STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–7.
Published online in International Academic Press (www.IAPress.org)

Efficient randomized response model tailored for estimating highly sensitive characteristics

Ahmad M. Aboalkhair

Department of Quantitative Methods, College of Business, King Faisal University, Saudi Arabia

Abstract When broaching extremely delicate subjects, individuals might offer inadequate or dishonest revelations, jeopardizing data precision. To counteract this challenge, this research proposes a new and effective randomized response structure crafted to enhance the assessment of highly sensitive characteristics. The proposed framework enhances Aboalkhair's (2025) model, which has emerged as a viable substitute for Mangat's frameworks. This investigation assesses the scenarios where the proposed method performs better than Mangat's method. Through theoretical scrutiny and numerical simulations—taking into consideration partial honest disclosures—the outcomes showcase the model's heightened effectiveness. Furthermore, the article quantifies the level of privacy safeguarding provided by this innovative approach.

Keywords Randomized response technique, Privacy safeguarding, Delicate subjects, Response error, Partial honest disclosure

AMS 2010 subject classifications 62D05, 62P15

DOI: 10.19139/soic-2310-5070-2879

1. Introduction

In survey contexts, questions of a sensitive nature frequently result in non-responses or untruthful answers stemming from concerns over privacy, leading to response bias. Warner [25] originally developed the randomized response method (RRT) in order to tackle this issue, emphasizing participant confidentiality. This method enables researchers to gather precise information on sensitive subjects while reducing bias. Through RRT, participants respond to questions chosen through a random process without disclosing the exact query they are answering, safeguarding the privacy of their true responses. By assuming honest engagement within this system, the collected data maintains its integrity for thorough statistical examination.

While Warner's method successfully safeguards privacy in sensitive data collection, its reliance on randomization increases variability in estimating the prevalence of the studied attribute. Subsequent research has advanced Warner's original framework to address this inefficiency, prioritizing two objectives: reducing estimator variance and improving operational effectiveness. Some refinements focus on optimizing parameter choices within Warner's model to lower variability, while others propose entirely new estimation techniques [8, 10, 13, 15, 16, 17, 22]. Some recent innovations emphasize structural redesigns of the RRT mechanism itself, enhancing its practicality and performance in real-world applications [1, 2, 3, 4, 5, 6, 7, 11, 12, 14, 20, 21, 23, 24, 26].

Aboalkhair et al. [7] introduced a two-stage design through structural redesigns, aimed at reliably estimating Sensitive Attributes. Their foundational work assumed complete honesty among respondents. However, in contexts involving highly sensitive topics—where truthful disclosure is often compromised—this assumption may not

*Correspondence to: Ahmad M. Aboalkhair (Email: aaboalkhair@kfu.edu.sa).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

hold. To address this gap, we refined Aboalkhair's framework by explicitly incorporating mechanisms to account for partial honest disclosure. This adaptation enhances the model's real-world applicability, enabling precise measurement of sensitive attributes while rigorously preserving participant confidentiality.

2. Pioneering models

2.1. Warner's model

Warner's approach [25] provides a method to estimate the proportion π of a population possessing sensitive attributes. In this framework, the estimator for π, with adjusted notation, is derived as:

$$\hat{\pi}_w = \frac{n'/n - 1 + p_1}{2p_1 + 1}, \quad p_1 \neq 0.5, \tag{1}$$

where n' denotes the count of 'yes' responses, and the variance is:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p_1(1-p_1)}{n(2p_1-1)^2}. \tag{2}$$

2.2. Mangat's model

Mangat [20] proposed a statistically efficient randomized response (RR) design. In this model, π is estimated as:

$$\hat{\pi}_m = \frac{n'/n - 1 + p_1}{p_1}, \tag{3}$$

with the variance calculated by:

$$V(\hat{\pi}_m) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p_1)}{np_1}. \tag{4}$$

Mangat also investigated scenarios involving partial honest reporting, demonstrating that the estimator $\hat{\pi}_m$ becomes biased under such conditions. The corresponding mean square error (MSE) is derived as follows

$$\text{MSE}(\hat{\pi}_M) = \frac{\pi H(1-\pi H)}{n(1-q_1)^2} + \frac{q_1(1-\pi)\big[1 - q_1(1-\pi) - 2\pi H\big]}{n(1-q_1)^2} + \frac{\pi^2(H-1)^2}{(1-q_1)^2}, \tag{5}$$

where H represents the honest-reporting probability.

2.3. Aboalkhair's model

Aboalkhair et al. [7] introduced a pragmatic and efficient (RR) design. In their approach, each participant receives sets of "Yes" and "No" cards along with a dual-stage randomization tool. Individuals opt for a "Yes" card if they possess the sensitive attribute; otherwise, they are instructed to employ the two-stage random tool. Initially, they use device R1, which offers:

(a) "I belong to group S" with a probability of p_2, or

(b) "Use device R2" with a probability of $1 - p_2$.

If directed to R2, they engage a Warner-style device which offers:

(a) "I belong group S" with a probability of p_1,

(b) "I do not belong to group S" with a probability of $1 - p_1$.

As outlined by Aboalkhair et al. [7], the probability of responding 'Yes' (α) is given by:

$$\alpha = \pi + (1-\pi)(1-p_1)(1-p_2), \tag{6}$$

and the estimator for π is:

$$\hat{\pi} = \frac{\hat{\alpha} - (1-p_1)(1-p_2)}{1-(1-p_1)(1-p_2)}, \tag{7}$$

where $\hat{\alpha}$ denotes the sample proportion of "Yes" answers.
In the scenario where all participants respond honestly, the estimator variance is

$$V(\hat{\pi}) \;=\; \frac{\pi(1-\pi)}{n} \;+\; \frac{(1-\pi)(1-p_1)(1-p_2)}{n\big[\,1-(1-p_1)(1-p_2)\,\big]}. \tag{8}$$

While Aboalkhair et al. [7] assumed perfect honest reporting, we subsequently examine the more realistic scenario of partial honest disclosure.

3. The proposed model

Expanding on Aboalkhair's methodology [7], the new model introduces a significant enhancement by considering partial honest reporting. Let H denote the probability that a respondent with the sensitive attribute (S) answers honestly. Respondents without S are assumed to have no incentive to falsify their responses. This means:

- A respondent with the sensitive attribute (S) tells the truth with probability H; and lies (say "No") with probability $1-H$.
- A respondent without (S) proceeds exactly as in Aboalkhair's model: they use the two-stage random device.

This change requires an adjustment to the formula for the probability of a 'Yes' answer (α) as follows:

$$\alpha' = \pi H + (1-\pi)(1-p_1)(1-p_2) \tag{9}$$

The ratio of respondents with the sensitive attribute, denoted as $\hat{\pi'}$, is estimated through the following formula:

$$\hat{\pi'} = \frac{\hat{\alpha'} - (1-p_1)(1-p_2)}{1-(1-p_1)(1-p_2)}, \tag{10}$$

where p_1, p_2 and $\hat{\alpha'}$ are as previously defined.

3.1. Statistical characteristics of the proposed model

The following theorems outline the bias, variance, and mean square error (MSE) of the estimator $\hat{\pi'}$ in Equation (10).
Theorem 1 (Bias). The bias of $\hat{\pi'}$ is:

$$\mathrm{B}(\hat{\pi'}) = \frac{\pi(H-1)}{1-(1-p_1)(1-p_2)}. \tag{11}$$

Proof. By definition, $\mathrm{B}(\hat{\pi'}) = \mathrm{E}\Big[\hat{\pi'} - \pi\Big] = \mathrm{E}(\hat{\pi'}) - \pi$. Since $n(\hat{\alpha'})$ follows a binomial distribution $\mathrm{Bin}(n, \alpha')$, substituting into Equation (10) gives:

$$\mathrm{B}(\hat{\pi'}) \;=\; \frac{\alpha' - \alpha}{1-(1-p_1)(1-p_2)}. \tag{12}$$

From prior results (Equations (9) and (6)), $\alpha' - \alpha = \pi(H-1)$. Inserting this into Equation (12) yields Equation (11).

Theorem 2 (Variance). The variance of $\hat{\pi}'$ is:

$$V(\hat{\pi}') = \frac{\pi H(1-\pi H)}{n\big[1-(1-p_1)(1-p_2)\big]^2} + \frac{(1-p_1)(1-p_2)(1-\pi)\Big[1-(1-p_1)(1-p_2)(1-\pi)-2\pi H\Big]}{n\big[1-(1-p_1)(1-p_2)\big]^2}. \quad (13)$$

Proof. Starting with Equation (10), the variance is:

$$V(\hat{\pi}') = V\left(\frac{\hat{\alpha}' - (1-p_1)(1-p_2)}{1-(1-p_1)(1-p_2)}\right) = \frac{V(\hat{\alpha}')}{\big[1-(1-p_1)(1-p_2)\big]^2}. \quad (14)$$

Given that $n\hat{\alpha}' \sim \text{Bin}(n, \alpha')$ and $V(\hat{\alpha}') = \alpha'(1-\alpha')/n$ substituting this into Equation (14) produces:

$$V(\hat{\alpha}') = \frac{\alpha'(1-\alpha')}{n\big[1-(1-p_1)(1-p_2)\big]^2}. \quad (15)$$

Expanding $\alpha'(1-\alpha')$ using Equation (9) leads to

$$\alpha'(1-\alpha') = \pi H(1-\pi H) \;+\; (1-p_1)(1-p_2)(1-\pi)\Big[1-(1-p_1)(1-p_2)(1-\pi)-2\pi H\Big] \quad (16)$$

When substituted into Equation (15), results in Equation (13).

Theorem 3 (MSE). The mean square error of $\hat{\pi}'$ is:

$$\begin{aligned}\text{MSE}(\hat{\pi}') =& \frac{\pi H(1-\pi H)}{n\big[1-(1-p_1)(1-p_2)\big]^2} + \frac{(1-p_1)(1-p_2)(1-\pi)\Big[1-(1-p_1)(1-p_2)(1-\pi)-2\pi H\Big]}{n\big[1-(1-p_1)(1-p_2)\big]^2} \\ &+ \frac{\pi^2(H-1)^2}{\big[1-(1-p_1)(1-p_2)\big]^2}.\end{aligned} \quad (17)$$

Proof. The MSE combines variance and squared bias:

$$\text{MSE}(\hat{\pi}') = V(\hat{\pi}') + \big[B(\hat{\pi}')\big]^2 \quad (18)$$

Substituting Equations (13) and (11) into Equation (18) directly yields Equation (17).

3.2. *Efficiency comparison*

The suggested approach enhances Aboalkhair's model [7], which itself is a proven advancement of Mangat's randomized response technique [20]. This research primarily assesses the effectiveness of the new method compared to Mangat's original framework.

The suggested estimator outperforms Mangat's estimator under partial truthfulness if:

$$\text{MSE}(\hat{\pi}') < \text{MSE}(\hat{\pi}_M)$$

Substituting Equations (17) and (5) and simplifying algebraically, this inequality reduces to:

$$(1-p_1)(1-\pi)\big[1+(1-p_2)\big] + 2\pi H < 1$$

This demonstrates the proposed method's consistent superiority over Mangat's approach for feasible parameter values. Empirical validation via 1 confirms this theoretical efficiency gain.

Figure 1 evaluates the performance of the proposed model against Mangat's method across feasible parameter ranges. The analysis employs a sample size of $n = 100$, varying proportions of the sensitive attribute $\pi = 0.01, 0.05, 0.10, 0.20$, honest-telling probabilities $H = 0.95, 0.90, 0.70, 0.50$, and randomization parameters $p_1, p_2 = 0.6, 0.7, 0.8, 0.9$. The results demonstrate consistently positive efficiency differences, validating the superiority of the proposed approach under all tested scenarios.

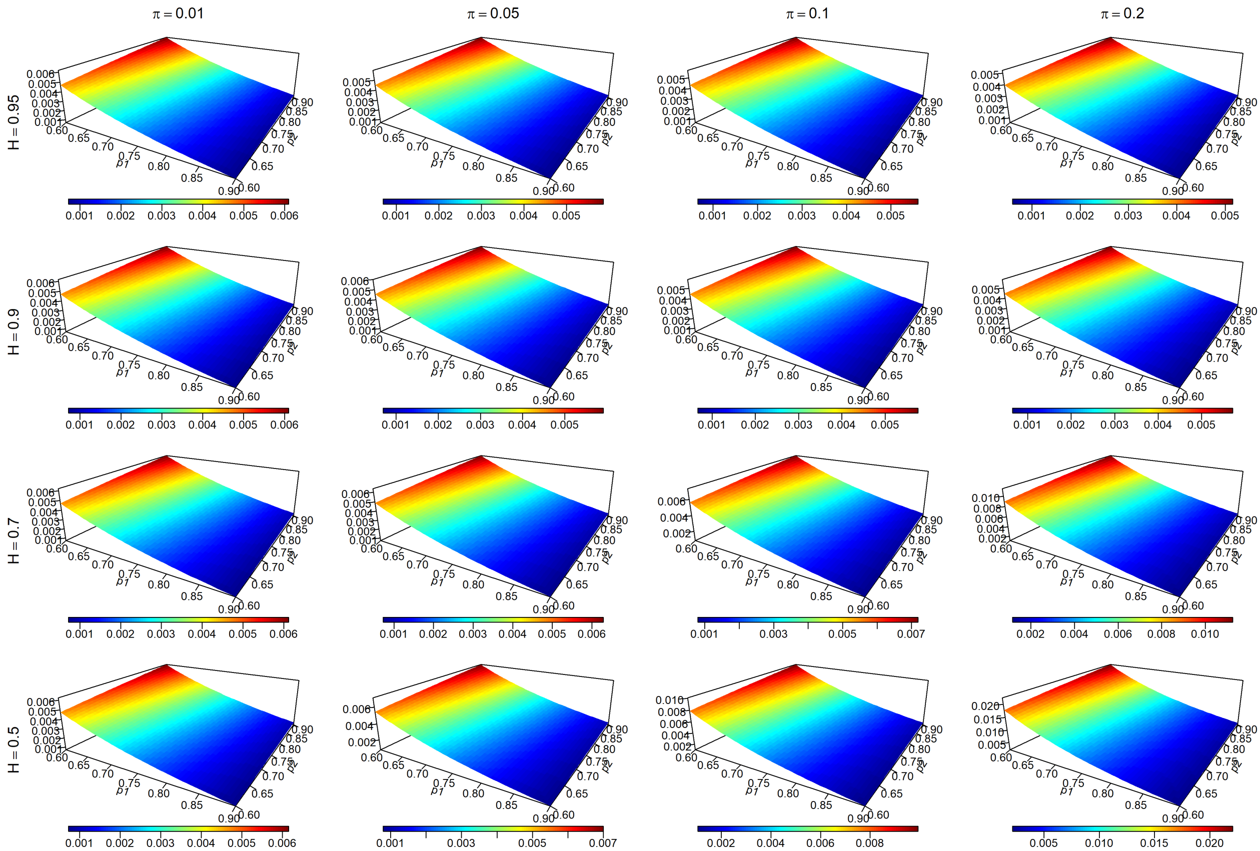

Figure 1. Efficiency comparison between the proposed model and Mangat's model under practical parameter configurations for p_1, p_2, H, and π.

Key Observations from Figure 1:

1. The proposed estimator outperforms Mangat's across all tested values of p_1, p_2, H, and π.
2. For fixed p_1, p_2, and H, the efficiency gap widens as π decreases from 0.2 to 0.01.
3. With constant p_1, p_2, and π, the difference increases as H declines from 0.95 to 0.50.
4. Lowering p_1 from 0.9 to 0.6 enhances the efficiency advantage, holding p_2, H and π constant.
5. Increasing p_2 from 0.6 to 0.9 amplifies the gap. This occurs because the proposed estimator's MSE decreases with higher p_2, while Mangat's MSE remains unaffected.

3.3. Privacy protection

Privacy protection is a cornerstone of all (RR) methodologies. Prior researches [9, 18, 19, 27] have introduced quantitative metrics to evaluate privacy levels within RR techniques. Building on Zhimin and Zaizai's approach

[27], the design probabilities are defined as:

$$P(\text{yes} \mid S) = H \quad \text{and} \quad P(\text{yes} \mid \overline{S}) = (1-p_1)(1-p_2)$$

$$P(\text{no} \mid S) = 1 - H \quad \text{and} \quad P(\text{no} \mid \overline{S}) = 1 - (1-p_1)(1-p_2)$$

and

$$P(S \mid \text{yes}) = \frac{\pi}{\pi + (1-\pi)\,(1-p_1)(1-p_2)/H}$$

$$P(S \mid \text{no}) = \frac{\pi}{\pi + (1-\pi)\,\big[1-(1-p_1)(1-p_2\big]/(1-H)}$$

The privacy metric $M_P(R)$ is formulated as:

$$M_P(R) = \left|\, 1 - \tfrac{1}{2}\big[\tau(\text{yes}) + \tau(\text{no})\big]\,\right|, \tag{19}$$

where

$$\tau(\text{yes}) \;=\; \frac{H}{(1-p_1)(1-p_2)} \quad \text{and} \quad \tau(\text{no}) \;=\; \frac{1-H}{1-(1-p_1)(1-p_2)}$$

As shown in [27], lower values of $M_P(R)$ correspond to stronger privacy safeguards. This negative correlation implies that minimizing $M_P(R)$ enhances respondent confidentiality in the RR framework.

4. Discussion

This research advances Aboalkhair's methodology [7] by incorporating an assumption of partial honest disclosure into the analytical framework. While this adaptation alters the estimator's statistical characteristics and the perceived sensitivity of the target variable, both the original and modified approaches employ the same randomization device. Thus, they share identical procedural demands in terms of implementation difficulty, participant effort, and validation rigor. The enhanced efficiency of the proposed model is most evident when estimating highly sensitive traits susceptible to underreporting—such as criminal activity, non-conventional sexual behaviors, substance abuse, mental health challenges, prejudicial attitudes, financial fraud, stigmatized medical conditions, or ethical breaches. These are domains where conventional surveys frequently produce biased or incomplete data due to respondents' hesitancy to disclose sensitive information candidly, often stemming from fear of social stigma or privacy violations.

The ethical implementation of the randomized response (RR) technique demands a careful equilibrium between collecting sensitive data and upholding participants' rights. Researchers must prioritize ensuring participants fully understand the method's purpose, mechanics, and voluntary nature, including their unrestricted right to withdraw. Transparency about how data will be used, analyzed, and disseminated is critical, as participants should be informed of their role in advancing the study's objectives. To mitigate potential distress, proactive assessment of the psychological impact of sensitive questions is essential, alongside safeguards such as anonymization and access to support resources. Crucially, robust privacy protections, including guarantees against re-identification of responses, must be emphasized to preserve confidentiality and foster participant trust. By addressing these ethical considerations holistically, researchers not only maintain integrity but also enhance the validity and reliability of sensitive data collection, bridging ethical rigor with meaningful research outcomes.

The selection of probabilities p_1 and p_2 must strike a balance between maximizing estimation accuracy and safeguarding participant privacy. This involves strategically optimizing these parameters to reduce the privacy metric (Equation 19), thereby minimizing potential confidentiality risks. By doing so, the design promotes truthful responses to sensitive questions while mitigating respondents' apprehensions, fostering trust and cooperation in the survey process.

5. Limitations and research opportunities

A primary limitation of this design is its effectiveness when dealing with highly sensitive attributes, particularly in cases where respondents may be reluctant to provide truthful information. Traditional randomized response models, which rely on the assumption of complete participant honesty, prove to be less effective in such contexts. However, when implementing Aboalkhair's model [7] in these scenarios, a limitation of the proposed framework becomes apparent: its estimator shows a higher Mean Squared Error (MSE) compared to Aboalkhair's method. This variation highlights the necessity for future research to devise more advanced randomized response techniques tailored specifically for extremely sensitive subjects, focusing on strategies to enhance precision through MSE reduction.

Acknowledgement

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU252126].

REFERENCES

1. A. M. Aboalkhair, A. M. Elshehawey, and M. A. Zayed, *A new improved randomized response model with application to compulsory motor insurance*, Heliyon, vol. 10, no. 5, e27252, 2024.
2. A. M. Aboalkhair, M. A. Zayed, A. H. Al-Nefaie, M. Alrawad, and A. M. Elshehawey, *A novel efficient randomized response model designed for attributes of utmost sensitivity*, Heliyon, vol. 10, no. 20, e39082, 2024.
3. A. M. Aboalkhair, M. A. Zayed, T. Elbayoumi, A. H. Al-Nefaie, M. Alrawad, and A. M. Elshehawey, *An innovative randomized response model based on a customizable random tool*, PLOS One, vol. 20, no. 4, e0319780, 2025.
4. A. M. Aboalkhair, E.-E. El-Hosseiny, M. A. Zayed, T. Elbayoumi, M. Ibrahim, and A. M. Elshehawey, *Estimating concealment behavior via innovative and effective randomized response model*, Statistics, Optimization & Information Computing, vol. 14, no. 1, pp. 183–192, 2025.
5. A. M. Aboalkhair, E.-E. El-Hosseiny, M. A. Zayed, T. Elbayoumi, M. Ibrahim, and A. M. Elshehawey, *Streamlined randomized response model designed to estimate extremely confidential attributes*, Statistics, Optimization & Information Computing, 2025. https://doi.org/10.19139/soic-2310-5070-2644.
6. A. M. Aboalkhair, E.-E. El-Hosseiny, M. A. Zayed, T. Elbayoumi, M. Ibrahim, and A. M. Elshehawey, *Enhanced-efficiency randomized response model: A simplified framework*, Statistics, Optimization & Information Computing, 2025. https://doi.org/10.19139/soic-2310-5070-2658.
7. A. M. Aboalkhair, *A two-stage design for superior efficiency in estimating sensitive attributes*, Statistics, Optimization & Information Computing, 2025. https://doi.org/10.19139/soic-2310-5070-2663.
8. A. Abul-Ela, and H. Dakrouri, *Randomized response model: a ratio estimator*. In Proceedings of the Survey Research Methods Section, American Statistical Association, USA, 1980.
9. H. Anderson, *Estimation of a proportion through randomized response*, International Statistical Review / Revue Internationale de Statistique, vol. 44, no. 2, pp. 213–217, 1976.
10. L. Barabesi, and M. Marcheselli, *Bayesian estimation of proportion and sensitivity level in randomized response procedures*, Metrika, vol. 72, no. 1, pp. 75–88, 2010.
11. F. Batool, J. Shabbir, and Z. Hussain, *An improved binary randomized response model using six decks of cards*, Communications in Statistics - Simulation and Computation, vol. 46, no. 4, pp. 2548–2562, 2016.
12. M. Bhargava, and R. Singh, *A modified randomization device for Warner's model*, Statistica, vol. 60, no. 2, pp. 315–322, 2000.
13. S. Ghufran, S. Khowaja, and M. J. Ahsan, *Compromise allocation in multivariate stratified sample surveys under two stage randomized response model*, Optimization Letters, vol. 8, no. 1, pp. 343–357, 2014.
14. B. G. Greenberg, A. L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz, *The unrelated question randomized response model: Theoretical framework*, Journal of the American Statistical Association, vol. 64, no. 326, pp. 520–539, 1969.
15. N. Gupta, S. Gupta, and M. Tanwir Akhtar, *Multi-choice stratified randomized response model with two-stage classification*, Journal of Statistical Computation and Simulation, vol. 92, no. 5, pp. 895–910, 2021.
16. S. H. Hsieh, S. M. Lee, and P. S. Shen, *Logistic regression analysis of randomized response data with missing covariates*, Journal of Statistical Planning and Inference, vol. 140, no. 4, pp. 927–940, 2010.
17. Z. Hussain, J. Shabbir, and M. Riaz, *Bayesian Estimation Using Warner's Randomized Response Model through Simple and Mixture Prior Distributions*, Communications in Statistics - Simulation and Computation, vol. 40, no. 1, pp. 147–164, 2010.
18. J. Lanke, *On the degree of protection in randomized interviews*, International Statistical Review / Revue Internationale de Statistique, vol. 44, no. 2, pp. 197–203, 1976.
19. F. W. Leysieffer, and S. L. Warner, *Respondent jeopardy and optimal designs in randomized response models*, Journal of the American Statistical Association, vol. 71, no. 355, pp. 649–656, 1976.

20. N. S. Mangat, *An improved randomized response strategy*, Journal of the Royal Statistical Society. Series B (Methodological), vol. 56, no. 1, pp. 93–95, 1994.
21. N. S. Mangat and R. Singh, *An alternative randomized response procedure*, Biometrika, vol. 77, no. 2, pp. 439–442, 1990.
22. N. J. Scheers, and C. M. Dayton, *Covariate randomized response models*, Journal of the American Statistical Association, vol. 83, no. 404, pp. 969–974, 1988.
23. G. N. Singh, and S. Suman, *A modified two-stage randomized response model for estimating the proportion of stigmatized attribute*, Journal of Applied Statistics, vol. 46, no. 6, pp. 958–978, 2018.
24. S. Singh, S. Horn, R. Singh, and N. S. Mangat, *On the use of modified randomization device for estimating the prevalence of a sensitive attribute*, Statistics in Transition, vol. 6, no. 4, pp. 515–522, 2003.
25. S. L. Warner, *Randomized response: A survey technique for eliminating evasive answer bias*, Journal of the American Statistical Association, vol. 60, no. 309, pp. 63–69, 1965.
26. Z. Zapata, S. A. Sedory, and S. Singh, *An innovative improvement in Warner's randomized response device for evasive answer bias*, Journal of Statistical Computation and Simulation, vol. 93, no. 2, pp. 298–311, 2022.
27. H. Zhimin, and Y. Zaizai, *Measure of privacy in randomized response model*, Quality & Quantity, vol. 46, no. 4, pp. 1167–1180, 2012.