International Journal of Mathematical, Engineering and Management Sciences
Vol. 10, No. 6, 1953-1966, 2025
https://doi.org/10.33889/IJMEMS.2025.10.6.090

Ram Arti Publishers

New Randomization Technique to Estimate the Population Mean of Quantitative Sensitive Variable

Chandraketu Singh
Decision Sciences Area,
Jaipuria Institute of Management Lucknow, Lucknow, 226010, Uttar Pradesh, India.
Corresponding author: chandraketu.singh@jaipuria.ac.in

Vijay Lakshmi Singh
OB & HR Management Area,
Jaipuria Institute of Management Lucknow, Lucknow, 226010, Uttar Pradesh, India.
E-mail: vijaylakshmi.singh@jaipuria.ac.in

Sanjay K. Singh
Economics & Business Environment Area,
Indian Institute of Management Lucknow, Lucknow, 226013, Uttar Pradesh, India.
E-mail: sanjay@iiml.ac.in

(Received on March 3, 2025; Revised on April 25, 2025 & June 20, 2025; Accepted on July 7, 2025)

Abstract
Sensitive topics such as workplace misconduct, corporate fraud, unethical decision-making, diversity and inclusion challenges, and leadership accountability often present significant obstacles in survey research. Respondents may hesitate to disclose truthful information due to fear of negative consequences, damage to their professional reputation, or concerns about confidentiality. Such apprehensions frequently lead to non-response or biased responses, compromising data quality. To address these challenges, this study introduces two scrambled randomized response techniques specifically designed to enhance the accuracy of estimates for population means involving quantitative sensitive variables. These methods are designed to enhance respondent confidentiality while ensuring greater accuracy in the collected data. Empirical studies have been conducted to validate theoretical findings and assess the efficiency of the techniques. The results demonstrate that the proposed techniques offer superior efficiency and stronger privacy protection compared to the existing method. The findings highlight the practical significance of these methods for researchers and practitioners working with sensitive survey topics and encourage for their adoption in future large-scale data collection initiatives.

Keywords- Sensitive characteristics, Poisson distribution, Privacy protection, Empirical study.

1. Introduction

Socio-economic indicators like quality of life, living conditions, availability of resources, and access to healthcare reflect the economic development status of an economy. Both the public and non-public agencies are involved in the preparation of these socio-economic development indicators by collecting primary data at the grassroots levels through sample surveys. These surveys are the only reliable source of information about the status of housing, poverty level, and living conditions of the people in a nation. The policy formulation by contemporary governments relies on the data collected through such sample surveys. For example, in India, public policy makers heavily utilize the socio-economic data collected by the National Sample Survey Office (NSSO) through its sample surveys conducted across the nation.

The collection of data through sample surveys relies on the ability and the efficiency of the field investigators. They collect information, using questionnaires, through interviews by recording the responses of the respondents. The sample survey questionnaires often include sensitive and personal questions. Such questions, if asked in the presence of family members, relatives and our neighbors, then the probability of getting incorrect responses becomes higher. Human behavior often makes respondents reluctant to share personal and sensitive information with surveyors they perceive as unfamiliar or untrustworthy. This hesitancy arises from concerns about privacy, fear of judgment, or potential misuse of the disclosed information.

The respondents might not be willing to share their financial information with unknown surveyors and hence they might hide their savings, investments, availability of bank accounts and usage of digital payments. The respondents might also hide involvement in unethical practices like tax evasion. They might hide earnings from clandestine sources, like theft, cheating or trade in illicit goods.

Furthermore, respondents, particularly women, tend to be sensitive about sharing information related to their health with unfamiliar surveyors. For instance, a woman respondent might feel uncomfortable answering questions about pregnancy, sexually transmitted diseases, or contraceptive usage. Similarly, topics that are often stigmatized in society, such as substance use, expenditure on addictions, or sexual orientation (e.g., homosexuality), may evoke discomfort or reluctance to respond truthfully. These sensitivities highlight the need for careful questionnaire design, culturally appropriate language, and creating a sense of trust and confidentiality during the survey process to improve response accuracy.

The socio-economic surveys conducted by prominent agencies such as the World Bank Organization for Economic Co-operation and Development (OECD), National Family Health Survey (NFHS) India, and the Socio-Economic Caste Census (SECC) include a significant number of sensitive questions. These questions often delve into personal, health-related, and socially stigmatized topics.

The objective of this paper is to develop a sampling method that facilitates the accurate collection of responses to sensitive questions in socio-economic surveys. The proposed randomization model for the estimation of the population mean of the quantitative sensitive variable is designed to address the challenges associated with respondent reluctance and social desirability bias. Furthermore, the paper provides a rigorous statistical analysis to demonstrate why the suggested method outperforms existing sampling methods in terms of privacy, response accuracy, and applicability in diverse socio-economic contexts. The remaining of the paper is structured as follows: the introduction is followed by a review of literature, and the methods are described through sections on proposed techniques and measure of privacy protection, followed by results described through the section on efficiency comparison, followed finally by a section on conclusion and recommendations.

2. Review of Literature

The collection of socio-economic data usually relies on interviewer-administered questionnaires. Sjöström and Holst (2002) argue that people tend to give answers to questionnaires more according to a social norm than to the actual situation. This is called social desirability. Warner (1965) introduced the Randomized Response Technique (RRT) to mitigate response bias in sensitive surveys, comparing its efficacy with direct questioning. Greenberg et al. (1969) developed a theoretical framework for RRT, comparing its efficiency with Warner's technique and discussing sample allocation methods. Greenberg et al. (1971) developed the Unrelated Question randomized response model for quantitative data, whereby participants respond to either a sensitive or an innocuous question based on a random mechanism, enabling unbiased estimation of the population mean. Goodstadt and Gruson (1975) compared randomized response and direct questioning for drug use and found that RRT not only reduced item nonresponse but also yielded significantly higher prevalence estimates. Their results indicate that direct questioning likely underestimates sensitive behavior, highlighting RRT's improved validity in such contexts. Pollock and Bek (1976) examined RRT techniques for quantitative data, emphasizing techniques involving addition or multiplication of random numbers. Horvitz et al. (1976) provided a comprehensive review of RRT, emphasizing its role in protecting respondent privacy. Begin and Boivin (1980) compared different survey methods, finding direct questionnaire and RRT data more valid. Clickner and Iglewicz (1980) extended Warner's randomized response framework by developing a model that simultaneously handles *two sensitive questions*, enabling efficient joint estimation of multiple sensitive traits. Eichhorn and Hayre (1983) introduced multiplicative RRT methods for quantitative responses. Singh et al. (1995) proposed an improved two-stage randomized response strategy for estimating proportions of sensitive attributes. Their method demonstrated greater efficiency than earlier RRT models in both theoretical and empirical evaluations.

Randomized response techniques (RRTs) encompass a wide array of models designed to protect respondent privacy while ensuring statistical validity. Classic models include Warner's original binary RRT, additive models such as those by Greenberg et al. (1969), cross-over models by Perri et al. (2016), and among others models allowing for misclassification or forced responses. Bar-Lev et al. (2004) introduced a generalized RR procedure that incorporates a design parameter—controlled by the experimenter—which expands upon the earlier work of Eichhorn and Hayre (1983). This framework not only enables a unified treatment of various RRTs but also allows for the optimization of estimator efficiency by reducing variance while maintaining privacy. Their approach demonstrates that, by appropriately choosing the design parameter, one can attain unbiased estimators with uniformly smaller variance than those of previous models. In the context of our proposed method, similar design flexibility can be explored by adjusting the distributional structure of the scrambling variables.

Christofides (2005) proposed a randomized response model using two independent randomization devices to estimate the prevalence of two sensitive characteristics, enabling unbiased joint estimation in dual-sensitive surveys. Kim and Elam (2007) proposed a stratified unrelated-question RRT model that integrates stratified sampling with unrelated-question randomization. Their model achieves lower variance and mean square error compared to its component models, enhancing efficiency under both truthful and imperfect reporting. Saha (2007) proposed a randomized response model using scrambling variables on quantitative sensitive data. This approach allows unbiased estimation of the population mean. Holbrook and Krosnick (2010) examined the validity of RRT for measuring voter turnout and found that many respondents did not follow the RRT instructions properly. Their findings raised concerns about the practical reliability of RRT in reducing misreporting in real-world surveys. Gupta et al. (2010) introduced an optional randomized response technique for quantitative data, allowing respondents to either report their true value or a scrambled response based on their sensitivity to the question. Diana and Perri (2010) introduced scrambled randomized response models for estimating the mean of a sensitive quantitative variable, using both multiplicative and additive scrambling approaches to balance respondent privacy and estimator efficiency. Ulrich et al. (2012) provided a statistical power analysis of various RRT techniques. Blair et al. (2015) provided a comprehensive review of RRT designs, including forced-response and unrelated-question models. Their work consolidated RRT methodology and enhanced its practical implementation. Gupta et al. (2018) proposed a unified measure of model quality for quantitative RRT techniques. Hsieh et al. (2018) proposed RRT model for multi-level categorical sensitive variables, where respondents report the absolute difference between their true category and a random integer. Singh et al. (2019) revisited the two-stage unrelated-question RRT model for rare sensitive attributes under Poisson population assumptions. Their method, which incorporates both simple random and stratified sampling, showed superior efficiency compared to previous two-stage models. Chu et al. (2020) focused on estimating mixed sensitive response types in RRT. Azeem et al. (2024b) proposed an efficient estimator of population variance of a sensitive variable with a new randomized response technique. Singh et al. (2021) proposed a quantitative randomized response model that uses scrambling variables and improves efficiency in estimating the population mean of sensitive variables. Ahmed and Shabbir (2023) addressed estimating sensitive subpopulation totals using RRT. Azeem et al. (2024a) introduced an enhanced quantitative RRT scrambling procedure that combines additive and multiplicative scramblers to improve both efficiency and respondent privacy. Their model demonstrated higher relative efficiency and a smaller joint privacy-efficiency metric than existing technique.

Several empirical studies have demonstrated the feasibility and success of implementing RRT techniques, including scrambled and mixed models, in real-world survey conditions (Shahzad et al., 2019, Zaman et al., 2023, 2024). For instance, Krumpal and Voss (2020) explored the influence of trust on respondent behavior under randomized response settings. Hsieh and Perri (2021) applied Christofides' model to sensitive demographic estimation. Perri et al. (2016) used a crossed-model RRT in a pilot survey to estimate both induced abortion among foreign women and irregular immigrant status. In the context of elite sports, Striegel et al. (2010) successfully utilized RRT to uncover the hidden prevalence of doping and drug use. These practical applications underline the adaptability and acceptance of RRT methods in complex field surveys and reinforce the viability of the proposed framework.

3. Method and Proposed Techniques

The scrambling technique has been used as an effective mechanism for increasing the cooperation of the respondents. Pollock and Bek (1976) proposed the additive model and another survey model with a quantitative response variable was taken into consideration by Eichhorn and Hayre (1983) who suggested an RR method for it. These techniques are especially helpful in surveys when a highly sensitive measured response variable is used. The multiplicative model that Eichhorn and Hayre (1983) considered involves the respondent multiplying his responses to the sensitive question by a random integer drawn from a predetermined distribution. Inspired by this technique Singh (1976) proposed techniques involving scrambled variables that can be used for dealing with highly sensitive issues such as how many abortions a woman has undergone, the number of drugs consumed by a teenager, etc. While thinking along these lines, in this paper we proposed two scrambled response techniques that will work more efficiently than the existing techniques in different situations. The technique's quality is measured by percent relative efficiency and measure of privacy protection.

A pre-survey demonstration suggested including sample exercises and visual aids, to guide respondents step-by-step. In addition, we suggest conducting a pilot survey before full deployment to assess comprehension, revise language instruction, and make context-specific adaptations if needed. Importantly, all arithmetic involved in the response mechanism is elementary in nature (basic addition, multiplication), ensuring that the cognitive load remains manageable even for respondents with limited formal education. Similar approaches using instruction cards and training aids have proven effective in earlier applications of scrambled response techniques (Saha, 2011; Singh and Gorey, 2017), supporting the feasibility of our proposed implementation strategy.

Suppose $Y \geq 0$ be a quantitative sensitive variable with unknown mean (μ_Y) and variance (σ^2_Y) respectively from a finite population with size N. Let W_1, W_2, W_3, S and U be the scrambling variables considered which is independent of the response variable Y with means μ_{W1}, μ_{W2}, μ_{W3}, μ_S and μ_U and variances σ^2_{W1}, σ^2_{W2}, σ^2_{W3}, σ^2_S, σ^2_U respectively. A simple random sampling with replacement strategy was used to select a random sample of n respondents from the population. Further, a random device having three types of cards containing statements would be provided for each selected respondent in the sample. The three types of cards with statements are:

1) Green cards: report the original sensitive variable Y
2) Red cards: report the scrambled response YW_1+W_2
3) Yellow cards: report the scrambled values $Z=\alpha\,(Y+U)+(1-\alpha)\,Y\,S$ (for technique 1),
$Z=W_3\,(\alpha Y+(1-\alpha)\,Y\,S)$ (for technique 2)

with probabilities P_1, P_2 and P_3 respectively such that $\Sigma^3_{i=1} P_i = 1$.

3.1 Technique 1

Suppose the response M_1 with distribution is given as:

$$M_1 = \begin{Bmatrix} Y \text{ with prob } P_1 \\ YW_1 + W_2 \text{ with prob } P_2 \\ \alpha(Y+U)+(1-\alpha)YS \text{ with prob } P_3 \end{Bmatrix}.$$

Hence, for the observed reported response

$$M_1 = Y\,(P_1 + P_2W_1 + \alpha P_3 + SP_3 - \alpha SP_3) + P_2W_2 + \alpha UP_3,$$

the mean and variance of quantitative sensitive variable Y are given by:

$$\mu\widehat{Y_1} = \frac{\overline{M_1} - P_2\mu_{W2} - P_3\alpha\mu_U}{P_1+P_2\mu_{W1}+P_3(\alpha+\mu_S-\alpha\mu_S)},$$

and

$$\mathrm{v}(\mu\hat{Y}1) = \frac{\sigma^2_{M1}}{\mathrm{n}[\mathrm{P}_1+\mathrm{P}_2\mu_{\mathrm{W}1}+\mathrm{P}_3(\alpha+\mu_{\mathrm{S}}-\alpha\mu_{\mathrm{S}})]^2}.$$

where,

$$\sigma^2_{M1} = P_1(\sigma^2_y+\mu^2_y) + P_2\{(\sigma^2_y+\mu^2_y)(\sigma^2_{W1}+\mu^2_{W1}) + (\sigma^2_{W2}+\mu^2_{W2})\} + P_3\,\{(\alpha((\sigma^2_y+\mu^2_y)+(\sigma^2_U+\mu^2_U))) + (1-\alpha)(\sigma^2_y+\mu^2_y)(\sigma^2_s+\mu^2_s)\} - \{P_1\mu_y + P_2(\mu_y\mu_{W1}+\mu_{W2}) + P_3\,(\alpha(\mu_y+\mu_U)+(1-\alpha)\mu_y\mu_s)\}^2.$$

3.2 Technique 2

Suppose the response M_2 with distribution is given as:

$$M_2 = \begin{Bmatrix} Y \text{ with prob } P_1 \\ YW_1 + W_2 \text{ with prob } P_2 \\ W_3[\alpha S+(1-\alpha)Y] \text{ with prob } P_3 \end{Bmatrix}.$$

Hence, for the observed reported response

$$M_2 = YP_1 + [YW_1 + W_2]\,P_2 + W_3\,[\alpha S + (1-\alpha)\,Y]\,P_3,$$

the mean sensitive variable Y is:

$$\mu\widehat{Y_2} = \frac{\overline{M_2} - P_2\mu_{W2} - \alpha P_3\mu_W\mu_S}{P_1+P_2\mu_{W1}+P_3(1-\alpha)\mu_{W3}},$$

and the variance of sensitive variable as:

$$\mathrm{v}(\mu\hat{\mathrm{Y}}2) = \frac{\sigma^2_{M2}}{\mathrm{n}[\mathrm{P}_1+\mathrm{P}_2\mu_{\mathrm{W}1}+\mathrm{P}_3((1-\alpha)\mu_{W3})]^2}.$$

where,

$$\sigma^2_{M2} = P_1(\sigma^2_y+\mu^2_y) + P_2\{(\sigma^2_y+\mu^2_y)(\sigma^2_{W1}+\mu^2_{W1}) + (\sigma^2_{W2}+\mu^2_{W2})\} + P_3\,(\sigma^2_{W3}+\mu^2_{W3})\left\{\left(\alpha(\sigma^2_S+\mu^2_S) + (1-\alpha)(\sigma^2_y\mu^2_y)\right)\right\} - \{P_1\mu_y + P_2(\mu_y\mu_{W1}+\mu_{W2}) + P_3\,(\alpha(\mu_s+(1-\alpha)\mu_Y)\}^2.$$

4. Analysis and Discussion

The degree of privacy protection plays a crucial role in the effectiveness of the Randomized Response Technique (RRT). When dealing with sensitive topics, ensuring the respondent's privacy is essential for obtaining honest and unbiased responses. If respondents feel their privacy is at risk, they may provide socially desirable answers rather than truthful ones, leading to biased estimates and reduced data reliability. Therefore, the success of RRT largely depends on the level of confidentiality it can guarantee.

The normalized measure of privacy protection introduced by Diana and Perri (2010) has become a standard benchmark. When this measure is closer to 1, it indicates a higher level of privacy protection, which in turn fosters greater respondent cooperation. Conversely, as the value nears 0, the level of privacy protection diminishes, making respondents more hesitant and less willing to provide truthful information.

In addition, Holbrook and Krosnick (2010) critically evaluated the practical implementation of RRT and found that inadequate comprehension of the protection mechanism by respondents can reduce its effectiveness, further emphasizing the importance of a design that is both mathematically sound and easy to understand. Christofides (2005) and Kim and Elam (2007) also proposed advanced randomized response models that incorporate privacy protection metrics alongside efficiency measures, thereby formalizing the now well-recognized privacy-efficiency trade-off.

Blair et al. (2015) provided a comprehensive review of RRT models and emphasized that privacy should be evaluated not only theoretically but also behaviorally, by observing whether respondents follow the instructions correctly. Their work strengthens the case for including respondent comprehension and practical usability as part of privacy assessment. Hsieh et al. (2018) extended this by proposing a model for multi-level sensitive traits using a single randomization device and demonstrated how privacy metrics can be preserved even when categorical data is involved, widening the applicability of RRT. Perri et al. (2016) demonstrated the application of a crossed-RRT model to jointly estimate abortion prevalence and irregular foreign presence, showing how well-designed privacy mechanisms can be used in high-stakes field applications.

Contributions, such as Singh et al. (2021), and Zhimin and Zaizai (2012), have continued this line of work by introducing scrambling-based models with strong privacy guarantees, showing how design choices impact both privacy and estimator performance. These works reinforce that privacy is not only a technical requirement but also a driver of respondent trust and data quality.

In our proposed model, the degree of privacy protection has been carefully evaluated using the normalized measure framework, ensuring that the technique meets the dual objectives of maintaining confidentiality and producing statistically efficient estimators. The design ensures that the privacy level remains high across varying parameter choices, demonstrating its practical viability for surveys involving sensitive quantitative variables.

4.1 Privacy Measure of Proposed Techniques

Based on the above works, the squared correlation coefficient for Technique 1 and Technique 2 are as follows:

$$S_1 = 1 - \frac{(P_1+P_2\mu_{W1}+P_3\alpha+P_3S-\alpha SP_3)^2}{\sigma_{M1}^2},$$

$$S_2 = 1 - \frac{(P_1+P_2\mu_{W1}+(1-\alpha)P_3\mu_{W3})^2}{\sigma_{M2}^2}.$$

To show the measure of privacy protection of the proposed models, we followed the normalized measure of privacy protection technique proposed by Diana and Perri (2010). We have mentioned the values for the measure of privacy protection for both the models in a graphical representation **Figures 1** to **4** as follows.

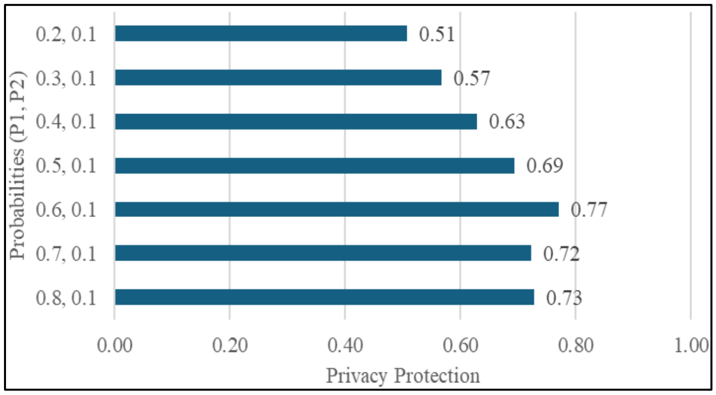

Figure 1. Measure of privacy protection (Model 1, Data A).

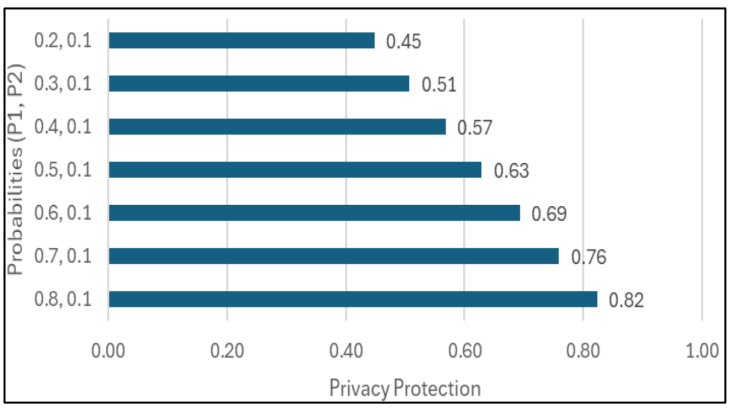

Figure 2. Measure of privacy protection (Model 1, Data B).

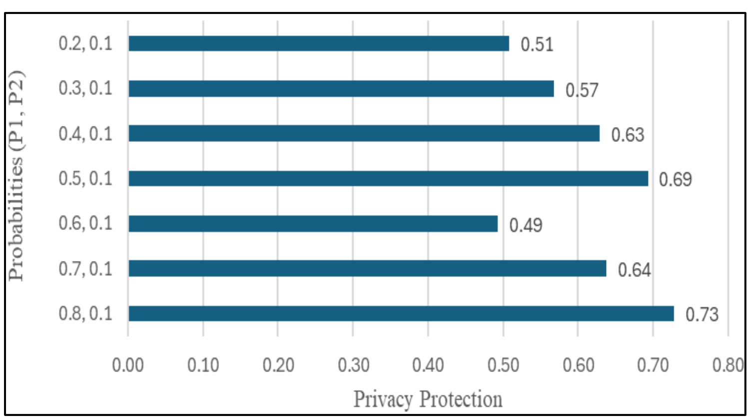

Figure 3. Measure of privacy protection (Model 2, Data A).

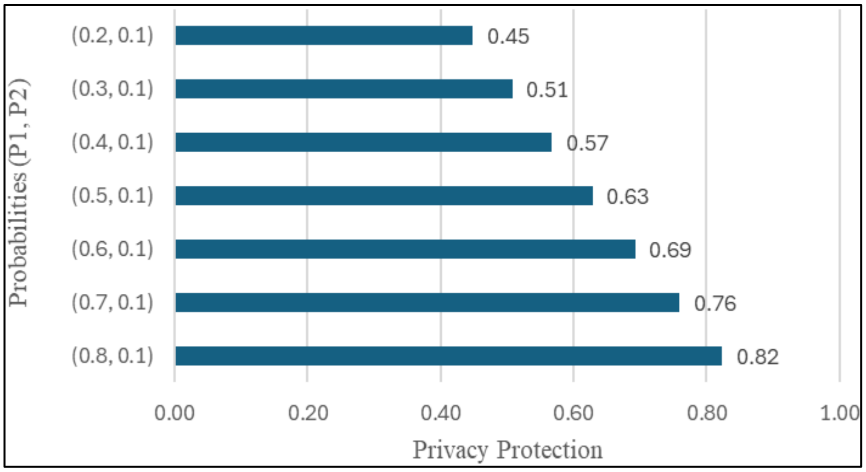

Figure 4. Measure of privacy protection (Model 2, Data B).

4.2. Efficiency Comparison

We have compared the performance of the proposed techniques to that of the Bar-Lev et al. (2004) technique in terms of PRE. For PRE (Percent Relative Efficiency) calculation, we used the following formula and assumed that the scrambling variables follow a Poisson distribution.

$$PRE = \frac{V(\mu\hat{Y}BL)}{V(\mu\hat{Y}i)} \times 100 \;\; for\; i = 1,2.$$

Table 1 represents the data sets used for empirical comparisons. The constant α can take any value between 0 and 1. For the empirical study, here we considered the value of α as 0.05.

Table 1. Data for efficiency comparison.

Set	Y	W_1	W_2	W_3	S	U	Z
A	4	3	5	4	2	3	4
B	7	4	8	5	2	1	4

Table 2. PRE of the suggested technique 1 for data set A with respect to Bar-Lev et al. (2004).

Probability		PRE of technique	Probability		PRE of technique
P1	P2	Technique 1	P1	P2	Technique 1
0.9	0.1	164.96	0.4	0.1	108.76
				0.2	119.95
				0.3	137.14
				0.4	162.41
				0.5	201.22
				0.6	266.37
0.8	0.1	166.62	0.3	0.2	108.75
	0.2	168.35		0.3	125.45
				0.4	149.86
				0.5	187.55
				0.6	251.83
				0.7	383.54

Table 2 continued…

0.7	0.1 0.2 0.3	151.67 159.22 174.96	0.2	0.2 0.3 0.4 0.5 0.6 0.7 0.8	114.66 138.14 149.86 174.74 238.45 374.86 864.99
0.6	0.1 0.2 0.3 0.4	135.48 145.76 162.88 189.12	0.1	0.3 0.4 0.5 0.5 0.6 0.7 0.8	104.61 127.14 162.7 174.74 226.24 369.91 993.30
0.5	0.1 0.2 0.3 0.4 0.5	120.90 132.32 149.75 175.70 215.57			

Table 3. PRE of the suggested technique 2 for data set A with respect to Bar-Lev et al. (2004).

Probability		PRE of technique	Probability		PRE of technique
P1	P2	Technique 2	P1	P1	Technique 2
0.9	0.1	164.96	0.4	0.1 0.2 0.3 0.4 0.5 0.6	300.69 282.93 267.17 255.01 250.34 266.37
0.8	0.1 0.2	190.25 168.35	0.3	0.2 0.3 0.4 0.5 0.6 0.7	317.74 304.29 293.40 285.89 286.17 304.16 383.54
0.7	0.1 0.2 0.3	225.22 197.79 174.96	0.2	0.2 0.3 0.4 0.5 0.6 0.7 0.8	331.95 322.97 316.77 315.18 322.09 347.51 427.44 864.99
0.6	0.1 0.2 0.3 0.4	255.27 229.81 206.69 189.12	0.1	0.3 0.4 0.5 0.5 0.6 0.7 0.8	339.51 338.70 343.61 358.87 395.96 493.49 993.30
0.5	0.1 0.2 0.3 0.4 0.5	280.14 258.28 238.35 222.24 215.57			

Table 4. PRE of the suggested technique 1 for data set B with respect to Bar-Lev et al. (2004).

Probability		PRE of technique	Probability		PRE of technique
P1	P2	Technique 1	P1	P1	Technique 1
0.9	0.1	118.58	0.4	0.1 0.2 0.3 0.4 0.5 0.6	207.90 203.69 201.70 203.50 212.59 238.85
0.8	0.1 0.2	132.72 127.47	0.3	0.1 0.2 0.3 0.4 0.5 0.6 0.7	221.78 220.67 222.24 228.25 242.46 274.61 359.88
0.7	0.1 0.2 0.3	155.16 145.54 139.78	0.2	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8	234.85 237.40 243.41 255.20 277.52 322.62 433.49 946.04
0.6	0.1 0.2 0.3 0.4	175.20 166.32 159.74 157.93	0.1	0.2 0.3 0.4 0.5 0.6 0.7 0.8	254.80 266.6 286.60 321.78 391.15 567.75 1646.16
0.5	0.1 0.2 0.3 0.4 0.5	192.59 185.75 180.96 179.89 186.76			

Table 5. PRE of the suggested technique 2 for data set B with respect to Bar-Lev et al. (2004).

Probability		PRE of technique	Probability		PRE of technique
P1	P2	Technique 2	P1	P2	Technique 2
0.9	0.1	118.58	0.4	0.1 0.2 0.3 0.4 0.5 0.6	106.04 120.59 143.48 179.24 238.85
0.8	0.1 0.2	134.30 127.47	0.3	0.1 0.2 0.3 0.4 0.5 0.6 0.7	111.76 134.62 170.71 232.74 359.88
0.7	0.1 0.2 0.3	129.65 128.89 139.78	0.2	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8	102.79 125.37 161.65 226.39 370.12 946.04

Table 5 continued…

0.6	0.1 0.2 0.3 0.4	119.75 123.6 136.00 157.93	0.1	0.2 0.3 0.4 0.5 0.6 0.7 0.8	115.88 152.22 220.13 387.39 1424.65
0.5	0.1 0.2 0.3 0.4 0.5	108.90 114.96 128.94 151.58 186.76			

Key insights from **Tables 2** to **5** and **Figures 1** to **4** are as follows:

1) The percent relative efficiencies of proposed techniques in comparison to the Bar-Lev et al. (2004) technique are always more than 100, as shown in **Tables 2** to **5**, indicating that the proposed techniques predominate over the Bar–Lev et al. (2004) technique.
2) From **Figures 1** to **4** it is visible that most of the values exceed 0.5 and are closer to 1. This indicates a higher level of privacy protection, which in turn suggests that respondents are more likely to cooperate with the proposed techniques.

5. Conclusion

The proposed randomized response techniques exhibit strong performance on two critical fronts, statistical efficiency and privacy protection, making them highly effective tools for estimating the population mean of quantitatively sensitive characteristics. The design leverages multiple types of scrambling cards involving additive, multiplicative, and mixed operations, which not only enhance statistical randomness but also strengthen perceived respondent anonymity. This nuanced structure reduces the likelihood of respondent discomfort and social desirability bias, factors often detrimental to data integrity in sensitive surveys.

From a privacy standpoint, the models offer a high degree of statistical confidentiality, aligning with contemporary privacy metrics such as the normalized measures proposed by Diana and Perri (2010). To ensure optimal effectiveness, however, it remains essential that researchers implement informed consent protocols, provide clear and respondent-friendly instructions, and maintain transparency in explaining the randomized nature of the technique. Ethical implementation is as important as statistical robustness, especially when dealing with sensitive data related to behaviors such as tax evasion, abortion, or criminal activity.

To illustrate the practical applicability of the proposed methods, we employed Poisson-distributed scrambling variables, which are particularly well-suited for modeling count-based and rare event data—frequently encountered in sensitive contexts. Empirical comparisons clearly indicate that Technique 1 and Technique 2 consistently outperform the benchmark Bar-Lev et al. (2004) method, with PRE values exceeding 100 across diverse probability configurations. This demonstrates not only improved accuracy but also the robustness and generalizability of the proposed methods. Overall, these techniques offer a statistically sound, ethically mindful, and practically versatile framework for randomized response surveys in sensitive research domains.

6. Implications

The implications of these results are significant for survey practitioners, offering actionable insights into enhancing the reliability of survey outcomes involving sensitive topics. By adopting the proposed randomized response techniques, practitioners can mitigate non-response bias and response inaccuracies inherent in conventional survey methodologies. Moreover, the compatible outperformance of the proposed techniques underscores their possibility for real-world applications, providing researchers with robust tools for estimating sensitive population means with greater precision and accuracy. Overall, the findings advocate for the widespread adoption of these innovative techniques in survey research, empowering practitioners to address the dare posed by sensitive survey topics while ensuring the integrity and representativeness of survey data.

7. Limitation & Future Research

The present study assumes simple random sampling with replacement, which is grounded in the assumption of a homogeneous population. In future research, we intend to extend our current framework to incorporate heterogeneous populations and analyze the performance of the proposed estimation techniques under more complex sampling designs. Further, the formal mathematical approaches used for scrambling (such as cards with mixed operations) might result in respondent confusion which could lead to non-compliance. However, to mitigate the risk of confusion, we suggest simplifying instructions wherever possible, using interviewer-assisted techniques (i.e., having trained personnel explain the procedure and guide respondents through an example), and pre-testing the survey instrument in a pilot study to ensure clarity. Additionally, graphical aids or digitized randomization tools (such as mobile apps or computer-assisted randomization) may improve compliance while preserving the integrity of the scrambling method. In future research, we will also focus more deeply on assessing respondents' psychological comfort, and perceived risk, possibly using empirical data from field trials and debriefing surveys.

Conflicts of Interest

No conflicts of interest have been associated with this publication.

Acknowledgments

We thank the anonymous reviewers for their valuable and enriching comments.

AI Disclosure

During the preparation of this work the author(s) used generative AI in order to improve the language of the article. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Ahmed, S., & Shabbir, J. (2023). On use of randomized response technique for estimating sensitive subpopulation total. *Communications in Statistics-Theory and Methods*, *52*(5), 1417-1430. https://doi.org/10.1080/03610926.2021.1928199.

Azeem, M., Hussain, S., Ijaz, M., & Salahuddin, N. (2024a). An improved quantitative randomized response technique for data collection in sensitive surveys. *Quality & Quantity*, *58*(1), 329-341. https://doi.org/10.1007/s11135-023-01652-5.

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

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.

Begin, G., & Boivin, M. (1980). Comparison of data gathered on sensitive questions via direct questionnaire, randomized response technique, and a projective method. *Psychological Reports*, *47*(3), 743-750. https://doi.org/10.2466/pr0.1980.47.3.743.

Blair, G., Imai, K., & Zhou, Y.Y. (2015). Design and analysis of the randomized response technique. *Journal of the American Statistical Association*, *110*(511), 1304-1319. https://doi.org/10.1080/01621459.2015.1050028.

Christofides, T.C. (2005). Randomized response technique for two sensitive characteristics at the same time. *Metrika*, *62*, 53-63. https://doi.org/10.1007/s001840400355.

Chu, A.M., So, M.K., Chan, T.W., & Tiwari, A. (2020). Estimating the dependence of mixed sensitive response types in randomized response technique. *Statistical Methods in Medical Research*, *29*(3), 894-910. https://doi.org/10.1177/0962280219847492.

Clickner, R.P., & Iglewicz, B. (1980). Warner's randomized response technique: the two sensitive questions case. *South African Statistical Journal*, *14*(2), 77-86. https://hdl.handle.net/10520/aja0038271x_664.

Diana, G., & Perri, P.F. (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, *37*(11), 1875-1890. https://doi.org/10.1080/02664760903186031.

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.

Goodstadt, M.S., & Gruson, V. (1975). The randomized response technique: a test on drug use. *Journal of the American Statistical Association*, *70*(352), 814-818. https://doi.org/10.1080/01621459.1975.10480307.

Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R., & Horvitz, D.G. (1969). The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association, 64*(326), 520-539. https://doi.org/10.1080/01621459.1969.10500991.

Greenberg, B.G., Kuebler Jr, R.R., Abernathy, J.R., & Horvitz, D.G. (1971). Application of the randomized response technique in obtaining quantitative data. *Journal of the American Statistical Association*, *66*(334), 243-250. https://doi.org/10.1080/01621459.1971.10482248.

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.

Gupta, S., Shabbir, J., & Sehra, S. (2010). Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference*, *140*(10), 2870-2874. https://doi.org/10.1016/j.jspi.2010.03.010.

Holbrook, A.L., & Krosnick, J.A. (2010). Measuring voter turnout by using the randomized response technique: evidence calling into question the method’s validity. *Public Opinion Quarterly*, *74*(2), 328-343.

Horvitz, D.G., Greenberg, B.G., & Abernathy, J.R. (1976). Randomized response: A data-gathering device for sensitive questions. *International Statistical Review/Revue Internationale de Statistique*, *44*(2), 181-196.

Hsieh, S.H., & Perri, P.F. (2021). Estimating the proportion of non-heterosexuals in Taiwan using Christofides’ randomized response model: a comparison of different estimation methods. *Social Science Research*, *93*, 102475.

Hsieh, S.H., Lee, S.M., & Tu, S.H. (2018). Randomized response techniques for a multi-level attribute using a single sensitive question. *Statistical Papers*, *59*, 291-306. https://doi.org/10.1007/s00362-016-0764-9.

Kim, J.M., & Elam, M.E. (2007). A stratified unrelated question randomized response model. *Statistical Papers*, *48*, 215-233. https://doi.org/10.1007/s00362-006-0327-6.

Krumpal, I., & Voss, T. (2020). Sensitive questions and trust: explaining respondents’ behavior in randomized response surveys. *SAGE Open*, *10*(3), 2158244020936223. https://doi.org/10.1177/2158244020936223.

Perri, P.F., Pelle, E., & Stranges, M. (2016). Estimating induced abortion and foreign irregular presence using the randomized response crossed model. *Social Indicators Research*, *129*, 601-618. https://doi.org/10.1007/s11205-015-1136-x.

Pollock, K.H., & Bek, Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association*, *71*(356), 884-886.

Saha, A. (2007). A simple randomized response technique in complex surveys. *Metron*, *65*(1), 59-66.

Saha, A. (2011). An optional scrambled randomized response technique for practical surveys. *Metrika*, *73*(2), 139-149. https://doi.org/10.1007/s00184-009-0269-9.

Shahzad, U., Perri, P.F., & Hanif, M. (2019). A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information. *Communications in Statistics-Simulation and Computation*, *48*(9), 2566-2585. https://doi.org/10.1080/03610918.2018.1458133.

Singh, C., Kamal, M., Singh, G.N., & Kim, J.M. (2021). Study to alter the nuisance effect of non-response using scrambled mechanism. *Risk Management and Healthcare Policy*, 1595-1613.

Singh, G.N., Singh, C., & Suman, S. (2019). Revisit of a randomized response model for estimating a rare sensitive attribute under probability proportional to size sampling using Poisson probability distribution. *Communications in Statistics-Theory and Methods*, *48*(11), 2766-2786. https://doi.org/10.1080/03610926.2018.1508718.

Singh, H.P., & Gorey, S.M. (2017). Efficient estimation of population mean of sensitive variable in presence of scrambled responses. *Communications in Statistics-Theory and Methods*, *46*(19), 9557-9565.

Singh, J. (1976). A note on the randomized response technique. In *Proceedings of the Social Statistics Section, American Statistical Association* (Vol. 772). American Statistical Association. Boston, USA.

Singh, R., Singh, S., Mangat, N.S., & Tracy, D.S. (1995). An improved two stage randomized response strategy. *Statistical Papers*, *36*, 265-271. https://doi.org/10.1007/bf02926038.

Sjöström, O., & Holst, D. (2002). Validity of a questionnaire survey: response patterns in different subgroups and the effect of social desirability. *Acta Odontologica Scandinavica*, *60*(3), 136-140.

Striegel, H., Ulrich, R., & Simon, P. (2010). Randomized response estimates for doping and illicit drug use in elite athletes. *Drug and Alcohol Dependence*, *106*(2-3), 230-232. https://doi.org/10.1016/j.drugalcdep.2009.07.026.

Ulrich, R., Schröter, H., Striegel, H., & Simon, P. (2012). Asking sensitive questions: a statistical power analysis of randomized response models. *Psychological Methods*, *17*(4), 623-641.

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.

Zaman, Q., Ijaz, M., & Zaman, T. (2023). A randomization tool for obtaining efficient estimators through focus group discussion in sensitive surveys. *Communications in Statistics-Theory and Methods*, *52*(10), 3414-3428.

Zaman, T., Shazad, U., & Yadav, V.K. (2024). An efficient Hartley–Ross type estimators of nonsensitive and sensitive variables using robust regression methods in sample surveys. *Journal of Computational and Applied Mathematics*, *440*, 115645.

Zhimin, H., & Zaizai, Y. (2012). Measure of privacy in randomized response model. *Quality & Quantity*, *46*, 1167-1180. https://doi.org/10.1007/s11135-012-9698-z.

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.