Modifications of Some Simple One-stage Randomized Response Models to Two-stage in Complex Surveys

Mohammad Rafiq
Department of Statistics, University of Malakand, Dir (Lower)
rafiq.state@gmail.com

Zawar Hussain
Department of Statistics, Quaid-e-Azam University, Islamabad
zhlangah@yahoo.com

Sohail Akhtar
Department of Statistics, University of Malakand, Dir (Lower)
s.akhtar@uom.edu.pk

Abstract

Warner (1965) introduced a Randomized Response Technique (RRT) to minimize bias due to non-response or false response. Thereafter, several researchers have made significant contribution in the development and modification of different Randomized Response Models. We have modified a few one-stage Simple Randomized Response Models to two-stage randomized response models in complex surveys and found that our developed models are more efficient.

Keywords: Randomized Response Technique, Sensitive Character, Unbiased Estimators and Complex Surveys.

1. Introduction

In survey sampling, the collection of information on sensitive variables like illegal income, accumulated savings, intentional tax evasion and usage of illegal drugs is very difficult to obtain and in case of any response they give false or evasive answers, when direct questioning is done. To overcome this problem Warner (1965) suggested a technique known as Randomized Response Technique.

There are some others articles which include, Chaudhuri and Mukerjee (1988), Bhargava and Singh (2000), Chaudhuri (2001a, 2001b), Horng et al. (2004), Chaudhuri (2004), Shabbir and Gupta (2005), Saha (2006), Hussain et al. (2007), Chaudhuri et al. (2009), Hussain et al. (2009), Rafiq et al. (2013). In this paper, we have modified Bhargava and Singh (2000), Horng et al. (2004), Shabbir and Gupta (2005) procedures to Two-stage Randomized Response procedures in complex sample surveys and observed that the Two-stage procedure is better than the conventional One-stage procedure in complex survey design.

2. Generalized Two-Stage Randomized Response Procedure

Suppose R_1 and R_2 are two independent randomization techniques. Randomization technique R_2 consists of a box which carries two types of balls that is black and white. These balls are kept in proportion q and $(1-q)$ respectively, where $0 < q < 1$, each selected individual in the sample is first provided with RR technique R_2 and is requested to drawn a ball from the box in the absence of the interviewer, if a black ball is chosen, the respondent is requested to disclose his / her true A or A^C character in terms of 1 or 0. If he/she select a white ball, then he / she is requested to use RR technique R_1 and report the RR say z_i depending on the outcome of the random experiment and the interviewers true A or A^C characteristics. Here R_1 may be one of the randomized response procedures.

Let w_i be the response from ith individual through a two stage RR procedure.

$$w_i = y_i \text{ with probability } q,$$

$$= z_i \text{ with probability } (1-q).$$

Let suppose that

$$E_{R_1}(z_i) = ay_i + b, \text{ and } V_{R_1}(z_i) = \mathrm{d}_i, \text{"}\ i \in U.$$

Suppose E_{R_1}, V_{R_1} are the expectation and variance respectively with respect to RR device R_1. Similarly E_{R_2}, V_{R_2} denote the expectation and variance with respect to RR device R_2. Let E_R, V_R represent the expectation and variance of the overall two-stage RR Procedure.

$$E_R(w_i) = \left[q + (1-q)a\right] y_i + b(1-q)$$

This implies that,

$$r_i = \frac{w_i - (1-q)b}{\left(q + (1-q)a\right)}, \tag{2.1}$$

Satisfying that, $E_R(r_i) = y_i, \text{"}\ i \in U.$

$$V_R(w_i) = (1-q)\mathrm{d}_i + q(1-q)\left[(1-a)y_i - b\right]^2. \tag{2.2}$$

Thus

$$V_R(r_i) = \frac{(1-q)\left[\mathrm{d}_i + q\left((1-a)y_i - b\right)^2\right]}{\left[q + (1-q)a\right]^2}.$$

$$V_R(r_i) = \frac{(1-q)\left[\mathrm{d}_1 + q(1-a-b)^2\right]}{\left[q + (1-q)a\right]^2} = V_{21i}, \text{ if } y_i = 1, \tag{2.3}$$

$$V_R\left(r_i\right)=\frac{\left(1-q\right)\left(\mathrm{d}_0+qb^2\right)}{\left[q+\left(1-q\right)a\right]^2}=V_{20i}\text{, if } y_i=0, \tag{2.4}$$

where $\mathrm{d}_i|_{yi=1}=\mathrm{d}_1$ and $\mathrm{d}_i|_{yi=0}=\mathrm{d}_0$.

Here it may be noted that for q = 0, (Which is the case of one stage RR procedure)

$$r_i=\frac{z_i-b}{a},$$

$$E_{R_1}\left(r_i\right)=\frac{E_{R_1}\left(z_i\right)-b}{a}=y_i,$$

and $$V_{R_1}\left(r_i\right)=\frac{V_{R_1}\left(z_i\right)}{a^2}=\frac{\mathrm{d}_i}{a^2}=V_{1i}.$$

$$V_{R_1}\left(r_i\right)=\frac{V_{R_1}\left(z_i\right)}{a^2}=\frac{\mathrm{d}_1}{a^2}=V_{11i}\text{, if } y_i=1,$$

and

$$V_{R_1}\left(r_i\right)=\frac{V_{R_1}\left(z_i\right)}{a^2}=\frac{\mathrm{d}_0}{a^2}=V_{10i}\text{, if } y_i=0,$$

Now we may two cases.

Case 1: if y_i =1

In this case the two-stage randomized response procedure will be more efficient than the conventional randomized response procedure if

$$V_{21i}<V_{11i},\quad "i\in U,$$

$$\frac{\left(1-q\right)\left[\mathrm{d}_1+q\left(1-a-b\right)^2\right]}{\left[q+\left(1-q\right)a\right]^2}<\frac{\mathrm{d}_1}{a^2}$$

Or $$q>1-\frac{\mathrm{d}_1}{\mathrm{d}_1\left(1-a\right)^2+a^2\left(1-a-b\right)^2}=d_1. \tag{2.5}$$

In case y_i =0 the two-stage randomized response procedure will be more efficient than the conventional randomized response procedure if

$$V_{20i}<V_{10i},\quad "i\in U,$$

or $$\frac{\left(1-q\right)\left(\mathrm{d}_0+qb^2\right)}{\left(q+\left(1-q\right)a\right)^2}<\frac{\mathrm{d}_0}{a^2}$$

Solving the above we get, $q > 1 - \dfrac{d_0}{d_0(1-a)^2 + a^2b^2} = d_2$ (2.6)

The two cases, when y_i =1 and y_i =0, discussed above, show that the Generalized two-stage RR procedure will be better than the usual *RR* procedure if

$$q > \max(d_1, d_2).$$

This condition is known as optimality condition.

3. Estimation Incorporating the Sample Designs

Suppose $U = (1, 2, ..., i, ..., N)$ represent a finite population with y_i as the value for its *ith* unit on a variable y such that $y_i = 1$, if *ith* individual carries a sensitive attribute A, $= 0$ if *ith* individual does not carry attribute A. The purpose is to obtain unbiased and precise estimator of the parameter $Y = \sum_{i=1}^{N} y_i$, on surveying a sample s of n units from U selected with sampling design $P(s)$. We will use E_R, V_R as expectation and variance operators with respect to any *RR* design.

4. Sample Selection and Estimation

Let E_P, V_P are the operators for expectation and variance with respect to any sampling design P. Let E, V are the operators for overall expectation and variance such that

$$E = E_P E_R = E_R E_P,\ V = E_P V_R + V_P E_R = E_R V_P + V_R E_P.$$

For a general sampling design, we shall write $t_b = \sum_{i=1}^{N} y_i b_{si} I_{si}$, an estimator for the population total Y, where b_{si}, s are constants free of $\underline{Y}$, $\underline{R}$, $I_{si} = 1$, if $i \in s$, and $= 0$, if $i \notin s$ such that $E_P(b_{si} I_{si}) = 1\ "\ i$.

In randomized response procedure y_i's are not directly ascertained, therefore, we consider an unbiased estimator for the population total Y, in two-stage randomization device as,

$$e_{2b} = \sum_{i=1}^{N} r_i b_{si} I_{si}, \quad \text{where} \qquad r_i = \frac{w_i - (1-q)b}{q + (1-q)a}.$$

then

$$E(e_{2b}) = E_P E_R(e_{2b}) = E_R E_P(e_{2b}) = Y.$$

also

$$V_R(r_i) = \frac{(1-q)\left[d_i + q\left((1-a)y_i - b\right)^2\right]}{\left(q + (1-q)a\right)^2} = V_{2i} \qquad (4.1)$$

$$V(e_{2b}) = E_P V_R(e_{2b}) + V_P E_R(e_{2b})$$

$$V(e_{2b}) = V_P(t_b) + E_P\left(\sum_{i=1}^{N} V_{2i} b_{si}^{\ 2} I_{si}\right).$$

If y_i's are not ascertainable, therefore, an unbiased estimator for the population total Y, using one stage randomization device is given by

$$e_{1b} = \sum_{i=1}^{N} r_i b_{si} I_{si}, \text{ where } \quad r_i = \frac{z_i - b}{a},$$

now $E(e_{1b}) = E_P E_R(e_{1b}) = E_R E_P(e_{1b}) = Y$. and

$$V(e_{1b}) = E_P V_R(e_{1b}) + V_P E_R(e_{1b})$$

$$V(e_{1b}) = V_P(t_b) + E_P\left(\sum_{i=1}^{N} V_{1i} b_{si}^{\ 2} I_{si}\right). \tag{4.2}$$

Writing

$V_P(t_b) = \sum_{i=1}^{N} y_i^2 c_i + \sum_{i \neq j}^{N} Y_i Y_j c_{ij}$, where $c_i = E_P\left(b_{si}^{\ 2} I_{si}\right) - 1$, $c_{ij} = E_P\left(b_{si} I_{si} - 1\right)\left(b_{sj} I_{sj} - 1\right)$, if c_{si}, c_{sij} are available free of $\underline{Y}, \underline{R}$, $I_{sij} = I_{si} I_{sj}$ Such that $E_P\left(c_{si} I_{si}\right) = c_i$ and $E_P\left(c_{sij} I_{sij}\right) = c_{ij}$.

5. Efficiency Comparisons

The two-stage randomize response model will be more efficient than its conventional RR model if $V(e_{1b}) - V(e_{2b}) \geq 0$

Or $\quad (V_{1i} - V_{2i}) \geq 0$ (5.1)

6. Bhargava and Singh (2000) Procedure

Bhargava and Singh (2000) randomization device consist of a deck having three types of cards, each card having different statements

i) I belong to the sensitive group A

ii) I do not belong to the sensitive group A

iii) Blank cards, with probabilities P_1, P_2 and P_3 respectively, where $P_1 \neq P_2$ and $P_1 + P_2 + P_3 = 1$

If the blank cards selected he / she respond *yes* irrespective of his or her actual status. The probability of yes response is given as,

$$E_{R_1}(I_i) = P_1 y_i + P_2(1 - y_i) + P_3,$$

and

$$V_{R_1}(I_i) = E_{R_1}(I_i)\left(1 - E_{R_1}(I_i)\right)$$

$$= \left[(P_1 - P_2)\, y_i + (P_2 + P_3)\right] - \left[(P_1 - P_2)\, y_i + (P_2 + P_3)\right]^2 .$$

were $\mathrm{d}_1 = (P_1 + P_3)\left(1 - (P_1 + P_3)\right)$ $\mathrm{d}_0 = (P_2 + P_3)\left(1 - (P_2 + P_3)\right)$

If y_i's are not ascertainable, therefore, an unbiased estimator for the population total Y, using one stage randomization device is given by

$$e_{1b} = \sum_{i=1}^{N} r_i b_{si} I_{si},\ r_i = \frac{\left[I_i - (P_2 + P_3)\right]}{(P_1 - P_2)},\ \text{satisfies that } E_R(r_i) = y_i .$$

$$V_{11i} = \frac{(P_1 + P_3)\left(1 - (P_1 + P_3)\right)}{(P_1 - P_2)^2} \text{ and } V_{10i} = \frac{(P_2 + P_3)\left(1 - (P_2 + P_3)\right)}{(P_1 - P_2)^2}$$

$$a = P_1 - P_2, b = P_2 + P_3, \mathrm{d}_0 = (P_2 + P_3)\left(1 - (P_2 + P_3)\right), \mathrm{d}_1 = (P_1 + P_3)\left(1 - (P_1 + P_3)\right)$$

In randomized response procedure y_i's are not directly ascertained, therefore, we consider an unbiased estimator for the population Y, total in two-stage randomization device as,

$$e_{2b} = \sum_{i=1}^{N} r_i b_{si} I_{si},\ \text{ where } r_i = \frac{w_i - (1-q)(P_2 + P_3)}{q + (1-q)(P_1 - P_2)}$$

$$V_{21i} = \frac{(1-q)\left(1 - (P_1 + P_3)\right)\left[(P_1 + P_3) + q\left(1 - (P_1 + P_3)\right)\right]}{\left[q + (1-q)(P_1 - P_2)\right]^2}$$

$$V_{20i} = \frac{(1-q)(P_2 + P_3)\left[\left(1 - (P_2 + P_3)\right) + q(P_2 + P_3)\right]}{\left[q + (1-q)(P_1 - P_2)\right]^2}$$

here

$$d_1 = \frac{(P_2 - P_1)(P_2 + P_1 + 2P_3)}{\left[(P_2 - P_1)(P_2 + P_1 + 2P_3) + (P_1 + P_3)\right]}$$

and

$$d_2 = \frac{(P_1 - P_2)\left((P_1 + P_2 + 2P_3 - 2)\right)}{\left[(P_1 - P_2)\left((P_1 + P_2 + 2P_3 - 2)\right) + (1 + P_2 - P_3 - 2P_1)\right]} .$$

Two-stage randomize response model will be more efficient than its conventional RR model if

$$V(e_{1b}) - V(e_{2b}) \geq 0$$

$$\frac{(P_1 + P_3)\left(1 - (P_1 + P_3)\right)}{(P_1 - P_2)^2} - \frac{(1-q)\left(1 - (P_1 + P_3)\right)\left[(P_1 + P_3) + q\left(1 - (P_1 + P_3)\right)\right]}{\left[q + (1-q)(P_1 - P_2)\right]^2} \geq 0$$

and

$$\frac{(P_2+P_3)\left(1-(P_2+P_3)\right)}{(P_1-P_2)^2}-\frac{(1-q)(P_2+P_3)\left[\left(1-(P_2+P_3)\right)+q(P_2+P_3)\right]}{\left[q+(1-q)(P_1-P_2)\right]^2}\geq 0$$

As it is apparent from the above inequality to derive an efficiency condition, so, we resort to an empirical study. The results based on empirical study are given in Table 1.

Table 1: *Values of* $V_{11i}, V_{10i}, V_{21i}, V_{20i}$, d_1 *and* d_2 *for different values of* P_1, P_2, P_3 *and* q

Rows	P_1	P_2	P_3	q	V_{11i}	V_{10i}	V_{21i}	V_{20i}	d_1	d_2
1	0.6	0.3	0.1	0.95	2.333	2.667	0.016	0.021	-0.891	-0.818
2	0.6	0.3	0.1	0.05	2.333	2.667	1.816	2.099	-0.891	-0.818
3	0.6	0.2	0.2	0.95	1.000	1.500	0.011	0.021	-1.500	-1.142
4	0.6	0.2	0.2	0.05	1.000	1.500	0.832	1.274	-1.500	-1.142
5	0.6	0.1	0.3	0.95	0.360	0.960	0.005	0.021	-2.600	-1.400
6	0.6	0.1	0.3	0.05	0.360	0.960	0.311	0.855	-2.600	-1.400
7	0.7	0.2	0.1	0.95	0.640	0.840	0.01	0.016	-2.200	-1.800
8	0.7	0.2	0.1	0.05	0.640	0.840	0.558	0.74	-2.200	-1.800
9	0.7	0.1	0.2	0.95	0.250	0.583	0.005	0.015	-4.000	2.181
10	0.7	0.1	0.2	0.05	0.250	0.583	0.224	0.53	-4.000	2.181
11	0.8	0.1	0.1	0.95	0.184	0.327	0.005	0.01	-5.923	-3.705
12	0.8	0.1	0.1	0.05	0.184	0.327	0.168	0.301	-5.923	-3.705

7. Shabbir and Gupta (2005) Procedure

Shabbir and Gupta (2005) randomized response model consists of a deck having three types of cards with different statements,

i) I belong to the sensitive group A,

ii) I do not belong to the sensitive group A,

iii) Blank cards, with probabilities P_1, P_2 and P_3 respectively, where $P_1 \neq P_2$ and $P_1 + P_2 + P_3 = 1$

If the blank card is selected by the respondent, he / she speak the truth.

The probability of *yes* response is given as,

$$E_{R_1}(I_i) = P_1 y_i + P_2(1-y_i) + P_3 y_i,$$

$$V_{R_1}(I_i) = E_{R_1}(I_i)\left(1 - E_{R_1}(I_i)\right)$$

$$= \left((P_1 - P_2 + P_3)y_i + P_2\right)\left[1 - \left((P_1 - P_2 + P_3)y_i + P_2\right)\right]$$

where $\mathrm{d}_1 = (P_1 + P_3)\left(1 - (P_1 + P_3)\right)$ and $\mathrm{d}_0 = P_2(1 - P_2)$

If y_i's are not ascertainable, therefore, an unbiased estimator for the population total Y, using one stage randomization device is given by

$$e_{1b} = \sum_{i=1}^{N} r_i b_{si} I_{si}, \text{ where } r_i = \frac{I_i - P_2}{P_1 - P_2 + P_3}$$

$$a = P_1 - P_2 + P_3, b = P_2, \mathrm{d}_1 = (P_1 + P_3)\left(1 - (P_1 + P_3)\right), \mathrm{d}_0 = P_2(1 - P_2)$$

$$V_{11i} = \frac{(P_1 + P_3)\left(1 - (P_1 + P_3)\right)}{(P_1 - P_2 + P_3)^2} \text{ and } V_{10i} = \frac{P_2(1 - P_2)}{(P_1 - P_2 + P_3)^2}$$

In randomized response procedure y_i's are not directly ascertained, therefore, we consider an unbiased estimator for the population total Y, in two-stage randomization device as,

$$e_{2b} = \sum_{i=1}^{N} r_i b_{si} I_{si}, \quad \text{where} \quad r_i = \frac{w_i - (1-q)P_2}{q + (1-q)(P_1 - P_2 + P_3)}$$

$$V_{21i} = \frac{(1-q)\left(1 - (P_1 + P_3)\right)\left[(P_1 + P_3) + q\left(1 - (P_1 + P_3)\right)\right]}{\left[q + (1-q)(P_1 - P_2 + P_3)\right]^2}$$

$$V_{20i} = \frac{(1-q)P_2\left[(1 - P_2) + qP_2\right]}{\left[q + (1-q)(P_1 - P_2 + P_3)\right]^2}.$$

$$d_1 = \frac{(P_2 + P_1 + P_3)(P_2 - P_1 - P_3)}{\left[(P_2 + P_1 + P_3)(P_2 - P_1 - P_3) + (P_1 + P_3)\right]}$$

$$d_2 = \frac{(P_1 - P_2 + P_3)(P_1 + P_2 + P_3 - 2)}{\left[(P_1 - P_2 + P_3)(P_1 + P_2 + P_3 - 2) + (1 - P_2)\right]}.$$

Two-stage randomize response model will be more efficient than its conventional RR model if

$$V(e_{1b}) - V(e_{2b}) \geq 0$$

or

$$\frac{(P_1 + P_3)\left(1 - (P_1 + P_3)\right)}{(P_1 - P_2 + P_3)^2} - \frac{(1-q)\left(1 - (P_1 + P_3)\right)\left[(P_1 + P_3) + q\left(1 - (P_1 + P_3)\right)\right]}{\left[q + (1-q)(P_1 - P_2 + P_3)\right]^2} \geq 0$$

and $$\frac{P_2(1-P_2)}{(P_1-P_2+P_3)^2}-\frac{(1-q)P_2\left[(1-P_2)+qP_2\right]}{\left[q+(1-q)(P_1-P_2+P_3)\right]^2}\geq 0$$

As it is apparent from the above inequality to derive an efficiency condition, so, we resort to an empirical study. The results based on empirical study are given in Table 2.

Table 2: *Values of* $V_{11i}, V_{10i}, V_{21i}, V_{20i}, d_1$ *and* d_2 *for different values of* $P_1, P_2, P_3,$ *and* q.

Rows	P_1	P_2	P_3	q	V_{11i}	V_{10i}	V_{21i}	V_{20i}	d_1	d_2
1	0.6	0.3	0.1	0.95	0.25	0.583	0.005	0.015	-4.0	-2.2
2	0.6	0.3	0.1	0.05	0.25	0.583	0.224	0.53	-4.0	-2.2
3	0.6	0.25	0.15	0.95	0.354	0.521	0.008	0.013	-3.5	-2.6
4	0.6	0.25	0.15	0.05	0.354	0.521	0.319	0.471	-3.5	-2.6
5	0.6	0.2	0.2	0.95	0.444	0.444	0.010	0.010	-3.0	-3.0
6	0.6	0.2	0.2	0.05	0.444	0.444	0.400	0.400	-3.0	-3.0
7	0.6	0.15	0.25	0.95	0.521	0.354	0.013	0.008	-2.6	-3.5
8	0.6	0.15	0.25	0.05	0.521	0.354	0.471	0.318	-2.6	-3.5
9	0.7	0.2	0.1	0.95	0.184	0.327	0.005	0.010	-5.9	-3.7
10	0.7	0.2	0.1	0.05	0.184	0.327	0.168	0.301	-5.9	-3.7
11	0.7	0.15	0.15	0.95	0.260	0.260	0.008	0.008	-4.7	-4.7
12	0.7	0.15	0.15	0.05	0.260	0.260	0.239	0.239	-4.7	-4.7
13	0.8	0.15	0.05	0.95	0.074	0.199	0.003	0.008	-12.6	-5.5
14	0.8	0.15	0.05	0.05	0.074	0.199	0.069	0.186	-12.6	-5.5
15	0.8	0.1	0.1	0.95	0.141	0.141	0.005	0.005	-8.0	-8.0
16	0.8	0.1	0.1	0.05	0.141	0.141	0.131	0.131	-8.0	-8.0
17	0.8	0.05	0.15	0.95	0.199	0.074	0.008	0.003	-5.5	-12.6
18	0.8	0.05	0.15	0.05	0.199	0.074	0.186	0.069	-5.5	-12.6
19	0.9	0.07	0.03	0.95	0.036	0.080	0.002	0.004	-27.5	-13.1
20	0.9	0.07	0.03	0.05	0.036	0.080	0.034	0.076	-27.5	-13.1
21	0.9	0.03	0.07	0.95	0.080	0.036	0.004	0.002	-13.1	-27.5
22	0.9	0.03	0.07	0.05	0.080	0.036	0.076	0.034	-13.1	-27.5

8. Horng *et. al*. (2004) Procedure

In this procedure the randomization device carries a deck having four types of cards, each having different statements,

i) I belong to the sensitive group A

ii) I do not belong to the sensitive group A

iii) Yes cards

iv) No cards, with probabilities P_1, P_2, P_3 and P_4 respectively, where $P_1 \neq P_2$ and $P_1 + P_2 + P_3 + P_4 = 1$

The probability of *yes* response is given by,

$$E_{R_1}(I_i) = P_1 y_i + P_2(1 - y_i) + P_3,$$

$$V_{R_1}(I_i) = E_{R_1}(I_i)\left(1 - E_{R_1}(I_i)\right)$$

$$= \left[(P_1 - P_2)y_i + (P_2 + P_3)\right] - \left[(P_1 - P_2)y_i + (P_2 + P_3)\right]^2.$$

$$\mathrm{d}_1 = (P_1 + P_3)\left(1 - (P_1 + P_3)\right) \text{ and } \mathrm{d}_0 = (P_2 + P_3)\left(1 - (P_2 + P_3)\right)$$

If y_i's are not ascertainable, therefore, an unbiased estimator for the population total Y, using one stage randomization device is given by

$$e_{1b} = \sum_{i=1}^{N} r_i b_{si} I_{si},\ r_i = \frac{I_i - (P_2 + P_3)}{P_1 - P_2}, \text{ satisfies that } E_R(r_i) = y_i.$$

$$a = P_1 - P_2, b = P_2 + P_3, \mathrm{d}_1 = (P_1 + P_3)\left(1 - (P_1 + P_3)\right), \mathrm{d}_0 = (P_2 + P_3)\left(1 - (P_2 + P_3)\right)$$

$$V_{11i} = \frac{(P_1 + P_3)\left(1 - (P_1 + P_3)\right)}{(P_1 - P_2)^2} \text{ and } V_{10i} = \frac{(P_2 + P_3)\left(1 - (P_2 + P_3)\right)}{(P_1 - P_2)^2}$$

In randomized response procedure y_i's are not directly ascertained, therefore, we consider an unbiased estimator for the population total Y, in two-stage randomization device as,

$$e_{2b} = \sum_{i=1}^{N} r_i b_{si} I_{si}, \text{ where } r_i = \frac{w_i - (1-q)(P_2 + P_3)}{q + (1-q)(P_1 - P_2)}.$$

$$V_{21i} = \frac{(1-q)\left(1 - (P_1 + P_3)\right)\left[(P_1 + P_3) + q\left(1 - (P_1 + P_3)\right)\right]}{\left[q + (1-q)(P_1 - P_2)\right]^2}$$

$$V_{20i} = \frac{(1-q)(P_2 + P_3)\left[\left(1 - (P_2 + P_3)\right) + q(P_2 + P_3)\right]}{\left[q + (1-q)(P_1 - P_2)\right]^2}.$$

$$d_1 = \frac{(P_2 - P_1)(P_2 + P_1 + 2P_3)}{\left[(P_2 - P_1)(P_2 + P_1 + 2P_3) + (P_1 + P_3)\right]}.$$

$$d_2 = \frac{(P_1 - P_2)(P_1 + P_2 + 2P_3 - 2)}{\left[(P_1 - P_2)(P_1 + P_2 + 2P_3 - 2) + (1 - P_1 - P_3)\right]}.$$

Two-stage randomize response model will be more efficient than its conventional RR model if

$$V(e_{1b}) - V(e_{2b}) > 0$$

$$\frac{(P_1+P_3)(1-(P_1+P_3))}{(P_1-P_2)^2} - \frac{(1-q)(1-(P_1+P_3))\left[(P_1+P_3)+q(1-(P_1+P_3))\right]}{\left[q+(1-q)(P_1-P_2)\right]^2} \geq 0$$

and $$\frac{(P_2+P_3)(1-(P_2+P_3))}{(P_1-P_2)^2} - \frac{(1-q)(P_2+P_3)\left[(1-(P_2+P_3))+q(P_2+P_3)\right]}{\left[q+(1-q)(P_1-P_2)\right]^2} \geq 0$$

As it is apparent from the above inequality to derive an efficiency condition, so, we resort to an empirical study. The results based on empirical study are given in Table 3.

Table 3: *Values of* $, V_{11i}, V_{10i}, V_{21i}, V_{20i}, d_1$ *and* d_2 *for different values of* $P_1, P_2,$ P_3, P_4 *and* q

Rows	P_1	P_2	P_3	P_4	q	V_{11i}	V_{10i}	V_{21i}	V_{20i}	d_1	d_2
1	0.6	0.2	0.1	0.1	0.95	1.313	1.313	0.016	0.016	-1.3	-1.3
2	0.6	0.2	0.1	0.1	0.05	1.313	1.313	1.102	1.102	-1.3	-1.3
3	0.6	0.1	0.2	0.1	0.95	0.640	0.840	0.01	0.016	-2.2	-1.8
4	0.6	0.1	0.2	0.1	0.05	0.640	0.840	0.558	0.739	-2.2	-1.8
5	0.6	0.1	0.1	0.2	0.95	0.840	0.640	0.016	0.010	-1.8	-2.2
6	0.6	0.1	0.1	0.2	0.05	0.840	0.640	0.739	0.558	-1.8	-2.2
7	0.7	0.1	0.1	0.1	0.95	0.444	0.444	0.010	0.010	-3.0	-3.0
8	0.7	0.1	0.1	0.1	0.05	0.444	0.444	0.400	0.400	-3.0	-3.0
9	0.8	0.1	0.05	0.05	0.95	0.260	0.260	0.006	0.008	-4.7	-4.7
10	0.8	0.1	0.05	0.05	0.05	0.260	0.260	0.239	0.239	-4.7	-4.7
11	0.9	0.05	0.03	0.02	0.95	0.090	0.102	0.005	0.004	12.0	-10.8
12	0.9	0.05	0.03	0.02	0.05	0.090	0.102	0.084	0.095	-12.0	-10.8

9. Discussion and Conclusion

For different values of P_1, P_2, P_3, P_4 and q the values, V_{10i}, V_{11i}, V_{20i}, V_{21i}, d_1 and d_2 are calculated (See Tables 1-3 in appendix). The values of d_1 and d_2 are calculated only to satisfy the optimality condition, i.e. $q > \max(d_1, d_2)$. We have observed that in all the procedures the variances V_{20i} and V_{21i} of the two stage procedures is smaller than the variances V_{10i} and V_{11i} of the conventional one stage procedure .It mean that two stage procedure is more precise than the one stage procedure in complex survey. We have noted that d_1 and d_2 for all the procedure are negative. This shows that for all choice of the parameter, the two stage procedure will be better than its conventional one-stage procedure. For almost all models discussed above, the two stage procedure is more precise than its one stage procedure for P_1 close to 0.6 and 0.7 and for higher value of P_1 close to 0.8 and 0.9 the two stage procedures performs slightly well.

References

1. Bhargava, M. and Singh, R. (2000). A modified randomization device for Warner's model. *Statistca*, 60, 315–321.
2. Chaudhuri, A. and Mukerjee, R. (1988). Randomized response theory and techniques. *Marcel Dekker*, Inc., *New York.*
3. Chaudhuri, A. (2001a). Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. *Journal of Statistical Planning and Inference* 94, 37-42.
4. Chaudhuri, A. (2001b). Estimating sensitive proportions from unequal probability sampling using randomize responses. *Pakistan Journal of statistics* 17, 259-270.
5. Chaudhuri, A. Christofied, T. C and Saha, A. (2009). Protection of privacy in efficient application of randomized response techniques, *Statistical Method and Application*, 18, 389-418.
6. Hussain, Z. and Shabbir, J. (2007). On estimation of mean of a sensitive quantitative variable. *Inter Stat*, *July #* 6.
7. Husain, Z. and Shabbir, J. (2009). On estimation of mean of a sensitive quantitative variable in complex surveys. *Pakistan Journal of statistics*. 25, 127-134.
8. Mangat, N. S. (1994). An improved randomized response strategy. *Journal of Royal Statistical Society*, 56, 93-95.
9. Rafiq, M., Hussain. Z., Zaman, Q., Iqbal, M., Wahab, S. (2013). Modified version of Kim and Warde mixed Randomized Response in to complex surveys, 29, 4, 591-598.
10. Shabbir, J. and Gupta, S. (2005). On modified randomization device for Warner's model. *Pakistan Journal of statistics*, 21,123-129.
11. Saha, A. (2006). A generalized two-stage randomized response procedure in complex sampling survey. *Australian and New Zealand Journal of Statistics*, 48, 429-443.
12. Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of American Statistical Association*, 60, 63-69.