STATISTICS IN TRANSITION new series, March 2019
Vol. 20, No. 1, pp. 67–86, DOI 10.21307/stattrans-2019-004

NONRANDOMIZED RESPONSE MODEL FOR COMPLEX SURVEY DESIGNS

Raghunath Arnab[1], **Dahud Kehinde Shangodoyin**[2], **Antonio Arcos**[3]

ABSTRACT

Warner's randomized response (RR) model is used to collect sensitive information for a broad range of surveys, but it possesses several limitations such as lack of reproducibility, higher costs and it is not feasible for mail questionnaires. To overcome such difficulties, nonrandomized response (NRR) surveys have been proposed. The proposed NRR surveys are limited to simple random sampling with replacement (SRSWR) design. In this paper, NRR procedures are extended to complex survey designs in a unified setup, which is applicable to any sampling design and wider classes of estimators. Existing results for NRR can be derived from the proposed method as special cases.

Key words: complex survey designs, parallel model, randomized response, probability proportional to size, varying probability sampling.

Mathematics Subject Classification: 62D05

1. Introduction

In epidemiological, medical and sociological surveys among others, information is often collected on highly sensitive issues such as induced abortion, HIV/AIDS, drug addiction, domestic violence and cheating in examination, etc. In such situations, direct response (DR) surveys where sensitive questions are asked directly to the respondents, the respondents often provide wrong answers, or refuse to answer because of social stigma and/or fear. Under such circumstances the randomized response (RR) techniques may be used to collect more reliable data, protect respondents' confidentiality and avoid unacceptable rate of nonresponse. The RR technique was introduced by Warner (1965). Warner's technique was later modified by Horvitz et al. (1967), Greenberg et al. (1969), Raghavrao (1978), Franklin (1989), Arnab (1990, 1996), Kuk (1990) and Rueda et al. (2015) to increase co-operations from respondents and improve efficiencies of the proposed estimators. The applications of the RR technique to real life situations were reported by many researchers: Greenberg et al. (1969)

[1] University of Botswana, Botswana and University of KwaZulu-Natal, South Africa. E-mail: arnabr@mopipi.ub.bw. ORCID ID: https://orcid.org/0000-0001-5755-5857.
[2] University of Botswana, Botswana. E-mail: shangodoyink@mopipi.ub.bw. ORCID ID: https://orcid.org/0000-0002-0449-9510.
[3] University of Granada, Spain. Email: arcos@ugr.es.

with regard to illegitimacy of offspring; Abernathy et al. (1970) concerning incidence of induced abortions; Goodstadt and Gruson (1975) concerning drug uses, Folsom et al. (1973) concerning drinking and driving; and Arnab and Mothupi (2015) concerning sexual habits of University students.

In all randomised response models proposed in the literature, respondents have to perform randomized response experiments using devices such as spinners, the drawing of cards and the drawing of random numbers. So, in a survey involving RR methods, the investigators have to describe the methods and supply randomized devices to the respondents, which make the survey more expensive and time consuming rather than the direct response surveys. Tan et al. (2009) pointed out a few further limitations of RR which include (i) lack of reproducibility in the sense that the same respondent may provide different response depending on the outcome of the RR trial, (ii) uneven implementation of RR devices, which make it difficult to convince the respondents that their privacy is protected, (iii) some of the questions are alternative to sensitive questions (e.g. Warner (1965) model) and (iv) unfeasible for mail questionnaire. To overcome some of the aforementioned difficulties, nonrandomized response (NRR) model was proposed by Tian et al. (2007), Yu et al. (2008), Tan et al. (2009), Tian (2014) among others. In the proposed NRR models, independent non-sensitive questions were used to obtain indirect answers on sensitive issues. Obviously, NRR models reduce costs and are feasible for mail questionnaire. Tan et al. (2009) and Tian (2014) reported that the NRR model is more efficient than the RR model for estimating population characteristics. NRR techniques in real life surveys were used by Tang et al. (2014) to investigate homosexual experience among college students; Tian (2014) to investigate sexual behaviour and on plagiarism; and Wu and Tang (2016) to investigate pre-marital sex experience.

All the NRR models available in the literature are limited to simple random sampling with replacement (SRSWR) sampling design only. However, in practice most surveys are complex and multi-character surveys. A sampling design other than simple random sampling is called a complex sampling design. Complex sampling often involves clustering, stratification and unequal probability sampling among others, while in multi-character surveys information of more than one character is collected at a time. Some of the characters are of a confidential nature and others are not. For example, Household Income and Expenditure Survey 2002/03 (HIES 2002/03) conducted by CSO (2004), Botswana, involved a selection of first stage units by inclusion probability proportional to size (IPPS) sampling design, and the second stage units by a systematic sampling procedure. The same survey design was used by Statistics South Africa (2005) for HIES 2005/06 survey, Botswana Aids Impact Surveys (BAIS (2008)) conducted by CSO (2009) to collect data relating to sensitive issues such as sexual behaviour along with non-sensitive items such as socio-economic conditions.

In this paper, we have extended Tian (2014) NRR model called “The parallel model” for estimating population characteristics when the data is collected using complex survey designs. The estimator of the population proportion, its variance and unbiased estimators of variances of the estimators are derived in a unified setup, which is applicable to any sampling design and estimators. The estimators of the population proportions, their variances and unbiased estimators of the variances for the existing NRR models can be obtained from the proposed

method as special cases. It was found that under the SRSWR sampling, both the estimator and variance of the estimator of the population proportion π_y for the Greenberg et al. (1969) and Tian (2014) are the same. However, for the simple random sampling without replacement (SRSWOR) estimators of π_y are the same while the variance of Greenberg et al. (1969) estimator is higher than the Tian (2014) estimator. Two pioneering RR techniques are described below.

1.1. Warner's model

In Warner's (1965) pioneering method, a sample of size n was selected from a population by SRSWR method. Each of the respondents selected in the sample was asked to draw a card at random from a pack of well scaffolded cards consisting of two types of cards with known proportions and identical in appearance. Card type 1, with proportion $P_1(\neq 1/2)$ contains the question "Do you belong to the sensitive group A?" while card type 2 with proportion $1-P_1$ contains the question "Do you belong to group $\overline{A}$?" where A is a sensitive group such as HIV positive and $\overline{A}$ is the complement of group A (HIV negative). The respondent will supply a truthful answer "Yes" or "No" for the question mentioned in the selected card. The experiment is performed in the absence of the interviewer and hence the privacy of the respondent is maintained because the interviewer will not know which of the two questions the respondent has answered (See Arnab, 2017).

1.2. Greenberg et al. model

Greenberg et al. (1969) modified Warner's method by incorporating a sensitive question (character y) along with a non-sensitive question (character x). In this method, a sample of n units is selected by SRSWR method and each of the respondents selected in the sample has to pick a card at random (unobserved by the interviewer) from a pack containing two types of identical cards with known proportions as in Warner's model. The type 1 cards bear the sensitive question "Do you belong to the sensitive group A?" with proportion $P_2(\neq 0)$ while card type 2 (with proportion $1-P_2$) bears a question of unrelated or non-sensitive characteristic B such as "Are you an African?". Here also, the respondent will supply a truthful answer "Yes" or "No" for the question mentioned in the selected card (See Arnab, 2017).

2. Tian's NRR model

Tian (2014) proposed the following NRR model called "The parallel model", where the respondents need not require RR devices to provide responses. In this parallel model, respondents fill the questionnaire themselves unobserved by the interviewer. The questionnaire is a mixture of sensitive and non-sensitive questions. The parallel method is described below.

2.1. Parallel method

Let A denote the group of people who possess a sensitive characteristic y (such as HIV positive) and $\bar{A}$ denotes the people who do not possess the sensitive characteristic y (HIV negative). Further, let x and w be two non-sensitive dichotomous variates, such that y, x and w are mutually independent. For example, $x = 1(0)$ if the respondent's birthday 1 to 15 (16-31) days of a month while $w = 1(0)$ if the respondent is born between July and December (January to June) of a year. Clearly x and w are independent of the HIV infection status y such that $\pi_x = \Pr ob(x = 1) \cong 0.5$ and $1 - p = \Pr ob(w = 1) \cong 0.5$. Here a respondent has to answer truthfully "Yes" or "No" the unrelated non-sensitive question $Q1$ if his/her birthday falls in the first half of the year, i.e. $(w = 0)$ or a sensitive question $Q2$ if his/her birthday falls within the second half of the year, i.e. $(w = 1)$. The respondent should provide the answer "Yes" or "No" without disclosing which question he/she has answered. Hence, the confidentiality of the respondent is maintained.

For example, the questions $Q1$ and $Q2$ are as follows:

$Q1$: Are you a vegetarian?

$Q2$: Are you HIV positive?

2.2. Sampling design and methods of estimation

Tian (2014) used SRSWR method of sampling for the selection of a sample. Let n be the number of respondents selected and n' be the number of respondents who answered "Yes". Here, the probability of obtaining "Yes" answer from a respondent is

$$\theta_t = \Pr ob\{w = 0 \cap x = 1\} + \Pr ob\{w = 1 \cap y = 1\}$$

$$= (1 - p)\pi_x + p\pi_y \tag{2.1}$$

Noting that n' follows binomial distribution, Tian (2014) obtained an unbiased estimator of π_y as

$$\hat{\pi}_{ty} = \frac{\hat{\lambda}_t - \pi_x(1 - p)}{p} \tag{2.2}$$

where $\hat{\lambda}_t = n'/n$ = proportion of "Yes" answers.

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

$$Var\left(\hat{\pi}_{ty}\right) = \frac{\pi_y(1 - \pi_y)}{n} + \frac{(1 - p)g(\pi_x \mid \pi_y, p)}{np^2} \tag{2.3}$$

where $g(\pi_x \mid \pi_y, p) = (p-1)\pi_x^2 + (1-2\pi_y p)\pi_x + \pi_y p$.

For $\pi_x = 1/2$, $g(\pi_x \mid \pi_y, p) = \frac{(p-1)}{4} + \frac{1}{2}$.

3. Parallel models for Complex survey designs

In this section we propose a methodology of estimating population proportion of a sensitive characteristic of a complex multi-character survey design where the data of the sensitive characteristic is collected by using the parallel method.

Consider a finite population $U = \{1,..,i,..,N\}$ of N units from which a sample s of size n units is selected with probability $p(s)$ using a sampling design $\mathcal{P}$. Let $\pi_i = \sum_{s \supset i} p(s)$ and $\pi_{ij} = \sum_{s \supset i,j} p(s)$ be the inclusion probabilities for the ith, and ith and jth $(i \neq j)$ units of the population. From each of the units in the sample s, the information on the sensitive characteristic y is obtained by using a parallel method. Let $B(\bar{B})$ be the group of respondents whose birthday falls between first half of a month i.e. 01 and 15 days (after 15th day of a month) of a month; $W(\bar{W})$ be the group of respondents born in the second half of the year, i.e. between July and December (January and June) and $A(\bar{A})$ be the group of respondents who do (do not) possess the sensitive characteristic y. Define

$$x_i = \begin{cases} 1 & \text{if the ith unit } \in B \\ 0 & \text{if the ith unit } \in \bar{B} \end{cases}, \quad w_i = \begin{cases} 1 & \text{if the unit i} \in W \\ 0 & \text{if the unit i} \in \bar{W} \end{cases}, \quad y_i = \begin{cases} 1 & \text{if the ith unit } \in A \\ 0 & \text{if the ith unit } \in \bar{A} \end{cases}$$

and $z_i = \begin{cases} 1 & \text{if the ith unit answers "}Yes\text{"} \\ 0 & \text{if the ith unit answers "}No\text{"} \end{cases}$

Under the parallel model, if a respondent belongs to the group $\bar{W}$, he/she answers the question $Q1$. In this case if the respondent's birthday falls between 01 and 15th day of a month, the respondent provides "Yes" answers with probability one. Otherwise if the respondent is born after 15th day of a month, the respondent supplies "No" answers with probability 1. Hence,

$$z_i = x_i \qquad \text{if } i \in \bar{W} \tag{3.1}$$

Similarly, if a respondent belongs to the group W, then the respondent answers the question $Q2$. In this case the respondent answers "Yes" with probability one if he/she belongs to the sensitive group A (HIV positive). On the other hand, if the respondent belongs to the complementary group $\bar{A}$ (HIV negative), then he/she supplies response answer "No" with probability one. Hence, in this cas

$$z_i = y_i \qquad i \in W \tag{3.2}$$

Equations (3.1) and (3.2) yield

$$z_i = w_i y_i + (1-w_i)x_i \tag{3.3}$$

and

$$\begin{aligned} Z &= \sum_{i\in U} z_i \\ &= \sum_{i\in \bar{W}} x_i + \sum_{i\in W} y_i \\ &= N_{\bar{W}B} + N_{WA} \end{aligned}$$

where $N_{\bar{W}B}\,(N_{WA})$ is the number of individuals of the population belonging to the groups $\bar{W}\cap B\ (W\cap A)$.

Assuming that the membership of an individual to the group $A(\bar{A})$, $W(\bar{W})$, and $B(\bar{B})$ is mutually independent, we make the following assumptions:

$$\pi_{WB} = p\pi_x;\ \pi_{\bar{W}B} = (1-p)\pi_x;\ \pi_{WA} = p\pi_y;\ N_{\bar{W}A} = (1-p)\pi_y \tag{3.4}$$

where

$$\pi_{WB} = \frac{N_{WB}}{N},\ \pi_{\bar{W}B} = \frac{N_{\bar{W}B}}{N},\ \pi_{WA} = \frac{N_{WA}}{N}, \pi_{\bar{W}A} = \frac{N_{\bar{W}A}}{N}\ , \pi_x = \frac{N_B}{N},\ \pi_y = \frac{N_A}{N}$$ and

$p = \frac{N_W}{N}$; N_F and N_{FG} denote the number of individuals belonging to the group F and $F\cap G$; $F, G = A, \bar{A}, B, \bar{B}, W, \bar{W}$.

Under the assumption (3.4), we have

$$\bar{Z} = Z/N = p\pi_y + (1-p)\pi_x \tag{3.5}$$

Here, we propose a linear homogeneous unbiased estimator of $\bar{Z}$ as

$$\hat{\bar{Z}} = \frac{1}{N}\sum_{i\in s} b_{si} z_i \tag{3.6}$$

Where $\sum_{i\in s}$ denotes the sum over distinct units in s, b_{si}'s are known constants satisfying the unbiasedness condition

$$\sum_{s\supset i} b_{si}\, p(s) = 1\,. \tag{3.7}$$

The variance of $\hat{\bar{Z}}$ is

$$V\left(\hat{\bar{Z}}\right) = V\left(\sum_{i\in s} b_{si} z_i\right)/N^2$$

$$= \left[E\left(\sum_{i \in s} b_{si} z_i \right)^2 - Z^2 \right] / N^2$$

$$= \frac{1}{N^2} E\left[\sum_s \left(\sum_{i \in s} b_{si}^2 z_i^2 + \sum_{i \neq} \sum_{j \in s} b_{si} b_{sj} z_i z_j \right) p(s) \right] - \bar{Z}^2$$

(where $p(s)$ is the probability of the selection of the sample s)

$$= \frac{1}{N^2} \left[\sum_{i \in U} \alpha_i z_i^2 + \sum_{i \neq} \sum_{j \in U} \alpha_{ij} z_i z_j \right] - \bar{Z}^2 \tag{3.8}$$

where

$$\alpha_i = \sum_{s \supset i} b_{si}^2 p(s) \text{ and } \alpha_{ij} = \sum_{s \supset i} b_{si} b_{sj} p(s).$$

The expression (3.8) yields

$$V\left(\hat{\bar{Z}}\right) = \sum_{i \in U} \alpha_i^* z_i^2 + \sum_{i \neq} \sum_{j \in U} \alpha_{ij}^* z_i z_j \tag{3.9}$$

where $\alpha_i^* = \frac{1}{N^2}(\alpha_i - 1)$ and $\alpha_{ij}^* = \frac{1}{N^2}(\alpha_{ij} - 1)$

From expression (3.9), we set an unbiased estimator of $V\left(\hat{\bar{Z}}\right)$ as

$$\hat{V}\left(\hat{\bar{Z}}\right) = \sum_{i \in s} c_{si} z_i^2 + \sum_{i \neq} \sum_{j \in s} c_{sj} z_i z_j \tag{3.10}$$

where c_{si} and c_{sij} are suitably chosen constants satisfying unbiasedness conditions

$$\sum_{s \supset i} c_{si} p(s) = \alpha_i^* \text{ and } \sum_{s \supset i} c_{sij} p(s) = \alpha_{ij}^* \tag{3.11}$$

We may choose c_{si} and c_{sij} in various ways. One of the obvious choices is $c_{si} = \alpha_i^* / \pi_i$ and $c_{sij} = \alpha_{ij}^* / \pi_{ij}$.

Substituting $z_i = w_i y_i + \bar{w}_i x_i$, $\bar{w}_i = 1 - w_i$ in equation (3.9) and noting that $w_i, \bar{w}_i, y_i$ and x_i are indicator variables, we have the following simplifications:

$$V\left(\hat{\bar{Z}}\right) = \sum_{i \in U} \alpha_i^* \left(w_i y_i + \bar{w}_i x_i \right) + \sum_{i \neq} \sum_{j \in U} \alpha_{ij}^* \left(w_i y_i + \bar{w}_i x_i \right) \left(w_j y_j + \bar{w}_j x_j \right)$$

$$= \left\{ \sum_{i \in W} \alpha_i^* y_i + \sum_{i \neq} \sum_{j \in W} \alpha_{ij}^* y_i y_j \right\} + \left\{ \sum_{i \in \bar{W}} \alpha_i^* x_i + \sum_{i \neq} \sum_{j \in \bar{W}} \alpha_{ij}^* x_i x_j \right\}$$

$$+\left\{\sum_{i\in W} y_i \sum_{j(\neq i)\in\bar{W}} \alpha_{ij}^* x_j + \sum_{j\in W} x_j \sum_{i(\neq j)\in\bar{W}} \alpha_{ij}^* y_j\right\}$$

$$=\left\{\sum_{i\in W\cap A} \alpha_i^* + \sum_{i\neq} \sum_{j\in W\cap A} \alpha_{ij}^*\right\}+\left\{\sum_{i\in \bar{W}\cap B} \alpha_i^* + \sum_{i\neq} \sum_{j\in \bar{W}\cap B} \alpha_{ij}^*\right\}$$

$$+\left\{\sum_{i\in W\cap A} \sum_{j(\neq i)\in \bar{W}\cap B} \alpha_{ij}^* + \sum_{j\in W\cap B} \sum_{i(\neq j)\in \bar{W}\cap A} \alpha_{ij}^*\right\}.$$

The above results lead to the following theorem.

Theorem 3.1.

Under assumptions (3.4),

(i) $\hat{\pi}_y = \dfrac{\hat{\bar{Z}} - (1-p)\pi_x}{p}$ is an unbiased estimator of π_y when the population proportion π_x is assumed to be known.

(ii) The variance of $\hat{\pi}_y$ is

$$V(\hat{\pi}_y) = \frac{1}{p^2}\left[\left\{\sum_{i\in W\cap A} \alpha_i^* + \sum_{i\neq} \sum_{j\in W\cap A} \alpha_{ij}^*\right\}+\left\{\sum_{i\in \bar{W}\cap B} \alpha_i^* + \sum_{i\neq} \sum_{j\in \bar{W}\cap B} \alpha_{ij}^*\right\}+\left\{\sum_{i\in W\cap A} \sum_{j(\neq i)\in \bar{W}\cap B} \alpha_{ij}^* + \sum_{j\in W\cap B} \sum_{i(\neq j)\in \bar{W}\cap A} \alpha_{ij}^*\right\}\right]$$

(iii) An unbiased estimator of $V(\hat{\pi}_y)$ is

$$\hat{V}(\hat{\pi}_y) = \frac{1}{p^2}\left[\sum_{i\in s} c_{si} z_i + \sum_{i\neq} \sum_{j\in s} c_{sij} z_i z_j\right].$$

We now present expressions of $\hat{\pi}_y$, $V(\hat{\pi}_y)$ and $\hat{V}(\hat{\pi}_y)$ for various sampling strategies as special cases of Theorem 3.1.

3.1. Arbitrary sampling design with Horvitz-Thompson estimator

For $b_{si} = 1/\pi_i$, we have $\alpha_i^* = \dfrac{1}{N^2}\left(\dfrac{1}{\pi_i} - 1\right)$, $\alpha_{ij}^* = \dfrac{1}{N^2}\left(\dfrac{\pi_{ij}}{\pi_i \pi_j} - 1\right)$ and the expression of the Horvitz- Thompson estimator for π_y as

$$\hat{\pi}_{hte} = \frac{\sum_{i\in s} \dfrac{z_i}{N\pi_i} - (1-p)\pi_x}{p} \tag{3.12}$$

The expression of the variance of and its unbiased estimators are obtained from the Theorem 3.1 as follows:

$$V\left(\hat{\pi}_{hte}\right)=\frac{1}{N^2p^2}\left[\begin{array}{l}\left\{\sum_{i\in W\cap A}\left(\frac{1}{\pi_i}-1\right)+\sum_{i\neq}\sum_{j\in W\cap A}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)\right\}+\left\{\sum_{i\in \bar{W}\cap B}\left(\frac{1}{\pi_i}-1\right)+\sum_{i\neq}\sum_{j\in \bar{W}\cap B}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)\right\}\\+\left\{\sum_{i\in W\cap A}\sum_{j(\neq i)\in \bar{W}\cap B}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)+\sum_{j\in W\cap B}\sum_{i(\neq j)\in \bar{W}\cap A}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)\right\}\end{array}\right] \tag{3.13}$$

and

$$\hat{V}\left(\hat{\pi}_{hte}\right)=\frac{1}{N^2p^2}\left[\sum_{i\in s}\frac{1}{\pi_i}\left(\frac{1}{\pi_i}-1\right)z_i+\sum_{i\neq j}\sum_{\in s}\frac{1}{\pi_{ij}}\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)z_iz_j\right] \tag{3.14}$$

3.2. Simple random sampling without replacement (SRSWOR)

For SRSWOR, $\pi_i=n/N$, $\pi_{ij}=n(n-1)/\{N(N-1)\}$, $\alpha=\left(\frac{1}{\pi_i}-1\right)=\frac{N-n}{n}$ and $\beta=\left(\frac{\pi_{ij}}{\pi_i\pi_j}-1\right)=-\frac{N-n}{n(N-1)}$. In this case, the expressions $\hat{\pi}_{hte}$, $V\left(\hat{\pi}_{hte}\right)$ and $\hat{V}\left(\hat{\pi}_{hte}\right)$ come out as follows:

$$\hat{\pi}_{wor}=\frac{\bar{z}_s-(1-p)\pi_x}{p} \tag{3.15}$$

where $\bar{z}_s=\sum_{i\in s}z_i/n=\lambda_s$ = proportion of "Yes" answers in the sample s.

$$V(\hat{\pi}_{wor})=\frac{1}{N^2p^2}\left[\left\{\alpha\sum_{i\in W\cap A}+\beta\sum_{i\neq}\sum_{j\in W\cap A}\right\}+\left\{\alpha\sum_{i\in \bar{W}\cap B}+\beta\sum_{i\neq}\sum_{j\in \bar{W}\cap B}\right\}+\beta\left\{\sum_{i\in W\cap A}\sum_{j(\neq i)\in \bar{W}\cap B}+\sum_{j\in W\cap B}\sum_{i(\neq j)\in \bar{W}\cap A}\right\}\right]$$

$$=\frac{1}{N^2p^2}\left[\left\{N_{WA}\alpha+N_{WA}\left(N_{WA}-1\right)\beta\right\}+\left\{\alpha N_{\bar{W}B}+\beta N_{\bar{W}B}\left(N_{\bar{W}B}-1\right)\right\}+\beta\left\{N_{WA}N_{\bar{W}B}+N_{WB}N_{\bar{W}A}\right\}\right]$$

$$=\frac{N-n}{Nnp^2}\left[\left\{p\pi_y-p\pi_y\left(Np\pi_y-1\right)\frac{1}{(N-1)}\right\}+\left\{(1-p)\pi_x-(1-p)\pi_x\left(N(1-p)\pi_x-1\right)\frac{1}{(N-1)}\right\}\right]$$

$$-2\frac{N-n}{n(N-1)}\frac{(1-p)}{p}\pi_x\pi_y$$

$$=\frac{N(1-f)}{n(N-1)p^2}\left[\left\{p\pi_y(1-p\pi_y)\right\}+\left\{(1-p)\pi_x\left(1-(1-p)\pi_x\right)\right\}-2p(1-p)\pi_x\pi_y\right]$$

$$\left(\text{where } f=n/N\right)$$

$$= \frac{N(1-f)}{(N-1)}\left[\frac{\pi_y(1-\pi_y)}{n} + \frac{(1-p)}{np^2}\left\{(p-1)\pi_x^2 + \left(1-2p\pi_y\right)\pi_x + p\pi_y\right\}\right] \tag{3.16}$$

From the expression (3.15), we set an unbiased estimator of $V(\hat{\pi}_{wor})$ as

$$\hat{V}(\hat{\pi}_{wor}) = \frac{1}{p^2}\frac{(1-f)}{n}\frac{1}{n-1}\sum_{i\in s}(z_i - \bar{z}_s)^2$$

$$= \frac{1}{p^2}\frac{(1-f)}{n-1}\lambda_s(1-\lambda_s) \tag{3.17}$$

3.3. Probability proportional to size with replacement (PPSWR)

Let a sample of size n be selected from the population by PPSWR method using normed size measure $p_i\left(>0, \sum p_i = 1\right)$ attached to the ith unit. Let $z(r)$ be the response obtained from the respondent selected at the r th $(r = 1,..,n)$ draw with probability $p(r)$ so that $z(r) = z_j$ and $p(r) = p_j$ if rth draw produces the jth unit. The Hansen-Hurwitz estimator of the population proportion π_y is given by

$$\hat{\pi}_{hh} = \frac{\frac{1}{N}\left(\frac{1}{n}\sum_{r=1}^{n}\frac{z(r)}{p(r)}\right) - (1-p)\pi_x}{p} \tag{3.18}$$

Noting that $E\left\{\frac{z(r)}{p(r)}\right\} = \sum_{i=1}^{N} z_i = \sum_{i=1}^{N}\left\{w_i y_i + \bar{w}_i x_i\right\} = N\left\{p\pi_y + (1-p)\pi_x\right\}$, we find that $\hat{\pi}_{hh}$ is an unbiased estimator of π_y.

The variance of $\hat{\pi}_{hh}$ is

$$V(\hat{\pi}_{hh}) = \frac{1}{N^2p^2}V\left(\frac{1}{n}\sum_{r=1}^{n}\frac{z(r)}{p(r)}\right)$$

$$= \frac{1}{N^2p^2n}\left(\sum_{i=1}^{N}\frac{z_i^2}{p_i} - Z^2\right)$$

$$= \frac{1}{N^2p^2n}\left[\sum_{i=1}^{N}\frac{\left\{w_i y_i + \bar{w}_i x_i\right\}}{p_i} - N^2\left\{p\pi_y + (1-p)\pi_x\right\}^2\right] \tag{3.19}$$

Further noting that $\frac{z(r)}{p(r)}$ are independently distributed random variables, we find an unbiased estimator of $\hat{\pi}_{hh}$ as

$$\hat{V}(\hat{\pi}_{hh}) = \frac{1}{N^2p^2n(n-1)}\sum_{r=1}^{n}\left\{\frac{z(r)}{p(r)} - \frac{1}{n}\sum_{r=1}^{n}\frac{z(r)}{p(r)}\right\}^2 \tag{3.20}$$

3.4. Simple random sampling with replacement (SRSWR)

The PPSWR sampling scheme reduces to SRSWR sampling scheme if $p_i = 1/N$ for $i = 1,...,N$. Substituting $p_i = 1/N$ in the expressions (3.18), we find an unbiased estimator of π_y for SRSWR sampling method as

$$\hat{\pi}_{swr} = \frac{\lambda_s - (1-p)\pi_x}{p} \tag{3.21}$$

The expression of the variance of $\hat{\pi}_{swr}$ and its unbiased estimator come out as follows:

$$\begin{aligned} V(\hat{\pi}_{swr}) &= \frac{1}{p^2 n}\left[\frac{1}{N}\sum_{i=1}^{N}\{w_i y_i + \bar{w}_i x_i\} - \{p\pi_y + (1-p)\pi_x\}^2\right] \\ &= \frac{1}{p^2 n}\left[\{p\pi_y + (1-p)\pi_x\} - \{p\pi_y + (1-p)\pi_x\}^2\right] \\ &= \frac{\pi_y(1-\pi_y)}{n} + \frac{(1-p)}{np^2}\left[(p-1)\pi_x^2 + (1-2p\pi_y)\pi_x + p\pi_y\right] \end{aligned} \tag{3.22}$$

and

$$\hat{V}(\hat{\pi}_{swr}) = \frac{\lambda_s(1-\lambda_s)}{p^2(n-1)} \tag{3.23}$$

Expressions (3.21), (3.22) and (3.23) are the same as those obtained by Tian (2014).

3.5. Stratified multi-stage sampling design

Consider a population comprising of H strata. The $h(=1,..,H)$th stratum consists of M_h first-stage units (fsus) and the ith fsu of the hth stratum consists of $M_{hi}\,(i = 1,..,M_h)$ second-stage units (ssus). The total number of ssus in the population is $\sum_{h=1}^{H}\sum_{i=1}^{Mh} M_{hi} = M$. From the h th stratum, a sample s_h of size n_h fsus is selected by using a suitable sampling scheme with $\pi_{i|h}$ and $\pi_{ij|h}$ as inclusion probabilities for the i th, and i th and $j\,(j \neq i)$th fsus. If the ith fsu is selected in the sample s_h, a sub-sample s_{hi} of size n_{hi} ssus is selected from it by using a suitable sampling scheme with inclusion probabilities $\pi_{k|hi}$ and $\pi_{kl|hi}$ for the k th, and k and $l(l \neq k)$th ssus. We denote the j th ssu of the i th fsu of the h th stratum as hij th unit. We define the following notations similar to the Section 3.

$$x_{hij} = \begin{cases} 1 & \text{if } hij\text{th unit} \in B \\ 0 & \text{if } hij\text{th unit} \in \bar{B} \end{cases}, \quad w_{hij} = \begin{cases} 1 & \text{if } hij\text{th unit} \in W \\ 0 & \text{if } hij\text{th unit} \in \bar{W} \end{cases}, \quad y_{hij} = \begin{cases} 1 & \text{if } hij\text{th unit} \in A \\ 0 & \text{if } hij\text{th unit} \in \bar{A} \end{cases},$$

$$z_{hij} = \begin{cases} 1 & \text{if } hij\text{th unit answers "}Yes\text{"} \\ 0 & \text{if } hij\text{th unit answers "}No\text{"} \end{cases}.$$

Now, writing $z_{hij} = w_{hij}y_{hij} + (1- w_{hij})x_{hij}$ and using the assumption similar to (3.4), we find that

$$Z = \sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{J=1}^{N_{hi}} z_{ijk} = M[p\pi_y + (1-p)\pi_x] \quad (3.24)$$

Further, noting that $\hat{Z}_{hte} = \sum_{h=1}^{H}\sum_{i\epsilon s_h}\frac{\hat{Z}_{i|h}}{\pi_{i|h}}$ with $\hat{Z}_{i|h} = \sum_{j\epsilon s_{hi}}\frac{z_{hij}}{\pi_{j|hi}}$ is an unbiased estimator of Z, we get the following theorem.

Theorem 3.2.

(i) $\hat{\pi}_y = \frac{1}{p}\left[\frac{\hat{Z}_{hte}}{M} - (1-p)\pi_x\right]$ is an unbiased estimator of π_y.

(ii) The variance of $\hat{\pi}_y$ is

$$V(\hat{\pi}_y) = \frac{1}{p^2M^2}\sum_{h=1}^{H}\left[\sum_{i\neq}^{M_h}\sum_{j=1}^{M_h}(\pi_{i|h}\pi_{j|h} - \pi_{ij|h})\left(\frac{Z_{i|h}}{\pi_{i|h}} - \frac{Z_{j|h}}{\pi_{j|h}}\right)^2 + \sum_{i=1}^{M_h}\frac{\sigma_{i|h}^2}{\pi_{i|h}}\right]$$

where

$Z_{i|h} = \sum_{j=1}^{M_{hi}} z_{hij}$ and $\sigma_{i|h}^2 = V(Z_{i|h}) = \sum_{k\neq}^{M_{hi}}\sum_{l=1}^{M_{hi}}(\pi_{k|hi}\pi_{l|hi} - \pi_{kl|hi})\left(\frac{Z_{hik}}{\pi_{k|hi}} - \frac{Z_{hil}}{\pi_{l|hi}}\right)^2$

(iii) An unbiased estimator of $V(\hat{\pi}_y)$ is

$$\hat{V}(\hat{\pi}_y) = \frac{1}{p^2M^2}\sum_{h=1}^{H}\left[\sum_{i\neq}\sum_{j\in s_h}\left(\frac{\pi_{i|h}\pi_{j|h} - \pi_{ij|h}}{\pi_{ij|h}}\right)\left(\frac{\hat{Z}_{i|h}}{\pi_{i|h}} - \frac{\hat{Z}_{j|h}}{\pi_{j|h}}\right)^2 + \sum_{i\in s_h}^{M_h}\frac{\hat{\sigma}_{i|h}^2}{\pi_{i|h}}\right]$$

where

$$\hat{\sigma}_{i|h}^2 = \sum_{k\neq}\sum_{l\in s_{hi}}\frac{(\pi_{k|hi}\pi_{l|hi} - \pi_{kl|hi})}{\pi_{kl|hi}}\left(\frac{z_{hik}}{\pi_{k|hi}} - \frac{z_{hil}}{\pi_{l|hi}}\right)^2$$

is an unbiased estimator of $\sigma_{i|h}^2$.

Proof:

(i) $E(\hat{\pi}_y) = \frac{1}{p}\left[\frac{E(\hat{Z}_{hte})}{M} - (1-p)\pi_x\right]$

$= \frac{1}{p}\left[\frac{Z}{M} - (1-p)\pi_x\right]$

$= \pi_y$

(ii) $V(\hat{\pi}_y) = \frac{1}{M^2p^2}\sum_{h=1}^{H} V(\hat{Z}_h)$

$$= \frac{1}{M^2p^2}\sum_{h=1}^{H}\left[V\{E(\hat{Z}_h|s_h)\} + E\{V(\hat{Z}_h|s_h)\}\right]$$

$$= \frac{1}{M^2p^2}\sum_{h=1}^{H}\left[V\left\{\sum_{i\in s_h}\frac{Z_{i|h}}{\pi_{i|h}}\right\} + E\left\{\sum_{i\in s_h}\frac{\sigma_{i|h}^2}{\pi_{i|h}^2}\right\}\right]$$

$$= \frac{1}{M^2p^2}\sum_{h=1}^{H}\left[\sum_{i\neq}^{M_h}\sum_{j=1}^{M_h}(\pi_{i|h}\pi_{j|h} - \pi_{ij|h})\left(\frac{Z_{i|h}}{\pi_{i|h}} - \frac{Z_{j|h}}{\pi_{j|h}}\right)^2 + \sum_{i=1}^{M_h}\frac{\sigma_{i|h}^2}{\pi_{i|h}}\right]$$

(iii) $E[\hat{V}(\hat{\pi}_y)]$

$$= \frac{1}{M^2p^2}\sum_{h=1}^{H} E\left[\sum_{i\neq}\sum_{j\in s_h}\left(\frac{\pi_{i|h}\pi_{j|h} - \pi_{ij|h}}{\pi_{ij|h}}\right)E\left\{\left(\frac{\hat{Z}_{i|h}}{\pi_{i|h}} - \frac{\hat{Z}_{j|h}}{\pi_{j|h}}\right)^2|s_h\right\} + E\left(\sum_{i\in s_h}\frac{\hat{\sigma}_{i|h}^2}{\pi_{i|h}}|s_h\right)\right]$$

$$= \frac{1}{M^2p^2}\sum_{h=1}^{H} E\left[\sum_{i\neq}\sum_{j\in s_h}\left(\frac{\pi_{i|h}\pi_{j|h} - \pi_{ij|h}}{\pi_{ij|h}}\right)\left\{\left(\frac{Z_{i|h}}{\pi_{i|h}} - \frac{Z_{j|h}}{\pi_{j|h}}\right)^2 + \frac{\sigma_{i|h}^2}{\pi_{i|h}^2} + \frac{\sigma_{j|h}^2}{\pi_{j|h}^2}\right\} + \sum_{i\in s_h}\frac{\sigma_{i|h}^2}{\pi_{i|h}}\right]$$

$$= \frac{1}{M^2p^2}\sum_{h=1}^{H}\left[\sum_{i\neq}^{M_h}\sum_{j=1}^{M_h}(\pi_{i|h}\pi_{j|h} - \pi_{ij|h})\left\{\left(\frac{Z_{i|h}}{\pi_{i|h}} - \frac{Z_{j|h}}{\pi_{j|h}}\right)^2 + \frac{\sigma_{i|h}^2}{\pi_{i|h}^2} + \frac{\sigma_{j|h}^2}{\pi_{j|h}^2}\right\} + \sum_{i=1}^{M_h}\sigma_{i|h}^2\right]$$

Now, noting that $\sum_{i=1}^{M_h}\pi_{i|h} = n_h$ and $\sum_{j(\neq i)=1}^{M_h}\pi_{ij|h} = (n_h - 1)\pi_{i|h}$, we find $E[\hat{V}(\hat{\pi}_y)] = V(\hat{\pi}_y)$.

4. Comparison with Greenberg RR model

Consider the Greenberg et al. (1969) model described in Section 1.2 with $P_2 = p$. Let $y_i = 1(0)$ if the ith unit does (does not) belong to the sensitive group A, $x_i = 1(0)$ if the ith unit possesses (does not possess) the non-sensitive characteristic B and $z_i = 1(0)$ if the ith respondent answers "Yes" ("No"). Denoting

$E_R(V_R)$ as expectation (variance) with respect to the RR model and noting x_i and y_i are indicator variables, one finds that

$$E_R(z_i) = py_i + (1-p)x_i = E_R(z_i^2) \tag{4.1}$$

$$V_R(z_i) = py_i + (1-p)x_i - \{py_i + (1-p)x_i\}^2$$
$$= p(1-p)(x_i + y_i - 2x_iy_i) \tag{4.2}$$

Let a sample s of size n be selected from the population using SRSWR method, $\lambda_s = \frac{1}{n}\sum_{i\in s} z_i$ be the proportion of "Yes" answers in the population and $\sum_{i\in s}$ denote the sum over the units in s with repetition. In this case we have the following theorem:

Theorem 4.1.

Under SRSWR sampling

(i) $\hat{\pi}_G = \frac{1}{p}[\lambda_s - (1-p)\pi_x]$ is an unbiased estimator of π_y when π_x is known.

(ii) The variance of $\hat{\pi}_G$ is

$$V(\hat{\pi}_G) = \frac{\pi_y(1-\pi_y)}{n} + \frac{1-p}{p^2n}\left[(p-1)\pi_x^2 + (1-2p\pi_y)\pi_x + p\pi_y\right]$$

(iii) An unbiased estimator of $V(\hat{\pi}_G)$ is

$$\hat{V}(\hat{\pi}_G) = \frac{1}{p^2n}\left[\frac{1}{n-1}\sum_{i\in s}(z_i - \lambda_s)^2\right] = \frac{\lambda_s(1-\lambda_s)}{(n-1)p^2}$$

Proof:

(i) $E(\hat{\pi}_G) = \frac{1}{p}[E(\bar{z}) - (1-p)\pi_x]$

$$= \frac{1}{p}\left[E_p\left\{\frac{1}{n}\sum_{i\in s} E_R(z_i)\right\} - (1-p)\pi_x\right]$$

$$= \frac{1}{p}\left[\frac{1}{N}\sum_{i\in U}\{py_i + (1-p)x_i\} - (1-p)\pi_x\right]$$

$$= \pi_y$$

(ii) $V(\hat{\pi}_G) = V_p[E_R(\hat{\pi}_G)] + E_p[V_R(\hat{\pi}_G)]$

$$= V_p\left[\frac{1}{np}\sum_{i\in s} E_R(z_i) - \frac{(1-p)}{p}\pi_x\right] + E_p\left[\frac{1}{(np)^2}\sum_{i\in s} V_R(z_i)\right]$$

$$= V_p\left[\frac{1}{np}\sum_{i\in s}\{py_i + (1-p)x_i\}\right] + E_p\left[\frac{1-p}{n^2p}\sum_{i\in s}(x_i + y_i - 2x_iy_i)\right]$$

$$= \frac{1}{np^2}\left[\frac{1}{N}\sum_{i\in U}\{py_i + (1-p)x_i\}^2 - \{p\pi_y + (1-p)\pi_x\}^2\right] + \frac{1-p}{npN}\sum_{i\in U}(x_i + y_i - 2x_iy_i)$$

$$= \frac{1}{np^2}\left[p^2\pi_y(1-\pi_y) + (1-p)^2\pi_x(1-\pi_x) + 2p(1-p)(\pi_{xy} - \pi_x\pi_y)\right] + \frac{1-p}{np}(\pi_x + \pi_y - 2\pi_{xy})$$

Noting that $\pi_{xy} = \pi_x\pi_y$, as x and y are independent, we obtain

$$V(\hat{\pi}_G) = \frac{\pi_y(1-\pi_y)}{n} + \frac{1-p}{p^2n}\left[(p-1)\pi_x^2 + (1-2p\pi_y)\pi_x + p\pi_y\right]$$

(iii) Further, z_i's, $i = 1,2,..,n$ are independent and identically distributed random variables, one finds that $E[\hat{V}(\hat{\pi}_G)] = E[V(\hat{\pi}_G)]$.

Here, we note that for the SRSWR sampling, the expressions $\hat{\pi}_G$ and $\hat{V}(\hat{\pi}_G)$ of the Greenberg et al. (1969) model are respectively the same as the expressions $\hat{\pi}_{swr}$ (Eq. 3.21) and $V(\hat{\pi}_{swr})$ (Eq. 3.22) in the Parallel model proposed by Tian (2014).

Consider the situation where a sample s of size n is selected by the SRSWOR method and from each of the selected respondents randomized responses were obtained by using Greenberg et al. (1969) RR technique. Let $\lambda_s = \bar{z}_s = \sum_{i\in s} z_i/n$ denote the proportion of "Yes" answers in the sample. In this case we have the following results:

Theorem 4.2.

Under SRSWOR sampling,

(i) $\hat{\pi}_G^* = \frac{1}{p}[\lambda_s - (1-p)\pi_x]$ is an unbiased estimator of π_y.

(ii) The variance of $\hat{\pi}_G^*$ is

$$V(\hat{\pi}_G^*) = \frac{N-n}{(N-1)n}\left[\pi_y(1-\pi_y) + \frac{1-p}{p^2}\pi_x(1-x)\right] + \frac{1-p}{np}(\pi_x + \pi_y - 2\pi_x\pi_y)$$

(iii) An unbiased estimator of $V(\hat{\pi}_G^*)$ is

$$\hat{V}(\hat{\pi}_G^*) = \frac{N-n}{p^2Nn}\frac{1}{n-1}\sum_{i\in s}(z_i-\bar{z})^2 + \frac{1-p}{p}(\hat{\pi}_G + \pi_x - 2\pi_x\hat{\pi}_G)$$

$$= \frac{N-n}{p^2N}\frac{\lambda_s(1-\lambda_s)}{(n-1)} + \frac{1-p}{p}(\hat{\pi}_G + \pi_x - 2\pi_x\hat{\pi}_G)$$

Proof:

(i) $E(\hat{\pi}_G^*) = \frac{1}{p}[E(\lambda_s) - (1-p)\,\pi_x\,]$

$$= \frac{1}{p}\left[E_p\left\{\frac{1}{n}\sum_{i\in s}E_R(z_i)\right\} - (1-p)\pi_x\right]$$

$$= \frac{1}{p}\left[\frac{1}{N}\sum_{i\in U}\{py_i + (1-p)x_i\} - (1-p)\pi_x\right]$$

$$= \pi_y$$

(ii) $V(\hat{\pi}_G^*) = V_p[E_R(\hat{\pi}_G^*)] + E_p[V_R(\hat{\pi}_G^*)]$

$$= V_p\left[\frac{1}{np}\sum_{i\in s}E_R(z_i) - \frac{(1-p)}{p}\pi_x\right] + E_p\left[\frac{1}{(np)^2}\sum_{i\in s}V_R(z_i)\right]$$

$$= \frac{N-n}{np^2}\left[\frac{1}{N}\sum_{i\in U}\{py_i + (1-p)x_i\}^2 - \left\{p\pi_y + (1-p)\pi_x\right\}^2\right]$$

$$+\frac{1-p}{npN}\sum_{i\in U}(x_i + y_i - 2x_iy_i)$$

$$= \frac{N-n}{np^2}\left[p^2\pi_y(1-\pi_y) + (1-p)^2\pi_x(1-\pi_x) + 2p(1-p)(\pi_{xy} - \pi_x\pi_y)\right]$$

$$+\frac{1-p}{np}(\pi_x + \pi_y - 2\pi_{xy})$$

Now, noting that, $\pi_{xy} = \pi_x\pi_y$ we find that

$$V(\hat{\pi}_G^*) = \frac{N-n}{(N-1)n}\left[\pi_y(1-\pi_y) + \frac{1-p}{p^2}\pi_x(1-\pi_x)\right] + \frac{1-p}{np}(\pi_x + \pi_y - 2\pi_x\pi_y)$$

(iii) $E\left[\hat{V}(\hat{\pi}_G^*)\right] = \frac{N-n}{p^2Nn}\frac{1}{n-1}E_p\left[\sum_{i\in S}E_R(z_i^2) - \frac{\sum_{i\in S}E_R(z_i^2) + \sum_{i\neq}\sum_{j\in S}E_R(z_i)E_R(z_j)}{n}\right]$

$$+\frac{1-p}{p}(\pi_x + \pi_y - 2\pi_x\pi_y)$$

$$= \frac{N-n}{p^2Nn}\left[\sum_{i\in U}\{E_R(z_i)\}^2 + \sum_{i\in U} V_R(z_i) - \frac{1}{N}\sum_{i\neq}\sum_{j\in U} E_R(z_i)\,E_R(z_j)\right]$$

$$+ \frac{1-p}{p}(\pi_x + \pi_y - 2\pi_x\pi_y)$$

$$= V(\hat{\pi}_G^*)$$

From the expressions of $V(\hat{\pi}_G^*)$ and (3.16), we find that

$$V(\hat{\pi}_G^*) - V(\hat{\pi}_{wor}) = \frac{n-1}{N-1}\frac{1-p}{np}\left[\pi_x(1-\pi_y) + \pi_y(1-\pi_x)\right]$$

$$\geq 0 \qquad (4.3)$$

From the Eq. (4.3), we conclude for the SRSWOR sampling, Tian's (2014) estimator $\hat{\pi}_{wor}$ based on NRR method is more efficient than the Greenberg et al.'s (1969) estimator $\hat{\pi}_G^*$ based on RR technique for estimating the population proportion π_y. However, for large N, both are equally efficient. The percentage relative efficiency of $\hat{\pi}_{wor}$ with respect to $\hat{\pi}_G^*$ under SRSWOR sampling assuming $\frac{N-1}{N} \cong 1$ is given by

$$\frac{V(\hat{\pi}_G^*)}{V(\hat{\pi}_{wor})} \times 100$$

$$= \frac{(1-f)\left[\pi_y(1-\pi_y) + \frac{1-p}{p^2}\pi_x(1-\pi_x)\right] + \frac{1-p}{p}(\pi_x + \pi_y - 2\pi_{xy})}{(1-f)\left[\pi_y(1-\pi_y) + \frac{1-p}{p^2}\{(p-1)\pi_x^2 + (1-2p\pi_y)\pi_x + p\pi_y\}\right]} \times 100 \qquad (4.4)$$

The percentage relative efficiency (E) for different values of π_x, π_y, p and f is given in the Table 4.1. For the given values of π_x,π_y, the efficiency increases with p until $p = 0.50$, then it decreases. Efficiency increases with the increase in the sampling fraction f. The maximum efficiency 148.6 is attained when

f = 0.40, π_x= 0.10, π_y = 0.75 and p= 0.40.

Table 4.1. Efficiency of $\hat{\pi}_G^*$ with respect to $\hat{\pi}_{wor}$

		f = 0.1						f = 0.2				
π_y	π_x	p						p				
		0.1	0.25	0.4	0.5	0.75		0.1	0.25	0.4	0.5	0.75
0.10	0.10	102.0	104.2	105.3	105.6	104.2		104.5	109.4	112	112.5	109.4
	0.25	101.7	103.7	105.2	105.8	105.3		103.8	108.4	111.7	113.0	111.9
	0.40	101.8	104	105.6	106.2	106.1		104.1	109.0	112.5	114.0	113.6
	0.50	102.0	104.3	105.9	106.6	106.5		104.5	109.8	113.4	114.9	114.6
	0.75	103.2	106.0	107.5	108.0	107.5		107.3	113.5	116.8	117.9	116.9

Table 4.1. Efficiency of $\hat{\pi}_G^*$ with respect to $\hat{\pi}_{wor}$ (cont.)

π_y	π_x	f = 0.1						f = 0.2				
		p						p				
		0.1	0.25	0.4	0.5	0.75		0.1	0.25	0.4	0.5	0.75
0.25	0.10	102.9	105.3	106.0	105.8	103.7		106.6	111.9	113.4	113.0	108.4
	0.25	102.0	104.2	105.3	105.6	104.2		104.5	109.4	112,0	112.5	109.4
	0.40	101.9	104.1	105.3	105.7	104.6		104.3	109.1	112.0	112.8	110.3
	0.50	102.0	104.2	105.6	105.9	104.8		104.5	109.5	112.5	113.3	110.9
	0.75	103.0	105.6	106.7	106.9	105.6		106.7	112.5	115.2	115.6	112.5
0.40	0.10	103.7	106.1	106.5	106.2	104.0		108.4	113.6	114.7	114.0	109.0
	0.25	102.3	104.6	105.6	105.7	104.1		105.2	110.3	112.6	112.8	109.1
	0.40	102.0	104.2	105.3	105.6	104.2		104.5	109.4	112.0	112.5	109.4
	0.50	102.0	104.2	105.4	105.6	104.3		104.5	109.4	112.1	112.6	109.6
	0.75	102.7	105.1	106.2	106.3	104.6		106.1	111.5	113.9	114.1	110.3
0.50	0.10	104.2	106.5	106.9	106.6	104.3		109.3	114.6	115.6	114.9	109.8
	0.25	102.5	104.8	105.9	105.9	104.2		105.6	110.9	113.2	113.3	109.5
	0.40	102.1	104.3	105.4	105.6	104.2		104.7	109.6	112.2	112.6	109.4
	0.50	102.0	104.2	105.3	105.6	104.2		104.5	109.4	112.0	112.5	109.4
	0.75	102.5	104.8	105.9	105.9	104.2		105.6	110.9	113.2	113.3	109.5
0.75	0.10	105.1	107.5	108.1	108.0	106.0		111.4	116.9	118.2	117.9	113.5
	0.25	103.0	105.6	106.7	106.9	105.6		106.7	112.5	115.2	115.6	112.5
	0.40	102.2	104.6	105.9	106.3	105.1		105.0	110.3	113.3	114.1	111.5
	0.50	102.0	104.2	105.6	105.9	104.8		104.5	109.5	112.5	113.3	110.9
	0.75	102.0	104.2	105.3	105.6	104.2		104.5	109.4	112.0	112.5	109.4
		f = 0.3						f = 0.4				
0.10	0.10	107.7	116.1	120.6	121.4	116.1		112.0	125.0	132.0	133.3	125.0
	0.25	106.4	114.4	120.1	122.3	120.3		110.0	122.4	131.2	134.6	131.6
	0.40	106.9	115.4	121.4	124.0	123.4		110.8	123.9	133.3	137.3	136.4
	0.50	107.8	116.7	122.9	125.5	125.1		112.1	126.0	135.7	139.7	139.1
	0.75	112.5	123.2	128.8	130.7	129.1		119.5	136.1	144.8	147.7	145.2
0.25	0.10	111.4	120.3	123.0	122.3	114.4		117.7	131.6	135.7	134.6	122.4
	0.25	107.7	116.1	120.6	121.4	116.1		112.0	125.0	132.0	133.3	125.0
	0.40	107.3	115.6	120.6	122.0	117.7		111.4	124.3	132.1	134.2	127.5
	0.50	107.7	116.3	121.4	122.9	118.7		112.0	125.4	133.3	135.6	129.1
	0.75	111.5	121.4	126.0	126.8	121.4		117.9	133.3	140.4	141.7	133.3
0.40	0.10	114.3	123.4	125.2	124.0	115.4		122.3	136.4	139.2	137.3	123.9
	0.25	108.9	117.7	121.6	122.0	115.6		113.9	127.5	133.7	134.2	124.3
	0.40	107.7	116.1	120.6	121.4	116.1		112.0	125.0	132.0	133.3	125.0
	0.50	107.7	116.1	120.7	121.6	116.4		112.0	125.1	132.2	133.7	125.6
	0.75	110.4	119.8	123.8	124.1	117.7		116.2	130.7	137.0	137.5	127.5
0.50	0.10	116.0	125.1	126.7	125.5	116.7		124.9	139.1	141.6	139.7	126.0
	0.25	109.7	118.7	122.6	122.9	116.3		115.0	129.1	135.2	135.6	125.4
	0.40	108.0	116.4	120.9	121.6	116.1		112.4	125.6	132.5	133.7	125.1
	0.50	107.7	116.1	120.6	121.4	116.1		112.0	125.0	132.0	133.3	125.0
	0.75	109.7	118.7	122.6	122.9	116.3		115.0	129.1	135.2	135.6	125.4
0.75	0.10	119.6	129.1	131.3	130.7	123.2		130.5	145.2	148.6	147.7	136.1
	0.25	111.5	121.4	126.0	126.8	121.4		117.9	133.3	140.4	141.7	133.3
	0.40	108.6	117.7	122.8	124.1	119.8		113.4	127.5	135.4	137.5	130.7
	0.50	107.7	116.3	121.4	122.9	118.7		112.0	125.4	133.3	135.6	129.1
	0.75	107.7	116.1	120.6	121.4	116.1		112.0	125.0	132.0	133.3	125.0

5. Conclusion

The Randomized Response technique was introduced by Warner (1965) to collect data on sensitive characteristics. In this technique, the respondents have to perform randomized response experiments using devices which make the survey more expensive and time-consuming than the direct response surveys. Apart from these limitations, the procedure may yield different response depending on the outcome of the RR trial and it is unfeasible for mail questionnaire. To overcome some of the aforementioned difficulties, nonrandomized response (NRR) model was proposed by Tian et al. (2007), Yu et al. (2008), Tan et al. (2009), Tian (2014), among others. All the proposed procedures are limited to SRSWR sampling design and are unusable in real life complex multi-character surveys. In this paper, NRR models have been extended to complex surveys in a unified setup, which is applicable to any sampling design and estimators. The estimators of the population proportions, their variances and unbiased estimators of the variances for the existing NRR models can be obtained from the proposed method as special cases. It has been found for the SRSWR sampling, expressions of the estimators of the population proportion π_y, its variance for the Greenberg et al. (1969) and Tian (2014) are the same. However, for the SRSWOR sampling, the variance of Tian (2014) estimator is smaller than that of the Greenberg et al. (1969) estimator. But for large population they are equal.

Acknowledgements

The authors are grateful to the anonymous reviewers, whose thoughtful suggestions led to the substantial improvement of the earlier version of the manuscript.

REFERENCES

ABERNATHY, J. R., GREENBERG, B. G., HORVITZ D. G., (1970). Estimates of induced abortion in urban North Carolina, Demography, 7, pp. 19–29.

ARNAB, R., (1990). On commutativity of design and model expectations in randomized response surveys. Communications in Statistics, Theory & Methods, pp. 3751–2757.

ARNAB, R., (1996). Randomized response trials: a unified approach for qualitative data, Commun. Statist. Theory & Methods 25 (6), p. 1173.

ARNAB, R., (2017). Survey Sampling Theory and Applications. Academic Press, Oxford.

ARNAB, R., MOTHUPI, T., (2015). Randomized response techniques: A case study of the risky behaviors' of students of a certain University, Model Assisted Statistics and Applications, 10, pp. 421–430.

CENTRAL STATISTICAL OFFICE, (2004). Household Income and Expenditure Survey 2002/03, Republic of Botswana.

CENTRAL STATISTICS OFFICE, (2009). Botswana Aids Impact Survey III (2008), Statistical Report.

FOLSOM, S. A., (1973). The two alternative questions randomized response model for human surveys. J. Amer. Statist. Assoc., 68, pp. 525-530.

FRANKLIN, L. A., (1989). A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations. Commun. Statist. Theory and methods 18, pp. 489–505.

GOODSTADT, M. S., GRUSON, V., (1975). The randomized response technique; a test on drug use. J. Amer. Statist. Assoc., 70, pp. 814–818

GREENBERG, B. G., ABUL-ELA, A. L. A., SIMMONS, W. R., HORVITZ, D. G., (1969). The unrelated question randomized response model: Theoretical framework. J. Amer. Statist. Assoc. 64, pp. 520–539

HORVITZ, D. G., SHAH, B. V., SIMMONS, W. R., (1967). The unrelated question randomized response model. Proceedings of Social Statistical section, Amer. Statist. Assoc. pp. 65–72.

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

RAGHAVRAO, D., (1978). On estimation problem in Warner's randomized response techniques. Biometrics 34, pp. 87–90.

RUEDA, M., COBO, B., ARCOS, A., (2015). Package 'RRTCS': Randomized Response Techniques for Complex Surveys, http://cran.r-project.org/web/packages/RRTCS/.

STATISTICS SOUTH AFRICA, (2005). Income and Expenditure of households 2005/2006, Republic of South Africa.

TAN, G.L., YU, J. W., TANG, M. L., (2009). Sample survey with sensitive questions: a non-randomized response approach, The American Statistician, 63, pp. 9–16.

TANG, M., WU, Q., TIAN, G., GUO, J., (2014). Two-sample Non Randomized Response Techniques for Sensitive Questions. Commun. Statist. Theory & Methods, 43, pp. 408–425.

TIAN, G. L., YU, J. W., TANG, M. L., GENG, Z., (2007). A new non-randomized model for analysing sensitive question with binary outcomes. Statistics in Medicine, 26, pp. 4238–4252.

TIAN, G. L., (2014). A new non-randomized response model: the parallel model. Statistica Neerlandica, 68, pp. 293–323.

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

WU, Q., TANG, M., (2016). Non-randomized response model for sensitive survey with noncompliance. Statistical Methods in Medical Research, 25, pp. 2827–2839.

YU, J. W., TIAN, G. L., TANG, M. L., (2008). Two new models for survey sampling with sensitive characteristics: Design and Analysis. Metrika, 67, pp. 251–263.