STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 2025, pp 0–7.
Published online in International Academic Press (www.IAPress.org)

Enhanced-Efficiency Randomized Response Model: A Simplified Framework

Ahmad M. Aboalkhair [1,6], El-Emam El-Hosseiny [2,*], Mohammad A. Zayed [3,*], Tamer Elbayoumi [1,4], Mohamed Ibrahim [5,6], A. M. Elshehawey [5]

[1]*Department of Applied Statistics and Insurance, Faculty of Commerce, Mansoura University, Egypt*
[2]*Department of Insurance and Risk Management, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia*
[3]*Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia*
[4]*Mathematics and Statistics Department, North Carolina A&T State University, 1601 East Market Street, Greensboro, NC 27411, USA*
[5]*Department of Applied, Mathematical & Actuarial Statistics, Faculty of Commerce, Damietta University, Egypt*
[6]*Department of Quantitative Methods, College of Business, King Faisal University, Saudi Arabia*

Abstract The key challenge in advancing randomized response techniques lies in enhancing efficiency while ensuring simplicity and ease of implementation. This research presents a fresh randomized response framework that achieves comparable effectiveness with fewer randomization devices, simplifying its real-world deployment. Compared to Aboalkhair's (2025) model that depend on two randomization tools, the suggested method delivers equivalent efficiency with only a single tool. The study evaluates the new model's superiority over existing approaches and establishes a measure of privacy protection. Through theoretical analysis and numerical comparisons, the results demonstrate distinct efficiency benefit of the suggested model.

Keywords Sample surveys, sensitive inquiries, randomized response technique, privacy protection, sensitive trait

AMS 2010 subject classifications 62D05

DOI: 10.19139/soic-2310-5070-2658

1. Introduction

Randomized Response (RR) serves as a statistical method employed to gather confidential and private data from individuals. Initially formulated by Warner [24], its purpose was to prompt honest responses to sensitive inquiries without exposing the identity of the respondent. This technique is particularly suitable in situations where individuals may hesitate to share such information due to concerns about societal judgment, potential targeting by specific groups, or the realization that legal consequences could follow if they admit to certain offenses. By introducing randomness into responses, the individual's identity remains protected, serving as a barrier to uphold the privacy and confidentiality of the participants.

Following Warner's introduction of the randomized response technique, numerous researchers have expanded the approach to enhance its efficiency and reduce the variance of estimates. Some scholars advocated for selecting parameter values that minimize estimator variance, while others proposed alternative estimation methods

*Correspondence to: El-Emam El-Hosseiny (Email: eaalhabashy@imamu.edu.sa). Department of Insurance and Risk Management, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia.,
Mohammad A. Zayed (Email: maazayed@imamu.edu.sa). Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia.

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

[6, 21, 14, 8, 16, 11, 15, 13, 26]. Many works have focused on boosting the method's efficiency via modifications in design [12, 20, 19, 10, 23, 9, 22, 25, 1, 2, 3, 5].

This paper presents a new randomized response model that is both efficient and user-friendly, achieved through design modifications and optimal parameter selection to minimize variance. The proposed model simplifies implementation by reducing the number of required randomization devices while maintaining efficiency. The model matches the statistical efficiency of Aboalkhair's technique [4] but reduces complexity by employing only one randomization device. As such, it serves as an efficient and streamlined alternative to the Mangat and Warner designs.

2. Formerly groundbreaking models

2.1. Warner's model

Warner's pioneering work [24] established an indirect estimation procedure for determining the occurrence rate of a sensitive trait A within a population (π), eliminating the need for direct disclosure of individual status (A or non-A). The methodology employs a randomization mechanism (e.g., a spinner device) that presents participants with two statements that are mutually exclusive: (a) "I have attribute A" (b) "I do not have attribute A".

These statements appear with predefined probabilities p and q, respectively. Participants privately select a statement through the randomizing device and provide a binary response ("yes" or "no") that reflects both the selected statement and their true status, without revealing which statement was chosen. Within this framework, Warner derived the estimate for π, which can be expressed in modified notation as:

$$\hat{\pi}_w = [\hat{\alpha} - q][1 - 2q]^{-1} \qquad\qquad q \neq 0.5 \tag{1}$$

where $\alpha = n'/n$ denoting the observed 'yes' response rate. The variance is calculated as:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{q[1-q][1-2q]^{-2}}{n} \tag{2}$$

2.2. Mangat's model

Mangat [19] proposed a simplified randomized response technique where participants possessing the sensitive attribute A are instructed to respond "yes" directly. Those without the attribute use Warner's randomization device instead. Within Mangat's model, the estimate for π is:

$$\hat{\pi}_M = [\hat{\alpha} - 1 + p][p]^{-1} \tag{3}$$

with the estimator variance expressed as:

$$V(\hat{\pi}_M) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)q[1-q]^{-1}}{n} \tag{4}$$

Mangat [19] proved this design outperforms Warner's model in efficiency when:

$$\pi > 1 - p^2[2p-1]^{-2} \tag{5}$$

This inequality holds true for all $p > \frac{1}{3}$.

2.3. Aboalkhair's design

Aboalkhair et al. [4] introduced a streamlined randomized response approach in which participants are instructed to respond "yes" if they possess the sensitive trait A. Otherwise, they are guided to utilize a two-step randomization

process. In the initial step, they either:
(a) Pick a 'No' card (with probability p) or
(b) Proceed to the second step (with probability q).
If they advance to the second step, they then choose between:
(a) Selecting a 'No' card (with probability p_1) or
(b) Selecting a 'Yes' card (with probability q_1).
As per Aboalkhair's methodology [5] (with modified notation), the estimator for π is:

$$\hat{\pi}_A = [\hat{\alpha} - q_1 q][1 - q_1 q]^{-1} \qquad q_1 q \neq 0.5 \tag{6}$$

with the estimator variance expressed as:

$$V(\hat{\pi}_A) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)q_1 q[1-q_1 q][1-q_1 q]^{-2}}{n} \tag{7}$$

Aboalkhair et al. [4] proved that their design outperforms Mangat's model in efficiency when:

$$q < 1$$

Section 3 proposes a novel design that achieves comparable efficiency to Aboalkhair's method [4], while offering the advantage of utilizing a single randomization mechanism instead of two.

3. The proposed RR model

To determine the proportion of individuals with the delicate attribute A, a random sample of 'n' participants is chosen. Each participant is given a set of 'Yes' cards, a set of 'No' cards, and a randomization tool. They are instructed to select a "yes" card if they have the sensitive trait; otherwise, they are directed to choose a card based on the outcome of the random device, which involves the following two options, depending on their true status concerning the delicate trait:
(a) Selecting a 'No' card (with probability $p(1+q)$) or
(b) Selecting a 'Yes' card (with probability q^2).
The probabilities $p(1+q)$ and q^2 are structured to sum to 1 (where $q = 1 - p$). The participant places the selected card into a container without revealing to the interviewer which card was picked or whether the random device was used.
The probability of a "Yes" card being submitted is:

$$\alpha = \pi + (1-\pi)q^2 \tag{8}$$

In this scenario, the estimation for $(\pi)$is:

$$\hat{\pi} = [\hat{\alpha} - q^2][1 - q^2]^{-1} \qquad q^2 \neq 0.5 \tag{9}$$

where $\hat{\alpha}$ is the ratio of 'yes' obtained from the sample.

3.1. Proposed estimator properties

Theorem 1. The proposed estimator $\hat{\pi}$ has a variance defined by:

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)q^2[1-q^2]^{-1}}{n} \tag{10}$$

Proof. Referring to Eq. (9),

$$V(\hat{\pi}) = V\left([\hat{\alpha} - q^2][1-q^2]^{-1}\right) = V(\hat{\alpha})[1-q^2]^{-2} \tag{11}$$

Considering $n\hat{\alpha} \sim Bin(n, \alpha)$,

$$V(\hat{\alpha}) = \frac{\alpha(1-\alpha)}{n} \tag{12}$$

Substituting Eq. (12) into Eq. (11) gives:

$$V(\hat{\pi}) = \frac{\alpha(1-\alpha)[1-q^2]^{-2}}{n} \tag{13}$$

Using Equation (8), the term $\alpha(1-\alpha)$ expands to:

$$\alpha(1-\alpha) = \pi(1-\pi)[1-q^2]^2 + (1-\pi)q^2[1-q^2] \tag{14}$$

Substituting Eq. into Eq. (13) yields the final variance expression (12). □

Theorem 2: The variance $V(\hat{\pi})$ has an unbiased estimator defined by:

$$\hat{V}(\hat{\pi}) = \frac{\hat{\alpha}(1-\hat{\alpha})[1-q^2]^{-2}}{(n-1)} \tag{15}$$

Proof. The proof is established by computing the expectation on both sides of Equation (15).

3.2. *Efficiency Comparison*

The suggested model offers a similar level of efficacy when compared to Aboalkhair's method [4] while utilizing fewer randomization devices, presenting an effective alternative to the seminal randomized response techniques introduced by Warner and Mangat. Previous efficiency analyses by Mangat [19] demonstrated superior performance of his model over Warner's approach [24]. Consequently, this study focuses on evaluating the efficiency of the proposed model relative to Mangat's framework [19].

Theorem 3. The introduced estimator shows superior efficiency compared to Mangat's approach in all scenarios.

Proof. The estimator $\hat{\pi}$ achieves superior efficiency to Mangat's estimator $\hat{\pi}_M$ iff

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

which simplifies to:

$$\frac{\pi(1-\pi)}{n} + \frac{(1-\pi)q^2[1-q^2]^{-1}}{n} < \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)q[1-q]^{-1}}{n}$$

Canceling common terms reduces the inequality to:

$$q^2[1-q^2]^{-1} < q[1-q]^{-1}$$

Simplifying further yields:

$$q < 1$$

which is inherently true for valid probability values of q. Thus, the introduced estimate always greater efficiency relative to Mangat's estimate. □

Figure 1 highlights the efficiency advantage of the proposed model over Mangat's model for every q value, where positive differences indicate superiority. It shows that:

- The introduced estimate always exhibits greater efficiency relative to Mangat's for every value of q.
- Although the introduced estimate attains its highest efficiency at $p = 0.9$, the optimal balance between variance and privacy is achieved at $p = 0.7$.
- As q gets larger, the efficiency advantage of the new estimator becomes even greater.

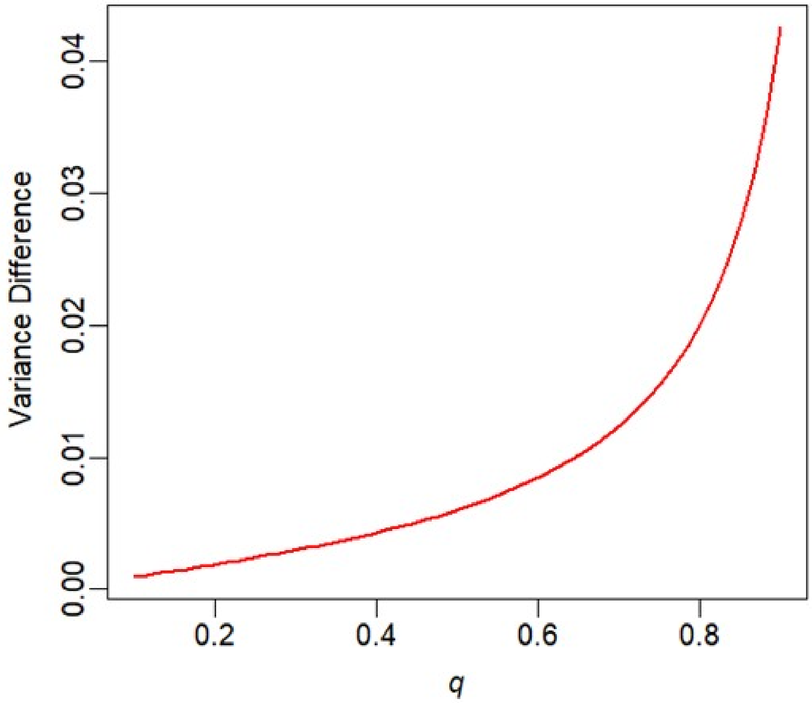

Figure 1. The variance difference between Mangat's estimator and the proposed estimator across q values

3.3. Privacy protection measure

Randomized response (RR) methodologies are inherently structured to preserve respondent confidentiality, a cornerstone of their design. Numerous strategies for enhancing privacy safeguards within RR frameworks have been explored in prior work [7, 17, 18, 27]. Adopting the formalism of [27], the privacy measure for Warner's model is:

$$M_w(R) = \frac{(1-2q)^2}{2q(1-q)} \tag{16}$$

Also, the measure of privacy protection for Mangat's model can be expressed as follows:

$$M_M(R) = \frac{2q-1}{2q} \tag{17}$$

And the privacy protection metric for the proposed model is derived as follows:

$$P(yes|A) = 1 \qquad \text{and} \qquad P(yes|\bar{A}) = (1-p)^2$$
$$P(no|A) = 0 \qquad \text{and} \qquad P(no|\bar{A}) = 1-(1-p)^2$$

and

$$P(A|yes) = \frac{\pi}{\pi + (1-\pi)P(yes|\bar{A})/P(yes|A)}$$
$$P(A|no) = \frac{\pi}{\pi + (1-\pi)P(no|\bar{A})/P(no|A)}$$

Hence,

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

where

$$\tau(yes) = \frac{P(yes|A)}{P(yes|\bar{A})} \qquad , \qquad \tau(no) = \frac{(no|A)}{P(no|\bar{A})}$$

then

$$M_P(R) = 1 - \frac{1}{2(1-p)^2} \tag{18}$$

As per Zhimin and Zaizai [27], the respondents privacy protection level improves when the privacy protection measure $M_P(R)$ (outlined in Eq. (18) tends towards zero.

4. Ethical Considerations

Implementing the randomized response technique (RRT), regardless of the specific model applied, necessitates thoughtful ethical oversight to ensure a balance between collecting sensitive information and safeguarding participant welfare. Key ethical requirements include obtaining informed consent, ensuring that participants clearly understand the purpose and procedure of the RRT, as well as their right to opt out. Researchers must also be transparent about the study's objectives, how the data will be used, and how findings will be shared. It is equally important to minimize any potential emotional or psychological discomfort that may arise from addressing sensitive topics. Prior approval from an ethics committee or institutional review board is essential to confirm that the study meets ethical guidelines and prioritizes participant protection. Lastly, absolute confidentiality must be maintained, with robust measures in place to ensure that individual responses remain completely anonymous and untraceable.

5. Discussion

The proposed model provides a simplified alternative to Aboalkhair's model [4], achieving a similar level of efficiency with fewer randomization components. By minimizing the complexity of randomization tools, the proposed model becomes more user-friendly and straightforward. This enhancement greatly boosts practical relevance in contrast to Aboalkhair's framework. The approach showcases wide-ranging versatility in addressing sensitive areas like mental health issues, substance abuse, discriminatory behaviors, financial fraud, traumatic events, stigmatized health conditions, unethical practices, and other socially sensitive contexts.

Aboalkhair's model [4] was utilized in an experimental study to estimate the rate of COVID-19 non-disclosure among university students. The study involved a randomly selected sample of Saudi undergraduate students who experienced the COVID-19 pandemic during their time at university. Given that the proposed model is a simplified variant of Aboalkhair's original model—already proven in real-world applications—its implementation is largely identical. The main distinction lies in the simplification of the procedure, where respondents follow a single-step randomization process instead of the original two-step method.

The choice of probabilities p should achieve a harmonious blend of statistical effectiveness and privacy protection. By reducing the privacy protection measure in Eq. (18), researchers can boost willingness to participate while safeguarding data utility. This fine-tuning decreases respondent distrust while optimizing the precision of collecting sensitive information.

6. Limitations and Future Research

Presenting the proposed model solely as a simplified replacement for Aboalkhair's model in scenarios where complete honesty prevails is a constraint. However, in instances concerning highly sensitive topics like sexual conduct, illicit activities, racial bias, unethical actions, or when respondents lack trust in the model, there is a potential for incomplete truthful reporting to occur [1]. This presents a potential avenue for future research: developing a tailored variant of the model that performs effectively even in the absence of full disclosure, thereby expanding its utility in handling profoundly sensitive attributes.

Author Contributions

"Conceptualization; methodology; visualization; validation; formal analysis; writing—original draft preparation; writing—review and editing; Ahmad M. Aboalkhair, El-Emam El-Hosseiny, Mohammad A. Zayed, Tamer Elbayoumi, Mohamed Ibrahim and A. M. Elshehawey; funding acquisition, El-Emam El-Hosseiny. All authors have read and agreed to the published version of the manuscript."

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The authors declare that the data supporting the findings of this study are available within the article.

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-2310-5070-2522.
6. 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.
7. H. Anderson, *Estimation of a proportion through randomized response*, International Statistical Review / Revue Internationale de Statistique, vol. 44, no. 2, pp. 213–217, 1976.
8. 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.
9. 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.
10. M. Bhargava, and R. Singh, *A modified randomization device for Warner's model*, Statistica, vol. 60, no. 2, pp. 315–322, 2000.
11. 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.
12. 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.
13. 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.

14. 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.
15. Z. Hussain, S. A. Cheema, and I. Hussain, *An Improved Two-stage Stratified Randomized Response Model for Estimating Sensitive Proportion*, Sociological Methods & Research, vol. 51, no. 3, pp. 1413–1441, 2019.
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. T. Zaman, U. Shazad, and V. K. Yadav, *An efficient Hartley–Ross type estimators of nonsensitive and sensitive variables using robust regression methods in sample surveys*, Journal of Computational and Applied Mathematics, vol. 440, 115645, 2024.
27. H. Zhimin, and Y. Zaizai, *Measure of privacy in randomized response model*, Quality & Quantity, vol. 46, no. 4, pp. 1167–1180, 2012.