Research Article GU J Sci 32(4): 1308-1320 (2019) DOI: 10.35378/gujs.479836

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

An Alternative Estimator for Unrelated Questions in Tripartite Randomized Response Technique

Olusegun S. EWEMOOJE[1,*]

[1]*Department of Statistics, Federal University of Technology, Akure, Nigeria.*

Highlights

- Unrelated question in Tripartite Randomized Response Technique (UTRRT) outperforms conventional ones.
- UTRRT is shown to be more efficient than the direct method.
- UTRRT is efficient in estimating prevalence of drug use disorders.

Article Info

Received: 07/11/2018
Accepted: 09/04/2019

Keywords

Hybrid Tripartite Randomized Response Technique
Relative Efficiency
Sensitive Characters
Unequal Probabilities

Abstract

This paper proposes an alternative randomized response technique by improving existing works on tripartite randomized response technique (TRRT) using unrelated questions. To improve the technique, this work studies TRRT under sampling with unequal probabilities with or without replacement. The study of percentage relative efficiency shows that the proposed estimator is far better than the Hybrid Tripartite Randomized Response Technique with increase in the sensitive character and that of the unrelated character. In addition, conducting survey using both the proposed technique and direct (traditional) technique shows that the proposed technique is more efficiency than the direct technique of data collection.

1. INTRODUCTION

The need to obtain accurate result from surveys is crucial with the growing need for precise data. However, the fear of being stigmatized has caused many not to response or to give false response when faced with sensitive questions such as addiction to drug, proneness to tax evasion, drunk driving, etc. This leads to bias in estimating the proportion of people involved in such acts. To reduce or eliminate this potential bias, Warner [1] proposed the Randomized Response Technique (RRT) as a survey technique to be used when asking questions about sensitive behaviors and beliefs. Since then, a lot of work has been done to improve Warner [1] work by modifying and developing new Randomized Response Techniques (RRT). These include; Mangat and Singh [2], Kim and Warde [3], Adebola *et al.* [4], Ewemooje and Amahia [5, 6], Singh and Gorey [7], Ewemooje [8], Ewemooje *et al.* [9,10] and so on. Adebola *et al.* [11] proposed a Hybrid Tripartite Randomized Response Technique (HTRRT) to take care of respondents who may not choose either "yes" or "no" which led to their third option called "undecided" response. In developing their model, they used three randomization devices, R_1, R_2 and R_3 and show that the HTRRT is more efficient than the Tripartite Randomized Response Technique (TRRT). This study aims at reducing the variance of estimator for proportion of persons belonging to the sensitive character while improving the privacy protection of the respondents. Therefore, an alternative Tripartite Randomized Response Technique is proposed using unrelated question and by modifying it design while examining its relative efficiency over the existing one. Also, the proposed technique is applied to a survey and then compare with direct method of data collection.

*e-mail: osewemooje@futa.edu.ng

2. THE HYBRID TRIPARTITE RANDOMIZED RESPONSE TECHNIQUE (HTRRT)

The Hybrid Tripartite Randomized Response Technique (HTRRT) uses three randomized devices R_1, R_2 and R_3, with each device consisting two unrelated questions. Three responses "yes", "no" and "undecided" were considered for the two questions. The probabilities of using randomized devices R_1, R_2 and R_3 were defined such that $q_1 = \frac{\alpha}{\alpha+\beta+\delta}, \alpha \neq \beta \neq \delta$ is the probability of using R_1; $q_2 = \frac{\beta}{\alpha+\beta+\delta}, \alpha \neq \beta \neq \delta$ is the probability of using R_2; $q_3 = \frac{\delta}{\alpha+\beta+\delta}, \alpha \neq \beta \neq \delta$ is the probability of using R_3 where $\alpha, \beta,$ and δ are positive real numbers with two unrelated questions each.

Their proposed unbiased estimate of the population proportion π_A was given as:

$$\hat{\pi}_{HTRRT} = \frac{\hat{\theta}(\alpha+\beta+\delta)+(\alpha P_1 - \delta P_1+\beta P_2 - \delta P_2-\alpha-\beta)\pi_U}{\alpha P_1-\delta P_1+\beta P_2 - \delta P_2+\delta}, \quad (1)$$

where P_1, P_2 and P_3 are the preset probabilities for sensitive question in each of the devices respectively and $\hat{\theta} = {n_0}/{n}$, n_0 is number of respondents that answered "yes" to sensitive character while n is the sample size. Also, π_U is the true proportion of respondents with the unrelated character.
The variance of the Hybrid Tripartite Randomized Response Technique was given as:

$$V(\hat{\pi}_{HTRRT}) = \left[\frac{\pi_A(2\pi_U-\pi_A)}{n} - \frac{\pi_U^2}{n}\right] + \left[\frac{(\alpha+\beta+\delta)[(\pi_U-\pi_U^2)(\alpha+\beta+\delta)+ (\pi_A-\pi_U)(1-2\pi_U)[(\alpha-\delta)P_1+(\beta-\delta)P_2+\delta]]}{n[(\alpha-\delta)P_1+(\beta-\delta)P_2+\delta]^2}\right], \quad (2)$$

where π_A is the true proportion of respondents with the sensitive character.

3. THE PROPOSED ALTERNATIVE TRIPARTITE RANDOMIZED RESPONSE TECHNIQUE

The newly proposed Tripartite Randomized Response Technique uses three randomized devices R_1, R_2 and R_3, each device consists of two unrelated questions as in other Tripartite Randomized Response Techniques. The probabilities of using randomized devices R_1, R_2 and R_3 are defined such that $q_1 = \frac{\alpha}{\alpha+\beta+\delta}, \alpha \neq \beta \neq \delta$ is the probability of using R_1; $q_2 = \frac{\beta}{\alpha+\beta+\delta}, \alpha \neq \beta \neq \delta$ is the probability of using R_2; $q_3 = \frac{\delta}{\alpha+\beta+\delta}, \alpha \neq \beta \neq \delta$ is the probability of using R_3 where $\alpha, \beta,$ and δ are positive real numbers with.

In selecting a sample from a finite population using simple random sampling with replacement, sensitive question "A" is being asked from each respondent in order to estimate proportion of respondents belonging to character "A". Each respondent is requested to answer "yes" if he/she belongs to character "A" if not, he/she is required to select one out of the three randomized devices R_1, R_2 and R_3, then respond "yes", "no" or "undecided" to a question drawn at random from the selected device.

If all respondents respond truthfully, the population proportion of "yes" answers for the procedure will be given by:

$$P(yes) = \theta = \frac{\alpha}{\alpha+\beta+\delta}[\pi_A + (1-P_1)\pi_U] + \frac{\beta}{\alpha+\beta+\delta}[\pi_A + (1-P_2)\pi_U] + \frac{\delta}{\alpha+\beta+\delta}[\pi_A + (1-P_3)\pi_U], \quad (3)$$

where π_A is the true proportion of people with the sensitive character and π_U is the true proportion of people with the non-sensitive character while P_1, P_2 and P_3 are the preset probabilities for sensitive question in each of the three devices respectively.

Therefore, equation (3) becomes:

$$\theta(\alpha+\beta+\delta) = \pi_A(\alpha+\beta+\delta) + [\alpha(1-P_1) + \beta(1-P_2) + \delta(1-P_3)]\,\pi_U\,),$$

Cross multiply and make π_A the subject of the formula

$$\pi_A = \frac{\theta(\alpha+\beta+\delta) - [\alpha(1-P_1) + \beta(1-P_2) + \delta(1-P_3)]\,\pi_U}{(\alpha+\beta+\delta)}$$

if $P_1 + P_2 + P_3 = 1$ and $P_3 = 1 - P_1 - P_2$, then;

The proposed unbiased estimate of the population proportion π_A is given as:

$$\hat{\pi}_A = \frac{\hat{\theta}(\alpha+\beta+\delta) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}. \quad (4)$$

3.1. Proof of Unbiasedness

$$E(\hat{\pi}_A) = E\left[\frac{\hat{\theta}(\alpha+\beta+\delta) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}\right]$$

$$E(\hat{\pi}_A) = \frac{(\alpha+\beta+\delta)E(\hat{\theta}) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}$$

$$E(\hat{\pi}_A) = \frac{\theta(\alpha+\beta+\delta) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)} \quad *$$

Substituting equation (3) in (*), we have:

$$E(\hat{\pi}_A) = \frac{\left\{\frac{\alpha}{\alpha+\beta+\delta}[\pi_A + (1-P_1)\pi_U] + \frac{\beta}{\alpha+\beta+\delta}[\pi_A + (1-P_2)\pi_U] + \frac{\delta}{\alpha+\beta+\delta}[\pi_A + (1-P_3)\pi_U]\right\}(\alpha+\beta+\delta)}{(\alpha+\beta+\delta)} - \frac{[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}$$

$$E(\hat{\pi}_A) = \frac{\pi_A(\alpha+\beta+\delta) + [\alpha(1-P_1) + \beta(1-P_2) + \delta(1-P_3)]\,\pi_U\,) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}$$

$$E(\hat{\pi}_A) = \frac{\pi_A(\alpha+\beta+\delta) + [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}$$

$$E(\hat{\pi}_A) = \frac{\pi_A(\alpha+\beta+\delta)}{(\alpha+\beta+\delta)}$$

$$E(\hat{\pi}_A) = \pi_A$$

Therefore, the proposed estimator, $\hat{\pi}_A$, is an unbiased estimator of the population parameter π_A.

3.2. Variance Estimation

The variance of the proposed estimator is as follows:

$$V(\hat{\pi}_A) = V\left[\frac{\hat{\theta}(\alpha+\beta+\delta) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\,\pi_U}{(\alpha+\beta+\delta)}\right]$$

$$V(\hat{\pi}_A) = \frac{(\alpha+\beta+\delta)^2 V(\hat{\theta})}{(\alpha+\beta+\delta)^2}$$

$$V(\hat{\pi}_A) = \frac{(\alpha+\beta+\delta)^2 \theta(1-\theta)}{n(\alpha+\beta+\delta)^2}.$$

Substituting equation (3), we have:

$$V(\hat{\pi}_A) = \frac{1}{n}\left\{\left[\frac{\alpha}{\alpha+\beta+\delta}[\pi_A + (1-P_1)\pi_U] + \frac{\beta}{\alpha+\beta+\delta}[\pi_A + (1-P_2)\pi_U] + \frac{\delta}{\alpha+\beta+\delta}[\pi_A + (1-P_3)\pi_U]\right]\left[1 - \left(\frac{\alpha}{\alpha+\beta+\delta}[\pi_A + (1-P_1)\pi_U] + \frac{\beta}{\alpha+\beta+\delta}[\pi_A + (1-P_2)\pi_U] + \frac{\delta}{\alpha+\beta+\delta}[\pi_A + (1-P_3)\pi_U]\right)\right]\right\}$$

$$V(\hat{\pi}_A) = \frac{1}{n}\left\{\left[\frac{(\alpha+\beta+\delta)\pi_A - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U}{(\alpha+\beta+\delta)}\right]\left[1 - \left(\frac{(\alpha+\beta+\delta)\pi_A - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U}{(\alpha+\beta+\delta)}\right)\right]\right\}$$

$$V(\hat{\pi}_A) = \frac{1}{n(\alpha+\beta+\delta)}\{(\alpha+\beta+\delta)^2\pi_A - (\alpha+\beta+\delta)^2\pi_A^2 - 2(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_A\pi_U + (\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]^2\pi_U^2\}$$

$$V(\hat{\pi}_A) = \frac{(\alpha+\beta+\delta)^2\pi_A(1-\pi_A)}{n(\alpha+\beta+\delta)} - \frac{2(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_A\pi_U}{n(\alpha+\beta+\delta)} + \frac{[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U\{(\alpha+\beta+\delta) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U\}}{n(\alpha+\beta+\delta)}$$

$$V(\hat{\pi}_A) = \frac{(\alpha+\beta+\delta)\pi_A(1-\pi_A)}{n} - \frac{2[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_A\pi_U}{n} + \frac{[\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U\{(\alpha+\beta+\delta) - [\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)]\pi_U\}}{n(\alpha+\beta+\delta)}. \quad (5)$$

3.3. Sampling with Unequal Probabilities with or Without Replacement Using the Proposed Alternative Tripartite Randomized Response Technique

Let $y_i = 1$, if respondent i bears a sensitive character A while $y_i = 0$, if he/she does not. Let the participation of sampled person be independent similar to that proposed by Ewemooje *et al.*[12]. Hence, the probability of "yes" becomes:

$$P(yes) = P(\theta_i = 1) = E_R(\theta_i) = \frac{\alpha}{\alpha+\beta+\delta}[y_i + (1-P_1)(1-y_i)] + \frac{\beta}{\alpha+\beta+\delta}[y_i + (1-P_2)(1-y_i)] + \frac{\delta}{\alpha+\beta+\delta}[y_i + (1-P_3)(1-y_i)] \quad (6)$$

$$P(yes) = y_i + (1-y_i)\left[\frac{\alpha(1-P_1) + \beta(1-P_2) + \delta(1-P_3)}{(\alpha+\beta+\delta)}\right].$$

Recall that $P_3 = 1 - P_1 - P_2$, therefore:

$$E_R(\theta_i) = y_i + (1-y_i)\left[\frac{\alpha+\beta+P_1(\delta-\alpha) + P_2(\delta-\beta)}{(\alpha+\beta+\delta)}\right]. \quad (7)$$

Let r_i be an unbiased estimator for y_i, then;

$$r_i = \frac{\theta_i(\alpha+\beta+\delta)-[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]}{\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)}, \tag{8}$$

where $E_R(r_i) = y_i$ for all i(s).

The variance of the unbiased estimator r_i is:

$$V_R(r_i) = \frac{(\alpha+\beta+\delta)^2 V_R(\theta_i)}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}$$

$$V_R(r_i) = \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2} \tag{9}$$

since $V_R(\theta_i) = E_R(\theta_i)[1-E_R(\theta_i)] = \frac{\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)}{\alpha+\beta+\delta}$.

In addition; $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$

$$E(\bar{r}) = E_p(\bar{y}) = \bar{Y} \text{ since } E_p(\bar{r}) = \bar{y}$$

$$V(\bar{r}) = V(\bar{y}) + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]}{n[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}$$

$$V(\bar{r}) = \frac{1}{n}\left[\theta(1-\theta) + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\right]$$

$$V(\bar{r}) = \frac{1}{n}\Big[(\alpha+\beta+\delta)\pi_A(1-\pi_A) - 2[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]\pi_A\pi_U + \frac{[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]\pi_U\{(\alpha+\beta+\delta)-[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]\pi_U\}}{(\alpha+\beta+\delta)} + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\Big]. \tag{10}$$

An unbiased estimator of $V(\bar{r})$ is given as:

$$\hat{V}(\bar{r}) = \frac{1}{(n-1)}\Big[(\alpha+\beta+\delta)\hat{\pi}_A(1-\hat{\pi}_A) - 2[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]\hat{\pi}_A\pi_U + \frac{[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]\pi_U\{(\alpha+\beta+\delta)-[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]\pi_U\}}{(\alpha+\beta+\delta)} + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\Big]. \tag{11}$$

Also, Chaudhuri and Christofides [13] suggested that the value of p chosen close to 0.5 will increase the respondent's faith in the procedure in protecting his/her privacy, nevertheless, the closer p is to 0.5, the higher the magnitudes of $\hat{V}(\bar{r})$ and the coefficient of variation of $\bar{r}$. Hence, an intelligent balance is needed in choosing p.

Therefore, agreeing with Chaudhuri [14], let a sample s be chosen in relation to the design of p with probabilities of inclusion such that;

$$\lambda_i = \sum_{s \in i} p(s) > 0 \quad for\ i \in U$$

and

$$\lambda_{ij} = \sum_{s \in i,j} p(s) > 0 \quad for\ i,j \in U \ (i \neq j)$$

$$e = \frac{1}{N}\sum_{i \in s}\frac{r_i}{\lambda_i} \equiv \bar{r} = \frac{1}{n}\sum_{i \in s} r_i$$

$$E_p(e) = \frac{1}{N}\sum_{i=1}^{N} r_i = \frac{R}{N} = \bar{R},$$

where

$$E_R(e) = \frac{1}{N}\sum_{i\in s}\frac{y_i}{\lambda_i}$$

$$E(e) = E_R E_p(e) = E_p E_R(e) = \theta.$$

Let $V_i = V_R(r_i) = \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}$, $for\ all\ i \in U$

$$V_R(e) = \frac{1}{N^2}\sum_{i\in s}\frac{V_i}{\lambda_i^2} = \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\frac{1}{N^2}\sum_{i\in s}\frac{1}{\lambda_i^2}$$

$$V_p(e) = \frac{1}{N^2}\sum_{i}\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)\left(\frac{r_i}{\lambda_i}-\frac{r_j}{\lambda_j}\right)^2$$

$$V_p(e) = \frac{N-n}{nN(N-1)}\sum_{i\in s}(r_i-\bar{r})^2$$

$$V(e) = E_p V_R(e) + V_p E_R(e)$$

$$V(e) = \frac{1}{N^2}\left[\frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\sum_{i=1}^{N}\frac{1}{\lambda_i} + \sum_i\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)\left(\frac{y_i}{\lambda_i}-\frac{y_j}{\lambda_j}\right)^2\right]$$

$$= E_R V_p(e) + V_R E_p(e)$$

$$= \frac{1}{N^2}\left[\sum_{i}\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)E_R\left(\frac{r_i}{\lambda_i}-\frac{r_j}{\lambda_j}\right)^2 + \frac{N(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\right]$$

$$V(e) = \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2 + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{N[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}. \tag{12}$$

Thus, the unbiased estimator for $V(e)$ is;

$$\hat{V}(e) = \frac{1}{N^2}\left[\sum_i\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)\left(\frac{r_i}{\lambda_i}-\frac{r_j}{\lambda_j}\right)^2 + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}\sum_{i\in s}\frac{1}{\lambda_i}\right]$$

$$\hat{V}(e) = \frac{N-n}{nN(n-1)}\sum_{i\in s}(r_i-\bar{r})^2 + \frac{(\alpha+\beta+\delta)[\alpha+\beta+P_1(\delta-\alpha)+ P_2(\delta-\beta)]}{[\delta-P_1(\delta-\alpha)-P_2(\delta-\beta)]^2}. \tag{13}$$

4. EFFICIENCY COMPARISON

The values of the percentage relative efficiencies were obtained to check if the proposed technique performs better than the Hybrid Tripartite Randomized Response Technique for different choices of the parameters when π_A = 0.4; π_U= 0.4; P_1= 0.5; P_2= 0.4; P_3= 0.1; α= 5; β= 10; δ= 25 and when π_A = 0.2; π_U= 0.7; P_1= 0.5; P_2= 0.4; P_3= 0.1; α= 5; β= 10; δ= 25 for varying sample sizes (n) as shown in Tables 1 & 2. Also, a numerical investigation was done using sample size of n = 200; $0.5 \leq P_1 \leq 0.8$; $0.1 \leq P_2 \leq 0.4$; and $P_3 = 0.1$ with a step of 0.1, $0.1 \leq \pi_A \leq 0.2$ with a step of 0.05 and $0.1 \leq \pi_U \leq 0.9$ with a step of 0.1 to ensure moderate confidentiality for the respondents. Using the formula:

$$\text{Percentage Relative Efficiency (PRE)} = \frac{\text{Var}(\hat{\pi}_{HTRRT})}{\text{Var}(\hat{\pi})} \times 100.$$

Tables 1 and 2 show that the proposed estimator is more efficient than the Hybrid Tripartite estimator and as the sample size increases, the variances of both the Hybrid tripartite RRT and the proposed Technique reduces; showing consistency of the two estimators (see Figure1). The variances due to the Hybrid Tripartite RRT reduce from 0.0957 to 0.0095 while that which is due to the proposed Technique reduce from 0.0228 to 0.0016. The Percentage Relative Efficiency (PRE) for Tables 1 and 2 are 595.57 and 420.75, respectively.

Table 1. *Percentage Relative Efficiency of Proposed Technique over Hybrid Tripartite Randomized Response Technique when π_A = 0.4; π_U= 0.4; P_1= 0.5; P_2= 0.4; P_3= 0.1; α= 5; β= 10; δ= 25 for varying sample sizes (n)*

n	Hybrid Tripartite Technique Variance	Proposed Technique Variance	Percentage Relative Efficiency
50	0.0948	0.0159	595.57
100	0.0474	0.0080	595.57
150	0.0316	0.0053	595.57
200	0.0237	0.0040	595.57
250	0.0190	0.0032	595.57
300	0.0158	0.0027	595.57
350	0.0135	0.0023	595.57
400	0.0119	0.0020	595.57
450	0.0105	0.0018	595.57
500	0.0095	0.0016	595.57

Table 2. *Percentage Relative Efficiency of Proposed Technique over Hybrid Tripartite Randomized Response Technique when π_A = 0.2; π_U= 0.7; P_1= 0.5; P_2= 0.4; P_3= 0.1; α= 5; β= 10; δ= 25 for varying sample sizes (n)*

n	Hybrid Tripartite Technique Variance	Proposed Technique Variance	Percentage Relative Efficiency
50	0.0957	0.0228	420.7459
100	0.0479	0.0114	420.7459
150	0.0319	0.0076	420.7459
200	0.0239	0.0057	420.7459
250	0.0191	0.0046	420.7459
300	0.0160	0.0038	420.7459
350	0.0137	0.0033	420.7459
400	0.0120	0.0028	420.7459
450	0.0106	0.0025	420.7459
500	0.0096	0.0023	420.7459

In Table 3, percentage relative efficiency ranges from 58.293 to 375.446 as P_1 increases. Also, the efficiency increases with increase in proportion of the unrelated character and then started decreasing after $\pi_U = 0.5$. It should be noted that for the proposed model to be more efficient, the proportion of the unrelated character, π_U, must be greater than the sensitive character, π_A. Tables 4 and 5 show that the percentage relative efficiency increase from 46.065 to 542.784 and 40.601 to 7125.000 respectively with increase in P_1 values. The variances of the proposed estimator also reduces with increase in the proportion of the sensitive character, π_A and the proportion of the unrelated character, π_U.

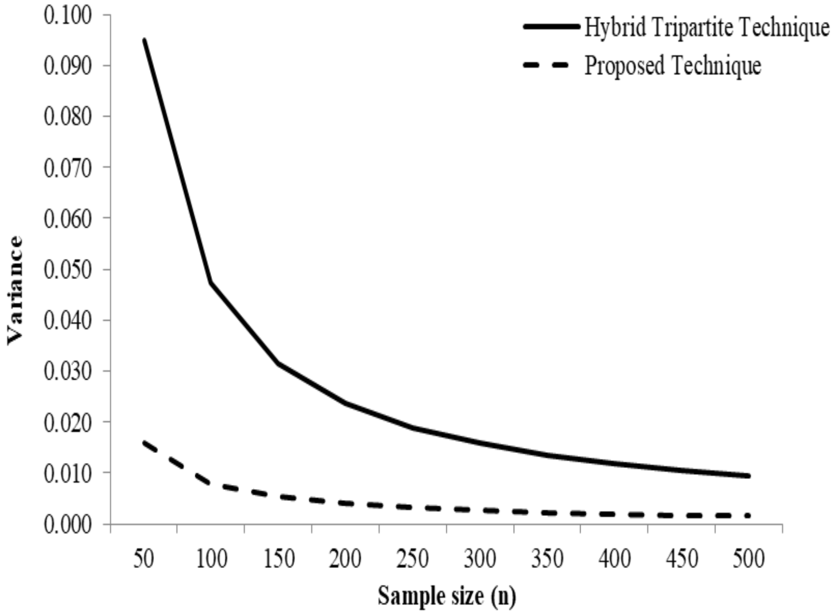

Figure 1. *Variances Comparison of the Hybrid Tripartite RRT with the Proposed Technique*

Table 3. *Percentage Relative Efficiency of Proposed Technique over Hybrid Tripartite Randomized Response Technique when n = 200;* $\pi_A = 0.1$*;* $\alpha = 5$*;* $\beta = 10$*;* $\delta = 25$ *for varying* π_U *and P*

P_1	P_2	P_3	π_U	**Hybrid Tripartite Technique Variance**	**Proposed Technique Variance**	**Percentage Relative Efficiency**
0.50	0.40	0.10	0.10	0.009	0.015	58.293
			0.20	0.014	0.013	109.277
			0.30	0.019	0.012	158.488
			0.40	0.022	0.011	196.068
			0.50	0.024	0.011	212.959
			0.60	0.025	0.012	206.407
			0.70	0.024	0.013	181.302
			0.80	0.023	0.016	146.171
			0.90	0.020	0.018	108.515
0.60	0.30	0.10	0.10	0.010	0.015	65.628
			0.20	0.016	0.013	124.663
			0.30	0.021	0.012	183.099
			0.40	0.025	0.011	229.778
			0.50	0.027	0.011	253.190
			0.60	0.028	0.011	248.255
			0.70	0.027	0.012	219.436
			0.80	0.025	0.014	176.854
			0.90	0.022	0.017	130.219
0.70	0.20	0.10	0.10	0.011	0.015	74.405
			0.20	0.018	0.013	143.245
			0.30	0.024	0.011	213.209
			0.40	0.028	0.010	271.802
			0.50	0.030	0.010	304.500
			0.60	0.031	0.010	302.810
			0.70	0.030	0.011	269.947
			0.80	0.028	0.013	217.780
			0.90	0.024	0.015	159.061
0.80	0.10	0.10	0.10	0.013	0.015	85.023
			0.20	0.021	0.013	165.940
			0.30	0.028	0.011	250.481
			0.40	0.032	0.010	324.895
			0.50	0.035	0.009	371.035
			0.60	0.036	0.009	375.446
			0.70	0.034	0.010	338.583
			0.80	0.032	0.012	273.961
			0.90	0.027	0.013	198.565

Table 4. *Percentage Relative Efficiency of Proposed Technique over Hybrid Tripartite Randomized Response Technique when n = 200; π_A = 0.15; α= 5; β= 10; δ= 25 for varying π_U and P*

P_1	P_2	P_3	π_U	**Hybrid Tripartite Technique Variance**	**Proposed Technique Variance**	**Percentage Relative Efficiency**
0.50	0.40	0.10	0.10	0.010	0.021	46.065
			0.20	0.015	0.018	85.952
			0.30	0.019	0.015	131.358
			0.40	0.022	0.012	178.526
			0.50	0.024	0.011	219.522
			0.60	0.025	0.010	243.143
			0.70	0.024	0.010	240.240
			0.80	0.022	0.011	210.514
			0.90	0.019	0.012	163.124
0.60	0.30	0.10	0.10	0.011	0.021	51.607
			0.20	0.017	0.017	97.752
			0.30	0.022	0.014	151.552
			0.40	0.025	0.012	209.832
			0.50	0.027	0.010	264.291
			0.60	0.028	0.009	300.905
			0.70	0.027	0.009	305.014
			0.80	0.025	0.009	271.713
			0.90	0.021	0.010	210.919
0.70	0.20	0.10	0.10	0.012	0.021	58.218
			0.20	0.019	0.017	111.976
			0.30	0.025	0.014	176.234
			0.40	0.028	0.011	249.015
			0.50	0.031	0.010	322.500
			0.60	0.031	0.008	380.018
			0.70	0.030	0.008	398.912
			0.80	0.028	0.008	364.611
			0.90	0.023	0.008	285.128
0.80	0.10	0.10	0.10	0.014	0.021	66.194
			0.20	0.022	0.017	129.315
			0.30	0.028	0.014	206.755
			0.40	0.032	0.011	298.736
			0.50	0.035	0.009	399.706
			0.60	0.036	0.007	492.063
			0.70	0.034	0.006	542.784
			0.80	0.031	0.006	517.384
			0.90	0.026	0.006	412.193

Table 5. *Percentage Relative Efficiency of Proposed Technique over Hybrid Tripartite Randomized Response Technique when n = 200; π_A = 0.2; α= 5; β= 10; δ= 25 for varying π_U and P*

P_1	P_2	P_3	π_U	**Hybrid Tripartite Technique Variance**	**Proposed Technique Variance**	**Percentage Relative Efficiency**
0.50	0.40	0.10	0.10	0.011	0.026	40.601
			0.20	0.016	0.021	75.268
			0.30	0.020	0.017	119.730
			0.40	0.023	0.013	176.955
			0.50	0.024	0.010	249.429
			0.60	0.025	0.007	335.574
			0.70	0.024	0.006	420.746
			0.80	0.022	0.005	466.154
			0.90	0.019	0.004	424.464
0.60	0.30	0.10	0.10	0.012	0.026	45.316
			0.20	0.018	0.021	85.426
			0.30	0.022	0.016	138.165
			0.40	0.025	0.012	209.241
			0.50	0.027	0.009	307.106
			0.60	0.028	0.006	442.596
			0.70	0.027	0.004	621.173
			0.80	0.024	0.003	807.512
			0.90	0.021	0.002	858.695
0.70	0.20	0.10	0.10	0.013	0.026	50.926
			0.20	0.020	0.020	97.656
			0.30	0.025	0.016	160.714
			0.40	0.029	0.012	250.000
			0.50	0.031	0.008	385.000
			0.60	0.031	0.005	609.375
			0.70	0.030	0.003	1041.667
			0.80	0.027	0.001	2125.000
			0.90	0.023	0.000	7125.000
0.80	0.10	0.10	0.10	0.015	0.026	57.677
			0.20	0.023	0.020	112.547
			0.30	0.029	0.015	188.621
			0.40	0.033	0.011	302.222
			0.50	0.035	0.007	493.250
			0.60	0.035	0.004	893.648
			0.70	0.034	0.001	2374.645
			0.80	0.031	0.001	6111.111
			0.90	0.025	0.002	1388.956

5. SURVEY DATA APPLICATION OF THE PROPOSED TECHNIQUE

A survey was conducted in Akure South Local Government Area of Ondo state, Nigeria between April and June 2016 in order to determine the proportion of people belonging to the sensitive character "drug use disorder" with unrelated question "were you born before 1990". The proposed technique was used to collect information from 200 respondents with assigned parameters α= 5; β= 10; δ= 25 and preset probabilities $P_1 = 1/3$; $P_2 = 1/3$; $P_3 = 1/3$ for the randomized devices.

In Table 6, the estimate of the proportion of respondents belonging to the sensitive character is calculated as 0.253 with variance and standard error; 0.046 and 0.213 respectively, using the proposed estimator while the direct technique estimate gave 0.1150 with 0.1018 and 0.319 as the variance and standard error respectively.

Table 6. *Comparative Analysis of the Proposed Technique with the Direct Technique*

Statistic	**Technique**	
	Direct	**Proposed**
$\hat{\pi}$	0.1150	0.2533
$V(\hat{\pi})$	0.1018	0.0455
$SE(\hat{\pi})$	0.3190	0.2133
$CV(\hat{\pi})$	277.42%	84.20%

The percentage relative efficiency (PRE) of the proposed technique over the direct techniques is calculated as 223.74%. This implies that the proposed technique is practically far more efficient in the estimation of the proportion of persons with sensitive character than the direct technique of data collection.

6. CONCLUSION

A new and more efficient estimator of the proportion of respondents belonging to a sensitive character in a population has been proposed. The proposed technique performs better as the proportion of the sensitive character and that of the unrelated character increases. The proposed technique was also considered when sampling with unequal probabilities with or without replacement. The proposed technique has also been shown to be more efficient than the direct technique of data collection by applying it to a survey data on drug use disorder.

ACKNOWLEDGMENTS

The author appreciates the useful contribution of Mr. Awogbemila Temitope and others for the data collection through survey.

CONFLICTS OF INTEREST

No conflict of interest was declared by the author.

REFERENCES

[1] Warner, S.L., “Randomized response: a survey technique for eliminating evasive answer bias”, Journal of the American Statistical Association, 60: 63-69, (1965).

[2] Mangat, N.S., and Singh, R., “An alternative randomized response procedure”, Biometrika, 77: 439-442, (1990).

[3] Kim, J-M., and Warde, W.D., “A stratified Warner's randomized response model”, Journal of Statistical Planning and Inference, 120 (1-2): 155-165, (2004).

[4] Adebola F.B., Ewemooje O.S., and Ayeni O.C., “Efficiency of Alternative Estimators in Randomized Response Technique”, Proceeding of 5Th Annual Conference of School of Sciences, SOS 2017, 346-356, (2017).

[5] Ewemooje, O.S., and Amahia, G.N., “Improved Randomized Response Technique for Two Sensitive Characters”, Afrika Statistika, 10 (2): 839-852, (2015).

[6] Ewemooje, O.S., and Amahia, G.N., “Improving the Efficiency of Randomized Response Technique for Two Sensitive Characters”, FUTA Journal of Research in Sciences, 12(1): 65-72, (2016).

[7] Singh, H.P., and Gorey, S.M., “An efficient stratified randomized response model”, Journal of Statistical Theory and Practice, 11(4): 790-809, (2017).

[8] Ewemooje, O.S., “Estimating two sensitive characters with equal probabilities of protection”, Cogent Mathematics, 4: 1-14, (2017).

[9] Ewemooje, O.S., Amahia, G.N., and Adebola, F.B., “Estimating prevalence of induced abortion and multiple sexual partners using improved randomized response technique for two sensitive characters”, Communications in Statistics: Case Studies, Data Analysis and Applications, 3(1-2): 21-28, (2017).

[10] Ewemooje, O.S., Adebola, F.B., and Adediran, A.A, “A Stratified Hybrid Tripartite Randomized Response Technique”, Gazi University Journal of Sciences, 31(4): 1246-1266, (2018).

[11] Adebola, F.B., Adediran, A.A., and Ewemooje, O.S., “Hybrid tripartite randomized response technique”, Communications in Statistics-Theory and Methods, 46(23): 11756-11763, (2017).

[12] Ewemooje, O.S., Adebola, F.B., and Amahia, G.N., “Alternative Estimator in Dichotomous Randomized Response Technique”, Communications in Mathematics and Statistics, DOI: 10.1007/s40304-018-0145-x, (2018).

[13] Chaudhuri A., and Christofides T.C., “Indirect Questioning in Sample Surveys”, Spinger-Verlag Berlin Heidelberg, New York, DOI: 10.1007/978-3-642-36276-7, (2013).

[14] Chaudhuri A., “Randomized Response and Indirect Questioning Techniques in Surveys”, Chapman & Hall/CRC, Taylor and Francis Group LLC, (2011).