Research Article GU J Sci 38(1): x-x (2025) DOI: 10.35378/gujs.1335499

Gazi University
Journal of Science
http://dergipark.gov.tr/gujs

Forced Response in Dichotomous Randomised Response Technique

Olusegun S. EWEMOOJE[1,*], Odunayo E. ADARANIWON[2], Wilford B. MOLEFE[1], Femi B. ADEBOLA[2]

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

Highlights

- This paper focuses on correcting biases due to collecting sensitive information.
- A robust method that combines dichotomous and force response designs was proposed.
- Higher precision and efficiency were obtained over conventional methods using proposed method.
- The applicability and efficiency of the proposed model was demonstrated over the direct method.

Article Info

Received: 03 Aug 2023
Accepted: 21 Sept 2024

Keywords

Cyber threat
Dichotomous variable
Forced response
Relative efficiency
Sensitive attribute

Abstract

In sample survey research when sensitive information such as illegal drug usage, rape, examination malpractices among students, induced abortion and the likes are to be obtained from individuals or respondents, false or no response is imminent. This paper proposes a new randomised response model to correct this bias by incorporating the force-response in a dichotomous technique. The unbiased estimate and variance of the proposed model were obtained. The proposed randomised response model is then compared with the Unbiased Estimator in the Unrelated Dichotomous Randomised Response Model, (UEUDRRT) and Alternative Estimator in Dichotomous Randomised Response Model, (AEDRRT). The numerical investigation revealed that the variance of the proposed model reduces while the V(UEUDRRT) and V(AEDRRT) increase as the sensitive issues increase. Hence, the proposed model has percentage relative efficiencies (PREs) as high as 1195.45% and 1977.20% over UEUDRRT and AEDRRT, respectively. Also, the application of the proposed method revealed that it is practically more efficient in estimating the proportion of respondents involved in sensitive character (cyber threat) than the direct method of data collection. Therefore, the proposed technique has been shown to be more efficient than conventional ones.

1. INTRODUCTION

Confiding in another person about one's top secret is one of the major problems experienced as a human. In statistics, when sensitive information such as illegal drug usage, rape, examination malpractices among students, induced abortion and the likes are to be obtained from individuals or respondents, false or no response is imminent. This has brought about biased responses from the respondents while answering these sensitive questions as many tend to give false responses to protect their image. It is widely recognized in various social research disciplines, especially when the investigation involves sensitive topics, that respondents may alter their answers to appear more socially acceptable (social desirability bias) or choose not to answer [1, 2]. To reduce this bias, Warner [3] in 1965 developed a model called Randomised Response Model (RRM) with the use of spinner.

Different authors have studied and modified Warner's model, some of which are Greenberg et al. [4], Boruch [5], Mangat [6], Kim and Warde [7], Hussain and Shabbir [8], Ewemooje and Amahia [9], Adebola et al. [10], Ewemooje [11], Ewemooje et al. [12], Adeniran et al. [13], Harriet et al. [14], Singh et al. [15] among others. Others also applied these techniques to test their applicability including Perri et al. [16], Ewemooje et al. [17], Adebola et al. [18], and so on. In 2020, Adediran et al. [19] proposed an unbiased estimator using unrelated questions in dichotomous randomised response, and in their research, sensitive

*Corresponding author, e-mail: ewemoojeos@ub.ac.bw

questions were asked in accordance with Ewemooje et al. [20]. Two randomised devices, denoted as R_1 and R_2 were used, these consist of two unrelated questions. The sensitive question "A" is of interest to the interviewer with probability p, while the non-sensitive attribute question "U" is unrelated to question "A" with probability $1-p$. Their proposed unbiased estimator model was then compared with Ewemooje et al. [20] which was found to be more efficient. However, the force-response model earlier proposed by Boruch [5] was shown to be relatively more efficient than the unrelated question model [14].

Therefore, this paper proposes an alternative Dichotomous Randomised Response Technique which combines the Force Response approach with dichotomous design to develop a more robust estimator. The estimate and variance of the proposed method are thereafter obtained and compared with the conventional ones using percentage relative efficiency as a criterion. This is then applied to detect the proportion of cyber threat involvement among adolescents and young adults.

2. MATERIAL AND METHOD

2.1. Alternative Estimator in Dichotomous Randomised Response Model (AEDRRT) by Ewemooje et al. [20]

In this model, respondents answer sensitive questions. If they reply positively, no randomizer is needed. If they respond negatively, two randomized devices, R_1 and R_2, will present two questions with varying selection probabilities. Ewemooje et al. [20] also introduced two constants, α and β, such that $q=\frac{\alpha}{\alpha+\beta}$ is the probability of using the first randomizer, R_1, which consist of two statements from Warner's device with probabilities p_1 and $1-p_1$, while $1-q=\frac{\beta}{\alpha+\beta}$ is the probability of using the second randomizer, R_2, which consist of two statements of Warner's device as well but with probabilities p_2 and $1-p_2$, respectively.

The unbiased estimate of the population proportion is provided as follows::

$$\hat{\pi}_{Ewe}=\frac{\hat{\theta}(\alpha+\beta)-p_2\alpha-p_1\beta}{p_1\alpha+p_2\beta} \quad (1)$$

where $\hat{\theta}=\frac{n_0}{n}$, n_0 is the number of "yes" responses to sensitive questions while n the sample size and $\hat{\pi}_{Ewe}$ is the unbiased estimate of the population proportion of respondents belonging to the sensitive character as proposed by Ewemooje et al. [20].

The variance of the estimate can also be expressed as follows:

$$V(\hat{\pi}_{Ewe})=\frac{\pi(1-\pi)}{n}+\frac{(1-\pi)(p_2\alpha+p_1\beta)}{n(p_1\alpha+p_2\beta)^2} \quad (2)$$

where π is the true probability of the sensitive character "A".

2.2. The Unbiased Estimator in Unrelated Dichotomous Randomised Response (UEUDRRT) by Adediran et al. [19]

In their study, participants were confronted with sensitive questions head-on, inviting them to share their thoughts openly and honestly. If the response is positive i.e., "yes", there will be no need for the randomised device to be used. But if the response is negative i.e., "no", the respondents are allowed to randomly select one of the two randomised devices, R_1 and R_2. However, R_1 contains two unrelated questions (the sensitive question "A" in which the interviewer is interested in with probability p_1, and the non-sensitive question "U" that is unrelated to the sensitive question A with probability $1-p_1$) while R_2 is also comprises two unrelated questions (the sensitive question "A" with probability p_2, and the non-sensitive question "U" with probability $1-p_2$). However, R_1 contains two unrelated questions: the sensitive question "A," which the interviewer is interested in with probability p_1, and the non-sensitive question "U," which is unrelated to question "A" with probability $1-p_1$. Additionally, R_2 contains another set of two unrelated questions that includes the sensitive question "A" with probability p_2, alongside the non-sensitive question "U," which has a probability, $1 - p_2$.

The estimate of their population proportion is given as:

$$\hat{\pi}_{Ade} = \frac{\hat{\theta}(\alpha+\beta)-\pi_u(\alpha+\beta)-\alpha p_1-\beta p_2}{(\alpha+\beta+\alpha p_1+\beta p_2)} \quad (3)$$

where $\hat{\theta} = \frac{n_0}{n}$, n_0 is the number of "yes" responses to sensitive questions while n is the total number of respondents in the survey. π_U, true probability of the unrelated question "U" and $\hat{\pi}_{Ade}$ is the unbiased estimate for the population proportion belonging to the sensitive attribute as proposed by Adediran et al. [19].

The variance of their unbiased estimator is given as:

$$V(\hat{\pi}_{Ade}) = \frac{\pi_A\{(\alpha+\beta)-\pi_A(\alpha+\beta+\alpha p_1+\beta p_2)\}}{n(\alpha+\beta+\alpha p_1+\beta p_2)} + \frac{\pi_u(\alpha+\beta-\alpha p_1-\beta p_2)(\alpha+\beta-2\pi_A(\alpha+\beta+\alpha p_1+\beta p_2))}{(n-1)(\alpha+\beta+\alpha p_1+\beta p_2)^2} \quad (4)$$

where π_A is the true probability for the sensitive attribute "A".

2.3. Proposed Model

In this paper, sensitive questions are asked directly to the respondents through a questionnaire. If a respondent answers "no," they are instructed to randomly choose one of two randomized devices to enhance their confidentiality. This exercise is being done without respondents revealing neither their questions nor answers to the interviewers, however, adequate information/training on how to interact with the devices is being given to the respondents before the survey. The two randomised devices, D_1 and D_2, consist of two questions with different selection probabilities each. One (D_1) comprise of two questions: the question on sensitive attribute with the probability, p_1 and a force response (yes) with the probability $1 - p_1$. Also, the second (D_2) consist of two statements: the question on the sensitive attribute with the probability, p_2 and a force response (yes) with the probability $1 - p_2$. A simple random sample (srs) with replacement was used to select the sample size n, with α and β representing any two positive real numbers for the selection of the randomizers such that $k = \frac{\alpha}{\alpha+\beta}$ is the probability of using the first randomised device, D_1 and $1 - k = \frac{\beta}{\alpha+\beta}$ as the probability of using the second randomised device, D_2. Figure 1 shows the tree diagram of the proposed force response dichotomous randomized response model.

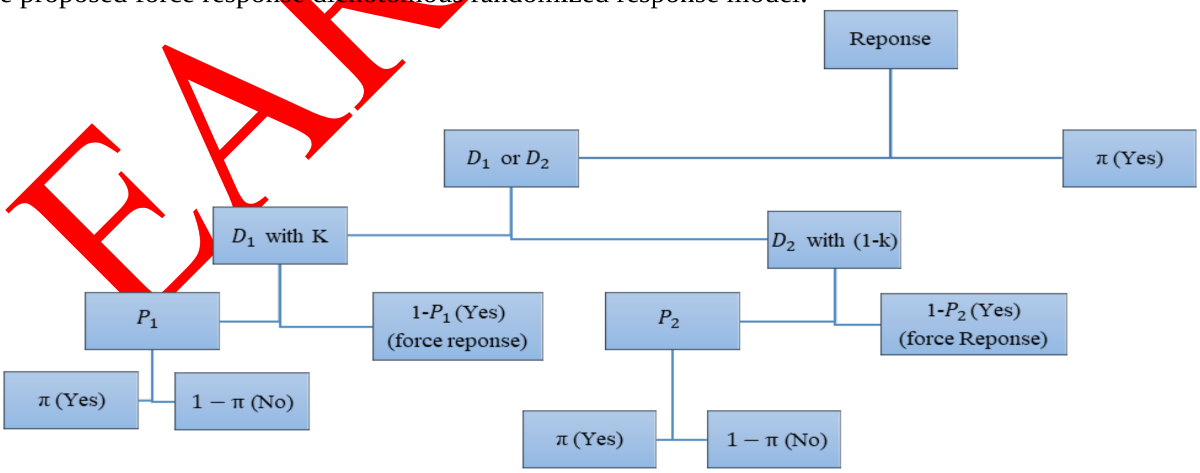

Figure 1. *The tree diagram of the proposed model*

The population proportion of respondents who answered "yes" is given as follows:

$$p(yes) = \theta = \pi + k[p_1\pi + (1 - p_1)] + (1 - k)[p_2\pi + (1 - p_2)] \quad (5)$$

where p_1 represents the probability of sensitive attribute in randomised devices D_1 while p_2 represents the probability of sensitive attribute in randomised devices D_2. Also, π represents the sensitive character's true probability.

Equation (5) also provides an estimate for the proportion of the sensitive attribute, which is expressed as follows:

$$\hat{\pi} = \frac{(\alpha+\beta)\hat{\theta}-(1-p_1)\alpha-(1-p_2)\beta}{\alpha+\beta+\alpha p_1+\beta p_2} \quad (6)$$

where $\hat{\theta} = \frac{n_o}{n}$, the number of respondents who answered "yes" to the sensitive question is denoted by n_o, while n represents the sample size.

Test for Unbiasedness

The estimator,$\hat{\pi}$, will be considered unbiased if $E(\hat{\pi}) = \pi$

$$E(\hat{\pi}) = E[\frac{(\alpha+\beta)\,\hat{\theta}-(1-P_1)\alpha-(1-P_2)\beta}{\alpha+\beta+\alpha P_1+\beta P_2}]$$

$$E(\hat{\pi}) = \frac{(\alpha+\beta)\mathrm{E}(\hat{\theta})-(1-P_1)\alpha-(1-P_2)\beta}{\alpha+\beta+\alpha P_1+\beta P_2} = \frac{(\alpha+\beta)\theta-(1-P_1)\alpha-(1-P_2)\beta}{\alpha+\beta+\alpha P_1+\beta P_2} \quad (7)$$

Since $\theta = \pi + \frac{\alpha}{\alpha+\beta}[P_1\pi + (1-P_1)] + \frac{\beta}{\alpha+\beta}[P_2\pi + (1-P_2)]$

$$\theta = \frac{\pi(\alpha+\beta+\alpha P_1+\beta P_2)+(1-P_1)\alpha+(1-P_2)\beta}{\alpha+\beta}.$$

Substituting θ into Equation (7), we have

$$E(\hat{\pi}) = \frac{\alpha+\beta\left[\frac{\pi(\alpha+\beta+\alpha P_1+\beta P_2)+(1-P_1)\alpha+(1-P_2)\beta}{\alpha+\beta}\right]-(1-P_1)\alpha-(1-P_2)\beta}{\alpha+\beta+\alpha P_1+\beta P_2}$$

$$E(\hat{\pi}) = \frac{\pi(\alpha+\beta+\alpha P_1+\beta P_2)+(1-P_1)\alpha+(1-P_2)\beta-(1-P_1)\alpha-(1-P_2)\beta}{\alpha+\beta+\alpha P_1+\beta P_2}$$

$$= \frac{\pi(\alpha+\beta+\alpha P_1+\beta P_2)}{\alpha+\beta+\alpha P_1+\beta P_2}$$

$$E(\hat{\pi}) = \pi\,. \quad (8)$$

Therefore, $E(\hat{\pi}) = \pi$. As a result, the estimator proposed is a reliable and unbiased estimate of the population proportion.

2.3.1. The estimation of variance for the proposed estimator

To obtain the variance,

$$v(\hat{\pi}) = v\left(\frac{(\alpha+\beta)\,\hat{\theta}-(1-p_1)\alpha-(1-p_2)\beta}{\alpha+\beta+\alpha p_1+\beta p_2}\right).$$

Note that $v(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$ as the proposed estimator follows a binomial distribution. Then we have

$$v(\hat{\pi}) = \frac{(\alpha+\beta)^2 v(\hat{\theta})}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\,. \quad (9)$$

From Equation (5), $\theta = \frac{\pi(\alpha+\beta+\alpha p_1+\beta p_2)+\alpha(1-p_1)+\beta(1-p_2)}{\alpha+\beta}$

$$\theta = \pi + \frac{(\alpha+\beta)-(1-\pi)(\alpha p_1+\beta p_2)}{\alpha+\beta}. \quad (10)$$

Then, $1-\theta = \frac{-\alpha\pi-\beta\pi-\alpha\pi p_1-\beta\pi p_2+\alpha p_1+\beta p_2}{(\alpha+\beta)}$

$$1-\theta = \frac{-\pi(\alpha+\beta)+(1-\pi)(\alpha p_1+\beta p_2)}{(\alpha+\beta)} \quad (11)$$

$$v(\hat{\theta}) = \frac{(\alpha\pi+\beta\pi+\alpha\pi p_1+\beta\pi p_2+\alpha-\alpha p_1+\beta-\beta p_2)(-\alpha\pi-\beta\pi-\alpha\pi p_1-\beta\pi p_2+\alpha p_1+\beta p_2)}{(\alpha+\beta)^2}$$

$$v(\hat{\theta}) = \frac{\begin{array}{c}-\pi^2[\alpha^2+2\alpha\beta+2\alpha^2p_1+2\alpha\beta p_2+\beta^2+2\alpha\beta p_1+2\beta^2p_2+2\alpha\beta p_1p_2+\alpha^2{p_1}^2+\beta^2{p_2}^2] \\ +\pi[\alpha\beta p_1+\beta^2{p_2}^2+2\alpha^2{p_1}^2+4\alpha\beta p_1p_2+2\beta^2{p_2}^2-\alpha^2-2\alpha\beta+\alpha^2p_1-\beta^2+\alpha\beta p_2] \\ +\alpha^2p_1+\alpha\beta p_2+\beta^2p_2-\alpha\beta p_1p_2-\alpha^2{p_1}^2+\alpha\beta p_1-\alpha\beta p_1p_2-\beta^2{p_2}^2\end{array}}{n(\alpha+\beta)^2}. \quad (12)$$

Substituting Equation (12) in (9) and simplified further, gives the variance as:

$$v(\hat{\pi}) = \frac{\pi[2(\alpha p_1+\beta p_2)-(\alpha+\beta)-\pi(\alpha+\beta+\alpha p_1+\beta p_2)]}{n(\alpha+\beta+\alpha p_1+\beta p_2)} + \frac{(\alpha p_1+\beta p_2)(\alpha+\beta-\alpha p_1-\beta p_2)}{n(\alpha+\beta+\alpha p_1+\beta p_2)^2}. \quad (13)$$

The estimate of the variance $(\hat{v}(\hat{\pi}))$ is given as:

$$\hat{v}(\hat{\pi}) = \frac{\hat{\pi}[2(\alpha p_1+\beta p_2)-(\alpha+\beta)-\hat{\pi}(\alpha+\beta+\alpha p_1+\beta p_2)]}{n(\alpha+\beta+\alpha p_1+\beta p_2)} + \frac{(\alpha p_1+\beta p_2)(\alpha+\beta-\alpha p_1-\beta p_2)}{n(\alpha+\beta+\alpha p_1+\beta p_2)^2} \quad (14)$$

where $\hat{\pi}$ is the estimate of the true probability of the sensitive character.

2.4. Exploring Forced Responses in Dichotomous Design: Unpacking Sampling with Unequal Probabilities With and Without Replacement

If we let $y_i = 1$, as i takes on a sensitive attribute "A" while $y_i = 0$, if i does not takes on a sensitive attribute "A" for a respondent considered i in the sample space, Ω = (1, 2, …, N). $Y = \sum y_i$, represents the total number of people bearing the attribute "A" to be estimated. Likewise, let every respondent sample partake in the proposed design independently.
Consequently, answers

$$\theta_i = \begin{cases} 1, & \text{as ith respondent responses "Yes"} \\ 0, & \text{if otherwise} \end{cases}$$

As earlier proposed by Chaudhuri & Stenger [21] and Chaudhuri [22], we write

$$P(yes) = P(\theta_i = 1) = E_R(\theta_i) = y_i + k[p_1y_i + (1-p_1)] + (1-k)[p_2y_i + (1-p_2)]$$

$$= y_i + \frac{(\alpha+\beta)-(1-y_i)(\alpha p_1+\beta p_2)}{\alpha+\beta}$$

$$P(no) = P(\theta_i = 0) = \frac{-y_i(\alpha+\beta)+(1-y_i)(\alpha p_1+\beta p_2)}{(\alpha+\beta)}.$$

This yield τ_i, an unbiased estimator for y_i

$$\tau_i = \frac{(\alpha+\beta)\theta_i-(1-p_1)\alpha-(1-p_2)\beta}{\alpha+\beta+\alpha p_1+\beta p_2}.$$

Since $E_R(\tau_i) = y_i \; \forall \; i$,

$$V_R(\tau_i) = \frac{(\alpha+\beta)^2 V_R(\theta_i)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}$$

where $V_R(\theta_i) = E_R(\theta_i)[1 - E_R(\theta_i)] = \frac{2(\alpha p_1+\beta p_2)}{(\alpha+\beta)}$

$$V_R(\tau_i) = \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{n(\alpha+\beta+\alpha p_1+\beta p_2)^2} \, . \qquad (15)$$

Also;

$$\bar{\tau} = \frac{1}{n}\sum_{i=1}^{n} \tau_i$$

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

$$V(\bar{\tau}) = V(\bar{y}) + \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{n(\alpha+\beta+\alpha p_1+\beta p_2)^2}$$

$$V(\bar{\tau}) = \frac{\pi(1-\pi)}{n} + \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{n(\alpha+\beta+\alpha p_1+\beta p_2)^2}$$

$$V(\bar{\tau}) = \frac{1}{n}\left[\pi(1-\pi) + \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\right]. \qquad (16)$$

The estimator $\bar{\tau}$ serves as an unbiased estimator for the parameter π whereas $\hat{V}(\bar{\tau})$ is an unbiased estimator for the parameter $V(\bar{y})$;

$$\hat{V}(\bar{\tau}) = \frac{1}{(n-1)}\left[\hat{\pi}(1-\hat{\pi}) + \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\right] . \qquad (17)$$

However, it is understood that selecting a value of p close to 0.5 will increase respondents' confidence in the effectiveness of privacy protection procedures. Furthermore, when the probability of the sensitive attribute, p, is closer to 0.5, both the size of the sample and the coefficient of variation increase.

Hence, consider a sample, s, selected through a general design *p* characterised by the probabilities of inclusion as recommended by Chaudhuri [22];

$$\lambda_i = \sum_{s \ni i} p(s) > 0 \qquad for \; i \in \Omega$$

and

$$\lambda_{ij} = \sum_{s \ni i,j} p(s) > 0 \qquad for \; i,j \in \Omega \; (\mathrm{i} \neq \mathrm{j})$$

$$e = \frac{1}{N}\sum_{i \in s}\frac{\tau_i}{\lambda_i} \;\; \equiv \;\; \bar{\tau} = \frac{1}{n}\sum_{i \in s}\tau_i \, .$$

For sampling without replacement;

$$E_p(e) = \frac{1}{N}\sum_{i=1}^{N}\tau_i = \frac{T}{N} = \bar{T}$$

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

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

Recall that $V_R(\tau_i) = \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2} = V_i$, for all $i \in \Omega$

$$V_R(e) = \frac{1}{N^2}\sum_{i\in s}\frac{V_i}{\lambda_i^2} = \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\frac{1}{N^2}\sum_{i\in s}\frac{1}{\lambda_i^2}$$

$$V_p(e) = \frac{1}{N^2}\sum_{i}\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)\left(\frac{\tau_i}{\lambda_i}-\frac{\tau_j}{\lambda_j}\right)^2 = \frac{N-n}{nN(N-1)}\sum_{i\in s}(\tau_i-\bar{T})^2$$

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

$$= \frac{1}{N^2}\left[\frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\sum_{i=1}^{N}\frac{1}{\lambda_i} + \sum_{i}\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)\left(\frac{y_i}{\lambda_i}-\frac{y_j}{\lambda_j}\right)^2\right]$$

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

$$= \frac{1}{N^2}\left[\sum_{i}\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)E_R\left(\frac{\tau_i}{\lambda_i}-\frac{\tau_j}{\lambda_j}\right)^2 + \frac{2N(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\right]$$

$$V(e) = \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2 + \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{N(\alpha+\beta+\alpha p_1+\beta p_2)^2} \ . \tag{18}$$

An unbiased estimator for $V(e)$ is;

$$\hat{V}(e) = \frac{1}{N^2}\left[\sum_{i}\sum_{j,j>i}\left(\frac{\lambda_i\lambda_j-\lambda_{ij}}{\lambda_{ij}}\right)\left(\frac{\tau_i}{\lambda_i}-\frac{\tau_j}{\lambda_j}\right)^2 + \frac{2N(\alpha+\beta)(\alpha p_1+\beta p_2)}{(\alpha+\beta+\alpha p_1+\beta p_2)^2}\sum_{i\in s}\frac{1}{\lambda_i}\right]$$

$$\hat{V}(e) = \frac{N-n}{nN(n-1)}\sum_{i\in s}(\tau_i-\bar{\tau})^2 + \frac{2(\alpha+\beta)(\alpha p_1+\beta p_2)}{n(\alpha+\beta+\alpha p_1+\beta p_2)^2} \ . \tag{19}$$

3. RESULTS AND DISCUSSION

3.1. Comparison Through Relative Efficiency

Comparing the proposed model, Forced Response in Dichotomous Randomised Response Technique (FRDRRT), with Unbiased Estimator in Unrelated Dichotomous Randomised Response Model, (UEUDRRT) by Adediran et al. [19] and Alternative Unbiased Estimator in Dichotomous Randomised Response Model, (AEDRRT) by Ewemooje et al. [20] using the percentage relative efficiency (PRE) given as:

$$\text{PRE} = \frac{\text{variance of conventional model}}{\text{variance of proposed model}} \times 100 \ .$$

The proposed model demonstrates superior efficiency compared to the two conventional models if and only if PRE > 100. The proposed model's PRE was evaluated against two conventional models, utilizing fixed sample sizes (n) and varying probabilities p_1 and p_2 by using the randomised devices at different values of π_A and π_U. For the Adediran et al. [19] and Ewemooje et al. [20] models, randomisation devices were employed at different values of π.

Figure 2 illustrates that as the value of π_A increases, the variance of the UEUDRRT by Adediran et al. [19] rises from 0.000345 to 0.000608. In contrast, the variance of our proposed model decreases from 0.000501 to 0.000145 as π_A increases, specifically within the range of $0.1 \leq \pi_A \leq 0.3$. The proposed model demonstrates greater efficiency than the approach by Adediran et al. [19], when $\pi_U \geq 0.4$ the PRE at $0.1 \leq \pi \leq 0.3$ increases largely from 68.8361 to 420.4709. This indicates that the proposed model outperforms Adediran et al. [19] at $0.1 \leq \pi_A \leq 0.3$ when $p_1 = 0.3$ and $p_2 = 0.7$, respectively.

As the p_1 increases to 0.4 and p_1 reduces to 0.6, the variance of Adediran et al. [19] increases from 0.000353 to 0.000634 as π_U increases while the variance of the proposed model decreases from 0.000504 to 0.000126 as π_A increases (See Figure 3). Here, the PRE increases largely from 70.0404 to 503.7252, which is an improvement on Figure 2 but the improved efficiency over Adediran et al. [19] still starts at $\pi_U \geq 0.4$.

Also, equal probabilities between the sensitive question and the forced response were considered, $p_1 = 0.5$ and $p_2 = 0.5$, in Figure 4 and the variance of Adediran et al. [19] increases from 0.000361 to 0.000661 as π_U increases while the proposed model's variance reduces from 0.000506 to 0.000106. This shows improvement in the PRE as the gap between proposed model and Adediran et al. [19] widens as shown in Figure 4. The model proposed starts to have better efficiency than Adediran et al. [19] when $\pi_U \geq 0.3$ and the PRE increases largely from 71.42857 to 626.3158.

In Figure 5, the gap between Adediran et al. [19] and the model proposed continues to widen as the probability, p_1, increases to 0.6 and $p_2 = 0.4$. The variance of Adediran et al. [19] is seen to increase from 0.00037 to 0.000690 as π_U increases while the variance of the proposed model still reduces further from 0.000506 to 0.0000838. This shows that the PRE increases largely from 73.01961 to 824.1353 which means that the model proposed has better efficiency compared to Adediran et al. [19].

Figure 6 demonstrates additional improvements in the proposed model as the value of p_1 increases to 0.7. Meanwhile, the variance reported by Adediran et al. [19] rises from 0.000379 to 0.000721. In contrast, the variance of the proposed model decreases further, going from 0.000506 to 0.0000603. This shows that the gap between the model proposed and Adediran et al. [19] further widens (see Figure 6) as PRE increases basically from 74.8366 to 1195.45 which also confirms the improved efficiency over Adediran et al. [19] model.

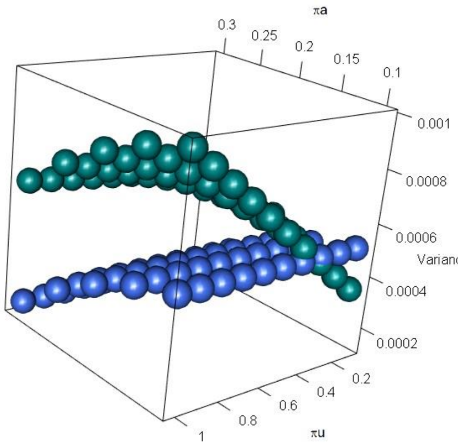

Figure 2. *Comparison of variance between the FRDRRT (Proposed) and UEUDRRT models by Adediran et al. [19] when* $p_1 = 0.3,\ p_2 = 0.7$.

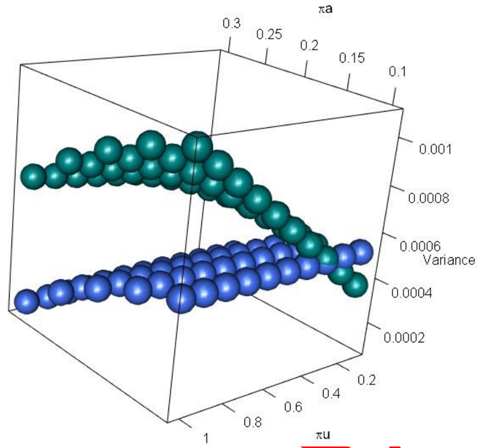

Figure 4. *Comparison of variance between the FRDRRT (Proposed) and UEUDRRT models by Adediran et al. [19] when* $p_1 = 0.5,\ p_2 = 0.5$.

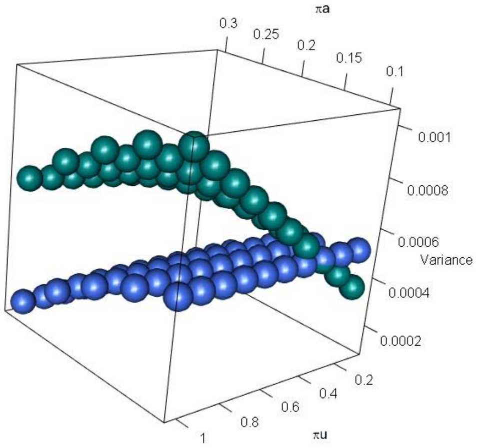

Figure 3. *Comparison of variance between the FRDRRT (Proposed) and UEUDRRT models by Adediran et al. [19] when* $p_1 = 0.4,\ p_2 = 0.6$.

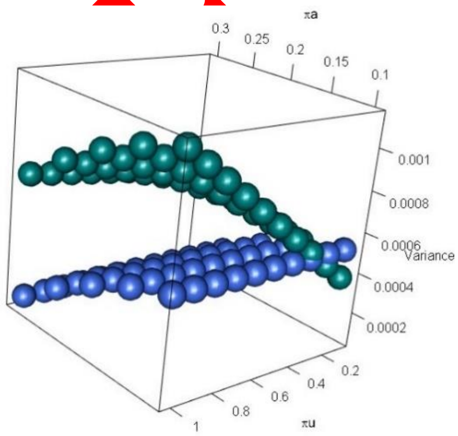

Figure 5. *Comparison of variance between the FRDRRT (Proposed) and UEUDRRT models by Adediran et al. [19] when* $p_1 = 0.6,\ p_2 = 0.4$.

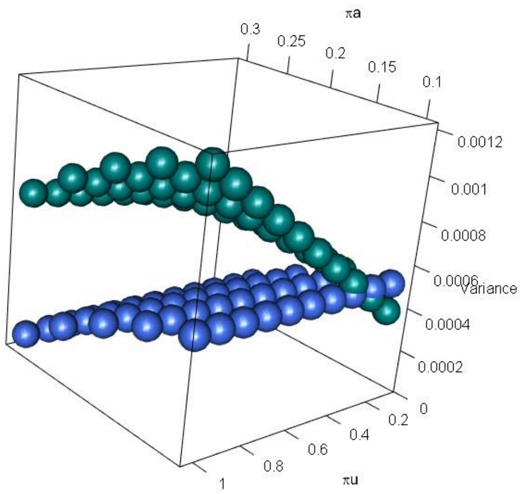

Figure 6. *Comparison of variance between the FRDRRT (Proposed) and UEUDRRT models by Adediran et al. [19] when* $p_1 = 0.7,\ p_2 = 0.3$.

Further analysis was carried out in comparing the proposed (FRDRRT) model with Ewemooje et al. [20] model at different levels of p_1 and p_2. The results show that when $p_1 = 0.3,\ p_2 = 0.7$, the Ewemooje et al. [20] (AEDRRT) model variance increases from 0.000573 to 0.001146 as π increases from 0.1 to 0.3 while the FRDRRT's variance reduces from 0.000501 to 0.000145. The PRE at $0.1 \leq \pi \leq 0.3$ is shown to increase from 114.3716% to 792.7756%, which means that the proposed model demonstrates superior efficiency compared to the method outlined by Ewemooje et al. [20] as shown in Table 1.

As the likelihood of encountering the sensitive question in the first randomizer increases, $p_1 = 0.4$ and $p_2 = 0.6$, the Ewemooje et al. [20]'s variance when $0.1 \leq \pi \leq 0.3$ is between 0.000586 and 0.001156 while the variance of the proposed model at $0.1 \leq \pi \leq 0.3$ is between 0.000504 and 0.000126. The PRE is shown to increase from 116.276% to 918.7924%. Also, when there are equal probabilities in the two randomizers, $p_1 = 0.5,\ p_2 = 0.5$, the variance reported by Ewemooje et al. [20] shows an increase from 0.0006 to 0.001167, whereas the variance of the proposed model demonstrates a decrease from 0.000506 to 0.000106 which shows increment in the PRE, from 118.6813% to 1105.263% (see Table 1). At $p_1 = 0.6,\ p_2 = 0.4$ the variance of Ewemooje et al. [20] is between 0.000616 and 0.001179 while FRDRRT's variance is from 0.000506 to 0.0000838 with PRE from 121.6558% to 1407.424%. Finally, at $p_1 = 0.7,\ p_2 = 0.3$ the Ewemooje et al. [20]'s variance moves from 0.000634 to 0.001193 while the variance of the proposed reduces from 0.000506 to 0.000603 with PRE from 125.2851% to 1197.202%, which indicate that the proposed model demonstrates superior efficiency compared to the approach developed by Ewemooje et al. [20].

Table 1. *A comparative analysis of percentage relative efficiency between the proposed model (FRDRRT) and the AEDRRT by Ewemooje et al. [20], specifically examining performance metrics across varying parameters when* $\alpha = 25;\ \beta = 35;\ n = 200$ *for varying* p_1, p_2 *and* π.

p_1	p_2	π	$v(\hat{\pi})$	$v(\hat{\pi}_{Ewe})$	PRE
0.3	0.7	0.1	0.000501	0.00057	114.372
		0.15	0.000449	0.00075	167.712
		0.2	0.000373	0.00091	243.945
		0.25	0.000271	0.00104	383.568
		0.3	0.000145	0.00115	792.776
0.4	0.6	0.1	0.000504	0.00059	116.276
		0.15	0.000447	0.00077	171.389
		0.2	0.000365	0.00092	252.396
		0.25	0.000258	0.00105	407.579
		0.3	0.000126	0.00116	918.792
0.5	0.5	0.1	0.000506	0.0006	118.681
		0.15	0.000443	0.00078	175.862
		0.2	0.000356	0.00093	262.500
		0.25	0.000243	0.00106	437.143
		0.3	0.000106	0.00117	1105.26
0.6	0.4	0.1	0.000506	0.00062	121.656
		0.15	0.000438	0.00079	181.263
		0.2	0.000345	0.00095	274.616
		0.25	0.000227	0.00108	474.121
		0.3	0.0000838	0.00118	1407.42
0.7	0.3	0.1	0.000506	0.00063	125.285
		0.15	0.000432	0.00081	187.759
		0.2	0.000333	0.00096	289.219
		0.25	0.000209	0.00109	521.315
		0.3	0.0000603	0.00119	1977.20

3.2. Implementation/Application of the Proposed Methodology

The proposed methodology was implemented through a survey targeting a sample of 300 adolescents and young adults, specifically those aged 15 to 26 years, residing in the FUTA South and North areas of the Akure South Local Government Area in Ondo State., Nigeria between June and July 2022 to detect the proportion of adolescents and young adults involved in the sensitive characteristic “cyber threat” as part of the research on Adaptive model of family planning through health promotion programs. This was done after proper education has been given to the respondents on how best to use the randomised devices with necessary demonstration. The direct (traditional) method was also used to obtain the same sensitive information from these 300 adolescents and young adults.

Table 2 shows the comparative analysis of the direct method with the proposed unbiased estimator in forced response dichotomous randomised response technique, the estimates of the proportion of respondents who are involved in cyber threat were calculated as 0.224 and 0.120 for the proposed and direct methods, respectively. The variances show that there is less variability using the proposed method (0.000250) than using the direct method (0.000353). This corroborated the use of RRT over the direct method as the coefficient of variations shows that there is 7.06% variation in using the proposed method while more than double variation (15.67%) was recorded using the direct method. This finding is corroborated by previous studies [23, 24] on substance use disorders, which demonstrate that randomised response techniques exhibit greater efficiency compared to conventional direct data collection methods. Additionally, the calculated percentage relative efficiency (PRE) of the proposed method compared to the direct method was determined to be 141.2%. The results suggest that the proposed methodology demonstrates superior efficacy in estimating the prevalence of individuals engaging in cyber threats compared to traditional direct data collection techniques.

Table 2. *Comparative analysis of the direct method with the proposed unbiased estimator in forced response dichotomous randomised response technique.*

Method	$\hat{\pi}$	$V(\hat{\pi})$	$S.E(\hat{\pi})$	$C.V(\hat{\pi})$
Proposed	0.224	0.000250	0.0158	7.06%
Direct	0.120	0.000353	0.0188	15.67%

4. CONCLUSION

The implementation of force response design has demonstrated significant enhancements in the efficacy of the randomised response technique, optimising its performance in maintaining confidentiality while collecting sensitive information. As the proportion of the sensitive attribute ($\hat{\pi}_A$) increases, the variance of the proposed model diminishes. This is in contrast to the AEDRRT model by Ewemooje et al. [20] and the UEUDRRT model by Adediran et al. [19], where the variance increases in relation to the increasing proportion of the sensitive attribute within the dataset. This enhancement results in a higher relative efficiency percentage of the proposed model compared to the two existing models. The newly proposed dichotomous forced response technique has been proven to be more effectively identifying respondents with the sensitive attribute. This method surpasses other models [19, 20] evaluated in this study, showcasing its superior ability to capture a greater proportion of individuals relevant to the sensitive characteristic. Also, the application of the proposed method showed that it is practically more efficient in estimatin proportion of respondents who were involved in the sensitive attribute (cyber threat) than the direct/traditional method of data collection. Consequently, it can be asserted that the proposed model demonstrates superior efficiency compared to traditional models.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Tourangeau, R., Yan, T., "Sensitive questions in surveys", Psychological Bulletin, 133(5): 859–883, (2007). DOI: https://doi.org/10.1037/0033-2909.133.5.859

[2] Krumpal, I., "Determinants of social desirability bias in sensitive surveys: a literature review", Quality & Quantity, 47: 2025-2047, (2013). DOI: https://doi.org/10.1007//s11135-011-9640-9

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

[4] Greenberg, B.G., Abul-Ela, L.A., Simmons, W.R., Horvitz, D.G., "The unrelated question randomized response model: theoretical framework", Journal of American Statistical Association, 64: 520-539, (1969).

[5] Boruch, R., "Assuring confidentiality of responses in social research: a note on strategies" American Sociologist, 6(4): 308-311, (1971).

[6] Mangat, N.S., "An improved randomized response strategy", Journal of the Royal Statistical Society: Series B (Methodological), 56: 93-95, (1994). DOI: https://doi.org/10.1111/j.2517-6161.1994.tb01962.x

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

[8] Hussain, Z., Shabbir, J., "Randomized use of Warner's randomized response model", Interstat, 1-21, (2007).

[9] Ewemooje, O.S., Amahia, G.N., "Improved Randomized Response Technique for Two Sensitive Attributes" Afrika Statistika, 10(2): 839-852, (2015). DOI: https://doi.org/10.16929/as/2015.639.78

[10] Adebola, F.B., Adediran, A.A., Ewemooje, O.S., "Hybrid tripartite randomized response technique", Communications in Statistics - Theory and Methods, 46(23): 11756-11763, (2017). DOI: https://doi.org/10.1080/03610926.2016.1277760

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

[12] Ewemooje, O.S., Adebola, F.B., Adediran, A.A, "A stratified hybrid tripartite randomized response technique" Gazi University Journal of Science, 31(4), 1246-1266, (2018).

[13] Adeniran, A.T., Sodipo, A.A., Udomboso, C.G., "A modified force randomized response model". Journal of the Nigerian Society of Physical Sciences, 2: 36-50, (2020).

[14] Harriet, I., Julia, P.G., Freya, A.V., "Asking sensitive questions in conservation using randomized response techniques" Biological conservation, 260: 109191, (2021). DOI: https://doi.org/10.1016/j.biocon.2021.109191

[15] Singh, N., Vishwakarma, G.K., Kumar, N. Singh, H.P., "Estimation of Mean of Sensitive Variable Using Multiplicative Scramble Variable Under Measurement Error", Journal of Statistical Theory and Practice, 16(46): 1-11, (2022). DOI: https://doi.org/10.1007/s42519-022-00272-w

[16] Perri, P.F., Elvira, P., Manuela, S., "Estimating Induced Abortion and Foreign Irregular Presence using the Randomized Response Crossed Model", Social Indicators Research, 129(2): 601–618, (2015). DOI: https://doi.org/10.1007/s11205-015-1136-x

[17] Ewemooje, O.S., Amahia, G.N., Adebola, F.B., "Estimating prevalence of induced abortion and multiple sexual partners using improved randomized response technique for two sensitive attributes", Communications in Statistics: Case Studies, Data Analysis and Applications, 3(1-2): 21-28, (2017). DOI: https://doi.org/10.1080/23737484.2017.1398057

[18] Adebola, O.G, Ewemooje, O.S., Adebola, F.B., "Plagiarism and Examination Malpractice: Persistent Examples of Academic Dishonesty", African Journal of Applied Statistics, 6(2): 723-739, (2019).

[19] Adediran, A.A., Ewemooje, O.S., Adebola, F.B., "Unbiased estimator model using unrelated question in dichotomous randomized response" Statistics in transition, 21(5): 119-132, (2020). DOI: https://doi.org/10.21307/stattrans-2020-058

[20] Ewemooje, O.S., Adebola, F.B., Amahia, G.N., "Alternative estimator in dichotomous randomized response technique", Communication in mathematics and statistics, 7: 383-400, (2019).

[21] Chaudhuri, A., Stenger, H. "Survey Sampling: Theory and Methods", 2nd Edition, CRC Press, Boca Raton, FL, (2005). DOI: https://doi.org/10.1201/9781420028638

[22] Chaudhuri, A., "Randomized Response and Indirect Questioning Techniques in Surveys", Chapman & Hall/CRC, Boca Raton, (2011).

[23] Ewemooje, O.S., "An Alternative Estimator for Unrelated Questions in Tripartite Randomized Response Technique", Gazi University Journal of Science, 32(4): 1308-1320, (2019).

[24] Ewemooje, O.S., Adebola, F.B., Awogbemila, A.T., "Substance Use disorder prevalence using tripartite randomized technique", Nigerian Journal of Technological Research, 14(1): 55-61, (2019). DOI: https://doi.org/10.4314/njtr.v14i1.8