2	597.6000	640.0000	576.0000
			15	447.2727	640.0000	349.0909
	10	15	25	708.0000	1200.0000	600.0000
			20	762.6667	1200.0000	666.6667
		10	30	449.6000	800.0000	320.0000
			2	746.9091	800.0000	727.2727
0.2	20	8	3	1177.0320	1280.0000	1156.1290
			5	1118.7880	1280.0000	1086.0610
		10	6	1369.3330	1600.0000	1333.3330
			4	1437.1760	1600.0000	1411.7650
	14	15	8	1363.4590	1680.0000	1316.7570
			9	1333.2630	1680.0000	1282.1050
		2	12	220.5714	224.0000	128.0000
			6	221.8182	224.0000	162.9091

Table 3 presents the weighted privacy-efficiency measure of the proposed model with respect to the Gjestvang & Singh (2009) model and Narjis & Shabbir (2023) model for different values of design parameters.

Table 3: Weighted Privacy-Efficiency Measure. $\sigma_y^2 = 2, \sigma_s^2 = 8, \mu_y = 5, n = 40$

w	α	β	γ	ω_1	ω_2	logϕ	
						$\log\phi_1$	$\log\phi_2$
0.8	5	3	1	0.2	0.8	0.04353	0.075075
			2	0.5	0.5	0.056687	0.069807
		6	1	0.8	0.2	0.084245	-0.025139
			2	0.2	0.8	0.019332	0.06047
	10	3	1	0.5	0.5	0.052945	0.050373
			2	0.8	0.2	0.087803	0.048696
		6	1	0.2	0.8	0.014119	0.027863
			2	0.5	0.5	0.056387	0.050486
0.5	8	6	2	0.2	0.8	0.07220	0.092053
			1	0.5	0.5	0.180341	0.170601
		10	2	0.8	0.2	0.279186	0.243451
			15	0.2	0.8	0.058448	0.12869
	10	15	25	0.5	0.5	0.0464876	0.279434
			20	0.8	0.2	0.427408	0.216553
		10	30	0.2	0.8	0.0072994	0.151819
			2	0.5	0.5	0.187058	0.172565
0.2	20	8	3	0.8	0.2	0.654509	0.614352
			5	0.2	0.8	0.259602	0.253403
		10	6	0.5	0.5	0.523876	0.468594
			4	0.8	0.2	0.66439	0.614881
	14	15	8	0.2	0.8	0.273959	0.253494
			9	0.5	0.5	0.541638	0.466632
		2	12	0.8	0.2	0.635965	0.434421
			6	0.2	0.8	0.255054	0.261024

4. Discussion of Results

The study proposed an efficient randomized response technique model designed to estimate the population mean of sensitive variable while protecting respondent's privacy and confidentiality. In evaluating the performance of the proposed model, a simulation study was carried out.

The results of efficiency comparison given in Table 1 show that the variance of the proposed model is smaller than those of Gjestvang & Singh (2009) model and Narjis & Shabbir (2023) model for all design parameters $(\alpha, \beta, \gamma, W)$ configurations. Likewise, the relative efficiencies are generally greater than one, showing that the proposed model gives more efficient estimates of the population mean. The improvement in precision can be attributed to the flexibility of the optional design, which allows the respondents to report truthfully response when the survey question is perceived nonsensitive thereby reducing the influence of false response. The introduction of multiplicative stochastic disturbance enhances variance reduction by balancing protection of privacy with estimation precision. The results further reveal that efficiency gains increase as the sensitivity level (W) approaches zero. Even as the sensitivity level approaches one, in that case more respondents perceived the survey question sensitive, the proposed model still maintains competitive performance confirming its robustness and suitability for various survey situations. Table 2 reveals that proposed model offers higher privacy protection level than Gjestvang & Singh (2009) and Narjis & Shabbir (2023) models for different values of design parameters.

The introduction of multiplicative stochastic disturbance substantially increases the variability between the true and reported responses. The model thus, ensures better protection of respondents' confidentiality without compromising data integrity. To broadly evaluate the performance of the proposed model, log weighted privacy-efficiency measure was computed. Table 3 shows predominantly positive values for log weighted privacy-efficiency measure for various design parameters configurations. This confirms that the proposed model outperforms the two models under consideration under different weight combinations. Notably, when higher weight values are assigned to privacy, the model maintained a strong balance, confirming that it effectively achieves twofold optimization - protecting respondents' privacy while maintaining high estimation precision.

These finding validate the theoretical expectations and demonstrate that the proposed model successfully addresses the long-standing challenge of balancing privacy protection and estimation efficiency in sensitive surveys.

5. Conclusion

The paper suggested an optional randomized response model that efficiently estimates the mean of sensitive variable while preserving privacy and confidentiality of the respondents. The proposed model is an enhanced alternative to Gjestvang & Singh (2009) and Narjis & Shabbir (2023) models.

Theoretical analysis established that the proposed estimator is unbiased and has lower variance compared to Gjestvang & Singh (2009) and Narjis & Shabbir (2023) models. The empirical evaluation further proves that the consistently results in higher relative efficiency and privacy protection level than the two models under consideration. The log weighted privacy-efficiency measure also confirmed the dominance of the model under various weighting combinations, suggesting its sturdiness in most practical situations.

Therefore, the proposed model will be a valuable alternative for obtaining reliable estimates of sensitive variables while safeguarding respondent privacy and confidentiality.

Acknowledgement

The authors are grateful to the reviewers for their suggestions that help to improve the quality of the paper.

References

[1] Azeem, M., Najma, S., Sundus, H., & Musarrat, I. (2024). An Efficient Estimator of population Variance of a Sensitive Variable with a New Randomized Response Technique. *Heliyon,* 10(224), e27488. https://doi.org/10.1016/j.heliyon.2024.e27488.

[2] Azeem, M. (2023). Introducing a Weighted Measure of Privacy and Efficiency for Comparison of Quantitative Randomized Response Models. *Pak. J. Statist*, 39(3), 377–383.

[3] Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A Note on Randomized Response Models for Quantitative Data. *Metrika*, 60, 255–260. https://doi.org/10.1007/s001840300308

[4] Bouza-Herrera, N. C., Juarez, P. O., Santiago-Moreno, A., & Sautto-Vallejo, M. J. (2022). A Two-Stage Scrambling Procedure: Simple and Stratified Random Sampling. An Evaluation of Covid19's Data in Mexico. *Revista Investigacion Operacional*, 43(4), 421–430.

[5] Eichhorn, B. H., & Hayre, S. L. (1983). Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data. *Journal of Statistical Planning and Inference*, 7, 307–316. https://doi.org/10.1016/0378-3758(83)90002-2.

[6] 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

[7] Greenberg, B. G., Roy, R. K., James, R. A., & Daniel, G. H. (1971). Application of the Randomized Response Technique in Obtaining Quantitative Data. *Journal of The American Statistical Association*, 66(3334), 243–250. https://doi.org/10.1080/01621459.1971.10482248

[8] 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.

[9] Gupta, S., Zhang, J., Khalil, S., & Sapra, P. (2022). Mitigating Lack of Trust in Quantitative Randomized Response Technique Models. *Communication in Statistics - Part B Simulation and Computation*, 51, 1–9.

[10] Gupta, S. N., Gupta, B. C., & Singh, S. (2002). Estimation of Sensitivity Level of Personal Interview Survey Questions. *Journal of Statistical Planning and Inference*, 100, 239–247. https://doi.org/10.1016/S0378-3758(01)00137-9

[11] Narjis, G., & Shabbir, J. (2023). An Efficient New Scrambled Response Model for Estimating Sensitive Population Mean in Successive Sampling. *Communication in Statistics - Simulation and Computation*, 52(11), 5317–5344. https://doi.org/10.1080/03610918.2021.1986528.

[12] Parker, M., Gupta, S., & Sadia, K. (2024). A Mixture Quantitative Randomized Response Model That Improves Trust in RRT Methodology. *Axioms*, 13(11), 1–13. https://doi.org/10.3390/axioms13010011

[13] Pollock, K. H., & Bek, Y. (1976). A Comparison of Three Randomized Response Models for Quantitative Data. *Journal of the American Statistical Association*, 71, 884–886. https://doi.org/10.1080/01621459.1976.10480963

[14] Saleem, I., Sanaullah, A., & Koyuncu, N. (2019). Estimation of Mean of a Sensitive Quantitative Variable in Complex Survey: Improved Estimator and Scrambled Randomized Response Model. *GU Journal of Science*, 32, 1021–1043. https://doi.org/10.35378/gujs.418681

[15] Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of The American Statistical Association*, 60, 63–69. https://doi.org/10.1080/01621459.1965.10480775

[16] 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

[17] Yan, Z., Wang, J., & Lai, J. (2009). An Efficiency and Protection Degree Based Comparison Among the Quantitative Randomized Response Strategies. *Communications in Statistics: Theory and Methods*, 38, 400–408. https://doi.org/10.1080/03610920802220785.