Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.13, 2014

A Modified Stratified Randomized Response Techniques

F.B. Adebola , O.O. Johnson , N.A. Adegoke *

Department of Statistics, Federal University of Technology Akure, P.M.B. 704, Nigeria.

* E-mail of the corresponding author: nurudeen.adegoke@yahoo.com

Abstract

This paper presents a modification of Kim and Warde's (2004) Stratified Warner's Randomized Response Technique (SWRRT) and Kim and Elam's (2007) Stratified Unrelated Question Randomized Response Technique with known distribution (SURRT) using optimal allocation. Our models further reduce the non-response bias by introducing the concept of sub-samples of non-respondent developed by Hansen and Hurwitz (1946) to the above mentioned models. In this paper we perform an empirical practice of our model and we also perform the empirical comparison of our models with both Kim and Warde (2004) and Kim and Elam (2007) models. We discovered a new note in the empirical comparative study of our models.

Keywords: Close supervision, Sensitive behavior, Non-respondent, Sub-sample, Randomized response techniques.

1. Introduction

In other to reduce biasness of respondent in surveys, most especially in surveys that the questions are so sensitive that might cause the respondent not to even answer the questions or falsely answer the questions; in this regard we appeal to Randomized Response Techniques. Randomized Response Techniques is the method of protecting the respondent's identity so as to collect reliable information on sensitive characteristics. Sensitive characteristics involve; rape, abortion, Sexually Transmitted Infection/HIV etc. Randomized Response Techniques help to reduce non-sampling error; non-sampling error involves response bias and non-response bias Warner (1965). Response bias is when the respondent gives a false answer and non-response bias is when the respondent refuses to answer the question.

Randomized Response Technique was introduced by Warner (1965) to estimate the proportion of the people in a given population who is involve in a particular social or legal stigmatized behavior. Warner uses a randomized device with known probability to encourage the respondent for truthful answer such that the researcher does not know which of the two questions he or she answers. Greenberg *et al.* (1969) proposed the unrelated question RR model that is a variation of Warner's (1965) RR model in which the respondent is presented with one sensitive question and one unrelated question (which is not related to the stigmatized behavior). Hong *et al*. (1994) presented a stratified randomized response technique using proportional allocation. Kim and Warde (2004) presented Stratified Warner's Randomized Response Model (SWRRT) using an optimal allocation. Kim and Elam (2007) presented the Stratified Unrelated Question Randomized Response Model (SURRT) using an optimal allocation.

This paper develops a Modified Stratified Randomized Response Model using optimal allocation. We developed for Warner (1965) and for Greenberg *et al*. (1969). As stated earlier that Randomized Response Techniques helps to reduce response and non-response bias. Our model further reduces the non-response bias by introducing the concept of sub-samples of non-respondent developed by Hansen and Hurwitz (1946). A comparative study of our two models was done as well so as to determine the most efficient.

2 Randomized Response Techniques

Warner (1965) gave a genius idea by using randomized device to encourage truthful answer from the respondent with respect to a sensitive behavior. The randomizing device, such as a spinning arrow, dice or coins is used to select one of the two questions; such as,

"*I have raped before*" (Presented with probability P)

"*I have not raped before*" (Presented with probability 1-P)

The respondents have the options "Yes" or "No" presented to him or her. The interviewer does not know which question any respondent has answered but knows the probability P and 1-P with which the two statements are presented. Here, with a random sample of n respondents, the interviewer records a binomial estimate

$$\hat{\theta} = \frac{y}{n}$$

of the proportion θ of "Yes" answers, where *y* is the number of yes answers.

If the questions are answered truthfully, the relation between θ_h and π_{Wh} in the population is given as:

$$\theta_h = \pi_{Wh} p_h + (1 - p_h)(1 - \pi_{Wh}) \tag{2.1}$$

Where π_{Wh} is the proportion of people with the stigmatized or sensitive behavior in stratum *h* using Warner's techniques and p_h is the probability of selecting the sensitive question in stratum *h*.

$$\hat{\pi}_{Wh} = \frac{\theta_h - (1 - p_h)}{(2p_h - 1)} \tag{2.2}$$

The Variance is given by

$$V(\hat{\pi}_{Wh}) = \frac{\pi_{Wh}(1 - \pi_{Wh})}{n_h} + \frac{p_h(1 - p_h)}{n_h(2p_h - 1)^2} \tag{2.3}$$

Horvitz *et al*. (1967) proposed the unrelated question design which is an alternative to the Warner's method. They discussed that combining one sensitive question and one unrelated question would enhance more cooperation of the respondents.

The unrelated question design was further improved by Greenberg *et al.* (1969). They examined the same design where the distribution of the non-sensitive question is known in advance. The unrelated question design with a known distribution is a simplified version of the original unrelated question design. To illustrate the pattern of the design;

Question 1: Do you have more than one girlfriend?

Question 2: Did you watch the 9.00pm cartoon yesterday?

By presenting the second question in this form gives the distribution of the unrelated question.

If all respond truthfully, the population proportion of "yes" answers is given by

$$\theta_h = p_h \pi_{Uh} + (1 - p_h)\, \pi_{Kh} \tag{2.4}$$

Where π_{Kh} is the proportion in the sampled population who watch the 9:00pm cartoon yesterday in stratum *h* which is always assume to be known.

If π_{Kh} is known, the obvious (and maximum likelihood) estimate of π_{Uh} is

$$\hat{\pi}_{Uh} = \frac{\hat{\theta}_h - (1 - p_h)\, \hat{\pi}_{Kh}}{p_h} \tag{2.5}$$

Where π_{Uh} is the proportion of people with the stigmatized or sensitive behavior in stratum *h* using Unrelated design techniques and p_h is the probability of selecting the sensitive question in stratum *h*.

The Variance is given by

$$V(\hat{\pi}_{Uk}) = \frac{\theta_h(1 - \theta_h)}{n_h p_h^2} \tag{2.6}$$

For a review, Hong *et al*. (1994) suggested a Stratified Randomized Response Techniques (RRT) under the proportional sampling assumption. Under Hong *et al.*'s (1994) proportional sampling assumption, it may be easy to derive the variance of the proposed estimator. However, it may come at a high cost in terms of time, effort, and money.

Mathematical Theory and Modeling www.iiste.org
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.13, 2014

To proffer solution to the deficiency of Hong *et al.* (1994), Kim and Warde (2004) and Kim and Elam (2005) both presented Stratified Randomized Response Techniques (SRRT) using an optimal allocation which are more efficient than the Hong *et al.* (1994) Stratified Randomized Response Techniques (SRRT) using a proportional allocation. (See Kim and Warde (2004))

They derived the probability θ_h of a "Yes" answer in stratum *h* for this procedure as: If all respond truthfully,

$$\theta_h = \pi_{Wh} p_h + (1-p_h)(1-\pi_{Wh}) \quad (2.7)$$

for h = 1, 2, . . ., L

Where π_{Wh} is the proportion of people with the sensitive trait in stratum *h*. The maximum likelihood estimate $\hat{\pi}_W$ of a sensitive proportion π_W is given by:

$$\hat{\pi}_W = \sum_{h=1}^{L} W_h \hat{\pi}_{Wh} \quad (2.8)$$

Where $W_h = \frac{N_h}{N}$ for h = 1,2,…, L so that $W = \sum_{h=1}^{L} W_h = 1$ (N is the number of unit in the population and N_h is the total number of units in stratum h). The minimum variance of the estimator $\hat{\pi}_W$ is given by (See Kim and Elam (2005)):

$$V(\hat{\pi}_W) = \frac{1}{n}\left[\sum_{h=1}^{L} W_h \left(\pi_{Wh}(1-\pi_{Wh}) + \frac{p_h(1-p_h)}{(2p_h-1)^2}\right)^{\frac{1}{2}}\right]^2 \quad (2.9)$$

Kim and Elam (2007) develop a Stratified Randomized Response Techniques (SRRT) model for Greenberg *et al.*'s (1969) Unrelated Question Randomized Response Techniques (URRT) model using an optimal allocation. Kim and Elam (2007) presented Stratified Randomized Response Techniques (SRRT) using an optimal allocation which are more efficient than the Kim and Warde (2004) and Kim and Elam (2005) Stratified Randomized Response Techniques (SRRT) (See Kim and Elam (2007))

They derived the probability θ_h of a "Yes" answer in stratum *h* for this procedure as: If all respond truthfully,

$$\theta_h = p_h \pi_{Uh} + (1-p_h)\pi_{Kh} \quad (2.10)$$

For h = 1,2, . . . L

Where π_{Kh} is the proportion in the sampled population who watch the 10:00pm news yesterday and it is known. Where π_{Uh} is the proportion of people with the sensitive trait in stratum *h*. The maximum likelihood estimate $\hat{\pi}_U$ of a sensitive proportion π_U is given by:

$$\hat{\pi}_U = \sum_{h=1}^{L} W_h \hat{\pi}_{Uh} \quad (2.11)$$

Where $W_h = \frac{N_h}{N}$ for *h = 1,2,..., L* so that $W = \sum_{h=1}^{L} W_h = 1$ (N is the number of unit in the population and N_h is the total number of units in stratum h)

The minimum variance of the estimator $\hat{\pi}_U$ given π_{Kh} is given by (See Kim and Elam (2007)):

$$V(\hat{\pi}_U \setminus \pi_{Kh}) = \frac{1}{n}\left[\sum_{h=1}^{L}\frac{w_h}{p_h}\sqrt{\theta_h(1-\theta_h)}\right]^2 \quad (2.12)$$

Here, we develop a Modified Stratified Warner's Randomized Response Technique (MSWRRT) and Modified Stratified Unrelated Question Randomized Response Technique with known distribution (MSURRT) using optimal allocation

3. Our Model

Hansen and Hurwitz (1946) provided a technique for dealing with the problem of the unit non-response in mail surveys. The technique is as follows:

The population considered is denoted by *N* and the population is stratified into *h* (*h* >1) strata. We have N_h the number of units in the population in stratum *h*. Stratification is the method of using auxiliary information to increase the precision of the estimate of the population characteristics. (See Okafor 2002).

A sample of size *n* was determined, the number of sample in each stratum was determined using the optimal allocation. Then we have the number of unit in each stratum as n_h such that $n = \sum_{h=1}^{L} n_h$. By Simple Random Sampling without Replacement; a questionnaire is mailed to each stratum and n'_{h1} and n'_{h2} is the number of unit responded and not responded in stratum *h* respectively. We have $n_h = n'_{h1} + n'_{h2}$

3.1 For the Respondent:

The interviewer records a binomial estimate $\hat{\theta}'_h = \frac{y}{n'_{h1}}$ of the proportion θ of "Yes" answers, where *y* is the number of yes answers.

For Warner (1965):

If the questions are answered truthfully, the relation between θ'_h and π'_{Wh} in the population is given as:

$$\theta'_h = \pi'_{Wh} p_h + (1-p_h)(1-\pi'_{Wh}) \quad (3.1)$$

Where π'_{Wh} is the proportion of people with the stigmatized or sensitive behavior in stratum *h* using Warner's techniques and p_h is the probability of selecting the sensitive question in stratum *h*.

$$\hat{\pi}'_{Wh} = \frac{\hat{\theta}'_h - (1-p_h)}{(2p_h - 1)} \quad (3.2)$$

The maximum likelihood estimate $\hat{\pi}'_{W1}$ of a sensitive proportion π'_{W1} for the respondent is given by:

$$\hat{\pi}'_{W1} = \sum_{h=1}^{L} W_h \hat{\pi}'_{Wh} \quad (3.3)$$

The Variance for stratum *h* is given by

$$V(\hat{\pi}'_{Wh}) = \frac{\pi'_{Wh}(1-\pi'_{Wh})}{n'_{h1}} + \frac{p_h(1-p_h)}{n'_{h1}(2p_h-1)^2} \quad (3.4)$$

Where $W_h = \frac{N_h}{N}$ for h = 1,2,…, L so that $W = \sum_{h=1}^{L} W_h = 1$ (N is the number of unit in the population and N_h is the total number of units in stratum h)

The minimum variance of the estimator $\hat{\pi}'_{W1}$ for the respondent is given by:

$$V(\hat{\pi}'_{W1}) = \frac{1}{n'_1}\left[\sum_{h=1}^{L} W_h \left(\pi'_{Wh}(1-\pi'_{Wh}) + \frac{p_h(1-p_h)}{(2p_h-1)^2}\right)^{\frac{1}{2}}\right]^2 \qquad (3.5)$$

Where $n'_1 = \sum_{h=1}^{L} n'_{h1}$

For Greenberg et al. (1969):

If all respond truthfully, the population proportion of "yes" answers is given by

$$\theta'_h = p_h \pi'_{Uh} + (1-p_h)\, \pi'_{Kh} \qquad (3.6)$$

Where π'_{Kh} is the proportion in the sampled population who watch the 9:00pm cartoon yesterday in stratum *h* which is always assume to be known.

If π'_{Kh} is known, the obvious (and maximum likelihood) estimate of π'_{Uh} is

$$\hat{\pi}'_{Uh} = \frac{\hat{\theta}'_h - (1-p_h)\, \hat{\pi}'_{Kh}}{p_h} \qquad (3.7)$$

Where π'_{Uh} is the proportion of people with the stigmatized or sensitive behavior in stratum *h* using Unrelated design techniques and p_h is the probability of selecting the sensitive question in stratum *h*.

The maximum likelihood estimate $\hat{\pi}'_{U1}$ of a sensitive proportion π'_{U1} for the respondent is given by:

$$\hat{\pi}'_{U1} = \sum_{h=1}^{L} W_h \hat{\pi}'_{Uh} \qquad (3.8)$$

Where $W_h = \frac{N_h}{N}$ for *h = 1,2,…, L* so that $W = \sum_{h=1}^{L} W_h = 1$ (N is the number of unit in the population and N_h is the total number of units in stratum h). The Variance for stratum *h* is given by

$$V(\hat{\pi}'_{Uh} \setminus \pi'_{Kh}) = \frac{\theta'_h(1-\theta'_h)}{n'_{h1} p_h^2} \qquad (3.9)$$

The minimum variance of the estimator $\hat{\pi}'_{U1}$ given π'_{Kh} is given by (See Kim and Elam (2007)):

$$V(\hat{\pi}'_{U1} \setminus \pi'_{Kh}) = \frac{1}{n'_1}\left[\sum_{h=1}^{L} \frac{W_h}{p_h}\sqrt{\theta'_h(1-\theta'_h)}\right]^2 \qquad (3.10)$$

Where $n'_1 = \sum_{h=1}^{L} n'_{h1}$

3.2 For the Non-Respondent:

A subsample of $\left(m_h = \frac{n'_{h2}}{k},\ k \geq 1\right)$ units is selected from the n'_{h2} non-respondents by SimpleRandom Sampling without Replacement in stratum *h*. The desired information is now collected from these m_h units by close supervision. This time around it is assumed that all the m_h units respond.

The interviewer records a binomial estimate $\hat{\theta}'_h = \frac{y}{m_h}$ of the proportion θ of "Yes" answers, where *y* is the number of yes answers and m_h is the sub-sample of the non-respondent in stratum *h*.

For Warner (1965):

If the questions are answered truthfully, the relation between θ'_h and π'_{Wh} in the population is given as:

$$\theta'_h = \pi'_{Wh} p_h + (1 - p_h)(1 - \pi'_{Wh}) \qquad (3.11)$$

Where π'_{Wh} is the proportion of people with the stigmatized or sensitive behavior in stratum *h* using Warner's techniques and p_h is the probability of selecting the sensitive question in stratum *h*.

$$\hat{\pi}'_{Wh} = \frac{\hat{\theta}'_h - (1 - p_h)}{(2p_h - 1)} \qquad (3.12)$$

The maximum likelihood estimate $\hat{\pi}'_{W2}$ of a sensitive proportion π'_{W2} for the respondent is given by:

$$\hat{\pi}'_{W2} = \sum_{h=1}^{L} W_h \hat{\pi}'_{Wh} \qquad (3.13)$$

The Variance for stratum *h* is given by

$$V(\hat{\pi}'_{Wh}) = \frac{\pi'_{Wh}(1 - \pi'_{Wh})}{m_h} + \frac{p_h(1 - p_h)}{m_h\ (2p_h - 1)^2} \qquad (3.14)$$

Where $W_h = \frac{N_h}{N}$ for h = 1,2,…, L so that $W = \sum_{h=1}^{L} W_h = 1$ (N is the number of unit in the population and N_h is the total number of units in stratum h)

The minimum variance of the estimator $\hat{\pi}'_{W2}$ for the respondent is given by:

$$V\left(\hat{\pi}'_{W1}\right)=\frac{1}{m}\left[\sum_{h=1}^{L}W_h\left(\pi'_{Wh}(1-\pi'_{Wh})+\frac{p_h(1-p_h)}{(2p_h-1)^2}\right)^{\frac{1}{2}}\right]^2 \qquad (3.15)$$

Where $m=\sum_{h=1}^{L}m_h$

For Greenberg et al. (1969):

If all respond truthfully, the population proportion of "yes" answers is given by

$$\theta'_h = p_h\pi'_{Uh}+\left(1-p_h\right)\pi'_{Kh} \qquad (3.16)$$

Where π'_{Kh} is the proportion in the sampled population who watch the 9:00pm cartoon yesterday in stratum h which is always assume to be known.

If π'_{Kh} is known, the obvious (and maximum likelihood) estimate of π'_{Uh} is

$$\hat{\pi}'_{Uh}=\frac{\hat{\theta}'_h-\left(1-p_h\right)\hat{\pi}'_{Kh}}{p_h} \qquad (3.17)$$

Where π'_{Uh} is the proportion of people with the stigmatized or sensitive behavior in stratum h using Unrelated design techniques and p_h is the probability of selecting the sensitive question in stratum h.

The maximum likelihood estimate $\hat{\pi}'_{U2}$ of a sensitive proportion π'_{U2} for the respondent is given by:

$$\hat{\pi}'_{U2}=\sum_{h=1}^{L}W_h\hat{\pi}'_{Uh} \qquad (3.18)$$

Where $W_h=\frac{N_h}{N}$ for $h = 1,2,..., L$ so that $W=\sum_{h=1}^{L}W_h=1$ (N is the number of unit in the population and N_h is the total number of units in stratum h)

The Variance for stratum h is given by

$$V\left(\hat{\pi}'_{Uh} \setminus \pi'_{Kh}\right) = \frac{\theta'_h\left(1-\theta'_h\right)}{m_h\ p_h^2} \tag{3.19}$$

The minimum variance of the estimator $\hat{\pi}'_{U2}$ given π'_{Kh} is given by (See Kim and Elam (2007)):

$$V\left(\hat{\pi}'_{U2} \setminus \pi'_{Kh}\right) = \frac{1}{m}\left[\sum_{h=1}^{L}\frac{W_h}{p_h}\sqrt{\theta'_h(1-\theta'_h)}\right]^2 \tag{3.20}$$

Where $m = \sum_{h=1}^{L} m_h$

3.3 The True Proportion of the Sensitive Behavior

The two independent estimates of the proportion of the sensitive behavior obtained from the respondent and non-respondent strata are suitably weighted to form the estimate of the population proportion of the sensitive behavior under study (See Hansen and Hurwitz (1946)).

Now, extending Hansen and Hurwitz (1946) work to randomized response techniques,

Then we have,

The unbiased estimator for proportion of the sensitive behavior is given by:

$$\hat{\pi}' = \frac{n'_1\pi'_1 + n'_2\pi'_2}{n} = w'_1\pi'_1 + w'_2\pi'_2 \tag{3.21}$$

π'_1 is the estimate of the proportion of the sensitive behavior of the n'_1 respondents at the first mailings.

π'_2 is the estimate of the proportion of the sensitive behavior of the *m* sub-sampled non-respondents at the second call.

Where, $w'_1 = \frac{n'_1}{n}$ *and* $w'_2 = \frac{n'_2}{n}$

The unbiased variance of the sensitive behavior given by:

$$\hat{V}\left(\hat{\pi}'\right) = \frac{N-n}{Nn(n-1)}\left[(n'_1-1)\frac{n'_1\pi'_1(1-\pi'_1)}{n'_1-1} + (n'_2-k)\frac{m\,\pi'_2(1-\pi'_2)}{m-1} + n'_1\pi_1'^2 + n'_2\pi_2'^2 - n\pi'^2\right]$$
$$+ \frac{N-1}{N(n-1)}w'_2(k-1)\frac{m\,\pi'_2(1-\pi'_2)}{m-1} \tag{3.22}$$

The unbiased variance of $\hat{\pi}'$ of π' is obtained upon replacing s_1^2 is replaced by $\frac{n_1'\pi_1'(1-\pi_1')}{n_1'-1}$, s_{2m}^2 by $\frac{m\pi_2'(1-\pi_2')}{m-1}$; $\bar{y}_1$ is replaced by π_1'; $\bar{y}_{2m}$ by π_2' and $\bar{y}_T$ by π' (See Hansen and Hurwitz (1946)) and Okafor (2002).

Equation (3.21) and (3.22) can be used for both Warner (1965) and Greenberg *et al.* (1969)

4 Our Study

Kim and Warde (2004) and Kim and Elam (2007) have presented Stratified Randomized Response Techniques (SRRT) and has done the efficiency comparison (See Kim and Elam (2007)), but we still have to compare our model with this two models. Here, we perform an empirical practice of our model and we also perform the empirical comparison of our models with both Kim and Warde (2004) and Kim and Elam (2007) models. We apply our model to sexual behavior among students in which we administered five different sensitive questions. The questions involve:

1. Have you contracted Sexually transmitted disease before?
2. Have you aborted for a lady or abort a pregnancy before?
3. Have you raped or being raped before?

These questions were considered sensitive because they are social undesirable attributes such that people do not feel comfortable to disclose information about it. For Greenberg et al (1969), we paired these questions with an unrelated question with known distribution.

4 1 Study Design

The population considered in the survey was the undergraduate students of the Federal University of Technology, Akure, Nigeria. The population was stratified into two in which we have the Male and the Female (*that is h= 1, 2*). Stratification is the method of using auxiliary information to increase the precision of the estimate of the population characteristics. (See Okafor 2002).

Two questionnaires were design to evaluate the performance of our model; one using the Warner (1965) design and the other using the Greenberg et al (1969) design. We asked some questions before asking the sensitive questions.

The total population of students in the university (N = 12051, (8782 males and 3269 females)), we chose a sample size of n=250 for each design, optimal allocation was used to determine the number of people in each stratum, we have $n_1 = 181$ males and $n_2 = 69$ females. Out of the 250 questionnaire sent out for the Stratified Warner's Randomized Response Techniques, only 183 (144 males and 39 females) was useable (respondent), 67 was recorded non-respondent and out of the 250 questionnaire sent out for Stratified Unrelated Question Randomized Response Techniques, only 200 (155 males and 45 females) was useable(respondent), 50 was recorded non-respondent.

We perform a sub-sample of non-respondent by choosing $k = 2$, this choice of k is to get a large sub-sample. As the procedure implies, the sub-sample was done with close supervision (this does not mean that we actually knew which question the respondent answered but we help to simplify the instruction so as to get more cooperation of the respondent). All the k sub-sample responses were now useable.

4 2 Results

Table 1. The Response and the Non-Response of the Survey

	*SWRRT				**SURRT			
	MALE	%	FEMALE	%	MALE	%	FEMALE	%
Number originally sent	181	100%	69	100%	181	100%	69	100%
Number unreturned	15	8.29%	10	14.49%	12	6.63%	6	8.70%
Number returned	166	91.71%	59	85.51%	169	93.37%	63	91.30%
Invalid response	22	12.15%	20	28.98%	14	7.73%	18	26.09%
Useable response	144	85.56%	39	56.52%	155	85.64%	45	65.21%
Non-respondent	37	14.44%	30	43.48%	26	14.36%	24	34.79%

* Stratified Warner's Randomized Response Techniques (Kim and Warde (2004))

** Stratified Unrelated Question Randomized Response Techniques (Kim and Elam (2007))

4.2.1 For Respondent

The table below presents the number of yes response and the proportion of student with the sensitive behavior alongside with the corresponding variances.

Table 2. For Male

	SWRRT*				SURRT**				
Type of Sexual behavior	Useable response	Yes Response	Proportion with the sensitive behavior	Variance of the sensitive behavior	Useable response	Yes Response	Proportion with the sensitive behavior	Variance of the sensitive behavior	Relative Efficiency: R.E
Q1	144	45	3.12%	0.0093	155	48	31.38%	0.0028	331%
Q2	144	56	22.22%	0.0103	155	66	47.97%	0.0032	320%
Q3	144	44	1.39%	0.0092	155	48	31.38%	0.0028	327%

Table 3. For Female

	SWRRT*				SURRT**				
Type of Sexual behavior	Useable response	Yes Response	Proportion with the sensitive behavior	Variance of the sensitive behavior	Useable response	Yes Response	Proportion with the sensitive behavior	Variance of the sensitive behavior	Relative Efficiency: R.E
Q1	39	13	8.33%	0.0356	45	11	22.06%	0.0084	425%
Q2	39	14	14.74%	0.0369	45	16	37.94%	0.0104	355%
Q3	39	12	1.92%	0.0341	45	18	44.29%	0.0109	314%

Table 4. Proportion of the Sensitive Behavior (Male and Female) For Respondents

	SWRRT*		SURRT**		
Type of Sexual behavior	Proportion with the sensitive behavior	Variance of the sensitive behavior	Proportion with the sensitive behavior	Variance of the sensitive behavior	Relative Efficiency: R.E
Q1	4.56%	0.0076	28.81%	0.0021	362%
Q2	20.16%	0.0082	45.20%	0.0025	328%
Q3	1.54%	0.0074	34.94%	0.0023	322%

4.2.2 Sub-Sample of Non-Respondent

Table 5. The Table Below Shows the Result of the Survey:

	SWRRT*				SURRT**			
	MALE	%	FEMALE	%	MALE	%	FEMALE	%
Total non-response	37		30		26		24	
Number sub-sampled	19	100%	15	100%	13	100%	12	100%
Number returned	19	100%	15	100%	13	100%	12	100%
Useable Response	19	100%	15	100%	13	100%	12	100%

The table below presents the number of yes response and the proportion of student with the sensitive behavior alongside with the corresponding variances for subsample of non-response.

Table 6. For Male

	SWRRT*				SURRT**				
Type of Sexual behavior	Useable response	Yes Response	Proportion with the sensitive behavior *	Variance of the sensitive behavior **	Useable response	Yes Response	Proportion with the sensitive behavior*	Variance of the sensitive behavior **	Relative Efficiency: R.E ***
Q1	19	9	17.11%	0.0765	13	5	42.09%	0.0372	206%
Q2	19	7	43.42%	0.0820	13	6	53.08%	0.0390	210%
Q3	19	5	3.95%	0.0711	13	6	53.08%	0.0390	182%

Table 7. For Female

	SWRRT				SURRT				
Type of Sexual behavior	Useable response	Yes Response	Proportion with the sensitive behavior	Variance of the sensitive behavior	Useable response	Yes Response	Proportion with the sensitive behavior	Variance of the sensitive behavior	Relative Efficiency: R.E
Q1	15	5	8.33%	0.1138	12	4	34.76%	0.0378	301%
Q2	15	7	41.67%	0.1195	12	6	58.57%	0.0425	281%
Q3	15	5	8.33%	0.1138	12	5	46.67%	0.0413	275%

Table 8. The Total Proportion of the Sensitive Behavior for the Non-Respondents

	SWRRT		SURRT		
Type of Sexual behavior	Proportion with the sensitive behavior	Variance of the sensitive behavior	Proportion with the sensitive behavior	Variance of the sensitive behavior	Relative Efficiency: R.E
Q1	13.18%	0.0462	38.57%	0.0188	246%
Q2	42.64%	0.0490	55.71%	0.0203	241%
Q3	5.91%	0.0445	50.00%	0.0201	222%

4.2.3 The true estimate (Combined Estimate)

Here we combine the estimates from the respondents and the non-respondents using the method developed by Hansen and Hurwitz (1946).

Table 9. The True Proportion of the Sensitive Behavior for the Respondents and Non-Respondents.

	Warner (1965)		Greenberg et al (1969)		
Type of Sexual behavior	Proportion with the sensitive behavior	Variance of the sensitive behavior	Proportion with the sensitive behavior	Variance of the sensitive behavior	Relative Efficiency: $RE = \frac{Var(\hat{\pi}_{AW})}{Var(\hat{\pi}_{AU})}$
Q1	6.87%	0.0004	30.76%	0.0010	40%
Q2	26.18%	0.0010	47.30%	0.0012	83%
Q3	2.71%	0.0002	37.96%	0.0011	18%

4.3 Comparing the Efficiency of our Model with Kim and Warde (2004) and Kim and Elam (2007) Model

Table 10. The Variance of the Models and the Relative Efficiency

QUESTIONS	SWRRT (2004) (A)	SURRT (2007) (B)	SWRRT (OUR MODEL) (C)	SURRT (OUR MODEL) (D)	* R.E 1 $= \frac{A}{C} \times 100$	** R.E 2 $= \frac{B}{D} \times 100$	***R.E3 $= \frac{C}{D} \times 100$
Q1	0.0076	0.0021	0.0004	0.0010	1900%	210%	40%
Q2	0.0082	0.0025	0.0010	0.0012	820%	208%	83%
Q3	0.0074	0.0023	0.0002	0.0011	3700%	209%	18%

*R.E 1 is the relative efficiency of Kim and Warde (2004) and our model for Warner (1965).

**R.E 2 is the relative efficiency of Kim and Elam (2007) model and our mode for Greenberg *et al.* (1969).

***R.E 3 is the relative efficiency of our model for Warner (1965) and our model for Greenberg *et al.* (1969).

Relative efficiency greater than 100% implies that denominator is better (or more efficient), otherwise the numerator is better (or more efficient).

From our results in the table above we can see that the Variances in column C and D (which are our models) are lesser than the variances in column A and B (which are the Kim and Warde (2004) and Kim and Elam (2007) models).

From R.E 1, our model for Warner (1965) is more efficient than the Kim and Warde (2004) considering the relative efficiency.

From R.E 2, our model for Greenberg et al (1969) is more efficient than Kim and Elam (2007) considering the relative efficiency.

We can also deduce from R.E 3 that the comparative study shows that our model for Warner (1965) is more efficient than our model for Greenberg *et al.* (1969).

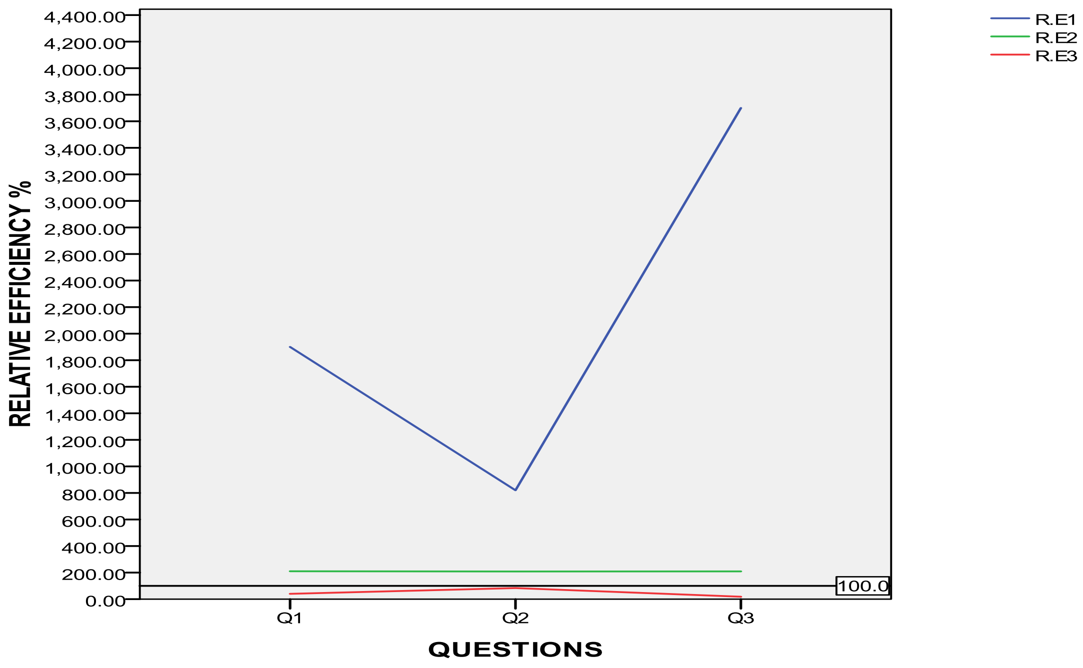

Fig 1. The relative efficiency of the variances.

5 Conclusion

We have successfully presented a model for Warner (1965) using optimal allocation and for Greenberg *et al.* (1969) using optimal allocation. The empirical comparative study as shown that our model is more efficient in terms of variance than the Kim and Warde (2004) and Kim and Elam (2007) models. Likewise, the comparative study of our two models shows that our model for Warner (1965) is more efficient in terms of variance than our model for Greenberg *et al.* (1969).

As discussed earlier that Greenberg *et al.* (1969) encourage more cooperation from the respondent because of the unrelated paired question, which make it more efficient than Warner (1965) (See Greenberg *et al.* (1969)), but we have been able to cater for the deficiency of Warner (1965) by introducing subsample of non-respondent and this gave birth to a new note in the comparative study.

Conclusively, as stated earlier that Randomized Response Techniques helps to reduce response and non-response bias, in addition, our model further reduces the non-response bias.

References

Adebola, F.B. & Adegoke, N.A. (2013), "A Survey of Examination Malpractices using the Randomized Response Technique", *Journal of the Nigerian Association of Mathematical Physics* 23, 375-388.

Greenberg, B.G., Abul-Ela, A.A., Simmons, W.R. & Horvitz, D.G. (1969), "The Unrelated Question Randomized Response: Theoretical Framework", *Journal of the American Statistical Association* 64, 520-539.

Hansen, M.H. & Hurwitz, W.N. (1946), "The Problem of Non-Response in Sample Surveys", *Journal of the American Statistical Association* 41, 517-529.

Hong, K., Yum, J. & Lee, H. (1994), "A Stratified Randomized Response Technique", *The Korean Journal of Applied Statistics* 7, 141-147.

Horvitz, D.G., Shah, B.U. & Simmons, W.R. (1967), "The Unrelated Question Randomized Response Model", Proceedings of the Social Statistics Section. *Journal of the American Statistical Association* 60, 63-69.

Kim, J-M. & Warde, W.D. (2004), "A Stratified Warner's Randomized Response Model", *Journal of Statistical Planning and Inference* 120 (1-2), 155-165.

Kim, J-M. &Elam, M.E. (2005), "A Two-Stage Stratified Warner's Randomized Response Model Using Optimal Allocation", *Metrika* 61, 1-7.

Kim J-M. & Elam, M.E. (2007), "A Stratified Unrelated Question Randomized Response Model", *Statistical Paper,* 48, 215-233.

Okafor, F.C. (2002), "*Sample Survey Theory with Applications",* First edition, Afro-Orbis publishers, Nsukka, Nigeria.

Warner, S.L. (1965), "Randomized response: A Survey Technique for Eliminating Evasive Answer Bias", *Journal of the American Statistical Association* 60, 63-69.