STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 2025, pp 0–8.
Published online in International Academic Press (www.IAPress.org)

A two-stage design for superior efficiency in estimating sensitive attributes

Ahmad M. Aboalkhair

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

Abstract Examining sensitive characteristics or data that individuals are hesitant to disclose in surveys poses a challenge due to the ethical duty to protect respondent privacy. Warner's randomized response (RR) technique, while enabling confidential estimation of such attributes' prevalence in populations, suffers from increased variance as the likelihood of directly probing sensitive questions rises. To address this limitation, we propose an innovative two-stage RR framework designed to enhance practicality and statistical efficiency compared to Mangat's model, while improving credibility in real-world applications. Privacy protection metrics were computed for the proposed models, with efficiency analyses consistently showing that the new model surpasses Mangat's model in efficiency.

Keywords Surveys, sensitive attributes, response errors, randomized response technique, privacy protection metrics

AMS 2010 subject classifications 62D05, 62P15

DOI: 10.19139/soic-2663

1. Introduction

Researchers commonly rely on surveys as a primary method to evaluate attitudes and behaviors across diverse scenarios and fields. However, challenges emerge when survey inquiries touch upon sensitive or uncomfortable topics such as drug usage, mental health conditions, dishonest behavior, insurance fraud, and more. Ethical responsibilities concerning respondent privacy often complicate research endeavors in such cases. Additionally, non-sampling errors like response bias, manifested through refusal to answer or dishonest responses, pose significant concerns. Before the advent of the randomized response technique (RRT), resolving these issues remained largely unaddressed.

The randomized response (RR) method is employed in surveys to mitigate response errors when investigating illegal activities or personal matters of a sensitive nature. This approach involves using a random mechanism where respondents select a question from multiple options, at least one of which pertains to sensitive content. The chosen question remains undisclosed to the interviewer. By ensuring uniform response types for each question, no respondent or response can definitively link back to the sensitive attribute. Therefore, RR inherently assumes that participant responses are truthful and sufficiently reliable for deriving accurate statistical estimates.

The inception of randomized response can be traced back to Warner in 1965 [24]. Warner's methodology hinges on the notion that respondents are more likely to cooperate if questions allow for less revealing answers to the interviewer. By relying on a random device that prompts respondents to provide information based on probabilities, Warner's technique facilitates the collection of data on sensitive topics while upholding confidentiality. However, the estimation of proportion of a sensitive topic incurs additional variance owing to the randomization process.

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

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

Following Warner's pioneering work, various authors have expanded the randomized response technique in different directions, aiming to reduce estimate variance and enhance model efficiency. These efforts have involved parameter selection strategies, alternative estimation methods, and modifications to the original Warner model to improve performance [1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26].

Mangat [19] proposed a model where respondents directly answer "yes" if they possess the sensitive trait (S); others use Warner's device. Building on this, our study introduces a novel RR model that adopts the randomization mechanism from Mangat and Singh's earlier framework rather than Warner's. This redesigned approach enhances efficiency compared to Mangat's original model, demonstrating superior statistical performance while maintaining respondent trust.

2. Materials and methods

2.1. Warner's model

Warner's original method [24] focuses on estimating the proportion π of individuals with a sensitive attribute S. Respondents use a randomized device to select one of two statements with probabilities p_1 and $1 - p_1$:

(a) "I belong to the sensitive group S"

(b) "I do not belong to the sensitive group S"

Respondents answer "yes" or "no", without revealing the chosen statement. The MLE for π is:

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

where n' is the number of 'yes' responses. Under truthful reporting, this estimator is unbiased and its 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 and Singh's model

Mangat & Singh [20] introduced a two-stage design. Each respondent first uses device R1, which presents:

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

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

If directed to R2, the respondent uses a Warner-type device with statements:

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

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

The MLE for π is:

$$\hat{\pi}_{\text{M\&S}} = \frac{n'/n - (1-p_2)(1-p_1)}{2p_1 - 1 + 2p_2(1-p_1)} \tag{3}$$

with variance:

$$V(\hat{\pi}_{\text{M\&S}}) = \frac{\pi(1-\pi)}{n} + \frac{(1-p_2)(1-p_1)\left[1-(1-p_2)(1-p_1)\right]}{n\left[2p_1 - 1 + 2p_2(1-p_1)\right]^2} \tag{4}$$

Mangat and Singh [20] showed this model outperforms Warner's when

$$p_2 > \frac{1-2p_1}{1-p_1} \tag{5}$$

2.3. Mangat's model

Mangat [19] introduced a design where respondents answer "yes" if they possess S. Otherwise, they use a Warner device. The MLE for π is:

$$\hat{\pi}_m = \frac{\hat{\alpha} - 1 + p_1}{p_1} \tag{6}$$

where $\hat{\alpha}$ is the observed "yes" proportion. The variance is:

$$V(\hat{\pi}_m) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p_1)}{np_1} \tag{7}$$

This framework demonstrates greater efficiency than Mangat & Singh's approach when:

$$\pi > 1 - \frac{p_1(1-p_2)\left[1-(1-p_1)(1-p_2)\right]}{\left[2p_1 - 1 + 2p_2(1-p_1)\right]^2} \tag{8}$$

and outperforms Warner's if

$$\pi > 1 - \frac{p_1^2}{(2p_1-1)^2} \tag{9}$$

which holds for $p_1 > \frac{1}{3}$. In the following section, we introduce a novel model that is more effective than the previously discussed randomized response models.

3. The proposed RR model

To estimate the proportion π of individuals with a sensitive attribute S, respondents in the selected sample are instructed to answer "yes" if they belong to S; otherwise, they use the Mangat & Singh two-stage procedure with devices R1 and R2. The probability of "Yes" answer is:

$$\alpha = \pi + (1-\pi)(1-p_1)(1-p_2) \tag{10}$$

The proposed estimator for π is:

$$\hat{\pi} = \frac{\hat{\alpha} - (1-p_1)(1-p_2)}{1-(1-p_1)(1-p_2)} \tag{11}$$

where $\hat{\alpha}$ is the observed "yes" proportion.

3.1. Theoretical properties

Theorem 1. The variance of $\hat{\pi}$ is:

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p_1)(1-p_2)}{n\left[1-(1-p_1)(1-p_2)\right]} \tag{12}$$

Proof. From $n\hat{\alpha} \sim \text{Bin}(n,\alpha)$

$$V(\hat{\alpha}) = \frac{\alpha(1-\alpha)}{n} \tag{13}$$

Substituting into

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

and expanding $\alpha(1-\alpha)$ using Eq. (10) yields Eq. (12).
Theorem 2. An unbiased variance estimator is:

$$\hat{V}(\hat{\pi}) = \frac{\hat{\alpha}(1-\hat{\alpha})}{(n-1)\left[1-(1-p_1)(1-p_2)\right]^2} \tag{14}$$

Proof: By calculating the expected value of Eq. (14), unbiasedness is confirmed.

3.2. Efficiency

This section provides a numerical assessment of the proposed model's efficiency through analyzing the variance (as defined in Eq. (12)) across different parameter combinations: a sample size of $n = 100$, population proportions of individuals with the sensitive attribute $\pi = 0.01,\ 0.05,\ 0.10,\ 0.20$ and randomization probabilities $0.6 \leq p_1, p_2 \leq 0.9$.

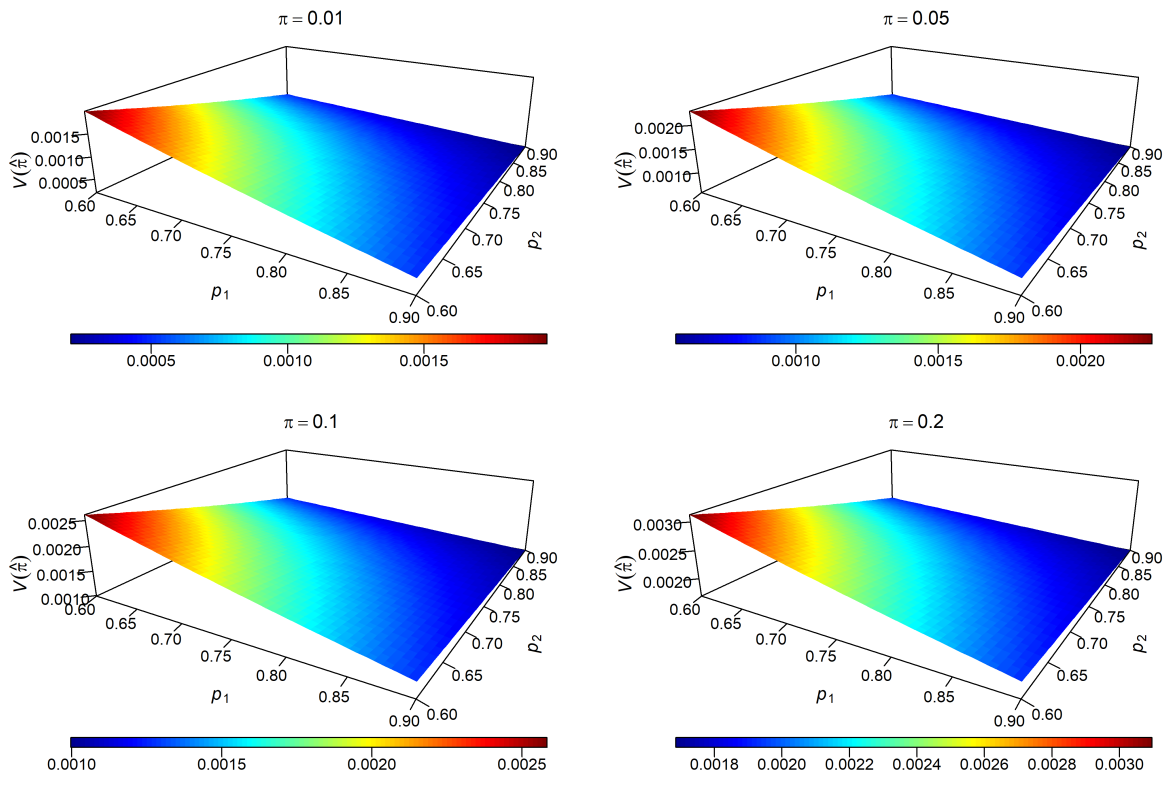

Figure 1. The variance $V(\hat{\pi})$ as a function of p_1 and p_2 $(0.6 \leq p_1, p_2 \leq 0.9)$ for some selected values of $\pi \in \{0.01,\ 0.05,\ 0.10,\ 0.20\}$.

The results, shown in Figure 1, reveal distinct trends. First, for fixed p_1 and p_2, the variance decreases as π decreases, reaching its minimum at $\pi = 0.01$. Second, higher values of p_1 and p_2 consistently reduce variance for any fixed π. For instance, when $\pi = 0.01$, increasing p_1 and p_2 from 0.6 to 0.9 lowers the variance from 0.001985 to 0.000199. These findings highlight two critical conclusions: (1) the model performs most efficiently for rare sensitive attributes (e.g., $\pi = 0.01$), making it particularly suitable for highly stigmatized topics, and (2) maximizing p_1 and p_2 enhances precision, though excessively high values risk undermining respondent trust. Thus, practitioners should prioritize higher p_1 and p_2 within a range that avoids raising suspicion, ensuring both accuracy and participant cooperation.

3.3. *Efficiency comparison*

The proposed model demonstrates greater efficiency than Mangat's model when

$$V(\hat{\pi}) < V(\hat{\pi}_m)$$

By applying Eqs. (12) and (7), this efficiency condition simplifies algebraically to

$$p_2 > 0$$

Since p_2 is inherently positive in all practical implementations, the model guarantees superior efficiency compared to Mangat's framework without requiring additional constraints.

To empirically validate this theoretical finding, Figure 2 presents a numerical comparison between the suggested estimator and Mangat's estimator. The analysis uses a sample size of $n = 100$, randomization probabilities $0.6 \leq p_1, p_2 \leq 0.9$, and population proportions $\pi = 0.01,\ 0.05,\ 0.10,\ 0.20$. Relative efficiency is computed as the ratio of the Mangat's variance to that of the suggested model, where values exceeding 1 indicate superior performance of the latter.

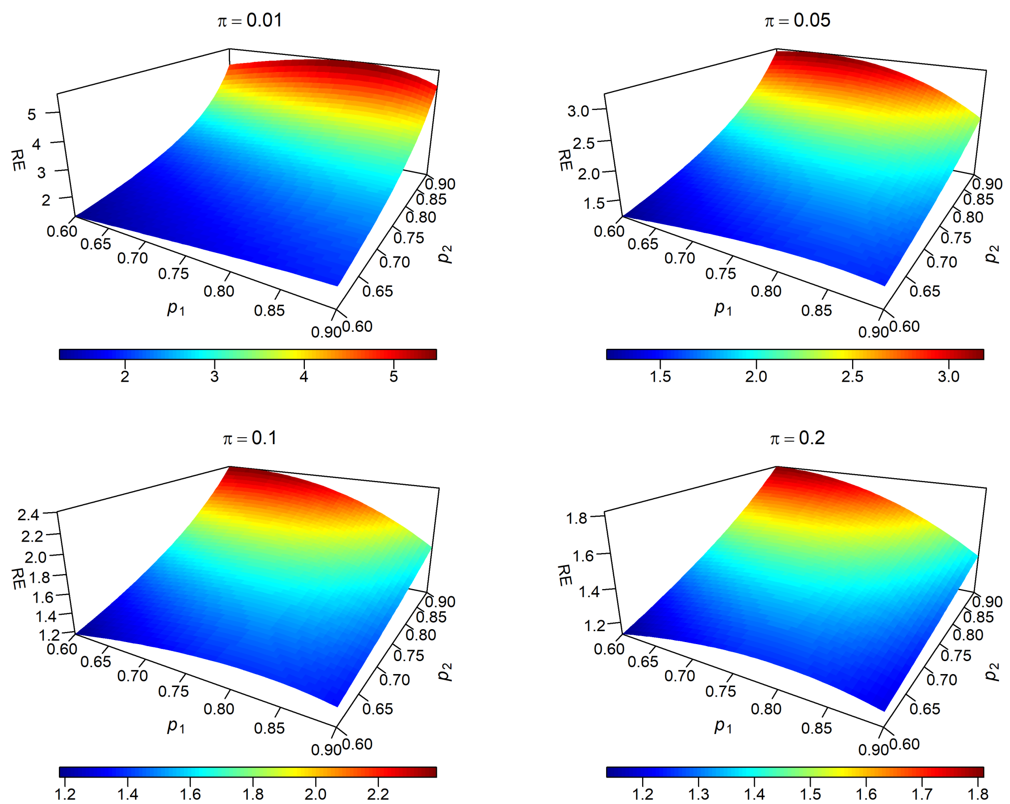

Figure 2. The relative efficiency (RE) of Mangat's model compared to the proposed model as a function of p_1 and p_2 $(0.6 \leq p_1, p_2 \leq 0.9)$ for some selected values of $\pi \in \{0.01,\ 0.05,\ 0.10,\ 0.20\}$.

Key findings include:

1. The proposed estimator outperforms Mangat's across all parameter combinations, with relative efficiency values ranging from 1.2874 (lowest) to 13.0968 (highest). For instance, at $\pi = 0.01$, $p_1 = 0.6$, and $p_2 = 0.9$ the proposed estimator is 13.09 times more efficient than Mangat's.

2. Efficiency gains increase as π decreases (e.g., from $\pi = 0.20$ to $\pi = 0.01$), highlighting the model's advantage for highly sensitive, rare attributes.

3. Higher p_1 improves efficiency (e.g., at $\pi = 0.01$, increasing p_1 from 0.6 to 0.9 boosts efficiency from 3.3753 to 6.0251 for $p_2 = 0.9$).

4. Lower p_2 enhances efficiency (e.g., at $\pi = 0.01$, reducing p_2 from 0.9 to 0.6 increases efficiency from 6.0251 to 13.0968 for $p_1 = 0.6$).

4. Measure of privacy

Randomized response (RR) models are inherently crafted to protect respondent confidentiality in surveys. Numerous methodologies have been proposed to quantify and enhance privacy in such models [8, 17, 18, 26]. Following the framework of [26], the design probabilities are:

$$P\left(yes|S\right) = 1 \quad \text{and} \quad P(yes|\overline{S}) = \left(1 - p_1\right)\left(1 - p_2\right)$$

$$P\left(no|S\right) = 0 \quad \text{and} \quad P\left(no\middle|\overline{S}\right) = 1 - \left(1 - p_1\right)\left(1 - p_2\right)$$

and

$$P\left(S|yes\right) = \frac{\pi}{\pi + (1-\pi)P(yes|\overline{S})/P(yes|S)}$$

$$P\left(S|no\right) = \frac{\pi}{\pi + (1-\pi)P(no|\overline{S})/P(no|S)}$$

Consequently, the privacy protection measure can be expressed as:

$$M_P(R) = \left|1 - \frac{1}{2}\ \{\tau(yes) + \tau(no)\}\right|$$

Thus

$$M_P(R) = 1 - \left[2(1-p_1)(1-p_2)\right]^{-1} \tag{15}$$

According to Zhimin and Zaizai [26], lower values of $M_P(R)$ indicate stronger privacy protection. As $M_P(R)$ approaches zero, respondents' confidentiality is better preserved, ensuring minimal linkage between their responses and sensitive attributes.

5. Guidelines for real-world application

To effectively apply the proposed Randomized Response (RR) model for estimating the proportion π of individuals with the sensitive attribute S, the following comprehensive guidelines should be followed. These are crafted to ensure a balance between statistical accuracy, protection of respondent privacy, and practical implementation.

1. Setup & probabilities: The survey administrator defines two probabilities: p_1 (probability of selecting the sensitive statement in the second stage) and p_2 (probability of selecting the sensitive statement in the first stage).

2. Sampling: A random sample of size n is selected.

3. Pre-experiment preparation: Several days before the experiment, participants were informed about and agreed to the location, date, and time.

4. Informed consent: Written informed consent should be obtained from every participant.

5. Initial briefing: At the start of the session, participants received a short presentation explaining the entire procedure and emphasizing how the design robustly protects their privacy.

6. Materials: The experiment utilized an empty box, separate sets of ’yes’ and ’no’ cards, and two spinner devices.

7. Participant procedure: Each respondent is instructed to select a ”yes” card and drop it in the box if he/she has a sensitive attribute S. If not, he/she is directed to use the two-stage random devices. The first random device set to show one of two options:

   (a) ”I have a sensitive attribute S” with a probability p_2, or

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

   So, if the first option appears, the experiment ends, and the interviewee places a ‘yes’ or a ‘no’ card into the container. If the second option appears and the second random device is used, the respondent answers one of the following two questions:

   (a) ”I have a sensitive attribute S” with a probability of p_1

   (b) ”I do not have a sensitive attribute S” with a probability of $1 - p_1$.

8. Privacy assurance: Participants complete the procedure individually behind a partition, ensuring no one else can observe their actions, before leaving the room.

9. Estimation: Using the collected sample data, estimate the proportion π of individuals possessing the sensitive attribute S and its variance, applying Eqs. (11) and (12).

By carefully adhering to these refined guidelines, researchers can successfully apply this tailored RRT model, increasing the likelihood of generating accurate and reliable estimates of sensitive trait prevalence, while maintaining high ethical standards and preserving respondent trust.

6. Ethical considerations

The randomized response (RR) technique requires a thoughtful ethical approach to balance the collection of sensitive information with the protection of participants’ rights. Key elements include ensuring that participants clearly understand the purpose of the method, voluntarily agree to take part, and are free to withdraw at any time. Researchers must emphasize transparency in how data is collected, used, and stored, while also considering the possible emotional effects of sensitive questions. Protective measures—such as anonymization and access to support resources—should be in place. Approval from ethics committees or Institutional Review Boards is essential to confirm adherence to ethical guidelines. Above all, strict privacy measures must be enforced to ensure that individual responses remain untraceable. Properly addressing these considerations allows for ethically sound and credible collection of sensitive data.

7. Discussion

Efficiency analyses demonstrate that the proposed model substantially outperforms Mangat's framework [19], particularly for rare or stigmatized traits and when utilizing higher values of p_1 paired with lower p_2. This positions the model as a reliable solution for surveys demanding high accuracy in sensitive contexts.

In randomized response techniques (RRT), perceived privacy plays a crucial role in shaping both response accuracy and participant willingness. When respondents feel that their privacy is strongly protected—meaning that their individual answers cannot be traced or inferred—they are more likely to provide truthful responses, especially when asked about stigmatized or sensitive behaviors. This increased trust enhances data quality and reduces evasive answering, thus improving the accuracy of the resulting estimates. Conversely, if the randomization process is too complex or poorly explained, respondents may not fully understand the mechanism or may doubt the promised confidentiality, leading to increased suspicion, reduced participation, or dishonest responses. This creates a practical trade-off: maximizing statistical efficiency often involves more complex designs or tighter parameter tuning, which may inadvertently reduce perceived privacy and deter honest participation. Therefore, striking a balance between privacy protection and efficiency is essential. Models that are both statistically efficient and easy to understand tend to perform better in real-world settings because they foster trust, encourage cooperation, and yield more reliable data.

Optimizing the probabilities p_1 and p_2 involves striking a balance between statistical precision and privacy safeguards. By minimizing the privacy protection measure outlined in Equation (15), researchers can enhance respondent willingness to participate while retaining data integrity. This adjustment mitigates distrust among participants, ensuring reliable data collection on sensitive topics while safeguarding respondent anonymity.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University (KFU), Saudi Arabia [Grant No. KFU252070].

Declaration of competing interest

The author declares no conflict of interest. KFU had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data availability statement

No new data was created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments

The author gratefully acknowledges KFU for technical and financial support.

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. Abul-Ela, and H. Dakrouri, *Randomized response model: a ratio estimator*. In Proceedings of the Survey Research Methods Section, American Statistical Association, USA, 1980.
8. H. Anderson, *Estimation of a proportion through randomized response*, International Statistical Review / Revue Internationale de Statistique, vol. 44, no. 2, pp. 213–217, 1976.
9. 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.
10. 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.
11. M. Bhargava, and R. Singh, *A modified randomization device for Warner's model*, Statistica, vol. 60, no. 2, pp. 315–322, 2000.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. N. S. Mangat and R. Singh, *An alternative randomized response procedure*, Biometrika, vol. 77, no. 2, pp. 439–442, 1990.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. H. Zhimin, and Y. Zaizai, *Measure of privacy in randomized response model*, Quality & Quantity, vol. 46, no. 4, pp. 1167–1180, 2012.