SORT 37 (2) July-December 2013, 189-210

An alternative to Kim and Warde's mixed randomized response model

Housila P. Singh and Tanveer A. Tarray*

Abstract

When open or direct surveys are about sensitive matters (e.g. gambling habits, addiction to drug and others intoxicants, alcoholism, proneness to tax invasion, induced abortions, drunken driving, history of past involvement in crimes, and homosexuality), non-response bias and response bias become serious problems because people oftentimes do not wish to give correct information. To reduce non-response and response bias, various alternative approaches have been proposed, for example a randomized response survey technique, or a mixed randomized response model using simple random sampling with a replacement sampling scheme that improves the privacy of respondents, proposed by authors Kim and Warde. In this paper we have suggested an alternative to Kim and Warde's mixed randomized response model to estimate the proportion of qualitative sensitive variable under the conditions presented in both the cases of completely truthful reporting and less than completely truthful reporting by the respondents. Properties of the proposed randomized response model have been studied along with recommendations. We have also extended the proposed model to stratified random sampling. Numerical illustrations and graphs are also given in support of the present study.

MSC: 62D05.

Keywords: Randomized response technique, Dichotomous population, Estimation of proportion, Privacy of respondents, Sensitive characteristics.

1. Introduction

Warner (1965) was first to introduce a randomized response (RR) model to estimate the proportion for sensitive attributes including homosexuality, drug addiction or abortion. Greenberg et al. (1969) proposed the unrelated question RR model that is a variation of Warner's (1965) RR model. Since the work by Warner (1965), a huge literature

* School of Studies in Statistics, Vikram University, Ujjain – 456010 – India. E-mail: tanveerstat@gmail.com
Received: February 2013
Accepted: September 2013

has emerged on the use and formulation of different randomization device to estimate the population proportion of a sensitive attribute in survey sampling. Mention may be made of the work of Tracy and Mangat (1996), Chudhuari and Mukherjee (1988), Ryu et al. (1993), Fox and Tracy (1986), Singh (2003), Singh and Tarray (2012, 2013a, b, c) and the references cited there in.

Mangat et al. (1997) and Singh et al. (2000) pointed out the privacy problem with the Moors (1971) model. Mangat et al. (1997) and Singh et al. (2000) have presented several strategies as an alternative to Moors model, but their models may lose a large portion of data information and require a high cost to obtain confidentiality of the respondents. Kim and Warde (2005) have suggested a mixed randomized response model using simple random sampling which rectifies the privacy problem.

In this paper we have suggested an alternative to Kim and Warde's (2005) mixed randomized response model and its properties are studied in simple random sampling with replacement (SRSWR) and Stratified random sampling in both the cases of completely truthful reporting and less than completely truthful reporting. Numerically we show that the proposed mixed randomized response model is better than Kim and Warde's (2005) estimator.

2. The suggested model

Let a random sample of size n be selected using simple random sampling with replacement (SRSWR). Each respondent from the sample is instructed to answer the direct question "I am a member of the innocuous group". If a respondent answers "Yes" to the direct question, then she or he is instructed to go to randomization device R_1 consisting of the statements (i) "I am the member of the sensitive trait group" and (ii) "I am a member of the innocuous trait group" with respective probabilities P_1 and $(1-P_1)$. If a respondent answers "No" to the direct question, then the respondent is instructed to use the randomization device R_2 consisting of the statements: (i) I belong to the sensitive group, (ii) "Yes" and (iii) "No" with known probabilities P, $(1-P)/2$ and $(1-P)/2$ respectively. For the second and third statements, the respondent is simply to report "Yes" or "No" as observed on the random device R_2 and it has no relevance to his actual status. It is to be mentioned that the randomization device R_2 is due to Tracy and Osahan (1999). The survey procedures are performed under the assumption that both the sensitive and innocuous questions are unrelated and independent in a randomization device R_1. To protect the respondent's privacy, the respondents should not disclose to the interviewer the question they answered from either R_1 or R_2.

Let n be the sample size confronted with a direct question and n_1 and n_2 $(=n-n_1)$ denote the number of "Yes" and "No" answers from the sample. Note that the respondents coming to R_1 have reported a "Yes" to the initial direct question, therefore $\pi_1 = 1$ in R_1, where π_1 is the proportion of "Yes" answers from the innocuous question.

Denote by 'Y' the probability of "Yes" from the respondents using R_1. Then

$$Y = P_1\pi_S + (1-P_1)\pi_1 = P_1\pi_S + (1-P_1), \tag{2.1}$$

where π_S is the proportion of "Yes" answers from the sensitive trait.

An unbiased estimator of π_S, in terms of the sample proportion of "Yes" responses $\hat{Y}$, becomes

$$\hat{\pi}_{a1} = \frac{\hat{Y}-(1-P_1)}{P_1}. \tag{2.2}$$

The variance of $\hat{\pi}_{a1}$ is

$$\begin{aligned} V(\hat{\pi}_{a1}) = \frac{Y(1-Y)}{n_1P_1^2} &= \frac{(1-\pi_S)\left[P_1\pi_S+(1-P_1)\right]}{n_1P_1} \\ &= \frac{1}{n_1}\left[\pi_S(1-\pi_S)+\frac{(1-\pi_S)(1-P_1)}{P_1}\right]. \end{aligned} \tag{2.3}$$

The proportion of "Yes" answers from the respondents using randomization device R_2 follows:

$$X = P\pi_S + \frac{(1-P)}{2} \tag{2.4}$$

An unbiased estimator of π_S, in terms of the sample proportion of "Yes" responses $\hat{X}$, becomes

$$\hat{\pi}_{b1} = \frac{\hat{X}-(1-P)/2}{P}. \tag{2.5}$$

The variance of $\hat{\pi}_{b1}$ is given by

$$V(\hat{\pi}_{b1}) = \frac{X(1-X)}{n_2P^2} = \left[\frac{\pi_S(1-\pi_S)}{n_2}+\frac{(1-P^2)}{4n_2P^2}\right]. \tag{2.6}$$

The estimator of π_S, in terms of the sample proportions of "Yes" responses $\hat{Y}$ and $\hat{X}$, is

$$\begin{aligned} \hat{\pi}_t &= \frac{n_1}{n}\hat{\pi}_{a1}+\frac{n_2}{n}\hat{\pi}_{b1} \\ &= \frac{n_1}{n}\hat{\pi}_{a1}+\frac{(n-n_1)}{n}\hat{\pi}_{b1}, \quad \text{for } 0 < \frac{n_1}{n} < 1. \end{aligned} \tag{2.7}$$

As both $\hat{\pi}_{a1}$ and $\hat{\pi}_{b1}$ are unbiased estimators, the expected value of $\hat{\pi}_t$ is

$$E(\hat{\pi}_t) = E\left[\frac{n_1}{n}\hat{\pi}_{a1} + \frac{n_2}{n}\hat{\pi}_{b1}\right] = \frac{n_1}{n}\pi_S + \frac{(n-n_1)}{n}\pi_S = \pi_S.$$

Thus the proposed estimator $\hat{\pi}_t$ is an unbiased estimator π_S.

Now the variance of $\hat{\pi}_t$ is given by

$$\begin{aligned}
V(\hat{\pi}_t) &= \left(\frac{n_1}{n}\right)^2 V(\hat{\pi}_{a1}) + \left(\frac{n_2}{n}\right)^2 V(\hat{\pi}_{b1}) \\
&= \left(\frac{n_1}{n}\right)^2 \frac{1}{n_1}\left[\pi_S(1-\pi_S) + \frac{(1-\pi_S)(1-P_1)}{P_1}\right] \\
&+ \left(\frac{n_2}{n}\right)^2 \frac{1}{n_2}\left[\pi_S(1-\pi_S) + \frac{(1-P^2)}{4P^2}\right] \\
&= \frac{n_1}{n^2}\left[\pi_S(1-\pi_S) + \frac{(1-\pi_S)(1-P_1)}{P_1}\right] + \frac{n_2}{n^2}\left[\pi_S(1-\pi_S) + \frac{(1-P^2)}{4P^2}\right]. \quad (2.8)
\end{aligned}$$

Since our mixed RR model also uses Simmon's (1967) method when $\pi_1 = 1$, we can apply Lanke's (1976) idea to our suggested model. Thus using Lanke's (1976) result for P with $\pi_1 = 1$, we get

$$P = \frac{1}{2-P_1}. \quad (2.9)$$

Putting $P = (2-P_1)^{-1}$ in (2.6), we get

$$\begin{aligned}
V(\hat{\pi}_{b1}) &= \frac{\pi_S(1-\pi_S)}{(n-n_1)} + \frac{(1-P_1)\,(3-P_1)}{4(n-n_1)} \\
&= \frac{1}{(n-n_1)}\left[\pi_S(1-\pi_S) + \frac{(1-P_1)(3-P_1)}{4}\right]. \quad (2.10)
\end{aligned}$$

Thus we established the following theorem.

Theorem 2.1 *The variance of $\hat{\pi}_t$ is given by*

$$V(\hat{\pi}_t) = \frac{\pi_S(1-\pi_S)}{n} + \frac{(1-P_1)\left[4\,\lambda(1-\pi_S) + (1-\lambda)P_1(3-P_1)\right]}{4nP_1} \quad (2.11)$$

for $n = n_1 + n_2$ and $\lambda = \dfrac{n_1}{n}$.

3. Efficiency comparisons

An efficiency comparison of the suggested model, under completely truthful reporting case, has been done with Kim and Warde's (2005) model.

From Kim and Warde's (2005) model, we have

$$V(\hat{\pi}_{kw}) = \frac{\pi_S(1-\pi_S)}{n} + \frac{(1-P_1)\left[\ \lambda P_1(1-\pi_S)+(1-\lambda)\right]}{nP_1^2}. \tag{2.12}$$

From (2.11) and (2.12) we have $V(\hat{\pi}_t) < V(\hat{\pi}_{kw})$ if

$$\frac{[4\lambda(1-\pi_S)+(1-\lambda)P_1(3-P_1)]}{4} < \frac{[\lambda P_1(1-\pi_S)+(1-\lambda)]}{P_1}$$

i.e. if $4-3P_1^2+P_1^3 > 0$ which is always true.

Thus the proposed model is always better than Kim and Warde's (2005) model.

An efficiency comparison of the proposed mixed randomized response technique to that of Kim and Warde's, we have computed the percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's estimator $\hat{\pi}_{kw}$ by using the formula:

$$\begin{aligned}\mathrm{PRE}(\hat{\pi}_t,\ \hat{\pi}_{kw}) &= \frac{V(\hat{\pi}_{kw})}{V(\hat{\pi}_t)} \times 100 \\ &= \frac{4\left[\pi_S(1-\pi_S)+\{(1-P_1)/P_1^2\}\{\lambda P_1(1-\pi_S)+(1-\lambda)\}\right]\pi_S(1-\pi_S)}{[4\pi_S(1-\pi_S)+\{(1-P_1)/P_1\}\{4\lambda(1-\pi_S)+(1-\lambda)P_1(3-P_1)\}]} \times 100\end{aligned}$$

for different values of P_1, n and n_1.

We have obtained the values of the percent relative efficiencies $\mathrm{PRE}(\hat{\pi}_t,\ \hat{\pi}_{kw})$ for $\lambda = 0.3, 0.5, 0.7$ and for different cases of π_S, n, n_1 and P_1. Findings are shown in Table 1 and its diagrammatic representation is given in Figure 1.

It is observed from Table 1 and Figure 1 that: The values of percent relative efficiencies $\mathrm{PRE}(\hat{\pi}_t,\ \hat{\pi}_{kw})$ are more than 100. We can say that the envisaged estimator $\hat{\pi}_t$ is always efficient than Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$. Figure 1 shows results for $\pi_S = 0.1$ and 0.6, $\lambda = 0.3, 0.5, 0.7$ and different values of P_1, n, n_1.

We note from Table 1 that the values of the percent relative efficiencies $\mathrm{PRE}(\hat{\pi}_t,\ \hat{\pi}_{kw})$ decrease as the value of P_1 increases. Also the values of the percent relative efficiencies $\mathrm{PRE}(\hat{\pi}_t,\ \hat{\pi}_{kw})$ increase as the value of λ decrease for fixed values of π_S and P_1.

We further note from the results of Figure 1 that there is large gain in efficiency by using the suggested estimator $\hat{\pi}_t$ over the estimator $\hat{\pi}_{kw}$ when the proportion of stigmatizing attribute is moderately large.

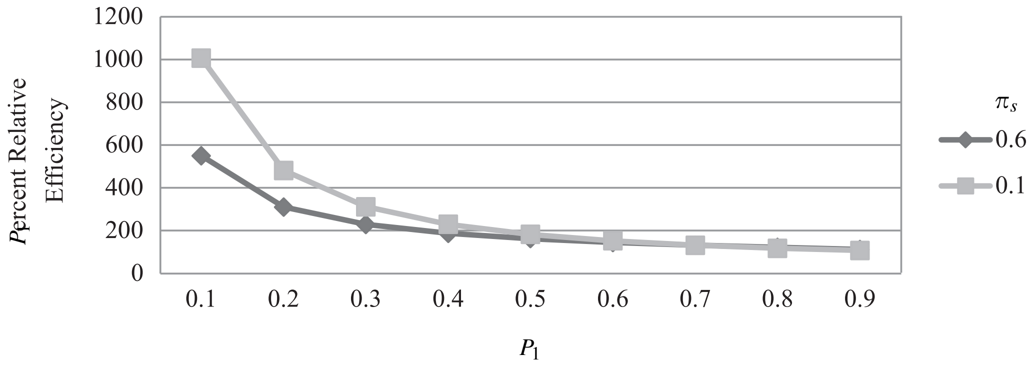

Figure 1: *Percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$.*

4. Less than completely truthful reporting

The problem of "Less than completely truthful reporting" in randomized response model has been tackled by several authors including Singh (1993), Mangat (1994 a,b), Tracy and Osahan (1999), Chang and Huang (2001), Kim and Warde (2004), Kim and Elam (2005), Nazuk and Shabbir (2010) and others. We write the proportion of "Yes" answers from the two randomization devices R_1 and R_2, incorporating the probability of truthful reporting. Let T_1 and T_2 be the probabilities of telling the truth regarding the stigmatizing question in the randomization device R_1 and R_2 respectively. The respondents in the innocuous trait have no reason to tell a lie, they may lie for the sensitive trait.

Note that the respondents coming to R_1 have reported a "Yes" to the initial direct question therefore $\pi_1 = 1$ in R_1. The probability of "Yes" answers from the respondents using R_1 is given by

$$Y^* = P_1 \pi_S T_1 + (1 - P_1). \tag{4.1}$$

An estimator for the true population proportion π_S of the sensitive trait is given by

$$\hat{\pi}_{a(1)} = \frac{\hat{Y}^* - (1 - P_1)}{P_1}, \tag{4.2}$$

where $\hat{Y}^*$ is the sample proportion of "Yes" response from the randomization device R_1.

Since $\hat{Y}^*$ follows Binomial distribution $\mathrm{B}(n_1, Y^*)$, therefore the bias and variance of the estimator $\hat{\pi}_{a(1)}$ are respectively given by

$$\mathrm{B}(\hat{\pi}_{a(1)}) = \pi_S (T_1 - 1) \tag{4.3}$$

Table 1: *Percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$.*

π_S	$n = 1000$		λ	P_1								
	n_1	n_2		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	700	300	0.7	550.05	309.93	228.78	187.50	162.21	144.90	132.03	121.65	112.11
0.1	500	500	0.5	1100.25	547.92	365.40	275.00	221.18	185.40	159.56	139.32	121.34
0.1	300	700	0.3	2201.05	971.19	586.67	406.25	304.10	239.21	194.28	160.44	131.68
0.2	700	300	0.7	596.76	327.10	236.25	190.24	162.21	143.17	129.20	118.24	108.94
0.2	500	500	0.5	1193.27	576.47	374.72	275.91	217.80	179.67	152.64	132.14	115.37
0.2	300	700	0.3	2352.39	1000.00	586.67	396.59	290.90	224.95	180.24	147.72	122.23
0.3	700	300	0.7	656.59	349.44	246.58	194.91	163.77	142.93	127.96	116.62	107.59
0.3	500	500	0.5	1311.90	614.28	389.25	280.64	217.80	177.31	149.27	128.72	112.89
0.3	300	700	0.3	2546.20	1043.68	596.11	394.73	284.93	217.71	173.16	141.75	118.39
0.4	700	300	0.7	735.81	379.31	260.93	202.12	167.16	144.14	127.96	116.07	107.03
0.4	500	500	0.5	1467.73	665.11	410.53	289.83	221.18	177.77	148.34	127.36	111.84
0.4	300	700	0.3	2799.61	1106.50	616.14	400.38	284.93	215.48	170.29	139.15	116.73
0.5	700	300	0.7	845.44	420.79	281.35	213.00	172.97	147.05	129.20	116.39	106.98
0.5	500	500	0.5	1680.67	735.29	441.50	304.87	228.57	181.15	149.58	127.55	111.66
0.5	300	700	0.3	3140.87	1195.65	649.41	414.43	290.90	217.71	170.75	138.88	116.37
0.6	700	300	0.7	1006.85	481.67	311.73	229.72	182.48	152.40	132.03	117.71	107.42
0.6	500	500	0.5	1987.95	836.36	487.37	328.57	241.50	188.28	153.37	129.34	112.29
0.6	300	700	0.3	3620.09	1323.74	701.08	439.39	304.10	224.95	174.68	140.85	117.12

and

$$V(\hat{\pi}_{a(1)}) = \frac{Y^*(1-Y^*)}{n_1 P_1^2} = \frac{(1-\pi_S T_1)[1-P_1(1-\pi_S T_1)]}{n_1 P_1}. \tag{4.4}$$

So the mean square error (MSE) of the estimator $\hat{\pi}_{a(1)}$ is given by

$$\mathrm{MSE}(\hat{\pi}_{a(1)}) = \left\{\frac{(1-\pi_S T_1)[1-P_1(1-\pi_S T_1)]}{n_1 P_1} + \pi_S^2(T_1-1)^2\right\}. \tag{4.5}$$

The proportion of "Yes" answers from the respondents using randomization device R_2 is

$$X^* = P\pi_S T_2 + \frac{(1-P)}{2}. \tag{4.6}$$

Thus an estimator of π_S is given by

$$\hat{\pi}_{b(1)} = \frac{\hat{X}^* - (1-P)/2}{P}, \tag{4.7}$$

where $\hat{X}^*$ is the sample proportion of "Yes" responses from the randomization device R_2.

Since $\hat{X}^*$ follows Binomial distribution $\mathrm{B}(n_1, X^*)$, therefore the bias and variance of the estimator $\hat{\pi}_{b(1)}$ are respectively given by

$$\mathrm{B}(\hat{\pi}_{b(1)}) = \pi_S(T_2-1) \tag{4.8}$$

and

$$V(\hat{\pi}_{b(1)}) = \frac{X^*(1-X^*)}{n_2 P^2} = \frac{[1-P^2(1-2\pi_S T_2)^2]}{4 n_2 P^2}, \tag{4.9}$$

where $n_1 + n_2 = n$.

Thus the mean square error (MSE) of the estimator $\hat{\pi}_{b(1)}$ is given by

$$\mathrm{MSE}(\hat{\pi}_{b(1)}) = \left\{\frac{[1-P^2(1-2\pi_S T_2)^2]}{4 n_2 P^2} + \pi_S^2(T_2-1)^2\right\}. \tag{4.10}$$

Now we propose the weighted estimator of π_S as

$$\hat{\pi}_t^* = \left[\left(\frac{n_1}{n}\right)\hat{\pi}_{a(1)} + \left(\frac{n_2}{n}\right)\hat{\pi}_{b(1)}\right]. \tag{4.11}$$

Since the two randomization devices are independent, we can derive the bias and MSE of $\hat{\pi}_t^*$ respectively as

$$\mathrm{B}\,(\hat{\pi}_t^*) = \pi_S \left[\left(\frac{n_1}{n} \right) (T_1 - 1) + \left(\frac{n - n_1}{n} \right) (T_2 - 1) \right] \tag{4.12}$$

and

$$\mathrm{MSE}\,(\hat{\pi}_t^*) = \left\{ \frac{\lambda(1 - \pi_S T_1)[1 - P_1(1 - \pi_S T_1)]}{nP_1} + \frac{(1-\lambda)[1 - P^2(1 - 2\pi_S T_2)^2]}{4nP^2} \right.$$
$$\left. + \pi_S^2[\lambda(T_1 - 1) + (1-\lambda)(T_2 - 1)]^2 \right\}. \tag{4.13}$$

Putting $P = (2 - P_1)^{-1}$ [see Lanke (1976)] in (4.13), we get the MSE of $\hat{\pi}_t^*$ as

$$\mathrm{MSE}\,(\hat{\pi}_t^*) = \left\{ \frac{\pi_S \{\lambda T_1 (1 - \pi_S T_1) + (1-\lambda) T_2 (1 - \pi_S T_2)\}}{n} \right.$$
$$+ \frac{(1 - P_1)\,[\,4\,\lambda(1 - \pi_S T_1) + (1-\lambda)P_1(3 - P_1)]}{4nP_1}$$
$$\left. + \pi_S^2[\lambda(T_1 - 1) + (1-\lambda)(T_2 - 1)]^2 \right\}. \tag{4.14}$$

Proceeding as above in a situation of "Less than completely truthful reporting" one can easily derive the following bias and MSE of Kim and Warde's estimator $\hat{\pi}_{kw}^*$ (say):

$$B(\hat{\pi}_{kw}^*) = \pi_S \left[\left(\frac{n_1}{n} \right) (T_1 - 1) + \left(\frac{n - n_1}{n} \right) (T_2 - 1) \right] \tag{4.15}$$

$$\mathrm{MSE}\,(\hat{\pi}_{kw}^*) = \left\{ \frac{\pi_S \{\lambda T_1 (1 - \pi_S T_1) + (1-\lambda) T_2 (1 - \pi_S T_2)\}}{n} \right.$$
$$+ \frac{(1 - P_1)\,[\,\lambda P_1(1 - \pi_S T_1) + (1-\lambda)]}{nP_1^2}$$
$$\left. + \pi_S^2[\lambda(T_1 - 1) + (1-\lambda)(T_2 - 1)]^2 \right\}. \tag{4.16}$$

From (4.14) and (4.16) we have

$$MSE(\hat{\pi}_{kw}^*) - MSE(\hat{\pi}_t^*) = \frac{(1 - P_1)\,(1-\lambda)(4 - 3P_1^2 + P_1^3)}{4nP_1^2} \tag{4.17}$$

which is always positive.

Table 2: *Percent relative efficiency of the proposed estimator $\hat{\pi}_t^*$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{kw}^*$.*

π_S	$n = 1000$					P_1								
	n_1	n_1	T_1	T_2	λ	0.35	0.4	0.45	0.50	0.55	0.60	0.65	0.7	0.75
0.1	700	300	0.6	0.5	0.2	228.01	191.20	166.08	148.36	135.54	126.09	119.02	113.67	108.05
0.1	500	500	0.7	0.6	0.3	251.84	210.96	182.19	161.33	145.87	134.21	125.31	118.45	103.20
0.1	300	700	0.8	0.7	0.4	273.47	231.72	201.04	177.88	160.05	146.09	135.05	126.23	101.48
0.1	700	300	0.9	0.8	0.5	278.84	242.34	214.36	192.31	174.55	159.98	147.86	137.64	102.29
0.2	500	500	0.6	0.5	0.2	139.08	127.19	119.30	113.86	110.02	107.24	105.20	103.69	108.05
0.2	300	700	0.7	0.6	0.3	153.60	137.65	126.93	119.48	114.16	110.29	107.42	105.29	103.20
0.2	700	300	0.8	0.7	0.4	178.79	156.50	141.12	130.19	122.23	116.34	111.92	108.57	101.48
0.2	500	500	0.9	0.8	0.5	220.81	190.76	168.86	152.49	140.01	130.36	122.81	116.87	102.29
0.3	300	700	0.6	0.5	0.2	118.12	112.54	108.86	106.34	104.56	103.29	102.35	101.66	108.05
0.3	700	300	0.7	0.6	0.3	125.82	117.95	112.72	109.13	106.59	104.76	103.42	102.42	103.20
0.3	500	500	0.8	0.7	0.4	141.32	129.02	120.75	115.00	110.90	107.91	105.71	104.06	101.48
0.3	300	700	0.9	0.8	0.5	178.43	156.81	141.70	130.83	122.85	116.89	112.38	108.94	102.29
0.4	700	300	0.6	0.5	0.2	110.35	107.15	105.04	103.60	102.59	101.86	101.33	100.94	108.05
0.4	500	500	0.7	0.6	0.3	114.97	110.36	107.32	105.24	103.77	102.72	101.95	101.38	103.20
0.4	300	700	0.8	0.7	0.4	124.81	117.28	112.27	108.81	106.37	104.60	103.30	102.34	101.48
0.4	700	300	0.9	0.8	0.5	152.61	137.34	126.93	119.62	114.34	110.47	107.59	105.43	102.29
0.5	500	500	0.6	0.5	0.2	106.67	104.61	103.24	102.32	101.66	101.20	100.85	100.60	108.05
0.5	300	700	0.7	0.6	0.3	109.72	106.71	104.74	103.38	102.43	101.75	101.25	100.88	103.20
0.5	700	300	0.8	0.7	0.4	116.40	111.37	108.04	105.76	104.15	102.99	102.14	101.52	101.48
0.5	500	500	0.9	0.8	0.5	136.96	125.94	118.53	113.39	109.72	107.05	105.08	103.62	102.29

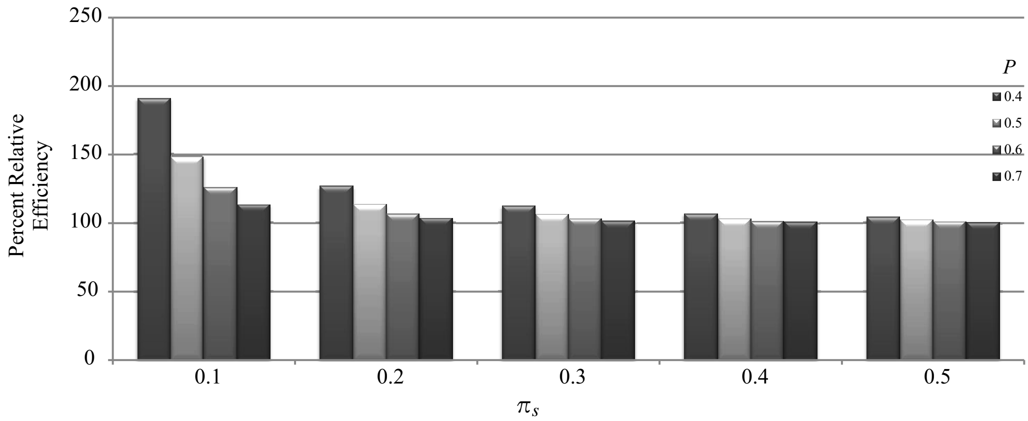

Figure 2: *Percent relative efficiency of the proposed estimator $\hat{\pi}_t^*$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{kw}^*$.*

Thus in the situation of "Less than completely truthful reporting" the proposed estimator $\hat{\pi}_t^*$ is more efficient than Kim and Warde's estimator $\hat{\pi}_{kw}^*$.

To have tangible idea about the performance of the proposed estimator $\hat{\pi}_t^*$ compared to estimator $\hat{\pi}_{kw}^*$, we have computed the percent relative efficiency of the proposed estimator $\hat{\pi}_t^*$ with respect to $\hat{\pi}_{kw}^*$ by using the formula:

$$\mathrm{PRE}(\hat{\pi}_t^*,\ \hat{\pi}_{kw}^*) = \frac{MSE(\hat{\pi}_{kw}^*)}{MSE(\hat{\pi}_t^*)} \times 100$$

We have obtained the values of the percent relative efficiencies $\mathrm{PRE}(\hat{\pi}_t^*,\ \hat{\pi}_{kw}^*)$ for $\lambda = 0.2, 0.3, 0.4, 0.5$, $n = 1000$ and for different cases of π_S, T_1, T_2 and P_1. Findings are shown in Table 2 and its diagrammatic representation is also demonstrated in Figure 2.

It is observed from Table 2 that the values of percent relative efficiencies PRE $(\hat{\pi}_t^*,\ \hat{\pi}_{kw}^*)$ are more than 100. We can say that the proposed estimator $\hat{\pi}_t^*$ is more efficient than Kim and Warde's estimator $\hat{\pi}_{kw}^*$. Figure 2 shows results for $\pi_S = 0.1, 0.2, 0.3, 0.4, 0.5$ and $P = 0.4, 0.5, 0.6, 0.7$, for $T_1 = 0.6$, $T_2 = 0.5$, and $n = 1000$.

Table 2 conceals that the values of the percent relative efficiency of the proposed estimator $\hat{\pi}_t^*$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{kw}^*$ decrease as the value of P_1 increases. Higher gain in efficiency is seen when the sample size n and π_S are small. However, the percent relative efficiency is more than 100 for all parametric values considered here; therefore the proposed estimator $\hat{\pi}_t^*$ is better than Kim and Warde's estimator $\hat{\pi}_{kw}^*$.

5. An alternative mixed randomized response model using stratification

5.1. An alternative to Kim and Warde's (2005) mixed stratified randomized response model

Stratified random sampling is usually obtained by partitioning the population into non-overlapping groups called strata and selecting a simple random sample from each stratum. A randomized response (RR) technique using a stratified random sampling yields the group characteristics associated to each stratum estimator. We also note that stratified sampling protects a researcher from the possibility of obtaining a poor sample. Hong et al. (1994) suggested a stratified RR technique using a proportional allocation. Kim and Warde (2004) suggested a stratified Warner's RR model using an optimal allocation which is more efficient than that using a proportional allocation. Kim and Elam (2005) have applied Kim and Warde's (2004) stratified Warner's RR model to Mangat and Singh's (1990) two-stage RR model. Further Kim and Elam (2007) have given a RR model that combines Kim and Warde's (2004) stratified Warner's RR technique using optimal allocation with the unrelated question randomized response model. Kim and Warde (2005) have suggested a mixed stratified RR model.

In the proposed model, the population is partitioned into strata, and a sample is selected by simple random sampling with replacement in each stratum. To get the full benefit from stratification, we assume that the number of units in each stratum is known. An individual respondent in a sample from each stratum is instructed to answer a direct question "I am a member of the innocuous trait group". Respondents reply the direct question by "Yes" or "No". If a respondent answers "Yes", then she or he is instructed to go to the randomization device R_{j1} consisting of the statements: (i) "I belong to the sensitive trait group" and (ii) "I belong to the innocuous trait group" with pre-assigned probabilities Q_j and $(1-Q_j)$, respectively. If a respondent answers "No", then the respondent is instructed to use the randomization device R_{j2} uses three statements: (i) " I belong to the stigmatizing group", (ii) " Yes" and (iii) "No" with known probabilities P_j, $(1-P_j)/2$ and $(1-P_j)/2$, respectively. For the second and third statements, the respondent is simply to report "Yes" or "No" as observed on the randomization device R_{j2}, and it has no relevance to his actual status. Let m_j denote the number of units in the sample from stratum j and n as the total number of units in samples from all strata. Let m_{j1} be the number of people answering "Yes" when respondents in a sample m_j were asked the direct question and m_{j2} be the number of people answering "No' when respondents in a sample m_j were asked the direct question so that $n = \sum_{j=1}^{L} m_j = \sum_{j=1}^{L} (m_{j1} + m_{j2})$. Under the supposition that these "Yes" or "No" reports are made truthfully, and Q_j and P_j are set by the researcher, then the proportion of "Yes" answers from the respondents using the randomization device R_{j1} will be

$$Y_j = Q_j \pi_{Sj} + (1-Q_j)\pi_{1_j} \quad \text{for } j = 1,2,\ldots,L, \tag{5.1}$$

where Y_j the probability of "Yes" response in stratum j, π_{Sj} is the proportion of respondents with the sensitive traits in stratum j, π_{1j} is the proportion of respondents with the innocuous trait in stratum j, and Q_j is the probability that a respondent in the sample stratum j is asked a sensitive question.

Since the respondent performing a randomization device R_{j1} answered "Yes" to the direct question of the innocuous trait, if he or she selects the same innocuous question from R_{j1}, then $\pi_{1j} = 1$, see Kim and Warde (2005, p. 217). Thus (5.1) reduces to

$$Y_j = Q_j \pi_{S_j} + (1-Q_j) \quad \text{for } j = 1,2,\ldots,L. \tag{5.2}$$

An unbiased estimator of π_{Sj} is given by

$$\hat{\pi}_{a_j} = \frac{\hat{Y}_j - (1-Q_j)}{Q_j} \quad \text{for } j = 1,2,\ldots,L, \tag{5.3}$$

where $\hat{Y}_j$ is the proportion of "Yes" answers in a sample in stratum j and $\hat{\pi}_{a_j}$ is the proportion of respondents with the sensitive trait in a sample from stratum j. The variance of $\hat{\pi}_{aj}$ is given by

$$V(\hat{\pi}_{a_j}) = \frac{(1-\pi_{Sj})\,[1-Q_j(1-\pi_{S_j})]}{m_{j1}Q_j} \quad \text{for } j = 1,2,\ldots,L. \tag{5.4}$$

The proportion of "Yes" responses from the respondents using randomization device R_{j2} will be

$$X_j = P_j\,\pi_{S_j} + (1-P_j)/2 \quad \text{for } j = 1,2,\ldots,L, \tag{5.5}$$

where X_j is the probability of "Yes" responses in stratum j. Thus an unbiased estimator of π_{Sj} is given by

$$\hat{\pi}_{b_j} = \frac{\hat{X}_j - (1-P_j)/2}{P_j} \quad \text{for } j = 1,2,\ldots,L, \tag{5.6}$$

where $\hat{X}_j$ is the proportion of "Yes" responses in a sample from a stratum j and $\hat{\pi}_{bj}$ is the proportion of respondents with the sensitive trait in a sample from stratum j.

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

$$V(\hat{\pi}_{b_j}) = \frac{\pi_{S_j}(1-\pi_{S_j})}{(m_j - m_{j1})} + \frac{(1-P_j^2)}{4(m_j - m_{j1})P_j^2} \quad \text{for } j = 1,2,\ldots,L. \tag{5.7}$$

Putting $P_j = (2-Q_j)^{-1}$ [see Lanke (1976)] for $j = 1,2,\ldots,L$ in (5.7) we get

$$V(\hat{\pi}_{b_j}) = \frac{\pi_{S_j}(1-\pi_{S_j})}{(m_j - m_{j1})} + \frac{(1-Q_j)(3-Q_j)}{4(m_j - m_{j1})} \quad \text{for } j = 1,2,\ldots,L. \tag{5.8}$$

Now we develop the unbiased estimator of π_{Sj}, in terms of sample proportion of "Yes" responses $\hat{Y}_j$ and $\hat{X}_j$,

$$\hat{\pi}_{mS_j} = \frac{m_{j1}}{m_j}\hat{\pi}_{a_j} + \frac{m_j - m_{j1}}{m_j}\hat{\pi}_{b_j} \quad \text{for } 0 < \frac{m_{j1}}{m_j} < 1. \tag{5.9}$$

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

$$V(\hat{\pi}_{mS_j}) = \frac{\pi_{S_j}(1-\pi_{S_j})}{m_j} + \frac{(1-Q_j)[4\lambda_j(1-\pi_{Sj}) + (1-\lambda_j)Q_j(3-Q_j)]}{4m_jQ_j}, \tag{5.10}$$

where $m_j = m_{j1} + m_{j2}$ and $\lambda_j = m_{j1}/m_j$.

The unbiased estimator of $\pi_S = \sum_{j=1}^{L} w_j \pi_{Sj}$ is given by

$$\hat{\pi}_S = \sum_{j=1}^{L} w_j \hat{\pi}_{mS_j} = \sum_{j=1}^{L} w_j \left\{ \frac{m_{j1}}{m_j}\hat{\pi}_{aj} + \frac{m_j - m_{j1}}{m_j}\hat{\pi}_{b_j} \right\} \tag{5.11}$$

where N is the number of units in the whole population, N_j is the total number of units in stratum j, and $w_j = N_j/N$ for $j = 1,2,\ldots,L$, so that $w = \sum_{j=1}^{L} w_j = 1$.

The variance of the estimator $\hat{\pi}_S$ is given by

$$V(\hat{\pi}_S) = \sum_{j=1}^{L} \frac{w_j^2}{m_j} \left\{ \pi_{S_j}(1-\pi_{S_j}) + \frac{(1-Q_j)[4\lambda_j(1-\pi_{Sj}) + (1-\lambda_j)Q_j(3-Q_j)]}{4Q_j} \right\} \tag{5.12}$$

Here, the requirement of doing the optimal allocation of a sample size n, is to know $\lambda_j = m_{j1}/m_j$ and π_{Sj}. In practice it is difficult to have information on $\lambda_j = m_{j1}/m_j$ and π_{Sj}. However if prior information about $\lambda_j = m_{j1}/m_j$ and π_{Sj} is available from past experience, it assists to derive the following optimal allocation formula.

Theorem 5.1 *The optimal allocation of n to n_1, $n_{2,\ldots}$, n_{L-1} and n_L to derive the minimum variance of the $\hat{\pi}_S$ subject to $n = \sum_{j=1}^{L} m_j$ is approximately given by*

$$\frac{m_j}{n} = \frac{w_j \left\{\pi_{S_j}(1-\pi_{Sj}) + \frac{(1-Q_j)[4\lambda_j(1-\pi_{Sj})+(1-\lambda_j)Q_j(3-Q_j)]}{4Q_j}\right\}^{1/2}}{\sum_{j=1}^{L} w_j \left\{\pi_{S_j}(1-\pi_{Sj}) + \frac{(1-Q_j)[4\lambda_j(1-\pi_{Sj})+(1-\lambda_j)Q_j(3-Q_j)]}{4Q_j}\right\}^{1/2}}, \tag{5.13}$$

where $m_j = m_{j1} + m_{j2}$ and $\lambda_j = m_{j1}/m_j$.

Thus the minimal variance of the estimator $\hat{\pi}_S$ is given by

$$V(\hat{\pi}_S) = \frac{1}{n}\left\{\sum_{j=1}^{L} w_j \left[\pi_{S_j}(1-\pi_{S_j}) + \frac{(1-Q_j)[4\lambda_j(1-\pi_{Sj})+(1-\lambda_j)Q_j(3-Q_j)]}{4Q_j}\right]^{1/2}\right\}^2, \tag{5.14}$$

where $n = \sum_{j=1}^{L} m_j$, $m_j = m_{j1} + m_{j2}$ and $\lambda_j = m_{j1}/m_j$.

5.2. Efficiency comparison

In this section we have made the comparison of proposed estimator $\hat{\pi}_S$ with the proposed mixed randomized estimator $\hat{\pi}_t$, Kim and Warde's (2005) mixed randomized response estimator $\hat{\pi}_m$ and Kim and Warde's (2005) stratified mixed randomized response estimator $\hat{\pi}_{mS}$. The comparisons are given in the form of following theorems.

Theorem 5.2 *Assume that there are two strata in the population (i.e. $L = 2$) and $\lambda_j = m_{j1}/m_j$. The proposed estimator $\hat{\pi}_S$ of a stratified mixed RR is more efficient than the estimator $\hat{\pi}_t$ of a mixed model, where $P_1 = Q_1 = Q_2$ and $\lambda = \lambda_1 = \lambda_2$.*

Proof. We denote by

$$a_1 = \pi_{S1}(1-\pi_{S1}), \qquad a_2 = \pi_{S2}(1-\pi_{S2}),$$

$$b_1 = \frac{\lambda(1-P_1)(1-\pi_{S1})}{P_1}, \quad b_2 = \frac{\lambda(1-P_1)(1-\pi_{S2})}{P_1}, \quad c = \frac{(1-\lambda)(1-P_1)(3-P_1)}{4}$$

Then for $L = 2$, $P_1 = Q_1 = Q_2$, $\lambda = \lambda_1 = \lambda_2$ and from (2.11) and (5.14) we have

$$V(\hat{\pi}_t) = \frac{1}{n}\left\{w_1\pi_{S1} + w_2\pi_{S2})(1 - w_1\pi_{S1} - w_2\pi_{S2}) + \frac{\lambda(1-P_1)(1-w_1\pi_{S1}-w_2\pi_{S2})}{P_1} + c\right\} \tag{5.15}$$

and

$$V(\hat{\pi}_S) = \frac{1}{n}\{w_1(a_1+b_1+c)^{1/2} + w_2(a_2+b_1+c)^{1/2}\}^2 \tag{5.16}$$

Now subtracting (5.16) from (5.15) we have

$$n[V(\hat{\pi}_t) - V(\hat{\pi}_S) = \ w_1 w_2 \ \left\{(\pi_{S1} - \pi_{S2})^2 + \left(\sqrt{(a_1+b_1+c)} - \sqrt{(a_2+b_2+c)}\right)^2\right\}$$

which is always positive.

Thus the proposed estimator $\hat{\pi}_S$ of stratified mixed RR is more efficient than the proposed estimator $\hat{\pi}_t$ (with $L = 2$) of a mixed model.

This proves Theorem 5.2.

Theorem 5.3 *Suppose there are two strata in the population and $\lambda_j = m_{j1}/m_j$. The proposed estimator $\hat{\pi}_S$ of a stratified mixed RR is more efficient than Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$ of a mixed model, where $P_1 = Q_1 = Q_2$ and $\lambda = \lambda_1 = \lambda_2$.*

Proof. For $L = 2$, $\pi_S = w_1\pi_{S1} + w_2\pi_{S2}$, $P_1 = Q_1 = Q_2$, $\lambda = \lambda_1 = \lambda_2$ and from Kim and Warde (2005, Eq (2.10), p. 213) we have

$$\begin{aligned} V(\hat{\pi}_m) = \frac{1}{n}\{& w_1\pi_{S1} + w_2\pi_{S2})(1 - w_1\pi_{S1} - w_2\pi_{S2}) \\ & + \frac{\lambda(1-P_1)(1-w_1\pi_{S1} - w_2\pi_{S2})}{P_1} + \frac{(1-\lambda)(1-P_1)}{P_1^2}\bigg\}. \end{aligned} \tag{5.17}$$

From (5.16) and (5.17) we have

$$\begin{aligned} n[V(\hat{\pi}_m) - V(\hat{\pi}_S) = \Bigg[& w_1 w_2 \left\{(\pi_{S1} - \pi_{S2})^2 + \left(\sqrt{(a_1+b_1+c)} - \sqrt{(a_2+b_2+c)}\right)^2\right\} \\ & + \frac{(1-\lambda)(1-P_1)(4-3P_1^2+P_1^3)}{4P_1^2}\Bigg] \end{aligned} \tag{5.18}$$

which is always positive.

Thus the proposed estimator $\hat{\pi}_S$ of a stratified mixed RR is more efficient than Kim and Warde's estimator $\hat{\pi}_m$ of a mixed model.

This proves the theorem.

Theorem 5.4 *Assume that there are two strata in the population (i.e. $L = 2$) and $\lambda_j = m_{j1}/m_j$. The proposed estimator $\hat{\pi}_S$ of a stratified mixed RR is more efficient than Kim and Warde's (2005) estimator $\hat{\pi}_{mS}$, $P_1 = Q_1 = Q_2$ and $\lambda = \lambda_1 = \lambda_2$.*

Proof. For $L = 2, P_1 = Q_1 = Q_2, \lambda = \lambda_1 = \lambda_2$ and from Kim and Warde (2005, Eq (4.12), p. 218) we have

$$V(\hat{\pi}_{mS}) = \frac{1}{n}\left\{w_1(a_1+b_1+c_1)^{1/2} + w_2(a_2+b_2+c_1)^{1/2}\right\}, \tag{5.19}$$

where $c_1 = (1-\lambda)(1-P_1)/P_1^2$.

From (5.16) and (5.19) we have

$$n[V(\hat{\pi}_{mS}) - V(\hat{\pi}_S) = (c_1 - c)\left[(w_1^2+w_2^2) + \frac{2w_1w_2(A_1+A_2^*)}{\sqrt{A_1A_2}+\sqrt{A_1^*A_2^*}}\right] \tag{5.20}$$

where

$$A_1 = (a_1+b_1+c_1),\ A_2 = (a_2+b_2+c_1), A_1^* = (a_1+b_1+c) \quad \text{and} \quad A_2^* = (a_2+b_2+c),$$

Since

$$(c_1 - c) = \frac{(1-\lambda)(1-P_1)(4-3P_1^2+P_1^3)}{4P_1^2} > 0,$$

therefore $n[V(\hat{\pi}_{mS}) - V(\hat{\pi}_S) > 0$.

It follows that the proposed estimator $\hat{\pi}_S$ of stratified mixed RR is more efficient than Kim and Warde's estimator $\hat{\pi}_{mS}$.

Thus the theorem 5.4 is proved.

If prior information on π_{S1}, π_{S2}, w_1, w_2, π_S and λ can be obtained and a researcher set Q_j, $j = 1,2$ then we can compute the percent relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect to Kim and Warde's estimator $\hat{\pi}_{mS}$ (for $L = 2$, $\lambda_1 = \lambda_2 = \lambda$) by using the formula:

$$\begin{aligned}\text{PRE}(\hat{\pi}_S, \hat{\pi}_{mS}) &= \frac{V(\hat{\pi}_{mS})}{V(\hat{\pi}_S)} \times 100 \\ &= \frac{\left(w_1\sqrt{B_1}+w_2\sqrt{B_2}\right)^2}{\left(w_1\sqrt{B_1^*}+w_2\sqrt{B_2^*}\right)^2} \times 100,\end{aligned}$$

where

$$B_1 = \left[\pi_{S_1}(1-\pi_{S1}) + \frac{(1-Q_1)[\lambda Q_1(1-\pi_{S1})+(1-\lambda)]}{Q_1^2}\right]$$

$$B_2 = \left[\pi_{S2}(1-\pi_{S2}) + \frac{(1-Q_2)[\lambda Q_2(1-\pi_{S2}) + (1-\lambda)]}{Q_2^2}\right]$$

$$B_1^* = \left[\pi_{S1}(1-\pi_{S_1}) + \frac{(1-Q_1)[4\lambda(1-\pi_{S1}) + (1-\lambda)Q_1(3-Q_1)]}{4Q_1}\right]$$

$$B_2^* = \left[\pi_{S2}(1-\pi_{S_2}) + \frac{(1-Q_2)[4\lambda(1-\pi_{S2}) + (1-\lambda)Q_2(3-Q_2)]}{4Q_2}\right]$$

We have computed $\mathrm{PRE}(\hat{\pi}_S, \hat{\pi}_{mS})$ for $n = 1000$, $\lambda = 0.2, 0.4, 0.6, 0.8$ and different values of w_1, w_2, Q_1, Q_2 , π_{S1} and π_{S2}. Findings are depicted in Table 3. Pictorial representation of $\mathrm{PRE}(\hat{\pi}_S, \hat{\pi}_{mS})$ is also given in Figure 3.

We have set eight different values of Q_j $(j = 1,2)$ and four different values of λ to verify the percent relative efficiency of the suggested estimator $\hat{\pi}_S$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{mS}$. Table 3 and Figure 3 show that the value of percent relative efficiency $\mathrm{PRE}(\hat{\pi}_S, \hat{\pi}_{mS})$ decreases as the values of Q_j $(j = 1,2)$ and λ increase.

The values of $\mathrm{PRE}(\hat{\pi}_S, \hat{\pi}_{mS})$ are greater than 100 for all values of π_{S1}, π_{S2}, w_1, w_2, Q_1, Q_2 and λ considered here. So we can say that the envisaged estimator $\hat{\pi}_S$ is more efficient than Kim and Warde's (2005) estimator $\hat{\pi}_{mS}$.

Figure 3 exhibits results from Tables 3 for $Q_1 = 0.1, 0.3, 0.5, 0.7$, $Q_2 = 0.2, 0.4, 0.6, 0.8$ and $\pi_S = 0.1, 0.2, 0.3, 0.4, 0.5$.

Remark 5.1. Proceeding as in Section 4 and the procedure adopted in Kim and Warde (2004) and Kim and Elam (2005, sec.4, p.4) the problem of "Less than completely truthful reporting" can be studied for the proposed mixed stratified RR model.

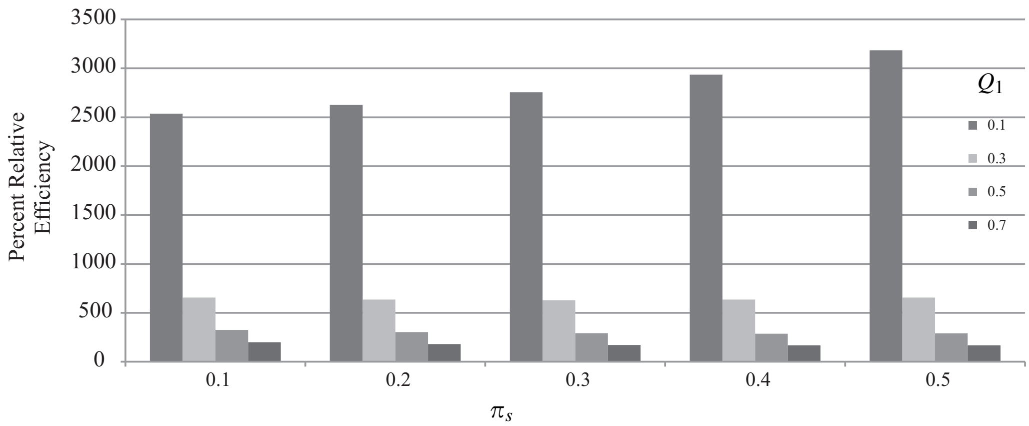

Figure 3: *Percent relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{mS}$.*

Table 3: *Percent relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{mS}$.*

π_S	π_{S1}	π_{S2}	λ	$Q_1=0.1$	$Q_1=0.2$	$Q_1=0.3$	$Q_1=0.4$	$Q_1=0.5$	$Q_1=0.6$	$Q_1=0.7$	$Q_1=0.8$
				$Q_2=0.2$	$Q_2=0.3$	$Q_2=0.4$	$Q_2=0.5$	$Q_2=0.6$	$Q_2=0.7$	$Q_2=0.8$	$Q_2=0.9$
0.1	0.08	0.13	0.2	2535.64	1106.56	655.86	444.83	326.11	251.15	199.36	159.99
0.1	0.08	0.13	0.4	1239.07	623.89	411.76	304.55	240.19	197.23	166.05	141.21
0.1	0.08	0.13	0.6	652.74	367.72	266.11	212.85	179.69	156.76	139.55	125.36
0.1	0.08	0.13	0.8	317.85	208.66	169.21	148.15	134.74	125.25	117.95	111.77
0.2	0.18	0.23	0.2	2626.11	1102.85	635.71	421.70	303.50	230.24	180.81	144.90
0.2	0.18	0.23	0.4	1322.53	644.81	414.52	299.86	232.04	187.53	155.97	131.97
0.2	0.18	0.23	0.6	702.79	385.05	272.40	213.78	177.60	152.88	134.67	120.29
0.2	0.18	0.23	0.8	339.78	217.44	173.24	149.66	134.70	124.17	116.19	109.68
0.3	0.28	0.33	0.2	2755.76	1117.51	629.28	410.36	291.63	219.37	171.69	138.19
0.3	0.28	0.33	0.4	1430.06	675.34	423.49	300.33	228.80	182.78	150.98	127.75
0.3	0.28	0.33	0.6	766.92	408.18	282.12	217.28	177.82	151.33	132.32	117.94
0.3	0.28	0.33	0.8	367.94	228.95	178.86	152.29	135.59	124.00	115.41	108.70
0.4	0.38	0.43	0.2	2935.97	1152.12	635.57	408.54	287.39	214.75	167.61	135.20
0.4	0.38	0.43	0.4	1571.60	718.58	439.77	306.03	229.77	181.61	149.06	125.93
0.4	0.38	0.43	0.6	851.54	439.50	296.41	223.84	180.38	151.75	131.68	116.99
0.4	0.38	0.43	0.8	405.31	244.45	186.75	156.40	137.54	124.67	115.36	108.35
0.5	0.48	0.53	0.2	3185.50	1211.03	655.53	415.93	289.83	215.08	167.11	134.53
0.5	0.48	0.53	0.4	1763.77	779.78	465.61	317.90	235.14	183.73	149.56	125.75
0.5	0.48	0.53	0.6	967.70	483.13	317.29	234.52	185.76	154.22	132.55	117.07
0.5	0.48	0.53	0.8	457.16	266.14	198.09	162.66	140.93	116.03	116.03	108.49

6. Discussion

In this article, we have proposed an alternative to Kim and Warde's (2005) mixed randomized response model to estimate the proportion of a qualitative sensitive characteristic under the conditions presented in both the cases of completely truthful reporting and less than completely truthful reporting by the respondents. We have also developed the proposed model to stratified sampling. It has been shown that the proposed mixed randomized response model is more efficient than Kim and Warde's (2005) mixed randomized response model.

7. Acknowledgements

The authors are grateful to the Editor-in-Chief and to the learned referee for their valuable suggestions regarding improvement of the paper.

References

Chang, H. J. and Huang, K. C. (2001). Estimation of proportion and sensitivity of a qualitative character. *Metrika*, 53, 269–280.

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response: Theory and Techniques*. Marcel-Dekker, New York, USA.

Cochran, W. G. (1977). *Sampling Technique*, 3rd Edition. New York: John Wiley and Sons, USA.

Fox, J. A. and Tracy, P. E. (1986). *Randomized Response: A Method of Sensitive Surveys*. Newbury Park, CA: SEGE Publications.

Greenberg, B., Abul-Ela, A., Simmons, W. R. and Horvitz, D. G. (1969). The unrelated question randomized response: Theoretical framework. *Journal of the American Statistical Association*, 64, 529–539.

Greenberg, B., Kubller, R. R., Abernathy, J. R. and Horvitz, D. G. (1971). Applications of the RR technique in obtaining quantitative data. *Journal of the American Statistical Association*, 66, 243–250.

Hong, K., Yum, J. and Lee, H. (1994). A stratified randomized response technique. *Korean Journal of Applied Statistics*, 7, 141–147.

Horvitz, D. G., Shah, B. V. and Simmons, W. R. (1967). The unrelated question randomized response model. *Proceedings of the Social Statistics Section, American Statistical Association*, 65–72.

Kim, J. M., Tebbs, J. M. and An, S. W. (2006). Extensions of Mangat's randomized response model. *Journal of Statistical Planning and Inference*, 136, 1554–1567.

Kim, J. M. and Warde, W. D. (2004). A stratified Warner randomized response model. *Journal of Statistical Planning and Inference*, 120, 155–165.

Kim, J. M. and Warde, W. D. (2005). A mixed randomized response model. *Journal of Statistical Planning and Inference*, 133, 211–221.

Kim, J. M. and Elam, M. E. (2005). A two-stage stratified Warner's randomized response model using optimal allocation. *Metrika*, 61, 1–7.

Kim, J. M. and Elam, M. E. (2007). A stratified unrelated randomized response model. *Statistical Papers*, 48, 215–233.

Lanke, J. (1976). On the degree of protection in randomized interview. *International Statistical Review* , 44, 80–83.

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

Mangat, N. S. (1994a). An improved randomized response strategy. *Journal of the Royal Statistical Society*, B, 56(1), 93–95.

Mangat, N. S. (1994b). An optional randomized response sampling technique. *Journal of the Indian Statistical Association*, 32(2), 71–75.

Mangat, N. S., Singh, R. and Singh, S. (1997). Violation of respondent's privacy in Moors model-its rectification through a random group strategy response model. *Communications in Statistics-Theory and Methods*, 26(3), 243–255.

Moors, J. A. (1971). Optimization of the unrelated question randomized response model. *Journal of the American Statistical Association*, 66, 627–629.

Nazuk, A. and Shabbir, J. (2010). A new mixed randomized response model. *International Journal of Business and Social Science*, 1, 186–190.

Ryu, J. B., Hong, K. H. and Lee, G. S. (1993). *Randomized Response Model*. Freedom. Academy, Seoul, Korea.

Singh, H. P. and Tarray, T. A. (2012). A stratified unknown repeated trials in randomized response sampling. *Communication of the Korean Statistical Society*, 19, (6), 751–759.

Singh, H. P. and Tarray, T. A. (2013a). Two-stage stratified partial randomized response strategies. Accepted in *Communication in Statistics Theory-Methods*.

Singh, H. P. and Tarray, T. A. (2013b). An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation. *Model Assisted Statistical Application-Special-Issue* (in press).

Singh, H. P. and Tarray, T. A. (2013c). An improved mixed randomized response model. *Model Assisted Statistical Application-Special-Issue* (in press).

Singh, R. and Mangat, N. S. (1996). *Elements of Survey Sampling*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Singh, S. (2003). *Advanced Sampling Theory with Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Sing, S., Singh, R. and Mangat, N. S. (2000). Some alternative strategies to Moor's model in randomized response model. *Journal of Statistical Planning and Inference*, 83, 243–255.

Tracy, D. S. and Mangat, N. S. (1996). Some developments in randomized response sampling during the last decade-A follow up of review by Chaudhuri and Mukherjee. *Journal of Applied Statistical Science*, 4(2/3), 147–158.

Tracy, D. S. and Osahan, S. S. (1999). An improved randomized response technique. *Pakistan Journal of Statistics*, 15(1), 1–6.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63–69.