The Evolution of Sensitive Data Estimation: Dual Response Frameworks for Enhanced Privacy

Li Honghe[1], Faiza Shah[1*], Muhammad Imran Shahid[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:

Faiza Shah*

Muhammad Imran Shahid *

Abstract

This research tackles the critical challenge of measuring stigmatized quantitative variables with high accuracy, while safeguarding respondent privacy and data integrity. We introduce a novel two-stage dual scrambling randomized response framework (DSRRF-II) that combines additive and subtractive scrambling techniques to optimize both privacy and statistical efficiency. By leveraging dual scrambling variables and employing simple random sampling with replacement (SRSWR), the framework generates unbiased, precise, and reliable estimates of means and sensitivity levels without burdening respondents. The mathematical formulations and extensive simulations demonstrate that our approach significantly outperforms traditional randomized response methods in terms of efficiency. Furthermore, DSRRF-II is adaptable to diverse data collection environments, ensuring robustness across various contexts. Our findings highlight the potential of these models to revolutionize the collection of sensitive data, providing more accurate estimates for public health and social research. The flexibility of DSRRF-II also sets the stage for future developments in privacy-preserving survey techniques, offering a powerful tool for addressing complex data measurement challenges.

Keywords: Stigmatized quantitative variables, Privacy protection, Randomized response frameworks, Dual scrambling, Data integrity, Efficiency enhancement.

1 Introduction

Measuring sensitive and stigmatized quantitative variables presents a unique challenge in research, particularly when ensuring both privacy and data integrity. These variables, often linked to socially undesirable topics, complicate the accuracy of data collection, leading to potential biases or unreliable estimates. Conventional methods of measurement fail to address these concerns, especially in sensitive domains like public health and social research. This study introduces a novel two-stage dual scrambling randomized response framework (DSRRF-II), which is designed to safeguard respondent privacy while improving statistical efficiency. By incorporating both additive and subtractive scrambling techniques, DSRRF-II offers a solution to improve the accuracy of estimates without placing an extra burden on participants. The proposed framework leverages simple random sampling with replacement (SRSWR) to ensure unbiased and precise data collection. Through mathematical derivations and simulations, this study demonstrates the framework's superior efficiency compared to traditional approaches. Ultimately, DSRRF-II offers a robust and adaptable method for addressing the challenges of measuring stigmatized data across various research settings.

To address these challenges, the concept of ambiguous response was initially introduced. Later, [6] proposed forced empirical models, and [15] developed a dual approach for averaging and estimating sensitivity levels, which was subsequently refined by [7] with the introduction of an additive scrambling technique. The methodologies for estimating sensitive quantitative data were further advanced by [9, 14], [16, 17, 25, 31], and [18, 26], contributing significantly to the field.

Building on these advancements, subsequent innovations in randomized response methods have effectively tackled issues like response bias and estimation efficiency. Kumar and Singh [21, 33] introduced advanced techniques for reducing bias in responses through randomized scattering approaches, while Gupta et al. [12] developed a variance estimator incorporating the randomization strategy of Diana and Perri [5]. Additionally, Saleem et al. [28] proposed enhanced estimators for population means. Drawing from previous works, Azeem [3] recently advanced the field by introducing a weighted metric that integrates privacy and efficiency, offering a comprehensive evaluation framework for randomized response methods. Siddiqui et al. [30] made further strides by analyzing mathematical models that account for correlations between variables. Significant contributions were also made by Azeem [2], Bhat [4], Gupta [11], and Zahid et al. [35], improving variance behavior and enhancing PRE performance across diverse sensitivity levels, as shown by Shahid et al. [32].

Recognizing the complexity of these challenges, we proposed an innovative solution through the DSRRF-II approach. Building on the methodology of Gupta and Shabbir [14], we developed two DSRRF-II aimed at obtaining accurate data while maintaining respondent confidentiality. The key novelty of our approach lies in the use of dual scrambling variables, P_i and Q_i, for each response in a two-fold response system. This strategy enabled the estimation of both the mean and sensitivity levels without imposing additional costs on the respondent. The two-response approach not only enhanced the reliability of the estimates but also preserved the integrity of the data collection process. Furthermore, this methodology reduces the likelihood of social desirability bias by encouraging more honest reporting. It also provides a robust framework for addressing stigmatized data, making it applicable in various fields such as public health and social research. By ensuring both privacy and accuracy, the DSRRF-II approach represents a significant advancement in survey methodology.

2 Innovative Approaches for Estimation Accuracy and Confidentiality

This section outlines the theoretical foundation behind the proposed methodologies and derives unbiased estimators along with their associated variances. We introduce a two-stage additive and subtractive DSRRF-II approach to estimate the mean μ_R and the sensitivity level W of a stigmatized variable R through a dual-response mechanism. The technique involves two steps of scrambling, where variables P_i and Q_i are sequentially added and subtracted from the true value of R. This procedure is designed to alter respondents' perceptions, encouraging them to provide more accurate responses to sensitive questions by mitigating both over-reporting and under-reporting tendencies. By offering improved privacy protection, the method enhances the reliability of the data collected. Respondents are informed in advance that their answers are anonymized, with the estimation process ensuring that individual responses cannot be traced back. This approach also clarifies the parameter estimation process, ensuring transparency about the study's objectives. These improvements reduce response bias and enhance estimator performance, making DSRRF-II particularly effective in complex survey settings where sensitivity levels vary.

2.1 Privacy-Preserving Response Approach on DSRRF-II

The DSRRF-II model introduces a structured responses framework designed to enhance respondent confidentiality and provide greater flexibility in data collection. The proposed strategy involves selecting respondents from a sample of size n, who are requested to provide two responses. In the first response, the interviewees are asked to report either the true response R or proceed to stage 2, with known probabilities T and $(1-T)$, respectively. At stage 2, the scrambling procedure requires each respondent to use two randomization devices that generate random numbers, denoted as P_1 and Q_1, drawn from pre-assigned Poisson distributions. These scrambling variables, P_1 and Q_1, are then added to the true response R. If the respondent does not perceive the question as sensitive, they provide the true response R. In the second response, the same procedure is followed, with the difference that the respondent subtracts the scrambling variables P_2 and Q_2 from the true response R. The two responses, denoted as V_1^* and V_2^*, are given as follows:

$$V_1^* = R\eta + (1-\eta)\left\{R(1-\alpha) + \alpha(R + P_1 + Q_1)\right\}, \tag{1}$$

$$V_2^* = R\eta + (1-\eta)\left\{R(1-\alpha) + \alpha(R - P_2 - Q_2)\right\}. \tag{2}$$

According to Fig. 1, the structural mechanism of DSRRF-II plays a pivotal role in the randomized response process, highlighting the key steps involved. With the response mechanism clearly defined, the next focus is on deriving unbiased and efficient estimators that leverage the two-stage randomization inherent to DSRRF-II, ensuring proper connectivity in the estimation process.

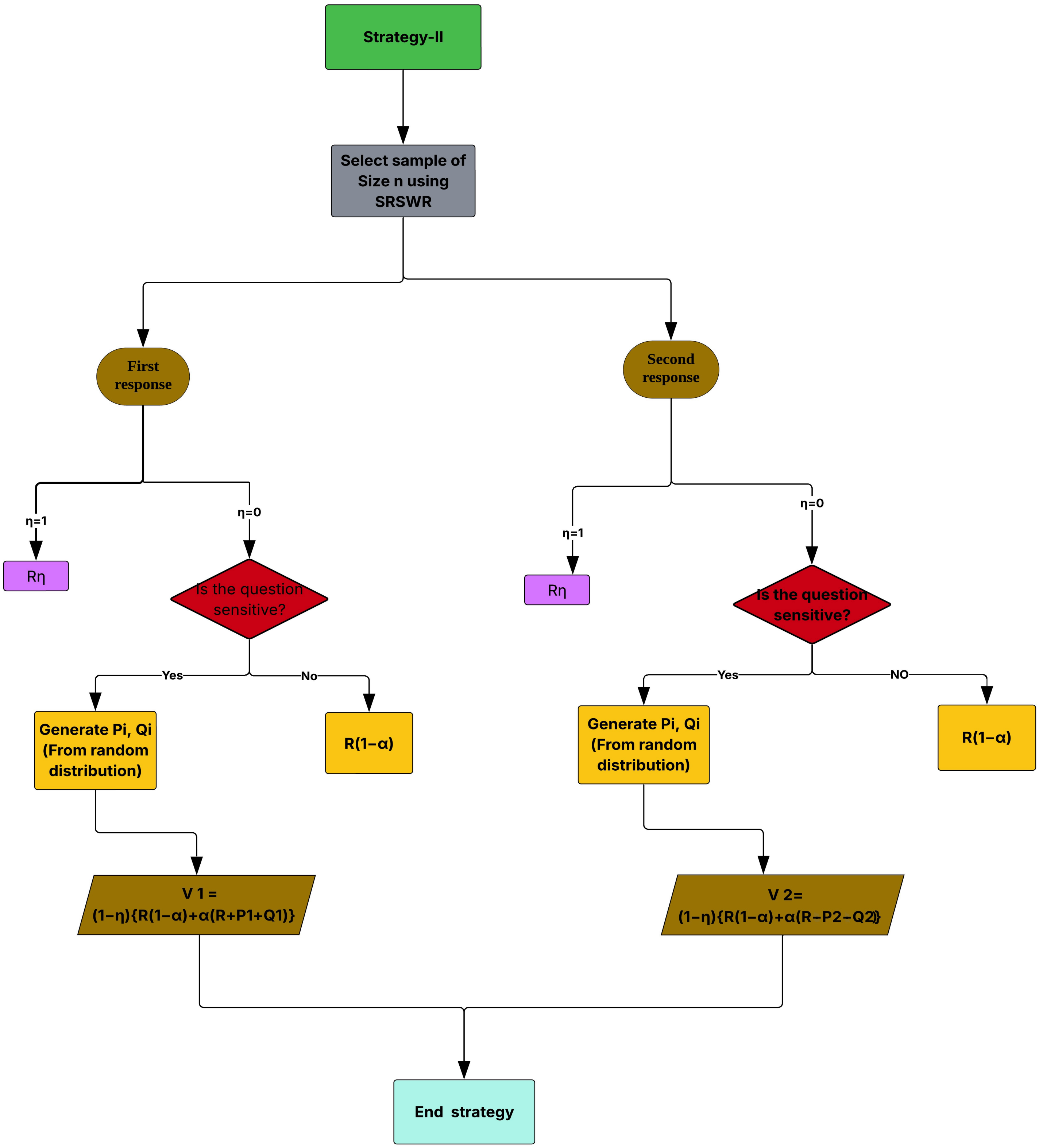

Figure 1: The structural mechanism for DSRRF-II in the process of randomized response.

the structural factors that might be influenced the unbiased and efficient estimators (UEE-II) and define their precise manifestations. The UEE-II utilizes a randomized response technique to ensure that the expected value of the scrambled response nearly approximates the true parameter, thereby minimizing variance and safeguarding respondent confidentiality. It is crucial to identify the unbiased and efficient estimators, as well as the key occurrences of spanning clusters. Notably, if $T = 0$, the method simplifies to a single-stage DSRRF-I. The final expressions for the expected responses are provided as follows:

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

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

The unbiased estimators of μ_{II} and W_{II} are presented by the preceding the following expressions:

$$\widehat{\mu}_{II} = \frac{\overline{V}_2^*(\mu_1+1) + \overline{V}_1^*(\mu_2+1)}{\mu_1+\mu_2+2}, \quad \text{for} \quad \mu_1+\mu_2 \neq -2, \tag{5}$$

$$\widehat{W}_{II} = \frac{1}{(1-T)}\left(\frac{\overline{V}_2^* - \overline{V}_1^*}{\mu_1+\mu_2+2}\right), \; \mu_1+\mu_2 \neq -2, \; T \neq 1. \tag{6}$$

Subsequent to the development of unbiased estimators, an in-depth analysis of their variance and covariance properties is critical for understanding the precision and robustness of the parameter estimates. The structural factors that could affect the dispersion of unbiased estimators (DUE-II) and provides their precise expressions. The DUE-II, implemented through the randomized response technique, minimizes variance, thereby enhancing the efficiency and reliability of the estimators. The variances associated with these estimators, as shown in equations (5) and (6), are expressed as:

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

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

So, for the resulting discussion, the following (7) and (8) expressions are required to estimate the variance between the first and second responses:

$$\mathrm{Var}(\overline{V}_1^*) = \frac{1}{n}\left[\sigma_R^2 + W(1-T)(\delta_1^2+\gamma_1^2) + W(1-T)(\mu_1+1)^2\left\{1-W(1-T)\right\}\right], \tag{9}$$

$$\mathrm{Var}(\overline{V}_2^*) = \frac{1}{n}\left[\sigma_R^2 + W(1-T)(\delta_2^2+\gamma_2^2) + W(1-T)(\mu_2+1)^2\left\{1-W(1-T)\right\}\right]. \tag{10}$$

Using expressions (9) and (10), the procedure for calculating the covariance between the responses is given by

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

In a nutshell by employing (9), (10), and (11) within (7) and (8), the consequent expressions for the variance of the estimators $\widehat{\mu}_{II}$ and $\widehat{W}_{II}$ might be acquired through

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

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

Finally, the variances of$\widehat{\mu}_{II}$ and $\widehat{W}_{II}$ for the stigmatizing variable are shown in equations (12) and (13). In these expressions, the variances of the estimators $\widehat{\mu}_{II}$ and $\widehat{W}_{II}$ reduce as the significance of T escalates. To validate the theoretical findings and evaluate practical performance, computational simulations employing Monte Carlo methods are conducted to optimize and benchmark the DSRRF-II estimators under varying conditions.

Maximizing Efficiency in DSRRF-II Simulation Processes

A simulation study using Monte Carlo methods to assess the computational optimization of the proposed DSRRF-II. The findings highlight the effectiveness of the model in improving estimation accuracy across a range of scenarios. The simulation results were based on 100,000 iterations, with sample sizes of n = 100, 500, and 1000. The performance of DSRRF-II was evaluated against the unbiased estimators of the mean and sensitivity level, $\widehat{\mu}_M$ and $\widehat{W}_M$, from Gupta and Shabbir [14], with respect to their PREs. In the first response, respondents incorporated 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 applied together with the actual response R_2, which follows a Poisson distribution. Table 1 presents an overview of the average, spread, and actual response parameters for Responses I and II, as applied in the original DSRRF-II model simulation.

For DSRRF-II, the parameters W and T were fixed, and comparisons were made with the estimators from 2-stage Gupta and Shabbir [14], ($\widehat{\mu}_{M2}$ and $\widehat{W}_{M2}$) based on PREs, using six values for W: 0.2, 0.3, 0.4, 0.5, 0.7, and 0.9, and three values for T: 0.2, 0.3, and 0.4. As W and T increased, the PREs improved, as shown in Table 1. To further assess the performance of the mean and sensitivity level estimators, we examined different parameter values and introduced new ones for optimization. The results, summarized in Table 2, demonstrate that DSRRF-II achieves greater efficiency under these conditions, improving the accuracy of estimations compared to Gupta and Shabbir [14].

The first Fig. 2(**a**) shows the PRE ($\widehat{W}_{II}$) for a sample size of 100. The DSRRF-II model (red) exhibits a distinct quadratic growth in efficiency as the W escalates, particularly from 0.6 to 1.0. In contrast, the Benchmark model (green) shows an analogous trend excluding a less sharp enlarge, indicating that DSRRF-II outperforms the singular model at all sensitivity levels. The polynomial fits for both replicas strengthen these observations, with DSRRF-II achieving sophisticated efficiency. In Fig. 2(**b**) for the sample size of n=500, the PRE $(\widehat{W}_{II})$ trends for both strategies are comparable to those in the first graph, but the difference between the models restricts slightly. DSRRF-II stationary shows convoluted implementation, particularly at W above 0.4.

Table 1: Overview of the average, spread, and the actual response parameters regarding Responses I & II utilized in the original simulation of the DSRRF-II model. The simulated estimates and PREs of the estimators for DSRRF-II, $\widehat{\mu}_{DSRRF-II}$ and $\widehat{W}_{DSRRF-II}$, 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_{R1}^2 (2)		
Response-II		μ_1, μ_2 (5, 1)		γ_1^2, γ_2^2 (2, 1)		μ_{R2} (6), σ_{R2}^2 (1)		
n	T	W	$\widehat{\mu}_{M2}$	$\widehat{W}_{M2}$	$\widehat{\mu}_{DSRRF-II}$	$\widehat{W}_{DSRRF-II}$	$PRE_{\mu_{DSRRF-II}}$	$PRE_{W_{DSRRF-II}}$
100	0.20	0.20	3.8991	0.2013	3.9287	0.2001	232.20	388.66
		0.30	3.9228	0.2996	3.9497	0.3004	341.92	393.00
	0.30	0.40	3.9496	0.4005	3.9793	0.3994	452.90	423.01
		0.50	3.9599	0.5021	3.9969	0.5016	563.95	644.26
	0.40	0.70	3.9994	0.6976	3.9997	0.6996	779.89	885.56
		0.90	3.9997	0.8992	4.0000	0.8999	800.53	999.37
500	0.20	0.20	3.9092	0.1992	3.9295	0.1999	305.49	424.06
		0.30	3.9512	0.2997	5.9054	0.2982	587.57	570.83
	0.30	0.40	3.9839	0.3985	5.9106	0.3987	684.83	890.32
		0.50	3.9939	0.5005	3.9982	0.5003	767.35	974.31
	0.40	0.70	3.9998	0.6969	5.8986	0.6997	111.32	5689.27
		0.90	4.0001	0.9029	4.0002	0.9004	1096.78	1325.11
1000	0.20	0.20	3.9404	0.1997	5.8985	0.2000	581.90	699.04
		0.30	3.9521	0.2997	3.9587	0.3002	664.72	763.27
	0.30	0.40	3.9898	0.4025	5.8983	0.4005	768.93	914.27
		0.50	3.9994	0.5000	3.9993	0.5001	859.63	1104.03
	0.40	0.70	3.9999	0.6976	5.8991	0.7000	1005.36	1381.71
		0.90	4.0001	0.8980	4.0007	0.9000	1269.91	1602.36

Whilst the comparison model also upturns in PRE $(\widehat{W}_{II})$ with W, it consequently accomplishes at a slower rate, further suggesting that DSRRF-II performance improves more significantly with higher truthfulness response probabilities. Fig. 2(**c**), for the sample size of 1000, reveals an even more manifest distinction between the two techniques. DSRRF-II continues to superior PRE$(\widehat{W}_{II})$ across all W, particularly at higher sensitivity values (0.6

to 1.0).

Table 2: Overview of the average, spread, and the actual response parameters of responses I & II utilized in the original simulation of the DSRRF-II model. The simulated estimates and PREs of the estimators for DSRRF-II, $\widehat{\mu}_{DSRRF-II}$ and $\widehat{W}_{DSRRF-II}$, 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 (3, 1)		δ_1^2, δ_2^2 (2, 1)		μ_{R1} (6), σ_{R1}^2 (2)		
Response-II		μ_1, μ_2 (4, 6)		γ_1^2, γ_2^2 (2, 2)		μ_{R2} (6), σ_{R2}^2 (1)		
n	T	W	$\widehat{\mu}_{M2}$	$\widehat{W}_{M2}$	$\widehat{\mu}_{DSRRF-II}$	$\widehat{W}_{DSRRF-II}$	$PRE_{\mu_{DSRRF-II}}$	$PRE_{W_{DSRRF-II}}$
100	0.20	0.20	5.8992	0.2003	5.9924	0.2001	325.90	469.20
		0.30	5.9828	0.2996	5.9987	0.3001	435.15	374.27
	0.30	0.40	5.9939	0.4000	5.9993	0.3999	544.66	602.86
		0.50	5.9950	0.5001	5.9995	0.5002	655.12	723.76
	0.40	0.70	6.0000	0.6998	5.9997	0.6999	865.64	962.46
		0.90	6.0002	0.8992	6.0000	0.8999	979.53	1170.82
500	0.20	0.20	5.9991	0.2000	5.9994	0.2001	425.65	562.27
		0.30	5.9992	0.2986	5.9997	0.2989	534.21	681.49
	0.30	0.40	5.9997	0.3995	6.0001	0.3998	844.45	901.30
		0.50	5.9989	0.5001	5.9991	0.4999	1070.89	1136.27
	0.40	0.70	5.9993	0.6998	5.9994	0.6999	1263.49	1358.35
		0.90	6.0003	0.9004	6.0000	0.9001	1378.61	1547.92
1000	0.20	0.20	5.9985	0.2000	5.9989	0.2001	525.98	666.21
		0.30	5.9992	0.2997	5.9994	0.3002	738.75	888.83
	0.30	0.40	5.9994	0.3996	5.9996	0.4000	1042.98	1241.57
		0.50	5.9996	0.5001	5.9997	0.5000	1380.61	1535.68
	0.40	0.70	5.9999	0.6989	6.0000	0.7000	1566.17	1473.04
		0.90	6.0004	0.8999	6.0001	0.9005	1682.20	2059.84

The polynomial fit for DSRRF-II shows a steeper increase compared to the original model, emphasizing DSRRF-II higher functioning as sensitivity rises. Fig. 2(**d**) illustrates the PRE($\widehat{\mu}_{II}$) for the sample size of 100. Both models demonstrate accumulative efficiency with developed sensitivity, but DSRRF-II consistently outperforms the original model, especially at W above 0.6. The polynomial fit for DSRRF-II is steeper, highlighting its better execution in response to increasing W. Fig. 2(**e**), for the sample size of 500, follows the same general pattern as the previous one. The efficiency of both models increases with W, but DSRRF-II consistently stipulates higher PRE($\widehat{\mu}_{II}$). The rate of expansion for DSRRF-II is remarkably steeper than that of the previous model, further reinforcing that DSRRF-II is more efficient at higher sensitivity levels. Fig. 2(**f**) for the sample size of 1000 indications a supplementary distinct differentiation between the models. DSRRF-II demonstrates a sharp rise in PRE($\widehat{\mu}_{II}$), particularly at the higher W, while the Original Model shows a more gradual increase. The polynomial fit on behalf of DSRRF-II is significantly steeper, emphasizing that DSRRF-II provides much higher efficiency than the standard strategy, especially at higher W.

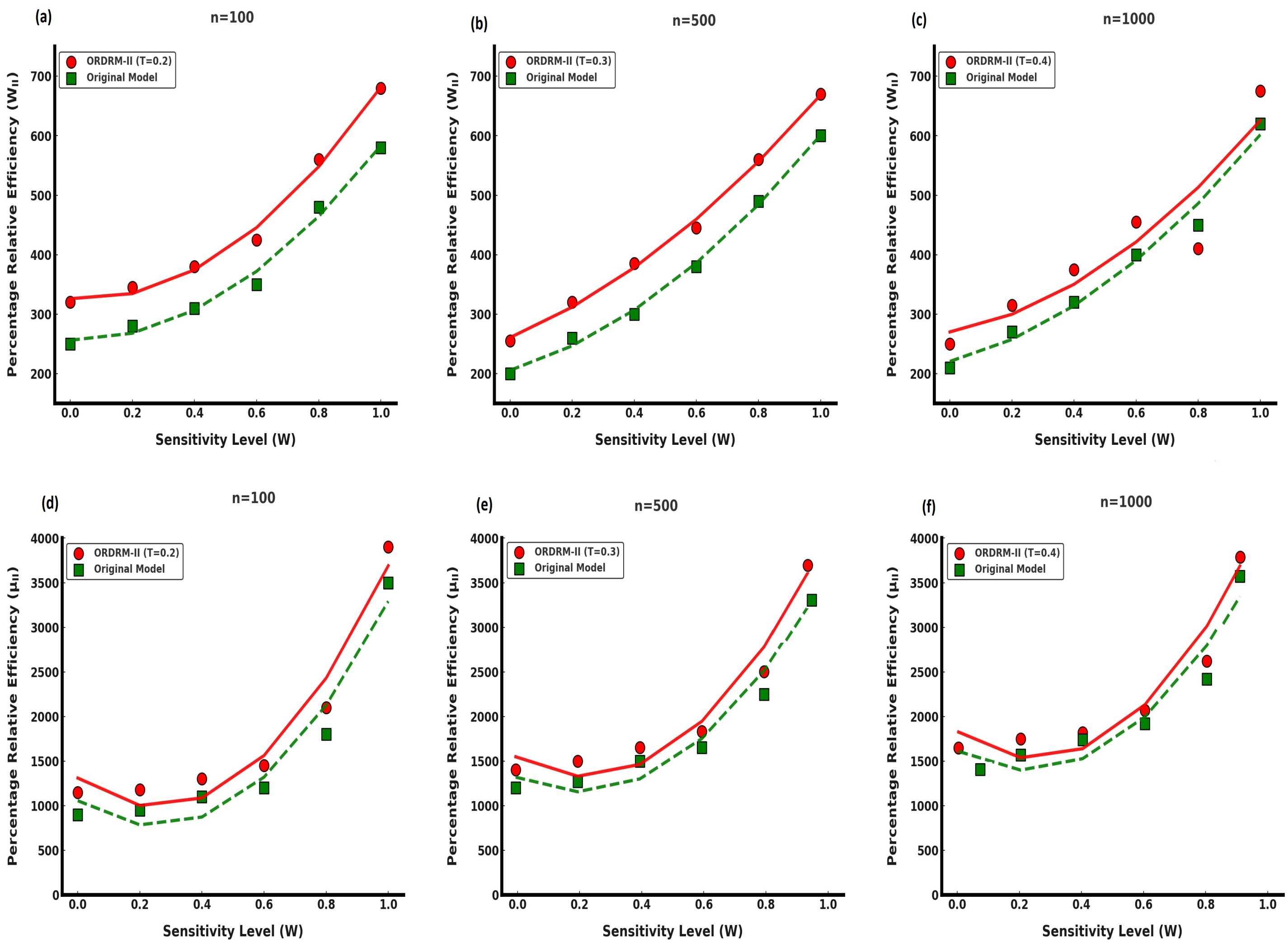

Figure 2: Comparison of DSRRF-II and the base model performance across sensitivity levels (W) at sample sizes (n = 100, 500, 1000). DSRRF-II consistently achieves higher PREs values ($\widehat{W}_{II}$ and $\widehat{\mu}_{II}$), especially at higher $\widehat{W}_{II}$.

The Fig. 3(**a**), mean estimator ($\widehat{\mu}_{II}$), reveals how the mean estimator varies with sensitivity level (W) and truthfulness (T), with the surface exhibiting a sinusoidal pattern inclined by both parameters. As $\widehat{\mu}_{II}$ upsurges, the estimator's level becomes more distinct, suggesting established accuracy in estimations with higher values of truthfulness. In Fig. 3(**b**) variance of the mean (Var($\widehat{\mu}_{II}$)), the variance is more prominent at higher sensitivity levels, emphasizing how variability in the model's output escalates with both W and T, leading to reduced predictability. The Fig. 3(**c**), PRE for the mean PRE ($\widehat{\mu}_{II}$), illustrates the performance enhancement in mean estimation, with the percentage drop in error aggregate as the sensitivity level and truthfulness enlargement, demonstrating improved model efficiency under these conditions. Fig. 3(**d**) shows the relationship between sensitivity level and truthfulness, where sophisticated truthfulness values contribute to increased sensitivity level estimator ($\widehat{W}_{II}$), thus enhancing the model's adaptability to variations in the dataset. The Fig 3(**e**) the variance of sensitivity level (Var($\widehat{W}_{II}$)) highlights how the sensitivity variance escalates with higher values of truthfulness, indicating that the model becomes more sensitive and less consistent at higher truthfulness levels. Lastly, Fig. 3(**f**) shows the percentage cutback in error for sensitivity leveln PRE ($\widehat{W}_{II}$), where improvements are observed as both sensitivity level and truthfulness increase, illustrating the model's ability to deliver more accurate sensitivity estimations under these conditions.

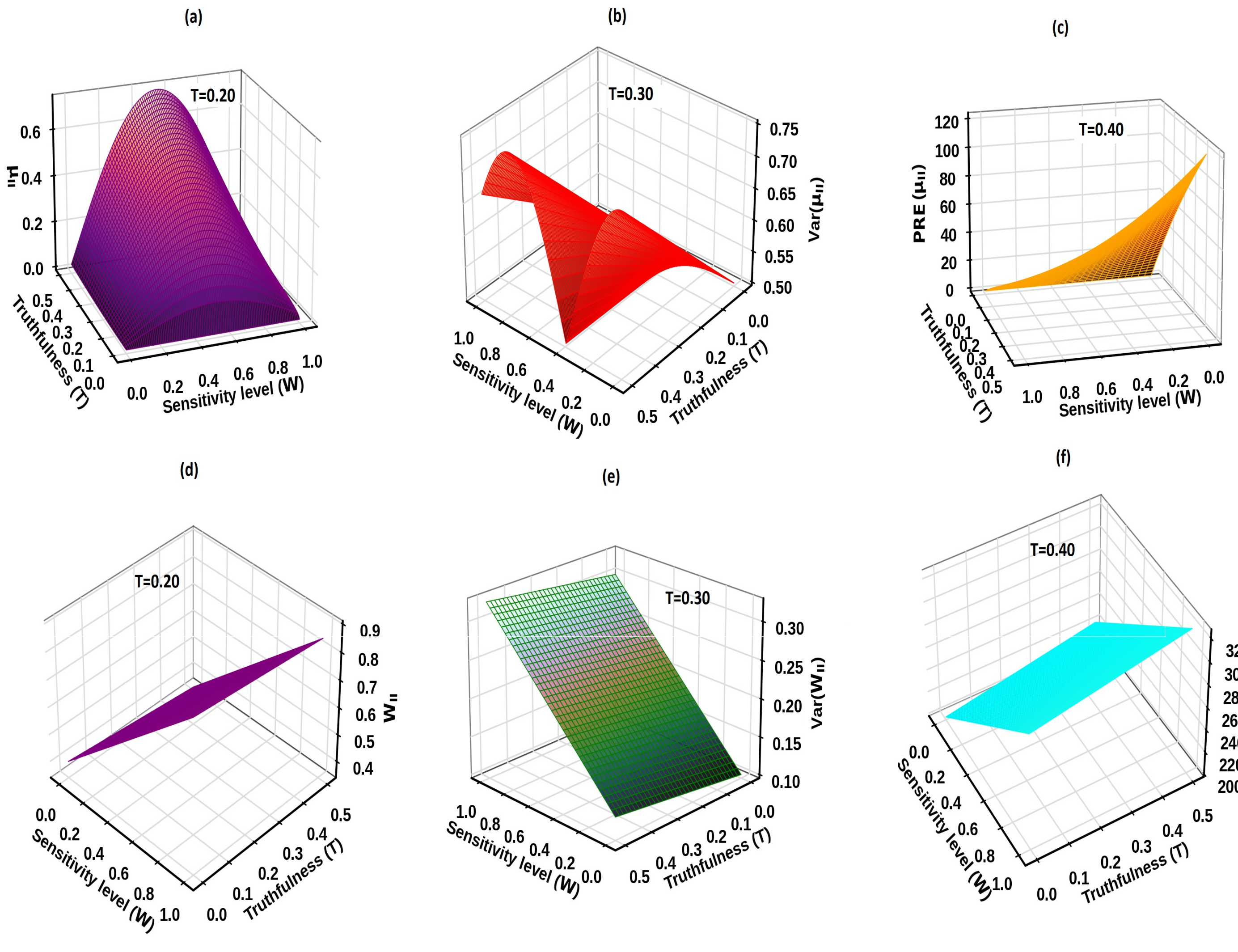

Figure 3: The figure shows 3D surface plots of DSRRF-II performance, illustrating estimators, variances, and PRE values for sensitivity levels and truthfulness.

3 Conclusion

This research introduced the innovative DSRRF-II framework, which combines additive and subtractive scrambling techniques to deliver unbiased, precise, and reliable estimates for the means and sensitivity levels of stigmatized quantitative variables, all while ensuring respondent privacy. Through rigorous theoretical analysis and extensive simulations, DSRRF-II demonstrated clear efficiency improvements over conventional randomized response methods. The comprehensive assessment of the framework highlighted its operational advantages, particularly the dual scrambling and staged randomization strategies that reduce response bias and improve estimator precision. Importantly, DSRRF-II enhances data quality without imposing additional burdens or costs on respondents, marking a significant advancement in survey methodologies. These models hold considerable value for policymakers and socioeconomic planners, offering more reliable and actionable data for informed decision-making. Furthermore, the adaptability of DSRRF-II suggests its potential for wide-ranging applications across fields such as healthcare and social research. The ease of implementation and cost-effectiveness of this methodology makes it a compelling choice for large-scale surveys. Future work could explore further refinements and adaptations of these models to various data collection environments, unlocking their full potential for diverse research contexts. This innovative approach sets the stage for more accurate, privacy-preserving data collection techniques that can address sensitive societal issues on a global scale.

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. (2006). An alternative to Warner's randomized response model. Journal of Modern Applied Statistical Methods, 5(2):328–331.

[14] 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.

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

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

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

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

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

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

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.