The Future of Sensitive Data Estimation: Privacy-Preserving Dual Response Models for Reliable Results

Han Wanqu[1], Muhammad Imran Shahid[1*], Faiza Shah[1*] Yasir Shah[2], Fadia Shah[3]

[1]College of Political Science and Public Administration, Henan Normal University, Xinxiang, 453000, China

[2]School of Business, Zhengzhou University, Zhengzhou, 450000, China

[3]Department of Computer Science, SZABIST University, Islamabad, 44000, Pakistan

Corresponding authors:

Muhammad Imran Shahid *

Faiza Shah*

Abstract

This research addresses the methodological challenges of reliably assessing stigmatized quantitative variables while protecting respondents' privacy and data integrity. We present innovative dual scrambling randomized response framework (DSRRF-I) which utilize additive and subtractive scrambling algorithms for optimizing both efficiency and privacy protection. By using dual scrambling variables and selecting respondents through simple random sampling with replacement (SRSWR), the models yield unbiased, precise, and meticulous estimates of means and sensitivity levels without increasing respondent burden. The mathematical derivations and thorough simulations demonstrate considerable efficiency enhancements over standard randomized response strategies. Furthermore, the proposed models adapt seamlessly to various data collection environments, ensuring robustness across diverse settings. Our findings underscore the potential of these models to transform the landscape of sensitive data collection, providing more reliable estimates in public health and social research. The flexibility of DSRRFs-I also paves the way for future advancements in privacy-preserving survey techniques, offering a versatile tool for tackling complex measurement issues.

Keywords: Stigmatized Quantitative Variables, Privacy Protection, Efficiency Enhancement, Survey Techniques, Sensitive Data Collection.

1 Introduction

Reliable measurement of stigmatized variables is a critical challenge in many research fields, including public health and social sciences. Respondents' reluctance to disclose sensitive information can lead to biases, underreporting, and compromised data integrity. Traditional methods, such as direct self-reporting, often fail to safeguard privacy and are prone to social desirability bias. This study introduces innovative privacy-preserving techniques, combining randomized response models with advanced scrambling algorithms, to ensure unbiased and trustworthy estimates of sensitive data while maintaining respondent confidentiality. In sample surveys, researchers frequently encountered significant challenges due to high refusal rates and false responses when collecting information on sensitive traits. Such sensitive characteristics included cheating during examinations, undeclared income, tax evasion, and induced abortions. In order to solve these issues, the ambiguous response approach was first proposed by [16], and [13] gave rough estimates of variance for previous techniques. Moreover, [6] formulated forced empirical models, and [17] devised simultaneous average and sensitivity level estimations, subsequently refined by [7] via an additive scrambling framework. The works of [9, 15], [18, 19, 27, 33], and [20, 28] have collectively enhanced the methodologies for estimating sensitive quantitative variables.

In view of these improvements, subsequent advances in randomized response algorithms have greatly addressed issues such as bias regarding responses and estimate efficiency. Kumar and Singh [23, 35] presented cutting-edge methods for eliminating bias between responses through randomized scattering techniques, whereas Gupta et al. [12] developed an aggregate variance estimator that utilizes Diana and Perri's randomization strategy [5] Furthermore, Saleem et al. [30] presented improved estimators for population means.

Based on earlier research, Azeem [3] recently evolved under discussion with a weighted metric that combines privacy along with efficiency. This gives us a complete way to judge how well randomized response approaches function. Siddiqui et al. [32] contributed additional progress in the area by analyzing mathematical models that take into account correlations between variables. Azeem [2], Bhat [4], Gupta [11], and Zahid et al. [37] also made important contributions. Consequently, from a mathematical perspective, the restructuring improves variance behavior and enhances PREs performance across varying sensitivity levels Shahid et al. [34].

We acknowledged the complexity of the challenge and envisioned innovative solutions to overcome it by presenting an approach known as DSRRF-I. Motivated by the methodology of Huang [18], we proposed DSRRF-I designed to acquire reliable information while preserving respondent anonymity. The novelty of the present study lay in employing DSRRF-I to estimate both the mean and sensitivity levels through a two-fold response approach, wherein respondents used two scrambling variables, P_i and Q_i, for each response. A key advantage of this methodology was that collecting the second response did not incur additional costs. This two-fold response strategy enhanced estimation reliability without compromising the integrity of the study. Moreover, the approach minimizes response bias, further strengthening the validity of the estimates. It also ensures that sensitive data is collected in a way that is both ethical and effective, contributing to the growing body of privacy-preserving survey methods. The application of this method holds great promise for a wide range of research fields requiring high levels of confidentiality and accurate estimations.

2 A methodical strategies for precision and confidentiality

In this section, we discuss the background theory of the proposed strategies and present expressions for unbiased estimators and their variances. We propose two additive and subtractive DSRRFs to estimate the mean μ_R and sensitivity level W of the stigmatizing variable R using a double response approach. In the scrambling procedure, two variables, P_i and Q_i, are added and subtracted from the actual response R in two separate steps. This method aims to influence respondents' perception, leading them to provide more reliable data on sensitive variables by balancing over-reporting and under-reporting tendencies. The process ensures greater privacy protection, improving the reliability of data collection. The extra burden of responding twice is addressed by informing interviewees beforehand that their responses are anonymized and that parameter estimation will not trace back to individual responses. Furthermore, interviewees would be clear about the estimation of the parameters in our study of interest.

2.1 Randomized response inquiry on DSRRF-I

To operationalize this conceptual framework, we now detail the prescribed response strategy, which forms the core mechanism enabling randomized yet informative data collection within DSRRF-I. The proposed model addresses these issues by introducing the simultaneous addition and subtraction of scrambling variables to the actual response in the scrambled response. The strategy involves selecting respondents from a sample of size n using simple random sampling with replacement (SRSWR) and requesting them to provide two responses. In the first response, if the respondent considers the question sensitive, they provide a scrambled response. The scrambling procedure entails providing each respondent with two randomization devices that generate random numbers, denoted as P_1 and Q_1, from pre-assigned Poisson distributions. The scrambling variables P_1 and Q_1 are then simultaneously added to the true response R. If the respondent does not consider the question sensitive, they provide the true response R. In the second response, the same procedure is applied, with the difference being that the respondent subtracts the scrambling variables P_2 and Q_2 from the actual response R. The two responses, V_1 and V_2, are expressed as:

$$V_1 = R(1-\alpha) + \alpha(R + P_1 + Q_1), \tag{1}$$

$$V_2 = R(1-\alpha) + \alpha(R - P_2 - Q_2). \tag{2}$$

As shown in Fig. 1, the structural mechanism of DSRRF-I during the randomized response process is depicted. With the response mechanism clearly defined, the next step is to derive unbiased and efficient estimators that utilize the randomized responses to accurately infer parameters, ensuring proper connectivity in the analysis.

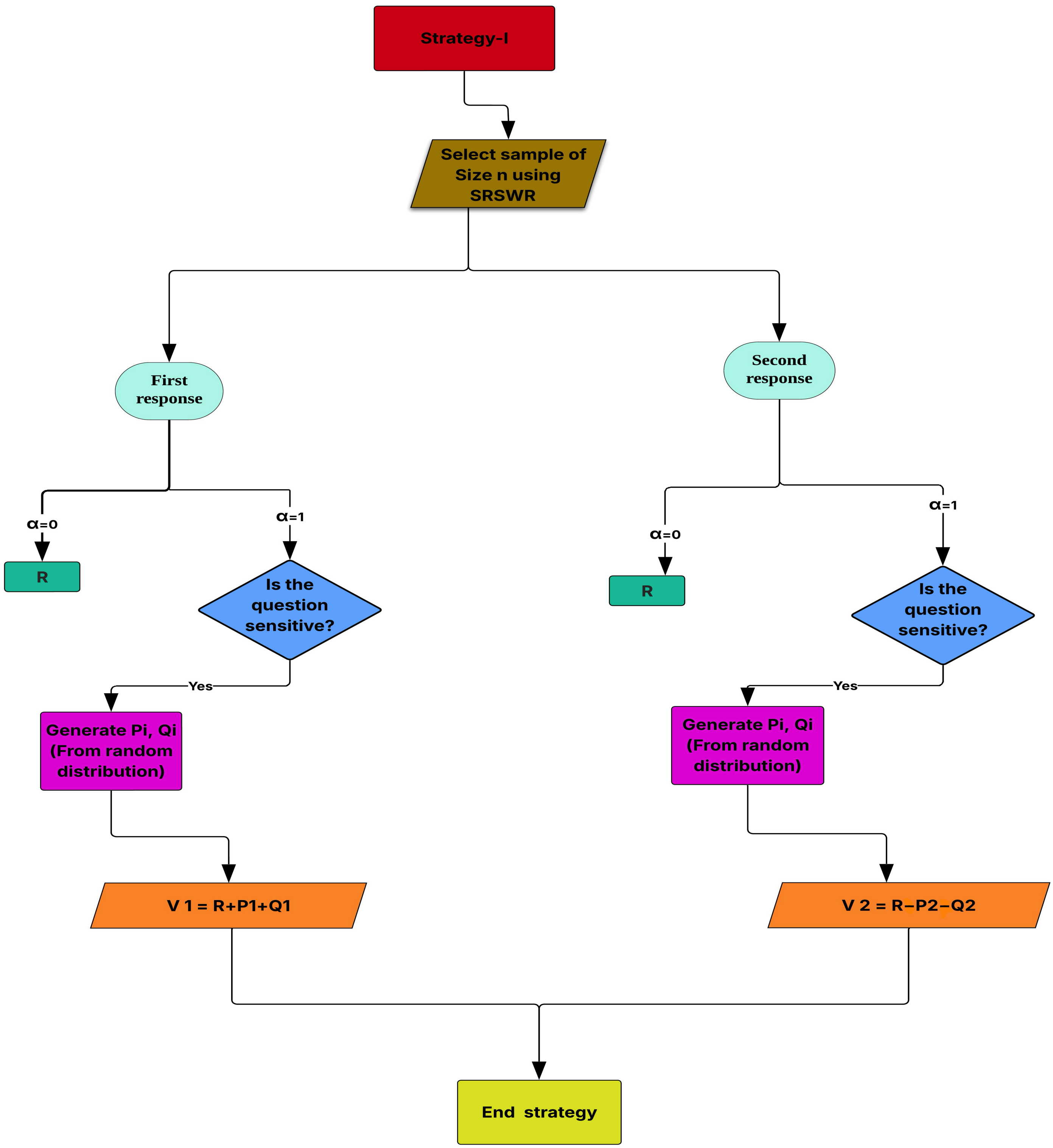

Figure 1: The structural mechanism for DSRRF-I in the process of randomized response.

Unbiasedness and dispersion of estimators for DSRRF-I

We discuss the structural aspects that could influence the unbiased and efficient estimators (UEE-I) and determine their precise manifestation. The UEE-I employs a randomized response technique to ensure that the expected value of the scrambled response closely approximates the true parameter, thereby minimizing variance and protecting the confidentiality of responders. The unbiased and efficient estimators and the essential occurrences of spanning clusters must be identified. The expected first and second responses, $E(V_1)$ and $E(V_2)$, in a sample of size n are given by

$$E(V_1) = (1 - W)\mu_R + W(\mu_R + \mu_1 + 1), \tag{3}$$

$$E(V_2) = (1 - W)\mu_R + W(\mu_R - \mu_2 - 1). \tag{4}$$

The unbiased estimators of μ_I and W_I are provided by the following expressions:

$$\widehat{\mu}_I = \frac{\overline{V}_2(\mu_1 + 1) + \overline{V}_1(\mu_2 + 1)}{\mu_1 + \mu_2 + 2}, \quad \text{under condition} \quad \mu_1 + \mu_2 \neq -2, \tag{5}$$

$$\widehat{W}_I = \frac{\overline{V}_2 - \overline{V}_1}{\mu_1 + \mu_2 + 2}, \quad \text{under condition} \quad \mu_1 + \mu_2 \neq -2. \tag{6}$$

Having established the estimators' unbiasedness, it is imperative to examine their statistical properties in depth by analyzing the dispersion and variance components that influence estimator precision. The dispersion of unbiased estimators (DUE-I) in the randomized response method minimizes variance, thereby enhancing the efficiency and reliability of the estimator. The variances of these estimators, $\widehat{\mu}_I$ and $\widehat{W}_I$, are expressed as follows:

$$\begin{aligned} \text{Var}(\widehat{\mu}_I) &= \frac{1}{(\mu_1 + \mu_2 + 2)^2}\left[(\mu_2 + 1)^2\left(\frac{\sigma^2_{V_1}}{n}\right) + (\mu_1 + 1)^2\left(\frac{\sigma^2_{V_2}}{n}\right)\right] \\ &+ \frac{1}{(\mu_1 + \mu_2 + 2)^2}\left[2\frac{(\mu_1 + 1)(\mu_2 + 1)\text{Cov}(\overline{V}_1, \overline{V}_2)}{n}\right], \end{aligned} \tag{7}$$

$$\text{Var}(\widehat{W}_I) = \frac{1}{(\mu_1 + \mu_2 + 2)^2}\left[\left(\frac{\sigma^2_{V_1}}{n}\right) + \left(\frac{\sigma^2_{V_2}}{n}\right) - 2\frac{\text{Cov}(\overline{V}_1, \overline{V}_2)}{n}\right], \tag{8}$$

as part of the subsequent analysis, the expressions (7) and (8) are essential for determining the variance between the first and second responses zare given by

$$\text{Var}(\overline{V}_1) = \frac{1}{n}\left\{\sigma^2_R + W(\delta^2_1 + \gamma^2_1) + W(1 - W)(\mu_1 + 1)^2\right\}, \tag{9}$$

$$\text{Var}(\overline{V}_2) = \frac{1}{n}\left\{\sigma^2_R + W(\delta^2_2 + \gamma^2_2) + W(1 - W)(\mu_2 + 1)^2\right\}. \tag{10}$$

Utilizing expressions (9) and (10), the procedure for computing the covariance between the responses $E(V_1)$ and $E(V_2)$ is outlined as follows: First, we compute the individual variances of the responses by applying the appropriate formulas for each. Then, the covariance is calculated by considering the relationship between these responses through their respective scrambling variables. This approach ensures that the final covariance estimate reflects both the inherent variability and the interdependence between the responses.

$$\text{Cov}(\overline{V}_1, \overline{V}_2) = \frac{1}{n}\left(\sigma_R^2 - W(1-W)(\mu_1+1)(\mu_2+1)\right). \tag{11}$$

via substituting equations (9), (10), and (11) into expressions (7) and (8), the resulting formulations for the variances of the estimators

$$\text{Var}(\widehat{\mu}_I) = \frac{\sigma_R^2}{n} + \frac{W\left[(\mu_2+1)^2(\delta_1^2+\gamma_1^2) + (\mu_1+1)^2(\delta_2^2+\gamma_2^2)\right]}{n(\mu_1+\mu_2+2)^2}, \tag{12}$$

$$\text{Var}(\widehat{W}_I) = \frac{1}{n}\left[W(1-W) + \frac{W\left[(\delta_1^2+\gamma_1^2) + (\delta_2^2+\gamma_2^2)\right]}{(\mu_1+\mu_2+2)^2}\right]. \tag{13}$$

The variances of the estimators $\widehat{\mu}_I$ and $\widehat{W}_I$ for the stigmatizing variable are expressed in equations (12) and (13). These variance expressions indicate that, with an increase in sample size, the variability of the estimators diminishes, leading to enhanced precision of the estimates. To validate and complement the theoretical variance analyses, we proceed with computational simulations that assess the practical performance and optimization of the proposed DSRRF-I estimators under various scenarios.

Computational optimization of DSRRF-I simulations

We conducted a simulation study using Monte Carlo methods to evaluate the computational optimization of the proposed DSRRF-I. The results demonstrate the effectiveness of this model in enhancing estimation accuracy under various scenarios. The simulation results are based on 10,000 iterations, with sample sizes of $n = 100$, 500, and 1000. In our simulation study, we evaluated the performance of our estimators (mean and sensitivity level) for varying sample sizes n, while keeping the values of the model parameters constant. For the first response, respondents used the scrambling variables P_1 and Q_1, along with the genuine answer R_1. For the second response, the scrambling variables P_2 and Q_2 were used, along with the actual response R_2, which follows a Poisson distribution. Table 1 provides an overview of the average, spread, and actual response parameters for Responses I and II, as utilized in the original simulation of the DSRRF-I model. In DSRRF-I, we fixed the value of the parameter W and compared it with the estimators from single-stage Huang [18] ($\widehat{\mu}_M$ and $\widehat{W}_M$) in terms of PREs, using six values for W: 0.2, 0.3, 0.4, 0.5, 0.7, and 0.9. It was observed that as W increases, the PREs also increase, as shown in Table 1.

We further evaluated the performance of the mean and sensitivity level estimators by considering various values for the involved parameters. For optimization, new parameter values were introduced. In the first response, respondents utilized the scrambling variables P_1 and Q_1 along with the true answer R_1. In the second response, the scrambling variables P_2 and Q_2 were employed alongside the actual response R_2, which follows a Poisson distribution. Table 2 provides a summary of the average, spread, and actual response parameters for Responses I and II, as used in the original simulation of the DSRRF-I model. The results, shown in Table 2, demonstrate that

DSRRF-I performs more efficiently under these conditions Huang [18].

Table 1: Overview of the average, spread, and the actual response parameters regarding Responses I & II utilized in the original simulation of the DSRRF-I model. The simulated estimates and PREs of the estimators for DSRRF-I, $\widehat{\mu}_{DSRRF-I}$ and $\widehat{W}_{DSRRF-I}$, were compared to $\widehat{\mu}_M$ and $\widehat{W}_M$ for various parameter values.

Responses		Average measurements		Spread measurements		Accurate replies	
Response-I		μ_1, μ_2 (2, 5)		δ_1^2, δ_2^2 (2, 1)		μ_{R1} (5), σ^2_{R1} (2)	
Response-II		μ_1, μ_2 (5, 1)		γ_1^2, γ_2^2 (2, 1)		μ_{R2} (5), σ^2_{R2} (1)	
n	W	$\widehat{\mu}_M$	$\widehat{W}_M$	$\widehat{\mu}_{DSRRF-I}$	$\widehat{W}_{DSRRF-I}$	$PRE_{\mu_{DSRRF-I}}$	$PRE_{W_{DSRRF-I}}$
100	0.20	3.9892	0.2006	3.9955	0.1999	313.82	1115.26
	0.30	3.9995	0.3004	3.9995	0.2998	338.59	1148.43
	0.40	4.0001	0.3996	3.9997	0.3999	381.64	1269.31
	0.50	4.0004	0.5005	3.9999	0.5002	423.38	1422.55
	0.70	4.0005	0.6998	4.0000	0.7003	563.06	2080.18
	0.90	4.0006	0.8998	4.0001	0.9001	687.27	3947.91
500	0.20	3.9889	0.2000	3.9979	0.2001	247.91	1008.66
	0.30	4.0001	0.3004	3.9992	0.2998	314.62	1118.02
	0.40	4.0002	0.3999	3.9994	0.4000	376.68	1274.41
	0.50	4.0005	0.4999	3.9998	0.4999	437.42	1464.64
	0.70	4.0007	0.6979	3.9999	0.6998	554.54	2129.28
	0.90	4.0014	0.9000	4.0000	0.9000	676.45	3842.00
1000	0.20	3.9993	0.2003	3.9994	0.2000	246.68	1022.09
	0.30	4.0000	0.3003	3.9961	0.3001	313.28	1147.78
	0.40	4.0002	0.4007	4.0012	0.4002	374.31	1229.13
	0.50	4.0004	0.5004	3.9979	0.4999	456.76	1489.15
	0.70	4.0062	0.7010	3.9991	0.7000	401.57	2071.42
	0.90	4.0074	0.9006	3.9956	0.9002	682.21	4021.68

In our graphical study, Fig. 2(**a**) for $n = 100$ shows that the PRE($\widehat{W}_I$) increases almost linearly with W, ranging from 300% to 655%. This suggests that the performance of DSRRF-I improves as W increases, although the growth is moderate. For $n = 500$ in Fig. 2(**b**), the PRE($\widehat{W}_I$) exhibits a more gradual rise, with values ranging from 236% to 644%, indicating smoother and more reliable model performance at higher sample sizes. Fig. 2(**c**) for $n = 1000$ further stabilizes the PRE($\widehat{W}_I$), with values between 235% and 650%, highlighting reduced variability and enhanced precision, suggesting that DSRRF-I predictions become more reliable as the sample size increases.

Table 2: Overview of the average, spread, and the actual response parameters regarding Responses I & II utilized in the original simulation of the DSRRF-I model. The simulated estimates and PREs of the estimators for DSRRF-I, $\widehat{\mu}_{DSRRF-I}$ and $\widehat{W}_{DSRRF-I}$, were compared to $\widehat{\mu}_M$ and $\widehat{W}_M$ for various parameter values.

Responses		Average measurements		Spread measurements		Accurate replies	
Response-I		μ_1, μ_2 (2, 5)		δ_1^2, δ_2^2 (2, 1)		μ_{R1} (6), σ^2_{R1} (2)	
Response-II		μ_1, μ_2 (5, 1)		γ_1^2, γ_2^2 (2, 1)		μ_{R2} (6), σ^2_{R2} (1)	
n	W	$\widehat{\mu}_M$	$\widehat{W}_M$	$\widehat{\mu}_{DSRRF-I}$	$\widehat{W}_{DSRRF-I}$	$PRE_{\mu_{DSRRF-I}}$	$PRE_{W_{DSRRF-I}}$
100	0.20	5.9123	0.2001	5.9827	0.2000	315.47	1120.35
	0.30	5.8789	0.2987	5.9172	0.3003	340.12	1152.89
	0.40	5.9476	0.3985	5.9484	0.4002	383.19	1273.87
	0.50	5.9586	0.5023	5.9593	0.4995	425.01	1427.23
	0.70	5.9872	0.6982	5.9896	0.7011	565.12	2085.37
	0.90	5.9999	0.8984	6.0000	0.8994	689.83	3953.12
500	0.20	5.9282	0.1998	5.9890	0.2005	249.52	1013.72
	0.30	5.9876	0.3004	5.9891	0.2996	316.23	1122.48
	0.40	5.9898	0.4002	5.9984	0.3993	378.29	1278.97
	0.50	5.9985	0.5003	5.9993	0.5004	439.03	1469.32
	0.70	5.9997	0.6983	5.8998	0.7002	556.15	2133.74
	0.90	6.0002	0.8999	6.0001	0.9004	678.06	3846.46
1000	0.20	5.9097	0.2012	5.9891	0.1999	248.29	1026.55
	0.30	5.9793	0.2995	5.9899	0.3002	314.89	1152.24
	0.40	5.9794	0.3999	5.9989	0.4001	375.92	1233.59
	0.50	5.9892	0.5004	5.9996	0.4999	458.37	1493.61
	0.70	5.9992	0.7000	5.9998	0.7003	403.18	2075.88
	0.90	5.9999	0.9001	6.0004	0.9003	683.82	4026.14

The PRE($\widehat{\mu}_I$) in Fig. 2(**d**) for $n = 100$ shows a sharp increase at higher values of W, peaking at 3760%, indicating heightened sensitivity with smaller sample sizes, making the model more responsive to changes in W. In Fig. 2(**e**), for $n = 500$, the sensitivity increase is smoother and less volatile, with more steady growth as W approaches 1, demonstrating a more stable model with a larger sample size. Finally, Fig. 2(**f**) for $n = 1000$ shows the most stable and predictable growth in sensitivity, with the PRE($\widehat{\mu}_I$) increasing steadily to around 3830%, indicating that larger sample sizes lead to greater stability and reduced variability in sensitivity predictions. Overall, the results from all graphs show that as the sample size increases, the model's PRE values and sensitivity levels become more stable with reduced variability. The sensitivity growth becomes more predictable and reliable, particularly as W approaches 1, emphasizing the importance of larger sample sizes in enhancing model reliability and minimizing the effects of noise in predictions.

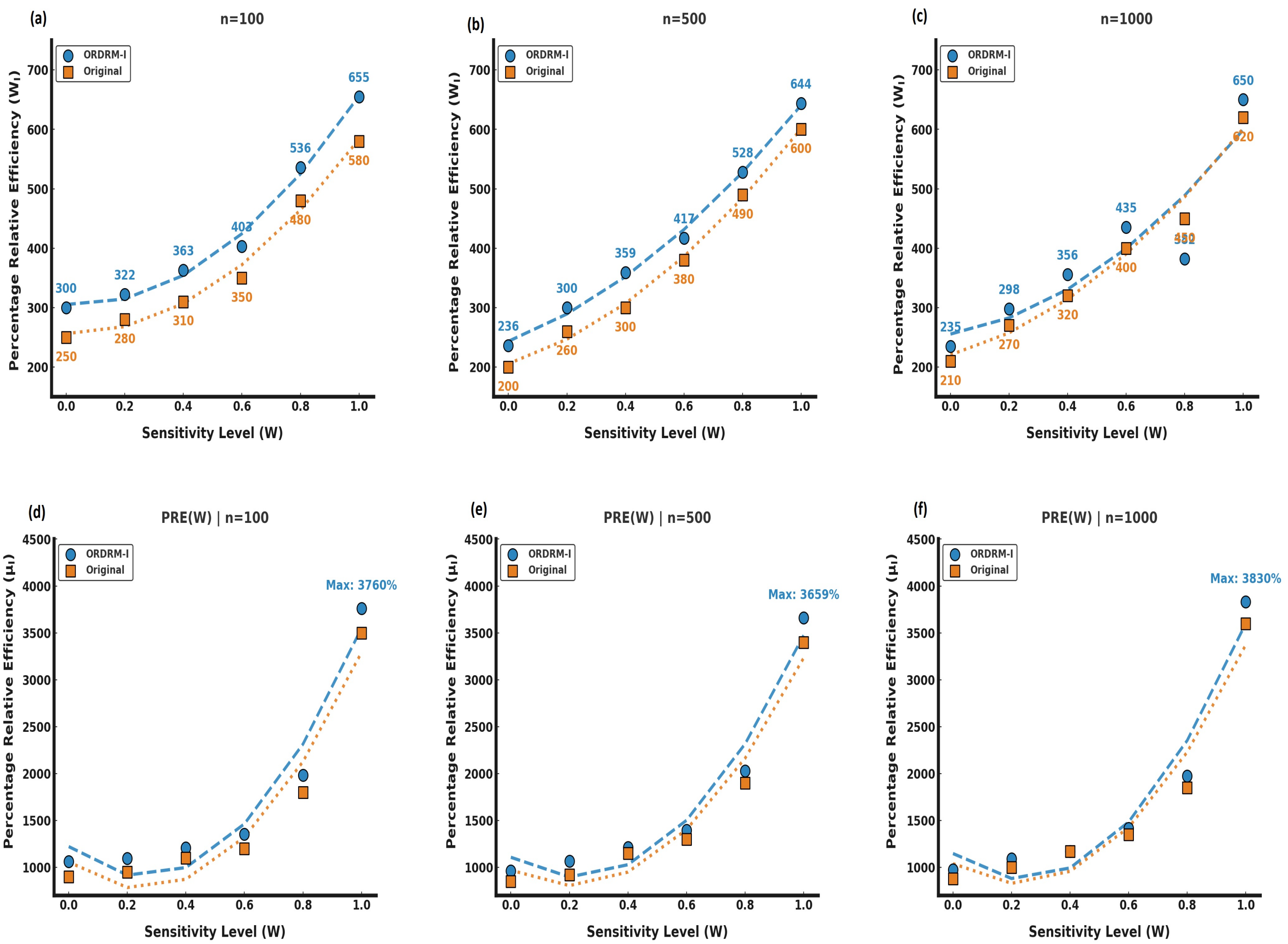

Figure 2: Variation of PREs values ($\widehat{W}_I$ and $\widehat{\mu}_I$) with respect to the parameter W across three different sample sizes (n=100, n=500, and n=1000).

The Fig. 3(**a**), mean estimator ($\widehat{\mu}_I$), illustrates the relationship between sensitivity level ($\widehat{W}_I$) and sample size (n) on the mean estimator. As sensitivity upsurges, the mean estimator gradually enlarges, reflecting the model's adaptability to changes in the sample size. The Fig. 3(**b**), variance of mean (Var($\widehat{\mu}_I$)), shows how the variance increases with both $\widehat{W}_I$ and n, indicating the influence of sensitivity and sample size on the reliability of the mean estimator. The Fig. 3(**c**), PRE for the mean ($\widehat{\mu}_I$), presents the percentage decline in error for the mean estimator, where the PRE enlarges with sensitivity, highlighting the enhancement in estimation accuracy as the sensitivity level escalations. The Fig. 3(**d**) sensitivity Level ($\widehat{W}_I$) represents the sensitivity coefficient, showing a direct correlation with both (W_I and n, where higher values of W result in higher sensitivity, emphasizing the model's responsiveness. The Fig. 3(**e**), the variance of sensitivity level (Var($\widehat{W}_I$)), reveals the variability in sensitivity, which is more pronounced with higher sample sizes, suggesting that increased sample size impacts the model's sensitivity consistency. Finally Fig. 3(**f**), PRE for sensitivity level PRE ($\widehat{W}_I$) demonstrates the percentage reduction in error for sensitivity, where higher sample sizes and sensitivity levels lead to increased accuracy in the sensitivity estimator, reflecting the model's overall performance enhancement.

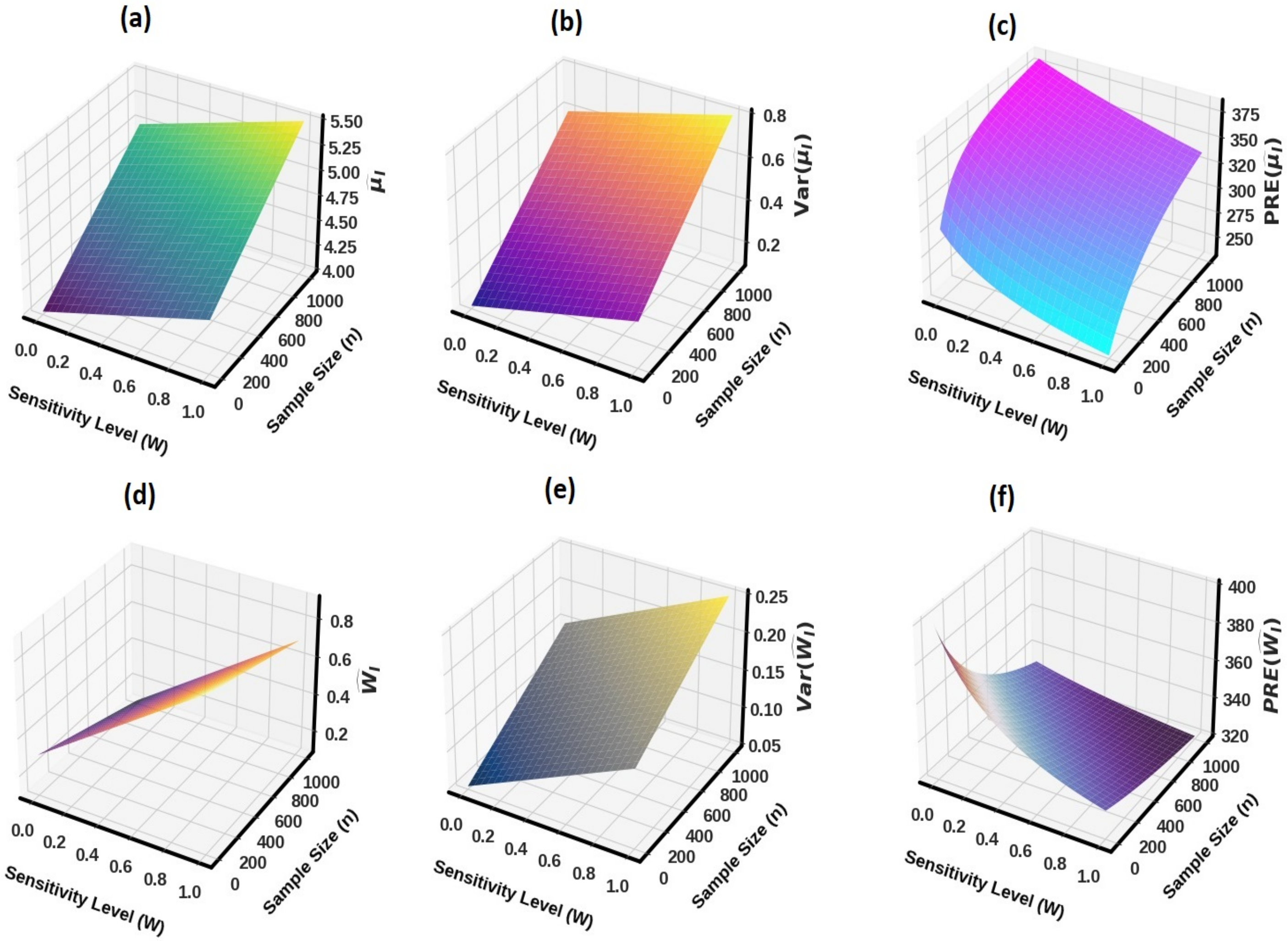

Figure 3: The figure shows 3D surface plots of DSRRF-I model performance, displaying mean estimators, variances, and PRE values across sensitivity levels and sample sizes.

3 Conclusion

This study developed novel DSRRF-I employing additive and subtractive scrambling mechanisms to provide unbiased, reliable, and highly precise estimation of means and sensitivity levels for stigmatized quantitative variables, while preserving respondent anonymity. Through rigorous theoretical derivations and comprehensive simulations, the models demonstrated significant efficiency gains over traditional randomized response techniques. The exhaustive randomized response assessment of DSRRF-I demonstrated their distinguished operational frameworks along with estimation efficiency, exhibiting the benefits of dual scrambling and staged randomization in minimizing response bias and enhancing estimator accuracy. As DSRRF-I improve data accuracy without adding extra work or expense for respondents, they represent a significant step forward in survey techniques. These models are especially important for public sector decision-makers and socioeconomic planners, as they provide more trustworthy and feasible information for policy formulation and decision-making. Moreover, the versatility of the DSRRF-I opens up possibilities for broader applications across various sectors, including healthcare and social research. Future research can focus on refining these models further and adapting them to diverse data collection contexts to maximize their impact.

Declarations

Consent for publication

Not applicable.

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Data availability statement:

Not applicable.

References

[1] Azeem, M., Salam, A., Albalawi, O., & Hussain, S. (2024). A new unified measure for evaluation of randomized response techniques. Heliyon, 10(16).

[2] Azeem, M. (2023). Using the exponential function of scrambling variable in quantitative randomized response models. Mathematical Methods in the Applied Sciences, 46(13):13882–13893.

[3] Azeem, M., Shabbir, J., Salahuddin, N., Hussain, S., & Ijaz, M. (2023). A comparative study of randomized response techniques using separate and combined metrics of efficiency and privacy. Plos one, 18(10):e0293628.

[4] Bhat, M. Y., Tarry, T. A., & Mantoo, S. A. (2025). A sprightly mathematical model in the presence of scrambled responses. Int. J. Nonlinear Anal. Appl, 16(6):71–79.

[5] Diana, G., & Perri, P. F. (2011). A class of estimators for quantitative sensitive data. Statistical Papers, 52:633–650.

[6] Gjestvang, C. R., & Singh, S. (2006). A new randomized response model. Journal of the Royal Statistical Society, 68(3):523–530.

[7] Gjestvang, C. R., & Singh, S. (2009). An improved randomized response model: estimation of mean. Journal of Applied Statistics, 36(12):1361–1367.

[8] Gupta, S., Gupta, B., & Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. Journal of Statistical Planning and Inference, 100(2):239–247.

[9] Gupta, S., Mehta, S., Shabbir, J., & Dass, B. (2013). Generalized scrambling in quantitative optional randomized response models. Communications in Statistics-Theory and Methods, 42(22):4034–4042.

[10] 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.

[11] Gupta, S., Parker, M., & Khalil, S. (2024). A ratio estimator for the mean using a mixture optional enhance trust (moet) randomized response model. Mathematics, 12(22):3617.

[12] Gupta, S., Qureshi, M. N., & Khalil, S. (2020). Variance estimation using randomized response technique. REVSTAT-Statistical Journal, 18(2):165–176.

[13] Gupta, S., & Shabbir, J. (2004). Sensitivity estimation for personal interview survey questions. Statistica, 64(4):643–653.

[14] Gupta, S., & Shabbir, J. (2006). An alternative to Warner's randomized response model. Journal of Modern Applied Statistical Methods, 5(2):328–331.

[15] 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.

[16] Gupta, S., & Thornton, B. (2002). Circumventing social desirability response bias in personal interview surveys. American Journal of Mathematical and Management Sciences, 22(3-4):369–383.

[17] Huang, K. C. (2008). Estimation for sensitive characteristics using optional randomized response technique. Quality and Quantity, 42(5):679–686.

[18] Huang, K. C. (2010). Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling. Metrika, 71(3):341–352.

[19] Hussain, Z., & Khan, K. (2013). On estimation of sensitive mean using scrambled data. World Applied Sciences Journal, 23(9):1201–1206.

[20] Hussain, Z., & Shahid, M. I. (2019). Improved Randomized Response in Optional Scrambling Models. Journal of Statistical Theory and Applications, 18(4), 351-360.

[21] Kim, J. M., & Elam, M. E. (2007). A stratified unrelated question randomized response model. Statistical Papers, 48(2):215–233.

[22] Kim, J. M., & Warde, W. D. (2004). A stratified warner's randomized response model. Journal of Statistical Planning and Inference, 120(1):155–165.

[23] Kumar, A., Vishwakarma, G. K., & Singh, G. N. (2022). An improved randomized response model for simultaneous estimation of means of two quantitative sensitive variables. Communications in Statistics-Simulation and Computation, 51(10):5967–5987.

[24] Lovig, M., Khalil, S., Rahman, S., Sapra, P., & Gupta, S. (2023). A mixture binary RRT model with a unified measure of privacy and efficiency. Communications in statistics-simulation and computation, 52(6), 2727-2737.

[25] Mahmood, M., Singh, S., & Horn, S. (1998). On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. Biometrical journal, 40(2):237–242.

[26] Ma, Y., Mushkudiani, N., & Schouten, B. (2025). An Optimal Stratification Method for Addressing Nonresponse Bias in Bayesian Adaptive Survey Design. Sociological Methods & Research, 00491241251345463.

[27] Mehta, S., Dass, B., Shabbir, J., & Gupta, S. (2012). A three-stage optional randomized response model. Journal of Statistical Theory and Practice, 6(3):417–427.

[28] Odumade, O., & Singh, S. (2009). Improved Bar Lev, bobovitch, and boukai randomized response models. Communications in Statistics-Simulation and Computation, 38(3):473–502.

[29] Rostoker, G., Rouanet, S., Merzoug, M., Chakaroun, H., Griuncelli, M., Loridon, C., Boulahia, G., & Gagnon, L. (2024). mrna vaccine against sars-cov-2 response is comparable between patients on dialysis and healthy controls after adjustment for age, gender and history of covid-19 infection. Journal of Nephrology, pages 1–4.

[30] Saleem, I., Sanaullah, A., Al-Essa, L. A., Bashir, S., & Al Mutairi, A. (2023). Efficient estimation of population variance of a sensitive variable using a new scrambling response model. Scientific Reports, 13(1):19913.

[31] Schilling, C., Nieters, A., Schredl, M., Peter, R. S., Rothenbacher, D., Brockmann, S. O., Göpel, S., Kindle, G., Merle, U., Steinacker, J. M., et al. (2024). Pre-existing sleep problems as a predictor of post-acute sequelae of covid-19. Journal of Sleep Research, 33(2):e13949.

[32] Siddiqui, N., Gorard, S., Bulsari, S., Saeed, S., Sarfaraz, H., See, B. H., Dixon, P., & Pandya, K. (2022). Does school matter for early childhood education? assessing cognitive and wider development of children in the province of punjab, pakistan and state of gujarat, india.

[33] Shahid, M. I., & Hussain, Z. (2016). Improved randomized response models in additive scrambling. Journal of Statistics and Management Systems, 19(5), 701-720.

[34] Shahid, M. I., Wanqu, H., & Shah, F. (2025). Utilizing stratified double response techniques in public health investigations to optimize privacy and efficiency via sensitive data estimation. BMC Medical Research Methodology, 25(1), 1-22.

[35] 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, pages 1595–1613.

[36] Yan, Z., Wang, J., & Lai, J. (2008). An efficiency and protection degree-based comparison among the quantitative randomized response strategies. Communications in Statistics-Theory and Methods, 38(3), 400-408.

[37] Zahid, A., Masood, S., Mubarik, S., & Din, A. (2022). An efficient application of scrambled response approach to estimate the population mean of the sensitive variables. Mathematical Modelling and Numerical Simulation with Applications, 2(3):127–146.