International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; SP-9(3): 95-98

https://www.mathsjournal.com
Received: 18-02-2024
Accepted: 23-03-2024

Bharti
ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India

Rahul Banerjee
ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India

Pankaj Das
ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India

Sarita Devi
Department of Basic Sciences, Dr. YSP UHF, CoHF Thunag, Himachal Pradesh, India

Ankita
Birsa Agricultural University, Kanke, Ranchi, Jharkhand, India

Bulbul Ahmed
School of Agriculture, Galgotias University, Greater Noida, Uttar Pradesh, India

Nitin Varshney
Department of Agricultural Statistics, Navsari Agricultural University, Navsari, Gujarat, India

Corresponding Author:
Rahul Banerjee
ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India

Randomized response technique in agricultural surveys

Bharti, Rahul Banerjee, Pankaj Das, Sarita Devi, Ankita, Bulbul Ahmed and Nitin Varshney

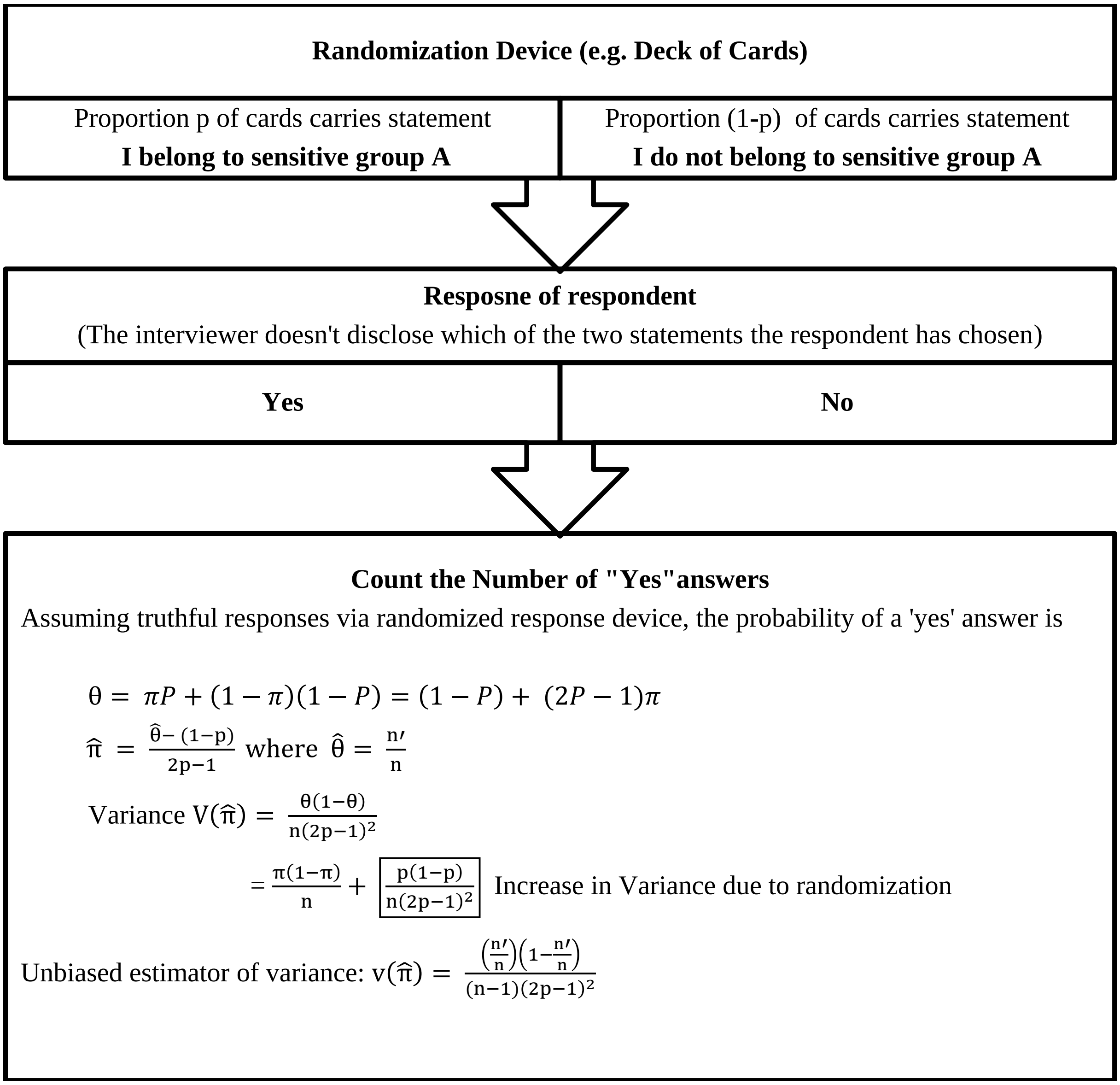
DOI: https://doi.org/10.22271/maths.2024.v9.i3Sb.1734

Abstract
Sample surveys rely on representative subsets to infer characteristics of entire populations. However, sensitive topics, like illegal activities in agriculture, can lead to biased responses or refusals. The Randomized Response Technique (RRT) addresses this challenge by guaranteeing respondent anonymity through randomization, encouraging truthful disclosure. This paper explores various RRT models and their specialized applications, particularly within the agricultural sector and beyond.

Keywords: Randomized response technique, agriculture, sensitive questions, privacy protection

1. Introduction
Sample surveys are extensively utilized to gather quality data in diverse fields such as agriculture, horticulture, forestry, public health research, and beyond. This data is used to estimate population parameters, enabling accurate statistical inferences about the target population. However, in real life, there are situations where surveys are conducted to gather data on sensitive subjects e.g., estimating the population proportion involving in illegal activities. Surveys that directly inquire about sensitive topics can often lead to the dissemination of inaccurate information or prompt individuals to refuse to respond.
The Randomized Response Technique (RRT) in such surveys offers a pathway to gather sensitive information while safeguarding the privacy of respondents. The random device (coins, chits, etc.) in the randomized response technique conceals individual responses and protects the privacy of respondents and respondents are more likely to give answers truthfully. The RRT operates on three main principles: randomization, privacy protection, and statistical analysis. It works by giving respondents a random way to answer sensitive questions, like flipping a coin or rolling a die. This disconnects their answers from their real behavior, giving them anonymity and encouraging honesty. In agriculture, where honest reporting can be hindered by fears of criticism or legal issues, the RRT is crucial. This article explores how the RRT functions in agricultural surveys, explaining its methods, advantages, and practical applications in different fields.

2. Understanding the Randomized Response Technique
2.1 Warner's Randomized Response Model: Warner (1965) [16] pioneered the randomized response technique for sensitive surveys, aiming to address biases from non-response and social desirability when sensitive questions are asked. He devised an innovative interviewing procedure to diminish evasive answers. The model built on the idea that respondents might be more cooperative when asked for information anonymously through a randomization device rather than through direct questioning. The randomization device could be a deck of cards, chits, or a spinning wheel. This device establishes a stochastic relationship between the question and the respondent's response, ensuring confidentiality of the respondent.
The procedure involves employing two statements, each dividing the population into mutually exclusive and complementary classes, say A and not A.



Warner's method has its limitations. The variance $v(\hat{\pi})$ consists of two components: (i) variance of estimator from the direct survey and (ii) variance component arises due to the use of randomization device. The efficiency of the randomized response technique is always less than the direct survey method, but it increases respondent cooperation for sensitive topics. If $p = 0$ or $p = 1$, Warner's model becomes equivalent to a direct survey method. However, If $p = ½$, although cooperation is m aximized, the estimator π is undefined. Hence, it is preferable to select p close to 1 or 0. Efforts has been made by researchers to overcome these limitations e.g. Horvitz *et al*. (1967) [1] and Greenberg *et al*. (1969) [8] combined a sensitive question of interest and another question that is completely unrelated to the sensitive topic to propose an unrelated-question RR design. Bhargava and Singh (2000) [2] introduced a modified randomization device for the Warner's RR design while Kim and Warde (2004) proposed stratified Warner's randomized response model.

2.2 Unrelated Question Model: Though the unrelated question model was initially proposed by Horvitz *et al*. (1967) [1], but its theoretical framework was further developed by Greenberg *et al*. (1969) [8]. Despite employing a randomization device, some respondents may feel uncomfortable answering questions in Warner's RR design. The confidence of the respondents, and thus the likelihood of truthful answers, might be increased by using two unrelated questions: one pertaining to the sensitive attribute, and the other to a non-sensitive characteristic. Horvitz *et al*. (1967) [1] and Greenberg *et al*. (1969) [8] modified Warner's method by including a non-sensitive question alongside a sensitive one, aiming to enhance the cooperation of the respondents and the accuracy of their responses. For example, a sensitive question could be: "Have to ever involved in encroachment activities?" (with probability p of selecting this question) and a non-sensitive question could be: "Is your favourite colour blue or red?" (with probability $1 - p$ of selecting this question). Respondents randomly select one of two unrelated questions, so that the mutually exclusive and complementary properties of the Warner technique no longer apply. Edgell *et al*. (1992) [5] found that compared to Warner's RRT, the unrelated-question model is more statistically efficient. This efficiency further improves when the population parameters of the non-sensitive question are known. There are two cases: the first is when the proportion of non-sensitive characteristic (π) in the population is known, and the second is when it is unknown.

Case I **π_y known**		**Case II** **π_y unknown**			
One randomized device		Two randomized device			
		R1		**R2**	
"Belong to sensitive group" (with prob p)	"Belong to non-sensitive group" (with prob 1-p)	"Belong to sensitive group" (with prob p_1)	"Belong to non-sensitive group" (with prob 1 - p_1)	"Belong to sensitive group" (with prob p_2)	"Belong to non-sensitive group" (with prob 1 - p_2)
$\hat{\pi}_g = \frac{\hat{\theta}\,(1-p)\pi_y}{p}$ $V(\hat{\pi}_g) = \frac{\theta(1-\theta)}{np^2}$ where $\theta = p\pi + (1-p)\pi_y$ $v(\hat{\pi}_g) = \frac{\hat{\theta}(1-\hat{\theta})}{(n-1)p^2}$		$\theta = P(yes) = \pi p_i + (1-p_i)\pi y$ $n_1' =$ probability of 'yes' answer in i − th sample (i = 1,2) $\hat{\pi}_g = \frac{1}{p_1-p_2}\left[(1-p_2)\hat{\theta}_1 - (1-p_1)\hat{\theta}_2\right], \hat{\theta}_1 = \frac{n_i'}{n_i}, i = 1,2$ $V(\hat{\pi}_g) = \frac{1}{(p_1-p_2)^2}\left[(1-p_2)^2\frac{\theta_1(1-\theta_1)}{n_1} - (1-p_1)^2\frac{\theta_2(1-\theta_2)}{n_2}\right]$ $v(\hat{\pi}_g) = \frac{1}{(p_1-p_2)^2}\left[(1-p_2)^2\frac{\theta_1(1-\theta_1)}{n_1-1} - (1-p_1)^2\frac{\theta_2(1-\theta_2)}{n_2-1}\right]$			
		For reducing $v(\hat{\pi}_g)$ following rules are helpful Choose p_i either close to 0 to 1 such that $p_1 + p_2 = 1$ Choose π_y close 0 or 1 according as $\pi < 0.5$ or $\pi > 0.5$ make $\lvert\pi_y - 0.5\rvert$ as maximum on either side. Take care about likelihood of cooperation of respondents when choosing π_y. When $n_1+n_2 = n$ is fixed, choose $\frac{n_1}{n_2} = \frac{(1-p_1)\sqrt{\theta_1(1-\theta_1)}}{(1-p_2)\sqrt{\theta_2(1-\theta_2)}}$			

2.3 Mangat and Singh's Two Stage Model

Mangat and Singh (1990) proposed a two-stage randomized response procedure involving two randomization devices, R_1 and R_2. Each interviewee is provided with two randomization devices R_1 and R_2.

R1		**R2**	
"Belong to sensitive group" (with prob T)	"Go to random device R_2" (with prob 1-T)	"Belong to sensitive group" (with prob p)	"Belong to non-sensitive group" (with prob 1-p)
$\theta = p(yes) = \pi T + (1-T)[\pi p + (1-\pi)(1-p)]$ $= \pi[T + (1-T)(2p-1)] + (1-T)(1-p)$ $\hat{\pi} = \frac{\hat{\theta} - (1-T)(1-p)}{(2p-1) + 2T(1-p)} = \frac{\frac{n'}{n} - (1-T)(1-p)}{(2p-1) + 2T(1-p)}$ $V(\hat{\pi}) = \frac{\theta(1-\theta)}{n\left((2p-1) + 2T(1-p)\right)^2}$ $= \frac{\pi(1-\pi)}{n} + \frac{(1-T)(1-p)[1 - (1-T)(1-p)]}{n\left((2p-1) + 2T(1-p)\right)^2}$ $v(\hat{\pi}) = \frac{\hat{\theta}(1-\hat{\theta})}{n\left((2p-1) + 2T(1-p)\right)^2}$ This strategy can be made more efficient than Warner's model by choosing T such that $T > \frac{1-2p}{1-p}$ This model is equivalent to warner's model with probability of cards p is equal to $T + (1-T)p = p + T(1-p)$			

Furthermore, numerous researchers have developed randomized response technique for various scenarios, expanding the technique's applicability. Christofides (2003) introduced the Generalized Randomized Response (GRR) design, allowing respondents to choose from more than two response options for a single sensitive question, thereby enhancing privacy protection. Abul-Ela *et al.* (1967) [1] enhanced Warner's design for trichotomous populations, while Hsieh *et al.* (2018) [14] extended the GRR design for more than two categories, using ML methods for estimation. Research has also focused on estimating the proportions of two sensitive characteristics simultaneously. Ewemooje and Amahia (2015, 2016) [6, 7] proposed new estimators for two related sensitive traits. Greenberg *et al.* (1971) [9] adapted the RRT for quantitative responses. Gupta *et al.* (2002) [10] estimated the mean of stigmatized variables using optional RR sampling. Hsieh *et al.* (2022) [13] used a two-stage multilevel RRT to estimate monthly income. Gupta *et al.* (2022) [11] introduced an optional enhanced trust model to improve respondent privacy and model efficiency, showing its superiority to Warner's traditional model.

3. Applications of Randomized Response Techniques in Agricultural Surveys

Randomized Response Techniques offer an alternative approach to conventional methods in agricultural surveys, particularly when handling sensitive data. Various applications of RRT in agriculture and its associated domains encompass.

- **Estimation of Illegal or Sensitive Activities:** Agricultural surveys often involve questions related to illegal activities such as the use of banned pesticides or unreported income from certain crops. Respondents may be hesitant to provide truthful answers due to legal implications. Randomized response techniques help researchers get more accurate data on these sensitive topics while maintain privacy of respondents' answers.

- **Evaluation of Yield and Crop Production:** Surveys that require information on crop yield, production techniques, and land use can be sensitive, especially if related to subsidies or taxation. RRT can help lessen biases in reporting, providing more reliable data and thus, planning resource allocation.
- **Assessment of Technology Adoption:** Surveys are often conducted to inquire about farmers' adoption of new technologies, such as precision agriculture tools or genetically modified crops. RRT can reduce response bias and provide more reliable insights into the extent and factors influencing technology adoption in agricultural practices.
- **Assessment of Environmental Practices:** Farmers' adherence to environmental regulations, such as waste disposal or use of environmentally harmful practices, may not be accurately reported in traditional surveys due to social desirability bias. RRT can help obtain more honest responses about their actual practices, aiding in environmental impact assessments and policy formulation. Chong *et al* 2019 [4] conducted survey to ask sensitive questions regarding illegal waste disposal in Public health research.
- **Farmer Demographics and Socioeconomic Status**: Collecting demographic and socioeconomic data from farmers can be challenging due to concerns about privacy and social stigma associated with certain characteristics. RRT can help alleviate these concerns and improve the accuracy of data related to farmers' demographics, education, income levels, and household composition.

Randomized response techniques offer a valuable tool for enhancing the validity and reliability of data collected in agricultural surveys, ultimately supporting evidence-based decision-making, policy formulation, and agricultural development initiatives.

4. Conclusion

In agricultural surveys, respondents may not always tell the truth about sensitive topics like subsidies or excessive use of chemicals and fertilizers. That is where the Randomized Response Technique emerges as a tool to encourage honest responses. It is ingenious combination of randomization, privacy protection, and statistical analysis transcends conventional survey methodologies, helping researchers and policymakers to explore sensitive topics accurately. In India, agricultural sector significantly influences society, environment, and economy. Thus, implementing RRT effectively in surveys can prioritize privacy and evidence-based decision-making.

5. Acknowledgement

The authors are thankful to the ICAR-IASRI for providing facilities for carrying out the present research. The authors are also thankful to editor of journal and the anonymous reviewers for their constructing insights in uplifting the quality of the study.

6. References

1. Abul-Ela ALA, Greenberg GG, Horvitz DG. A multi-proportions randomized response model. J Am Stat Assoc. 1967;62:990-1008.
2. Bhargava M, Singh R. A modified randomization device for Warner’s model. Statistica. 2000;60:315-322.
3. Chaudhuri A, Mukerjee R. Optionally randomized response techniques. Calcutta Stat Assoc. Bull. 1985;34:225-230.
4. Chong ACY, Chu AMY, So MKP, Chung RSW. Asking sensitive questions using the randomized response approach in public health research: an empirical study on the factors of illegal waste disposal. Int. J Environ Res Public Health. 2019;16(6):970.
5. Edgell SE, Duchan KL, Himmelfarb S. An empirical test of the unrelated question randomized response technique. Bull Psychon Soc. 1992;30:153-156.
6. Ewemooje OS, Amahia GN. Improved randomized response technique for two sensitive attributes. Afr. Stat. 2015;10:839-852.
7. Ewemooje OS, Amahia GN. Improving the efficiency of randomized response technique for two sensitive characters. FUTA J Res Sci. 2016;12:65-72.
8. Greenberg BG, Abul-Ela ALA, Simmons WR, Horvitz DG. The unrelated question randomized response model: theoretical framework. J Am Stat Assoc. 1969;64:520-539.
9. Greenberg BG, Kuebler Jr. RR, Abernathy JR, Horvitz DG. Application of the randomized response technique in obtaining quantitative data. J Am Stat Assoc. 1971;66:243-250.
10. Gupta S, Gupta B, Singh S. Estimation of sensitivity level of personal interview survey questions. J Stat Plan Inference. 2002;100:239-247.
11. Gupta S, Zhang J, Khalil S, Sapra P. Mitigating lack of trust in quantitative randomized response technique models. Commun. Stat Simul. Comput.; c2022. p. 1-9.
12. Horvitz DG, Shah BV, Simmons WR. The unrelated question randomized response model. Proc. Soc. Stat Sect Am Stat Assoc. 1967;62:65-72.
13. Hsieh SH, Lee SM, Li CS. A two-stage multilevel randomized response technique with proportional odds models and missing covariates. Sociol. Methods Res. 2022;51:439-467.
14. Hsieh SH, Lee SM, Tu SH. Randomized response techniques for a multi-level attribute using a single sensitive question. Stat Pap. 2018;59:291-306.
15. Kim JM, Warde WD. A stratified Warner’s randomized response model. J Stat Plan Inference. 2004;120:155-165.
16. Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc. 1965;60:63-69.