Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(2) 2016 131–138

Generalization of Randomized Response Techniques for Complex Survey Designs

Raghunath Arnab, D.K. Shangodoyin and V.K. Dwivedi
Department of Statistics, University of Botswana, Botswana

Received 11 May 2016; Revised 19 May 2016; Accepted 30 May 2016

SUMMARY

Lee *et al*. (2015) proposed a randomized response model which possesses a Cramer-Rao lower bound of variance than that of suggested by Singh and Sedory (2011) at equal or greater protection of respondents. The proposed method Lee *et al*. (2015) does not provide explicit expression of the estimator of the population proportion of sensitive attribute π and it is limited only to simple random sampling with replacement (SRSWR) sampling design. In this paper, we propose the generalization of Lee *et al*. (2015)'s method for complex survey designs in a unified set up. This is applicable for wider classes of sampling design and estimators. Explicit expressions of the unbiased estimators, variances and their unbiased estimators are provided under wider classes of sampling designs.

Keywords: Complex sampling designs, Randomized response, Relative efficiency.

1. INTRODUCTION

While collecting information directly from respondents that relates to sensitive issues such as induced abortion, drug addiction, duration of suffering from Aids and so on, the respondents very often report untrue values or even refuse to respond. Warner (1965) introduced an ingenious technique known as the randomized response technique (RR) for estimating π, the proportion of the population possessing a certain stigmatizing characteristic A (say) that protects the privacy of respondents and so prevents an unacceptable rate of non-response. Warner's (1965) technique was modified by Horvitz *et al*. (1967), Greenberg *et al*. (1969), Kim (1978), Franklin (1989), Mangat and Singh (1990), Kuk (1990), Gjestvang and Singh (2006), Singh *et al*. (2006), Arnab (2004), Arnab and Mothupi (2015), Lee *et al*. (2015) among other researchers to improve co-operation and efficiency.

A generalization of Lee *et al*. (2015) RR model for complex survey designs is presented in this study. Complex sampling involves clustering, stratification and unequal probability of selection of samples among others. Most of the surveys in practice are complex and multi-character in nature where information on more than one character is collected at a time. Some of them are of a confidential nature while the others are not. In this paper, we will propose methodology of estimating a population proportion of a sensitive characteristic for a complex multi-character survey design where the data for sensitive characteristic is collected by RR technique and varying probability sampling method was used for selection of sample. Expression of unbiased estimator of the population proportion, its variance and unbiased estimator of the variance are derived. We will describe some of randomized response (RR) techniques relevant to the present study.

Corresponding author: Raghunath Arnab
E-mail address: arnabr@mopipi.ub.bw

1.1 Warner's Technique

In this method, a sample s of size n units is selected from the population by the SRSWR method and information on a sensitive characteristic is collected by using a randomization device. The randomization device consists of a deck of cards with each card having one of the following two statements: (i) "I belong to the sensitive group A"; (ii) "I do not belong to the sensitive group A". The statements occur with known relative frequencies P_0 and $(1 - P_0)$, respectively, in the deck of cards. Each respondent in the sample is asked to select one card at random from the well-shuffled deck. Without showing the card to the interviewer, the interviewer answers the question, "Is the statement true?" The respondent answers "Yes" or "No" to the interviewer. Confidentiality of the respondent is maintained because the interviewer will not know which question the respondent has answered. Let n_1 be the total number of "Yes" answers obtained from the sampled respondents. Warner (1965) proposed the following unbiased estimator for π:

$$\hat{\pi}_w = \frac{\hat{p}_w - (1 - P_0)}{2P_0 - 1} \tag{1.1}$$

where $\hat{p}_w = n_1/n$ = proportion of "Yes" answer and $P_0 \neq 0.5$. The variance of the estimator $\hat{\pi}_w$ is given by:

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{P_0(1-P_0)}{n(2P_0 - 1)^2} \tag{1.2}$$

1.2 Mangat and Singh's Technique

In the Mangat and Singh RR (1990) technique, the randomization was done in two stages. In the first stage, each of the respondents selected in the sample was asked to use the randomization device R_1. This device consisting of two types of cards: type I of known proportion T_0, on which is written, "I belong to the sensitive group A", and type II of proportion $(1 - T_0)$ on which is written "Go to the randomization device R_2". In the second stage, if so directed by the outcome of R_1, the respondent is requested to use R_2 which is the same as the Warner (1965) device. Under this two-stage RR technique an unbiased estimator of the population proportion π is given by:

$$\hat{\pi}_{ms} = \frac{\hat{p}_{ms} - (1-P_0)(1-T_0)}{(2P_0 - 1) + 2T_0(1-P_0)} \tag{1.3}$$

where $\hat{p}_{ms}$ is the proportion of "Yes" answers obtained from a selected sample of size n selected by SRSWR. The variance of the estimator $\hat{\pi}_{ms}$ is given by:

$$V(\hat{\pi}_{ms}) = \frac{\pi(1-\pi)}{n} + \frac{(1-P_0)(1-T_0)\left[1-\left(1-P_0\right)\left(1-T_0\right)\right]}{n\left\{\left(2P_0-1\right)+2T_0\left(1-P_0\right)\right\}^2} \tag{1.4}$$

1.3 Odumade and Singh's Technique

Odumade and Singh (2009) selected a sample of size n by SRSWR. Each of the selected respondents in the sample is asked to select two cards, one card from Deck-I and the other from Deck-II. Each of the decks consists of two types of cards as in the Warner (1965) model. The proportion of cards bearing the statement "I belong to the sensitive group A" in Deck-I and Deck-II are P and T respectively. The respondents are asked to report his/her response as (X, Y) where X indicates response from the card selected from Deck-I while Y indicates response from the card selected from Deck-II. For example if a respondent selects a card written "I belong to the sensitive group A" from the Deck-I and selects the other card written "I do not belong to the sensitive group A" from the Deck-II, then he/she will supply with a response (Yes, No) if he/she belongs to the sensitive group A. On the other hand if the respondent do not belongs to the group A, he/she will supply (No, Yes) as his/her response. From the n responses, let n_{11}, n_{10}, n_{01} and n_{00} denote, respectively, the frequencies of the responses (Yes, Yes), (Yes, No), (No, Yes) and (No, No). Odumade and Singh (2009) proposed an unbiased estimator for the population proportion π given by:

$$\hat{\pi}_{os} = \frac{1}{2} + \frac{(P+T-1)(n_{11}-n_{00}) + (P-T)(n_{10}-n_{01})}{2n\left\{(P+T-1)^2 + (P-T)^2\right\}} \quad (1.5)$$

The variance of the estimator $\hat{\pi}_{os}$ and an unbiased estimator of the variance of $\hat{\pi}_{os}$ are, respectively, given by:

$$V(\hat{\pi}_{os}) = \frac{(P+T-1)^2\{PT+(1-P)(1-T)\} + (P-T)^2\{T(1-P)+P(1-T)\}}{4n[(P+T-1)^2+(P-T)^2]^2} - \frac{(2\pi-1)^2}{4n} \quad (1.6)$$

and

$$\hat{V}(\hat{\pi}_{os}) = \frac{1}{4(n-1)}\left[\frac{(P+T-1)^2\{PT+(1-P)(1-T)\} + (P-T)^2\{T(1-P)+P(1-T)\}}{[(P+T-1)^2+(P-T)^2]^2} - (2\hat{\pi}_{os}-1)^2\right] \quad (1.7)$$

Recently Singh and Sedory (2011) developed a lower bound on the variance for (1.9) as

$$V(\hat{\pi}_{ss}) \geq \frac{1}{n\left\{(P+T-1)^2\left(\frac{1}{\theta_{11}^*}+\frac{1}{\theta_{00}^*}\right) + (P-T)^2\left(\frac{1}{\theta_{10}^*}+\frac{1}{\theta_{01}^*}\right)\right\}} \quad (1.8)$$

1.4 Lee *et al.* (2015)

Lee *et al*. (2015) selected a sample s of size n by SRSWR method. If the selected respondent in the sample s does not belong to the sensitive group A^C, he or she is directed to draw one card from each of the two decks D1 and D2 of the Pair I (shown in Fig. 1) and was asked to report one of the following four responses: (Yes, Yes) or (Yes, No) or (No, Yes) or (No, No). If a respondent belongs to the sensitive group A, he or she is directed to perform similar randomized device using the Pair type II with decks cards D3 and D4.

Pair I	
Deck-DI	**Deck-D2**
Q_1 : I belong to the group A with probability P_1 Q_2 : I belong to the group A^c with probability $(1-P_1)$	Q_1 : I belong to the group A with probability P_2 Q_2 : I belong to the group A^c with probability $(1-P_2)$
Pair II	
Deck-D3	**Deck-D4**
Q_1 : I belong to the group A^c with probability T_1 Q_2 : I belong to the group A with probability $(1-T_1)$	Q_1 : I belong to the group A^c with probability T_2 Q_2 : I belong to the group A^c with probability $(1-T_2)$

Fig. 1. Representation of Deck's Combination

The probability of obtaining the four randomized responses (Yes, Yes) or (Yes, No) or (No, Yes) or (No, No) from a respondent was obtained by Lee *et al*. (2015) as follows:

$$\theta_{11}^* = \pi(P_1P_2 - T_1T_2) + T_1T_2,$$

$$\theta_{10}^* = \pi\{P_1(1-P_2) - T_1(1-T_2)\} + T_1(1-T_2)$$

$$\theta_{01}^* = \pi\{P_2(1-P_1) - T_2(1-T_1)\} + T_2(1-T_1),$$

$$\theta_{00}^* = \pi\{(1-P_1)(1-P_2) - (1-T_1)(1-T_2)\} + (1-T_1)(1-T_2)$$

Let $n_{11}^*, n_{10}^*, n_{01}^*, n_{00}^*$ be the total number of (Yes, Yes), (Yes, No), (No, Yes) and (No, No) answers. Then θ_{ij}^* can be unbiasedly estimated by the proportion $\hat{\theta}_{ij}^* = n_{ij}^*/n; i,j = 0,1$. The likelihood function of θ_{ij}^* is given by

$$L^* = \binom{n}{n_{11}^* \quad n_{10}^* \quad n_{01}^* \quad n_{00}^*} \left(\theta_{11}^*\right)^{n_{11}^*}\left(\theta_{10}^*\right)^{n_{10}^*}\left(\theta_{01}^*\right)^{n_{01}^*}\left(\theta_{00}^*\right)^{n_{00}^*}$$

Lee *et al*. (2015) proposed the maximum likelihood estimator of the population proportion π as the solution of the equation

$$\frac{\partial Log\left(L^*\right)^{n_{01}^*}}{\partial \pi} = 0,$$ given as

$$\frac{\hat{\theta}_{11}^*(P_1P_2-T_1T_2)}{\theta_{11}^*}+\frac{\hat{\theta}_{10}^*\{(P_1(1-P_2)-T_1(1-T_2)\}}{\theta_{10}^*}$$
$$+\frac{\hat{\theta}_{01}^*\{(1-P_1)P_2-(1-T_1)T_2\}}{\theta_{01}^*}$$
$$+\frac{\hat{\theta}_{00}^*\{(1-P_1)(1-P_2)-(1-T_1)(1-T_2)\}}{\theta_{00}^*}=0$$

Lee *et al*. (2015) did not provide any explicit expression of the Maximum Likelihood estimator of π. However they derived the Cramer-Rao lower bound of the variance of the maximum likelihood estimator $\hat{\pi}_{mle}$ of π as

$$V(\hat{\pi}_{mle})\geq \frac{1}{n\left\{\frac{(P_1P_2-T_1T_2)^2}{\theta_{11}^*}+\frac{\{(P_1(1-P_2)-T_1(1-T_2)\}^2}{\theta_{10}^*}+\frac{\{(1-P_1)P_2-(1-T_1)T_2\}^2}{\theta_{01}^*}+\frac{\{(1-P_1)(1-P_2)-(1-T_1)(1-T_2)\}^2}{\theta_{00}^*}\right\}}$$

Finally Lee *et al*. (2015) compared the estimate of the standard error of the lower bound of the variance estimator of the proposed estimator $\hat{\pi}_{mle}$ with the standard error of the lower bound of Odumonde and Singh obtained by Singh and Sedory (2011, 2012).

The strategy proposed by Lee *et al*. (2015) mentioned above can be used only for an SRSWR sampling design. Most of the surveys in practice are complex surveys and information on more than one character is collected at a time. Some of them are sensitive and others are not. To alleviate the aforementioned problem, we have formulated general method of estimation of the population π, variance of the proposed estimator and its unbiased estimator for any complex survey designs. The details are given below.

2. PROPOSED STRATEGY

In this strategy, an initial sample s of size n is selected from a finite population U = (U_1, U_2, ..., U_N) of N identifiable units by some arbitrary sampling design p. Let π_i (> 0) and π_{ij} (>0) be the inclusion probabilities for the i^{th} and i^{th} and j^{th} ($i \neq j$) unit, respectively. Then each of the selected respondent in the sample s is asked to use the randomization device suggested by Odumade and Singh (2009). From each of the respondents one of the answers (Yes, Yes), (Yes, No), (No, Yes) and (No, No) is obtained as his/her randomize response. Let us denote:

$$x_{11}(i)=\begin{cases}1 & \text{if the answer from the } i^{th} \text{ unit from the deck D1 is "YES"}\\ 0 & \text{otherwise}\end{cases} \quad (2.1)$$

$$x_{12}(i)=\begin{cases}1 & \text{if the answer from the } i^{th} \text{ unit from the deck D2 is "YES"}\\ 0 & \text{otherwise}\end{cases} \quad (2.2)$$

$$x_{21}(i)=\begin{cases}1 & \text{if the answer from the } i^{th} \text{ unit from the deck D3 is "YES"}\\ 0 & \text{otherwise}\end{cases} \quad (2.3)$$

and

$$x_{22}(i)=\begin{cases}1 & \text{if the answer from the } i^{th} \text{ unit from the deck D4 is "YES"}\\ 0 & \text{otherwise}\end{cases} \quad (2.4)$$

Further, let us denote an indicator variable attached to the i^{th} unit as:

$$y_i=\begin{cases}1 & \text{if the } i^{th} \text{ unit } \in \text{A}\\ 0 & \text{if the } i^{th} \text{ unit } \notin \text{A}\end{cases} \quad (2.5)$$

Probability of getting answer "Yes" from the deck D_1, D_2, D_3 and D_4 are respectively given by

$$\theta_{11}(i)=y_iP_1+(1-y_i)(1-P_1);$$
$$\theta_{12}(i)=y_iP_2+(1-y_i)(1-P_2);$$
$$\theta_{21}(i)=y_iT_1+(1-y_i)(1-T_1) \text{ and}$$
$$\theta_{22}(i)=y_iT_2+(1-y_i)(1-T_2)$$

Noting that $x_{11}(i)$ is a Bernoulli variable with parameter $\theta_{11}(i)$, we have

$$E(x_{11}(i))=\theta_{11}(i)=y_iP_1+(1-y_i)(1-P_1) \quad (2.6)$$

and

$$V(x_{11}(i))=\theta_{11}(i)(1-\theta_{11}(i)) \quad (2.7)$$

The equation (2.6) yields an unbiased estimator of y_i based on the observation $x_{11}(i)$ obtained from the deck $D1$ is

$$\hat{y}_i(1) = \frac{x_{11}(i) - (1-P_1)}{(2P_1 - 1)} \tag{2.8}$$

The variance of $\hat{y}_i(1)$ is

$$V\left(\hat{y}_i(1)\right) = \frac{V\left(x_{11}(i)\right)}{(2P_1-1)^2} = \frac{\theta_{11}(i)\left(1-\theta_{11}(i)\right)}{\left(2P_1-1\right)^2}$$

$$= \frac{P_1\left(1-P_1\right)}{\left(2P_1-1\right)^2} = \phi_1\left(\text{say}\right) \tag{2.9}$$

Similarly, estimators of y_i and their variances based on decks D_2, D_3 and D_4 are given as follows:

$$\hat{y}_i(2) = \frac{x_{12}(i) - (1-P_2)}{(2P_2-1)},$$

$$V\left(\hat{y}_i(2)\right) = \frac{P_2(1-P_2)}{(2P_2-1)^2} = \phi_2$$

$$\hat{y}_i(3) = \frac{x_{21}(i) - (1-T_1)}{(2T_1-1)},$$

$$V\left(\hat{y}_i(3)\right) = \frac{T_1(1-T_1)}{(2T_1-1)^2} = \phi_3$$

$$\hat{y}_i(4) = \frac{x_{22}(i) - (1-T_2)}{(2T_2-1)},$$

$$V\left(\hat{y}_i(4)\right) = \frac{T_2(1-T_2)}{(2T_2-1)^2} = \phi_4 \tag{2.10}$$

Since $\hat{y}_i(1)$, $\hat{y}_i(2)$, $\hat{y}_i(3)$ and $\hat{y}_i(4)$ are independently distributed, we find the optimum estimator of y_i from the i$^{\text{th}}$ respondent based on the data obtained from four decks of cards as

$$\hat{y}_i = \sum_{j=1}^{4} w_j \hat{y}_i(j) \tag{2.11}$$

where $w_j = \dfrac{1}{\phi_j}\left(\displaystyle\sum_{k=1}^{4}\frac{1}{\phi_k}\right)^{-1}$

The variance of $\hat{y}_i$ is

$$V\left(\hat{y}_i\right) = \overline{\phi} = \left(\sum_{k=1}^{4}\frac{1}{\phi_k}\right)^{-1} \tag{2.12}$$

The above findings yield the following results:

Theorem 1:

1. $E_R(\hat{y}_i) = y_i$
2. $V_R(\hat{y}_i) = \overline{\phi}$ where $\hat{y}_i$ and $\overline{\phi}$ are given in (2.11) and (2.12) respectively.

It should be noted that the variance $\overline{\phi}$ is known and independent of y_i's. In case $P_1 = P_2$ and $T_1 = T_2$, Lee *et al*. (2015) RR technique reduces to Odumade and Singh's (2009) RR technique where each respondent supplies two independent randomized responses. Further, if $P_1 = P_2 = T_1 = T_2 = P$, Lee *et al*. (2015) RR model reduces to Warner's (1965) model where each respondent supplies four independent RR responses.

In this case,

$$\phi_i = \frac{P\left(1-P\right)}{\left(2P-1\right)^2} = \overline{\phi} \quad \text{and} \quad \hat{y}_i = \frac{\overline{z}_i - (1-P)}{(2P-1)}$$

where $\overline{z}_i = \dfrac{1}{4}\displaystyle\sum_{i=1}^{4} x_{11}(i)$ = proportion of "Yes" answers obtained by i^{th} respondents in four independent RR trials.

2.1 Unbiased Estimation of Population Proportion π

Consider a linear homogeneous unbiased estimator of the population proportion π as:

$$\hat{\pi}_{as} = \frac{1}{N}\left(\sum_{i\in s} b_{si}\hat{y}_i\right) \tag{2.13}$$

where b_{si}'s are known constants and satisfy unbiasedness condition:

$$\sum_{s\supset i} b_{si}\, p(s) = 1 \tag{2.14}$$

and $\sum_{i\in s}$ denotes the sum over distinct units in the sample s.

Now using method proposed in Arnab and Shangodoyin (2015), we get the following theorem:

Theorem 2:

1. The estimator $\hat{\pi}_{as}$ is unbiased for the population proportion π.

2. The variance of the estimator $\hat{\pi}_{as}$ is given by:

$$V(\hat{\pi}_{as}) = \frac{1}{N^2}\left[\left\{\sum_{i\in A}(a_i - 1) + \sum_{i\neq}\sum_{j\in A}(a_{ij}-1)\right\} + \left(\sum_{i=1}^{N} a_i\right)\bar{\phi}\right]$$

where

$$a_i = \sum_{s\supset i} b_{si}^2 p(s)$$

$$a_{ij} = \sum_{s\supset i,j} b_{si} b_{sj} p(s) \quad (2.15)$$

and $p(s)$ is the probability of selection of the sample s.

3. An unbiased estimator of $V(\hat{\pi}_{as})$ is given by:

$$\hat{V}(\hat{\pi}_{as}) = \frac{1}{N^2}\left[\sum_{i\in s} c_{si}\hat{y}_i^2 + \sum_{i\neq}\sum_{j\in s} c_{sij}\hat{y}_i\hat{y}_j\right] + \frac{\bar{\phi}}{N}$$

where the constants c_{si} and c_{sij} satisfy the conditions:

$$\sum_{s\supset i} c_{si} p(s) = a_i - 1$$

and $\sum_{s\supset i,j} c_{sij} p(s) = a_{ij} - 1$ for $i \neq j$ (2.16)

Further, using the Threorem 2 and method proposed in Arnab and Shangodoyin (2015), we have the following results:

2.2 Horvitz-Thompson Estimator

The estimator $\hat{\pi}_{as}$ reduces to the Horvitz-Thompson estimator:

$$\hat{\pi}_{ht} = \frac{1}{N}\sum_{i\in s}\frac{\hat{y}_i}{\pi_i} \quad (2.17)$$

for the population proportion π when $b_{si} = 1/\pi_i$

$$V(\hat{\pi}_{ht}) = \frac{1}{N^2}\left\{\sum_{i\in A}\left(\frac{1}{\pi_i} - 1\right) + \sum_{i\neq}\sum_{j\in A}\left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right)\right\} + \frac{1}{N^2}\left(\sum_{i=1}^{N}\frac{1}{\pi_i}\right)\bar{\phi} \quad (2.18)$$

$$\hat{V}(\hat{\pi}_{ht}) = \frac{1}{N^2}\left[\left\{\sum_{i\in s}\frac{1}{\pi_i}\left(\frac{\hat{y}_i}{\pi_i} - 1\right) + \sum_{i\neq}\sum_{j\in A}\frac{1}{\pi_{ij}}\left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right)\hat{y}_i\hat{y}_j\right\}\right] + \frac{\bar{\phi}}{N} \quad (2.19)$$

It can be shown that for a fixed effective size n design an alternative Yates-Grundy (1953) analogue estimator of $V(\hat{\pi}_{ht})$ is given by:

$$\hat{V}_{hg} = \frac{1}{N^2}\sum_{i<}\sum_{j\in s}\left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{\hat{y}_i}{\pi_i} - \frac{\hat{y}_j}{\pi_j}\right)^2 + \frac{\bar{\phi}}{N} \quad (2.20)$$

2.3 SRSWOR Sampling

For the SRSWOR sampling design, $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/\{N(N-1)\}$. In this case $\hat{\pi}_{ht}$ reduces to the estimator:

$$\hat{\pi}_{wor} = \frac{1}{n}\sum_{i\in s}\hat{y}_i \quad (2.21)$$

The variance and unbiased estimator of the variance of $\hat{\pi}_{wor}$ are given as follows:

$$V(\hat{\pi}_{wor}) = (1-f)\frac{N}{n(N-1)}\pi(1-\pi) + \frac{\bar{\phi}}{N} \quad (2.22)$$

$$\hat{V}(\hat{\pi}_{wor}) = (1-f)\frac{s_{\hat{y}}^2}{n} + \frac{\bar{\phi}}{N} \quad (2.23)$$

where $s_{\hat{y}}^2 = \frac{1}{n-1}\sum_{i\in s}(\hat{y}_i - \hat{\pi}_{wor})^2$.

2.4 PPSWR Sampling Scheme

For a probability proportional to size with replacement (PPSWR) sampling scheme, we select a sample of size n using normed size measure $p_i(>0, \sum_{i=1}^{N} p_i = 1)$ attached to the i^{th} unit (i = 1, ..., N).

Let $U_{(r)}$ be the unit selected at the r^{th} draw (r = 1, ..., n) with probability $p_{(r)}$ and $\hat{y}_{(r)}$ be an unbiased estimator of $y_{(r)}$ obtained from the unit $U_{(r)}$ by using the randomized device proposed by Lee *et al.* (2015). In this case the Hansen-Hurwitz estimator for the population proportion is given by

$$\hat{\pi}_{hh} = \frac{1}{Nn}\sum_{r=1}^{n}\frac{\hat{y}_{(r)}}{p_{(r)}} \tag{2.24}$$

The expression of the variance and unbiased estimator of the variance are respectively given by

$$V(\hat{\pi}_{hh}) = \frac{1}{n}\left[\left\{\frac{1}{N^2}\sum_{i\in A}\frac{1}{p_i} - \pi^2\right\} + \bar{\phi}\,\frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{p_i}\right] \tag{2.25}$$

and $$\hat{V}(\hat{\pi}_{hh}) = \frac{1}{N^2 n(n-1)}\sum_{r=1}^{n}\left(\frac{y_{(r)}}{p_{(r)}} - \hat{\bar{y}}\right)^2 \tag{2.26}$$

where $\hat{\bar{y}} = \frac{1}{n}\sum_{r=1}^{n}\frac{\hat{y}_{(r)}}{p_{(r)}}$

2.5 SRSWR Sampling Design

The PPSWR design reduces to the SRSWR design when $p_i = 1/N$ for every $i = 1,..,N$. Hence substituting $p_i = 1/N$ into $\hat{\pi}_{hh}$, we find an unbiased estimator for π under SRSWR as:

$$\hat{\pi}_{wr} = \frac{1}{n}\sum_{r=1}^{n}\hat{y}_{(r)} \tag{2.27}$$

Finally substituting $p_i = 1/N$ in (2.25), we obtain

$$V(\hat{\pi}_{wr}) = \frac{\pi(1-\pi)}{n} + \frac{\bar{\phi}}{n} \text{ and}$$

$$\hat{V}(\hat{\pi}_{wr}) = \frac{s_{\hat{y}}^2(r)}{n}$$

where $s_{\hat{y}}^2(r) = \frac{1}{(n-1)}\sum_{r=1}^{n}(\hat{y}_{(r)} - \hat{\pi}_{wr})^2$

3. COMPARISON OF EFFICIENCIES

For the DR (direct response) surveys the values of y_i are directly obtained by respondent. In this case an unbiased estimator of the population proportion is obtained by replacing $\hat{y}_i$ by its true value y_i in (2.13). The expression of the proportion π comes out as

$$\hat{\pi} = \frac{1}{N}\sum_{i\in s} b_{si} y_i \tag{3.1}$$

The variance $\hat{\pi}$ is given by

$$V(\hat{\pi}) = V\left(\frac{1}{N}\sum_{i\in s} b_{si} y_i\right)$$

$$= \frac{1}{N^2}\left\{\sum_{i\in A}(a_i - 1) + \sum_{i\neq}\sum_{j\in A}(a_{ij} - 1)\right\} \tag{3.2}$$

From Theorem 2 and (3.8), we note that the estimator of the population proportion for the direct response (DR) survey provide much smaller variance than that of the RR response surveys. The amount of loss of efficiency is given by

$$E = \frac{V(\hat{\pi}_{as})}{V(\hat{\pi})} - 1 = \frac{\left(\sum_{i=1}^{N} a_i\right)\bar{\phi}}{\left\{\sum_{i\in A}(a_i - 1) + \sum_{i\neq}\sum_{j\in A}(a_{ij} - 1)\right\}} - 1 \tag{3.3}$$

It is well known that Horvitz-Thopmson estimator based on an inclusion probability proportional to size (πps) sampling design where $\pi_i = np_i$ for i = 1, ..., N always produce lower variance than the Hansen-Hurwitz estimator based on the same normed size measures *pi's.* So, the expression of the first component of (2.18),

$$\frac{1}{N^2}\left\{\sum_{i\in A}\left(\frac{1}{\pi_i} - 1\right) + \sum_{i\neq}\sum_{j\in A}\left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right)\right\}$$

is always less than the first component of (2.25)

$$\frac{1}{n}\left(\frac{1}{N^2}\sum_{i\in A}\frac{1}{p_i} - \pi^2\right).$$

The second components of (2.18) and (2.25) are the same. Hence for a given RR model, the Horvitz-Thompson estimator based on πps sampling design produces smaller variance than the Hansen-Hurwitz estimator based on a PPSWR sampling design.

Since the πps reduces to SRSWOR sampling design when p_i s are equal to $1/N$. Hence SRSWOR sampling scheme always yield lower variance than that of SRSWR sampling procedure under a given RR technique.

For the DR surveys, the variance of the Horvitz-Thomson estimator based on πps sampling design becomes smaller than that of the expansion estimator based on SRSWOR sampling design when the study variable is well related to the auxiliary variable. But in RR response surveys when the study variable y is an indicator variable taking values 0 or 1, y is seldom proportional to the auxiliary variable. Hence, the comparison between the Horvitz-Thompson estimator and the sample mean based on a SRSWOR sampling is not straight forward.

4. CONCLUSION

Method of estimation of population proportion of sensitive characteristics based on RR technique was proposed by Lee *et al*. (2015) under SRSWR sampling. In this paper generalizations of the method of estimation of Lee *et al*. (2015) for the complex survey designs have been proposed. The expressions of unbiased estimators for the population proportion, the variances of and unbiased estimators of variances under various sampling strategies are also proposed. It is further noted that the Lee *et al*. (2015) RR model reduces to Odumande and Singh (2009) and Warner's (1965) model under special cases where each respondent supplies independently two and four RR responses respectively.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for their valuable suggestions which lead improvement of the earlier version of this paper.

REFERENCES

Arnab, R. (2004). Optional randomized response techniques for complex survey designs. *Biom. J.,* **46**, 114-124.

Arnab, R. and Mothupi, T. (2015). Randomized response techniques: A case study of the risky behaviors' of students of a certain University. *Model Assist. Statist. Appl.,* **10**, 421-430.

Arnab, R. and Shangodoyoin, D.K. (2015). Randomized response techniques using maximum likelihood estimator. *Comm. Statist.—Theory Methods,* **44**, 3340-3349.

Franklin, L.A. (1989). A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations. *Comm. Statist.-Theory Methods* **18(2)**, 489-505.

Gjestvang, C.R. and Singh, S. (2006). A new randomized response model. *J. Roy. Statist. Soc*. B, **68**, 523-530.

Greenberg, B.D., Abul-Ela, A.L.A., Simmons, W.R. and Horvitz, D.G., (1969). The unrelated question randomized response model: Theoretical framework. *J. Amer. Statist. Assoc*., **64**, 520-539.

Horvitz, D.G., Shah, B.V., and Simmons, W.R., (1967). The unrelated question randomized response model. *Proceedings of Social Statistics Section, Amer. Statist. Assoc.*, 65-72.

Kim, J.I. (1978). Randomized response technique for surveying human populations. *Ph.D. Dissertation, Temple University,* Philadelphia, USA.

Kuk, A.Y.C. (1990). Asking sensitive question indirectly. *Biometrika*, **77**, 436-438.

Lee, C. (2015). Comparison of Cramer–Rao lower bounds of variances for at least equal protection of respondents

Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika,* **77**, 439-442.

Odumade, O. and Singh, S. (2009). Efficient use of two deck of cards in randomized response sampling. *Comm. Statist.-Theory Methods,* **38**, 439-446.

Singh, S., Singh, R. and Mangat, N.S., (2006). Some alternative strategies to Moors' model in randomized response sampling. *J. Statist. Plann. Inf.*, **83**, 243-255.

Singh, S. and Sedory, S.A. (2011). Cramer-Rao lower bound of variance in randomized response sampling. *Socio. Methods Res.*, **40(3)**, 536-546.

Singh, S. and Sedory, S.A. (2012). A true simulation study of three estimators at equal protection of respondents in randomized response sampling, Statistic *Neerlandica*, **66**, 442–451.

Warner, S.L., (1965). Randomize response: a survey technique for eliminating evasive answer *bias. J. Amer. Statist. Assoc.,* **60**, 63-69.

Yates, F. and Grundy, P.M. (1953). Selection without replacement within strata with probability proportional to size. *J. Roy. Statist. Soc.,* B, **15**, 253-261.