Model Assisted Statistics and Applications 9 (2014) 73–87
DOI 10.3233/MAS-130279
IOS Press

An improved mixed randomized response model

Housila P. Singh* and Tanveer A. Tarray
School of Studies in Statistics, Vikram University, Ujjain, India

Abstract. The quest of this paper is to propose an improved mixed randomized response model for estimating the proportion of sensitive attribute π_S in the population. Mangat et al. [17] and Singh et al. [28] pointed out the privacy problem with the Moors [18] model. However their models may lose a large portion of data information and require a high cost to obtain confidentiality of the respondents as pointed by Kim and Warde's [11]. This led Kim and Warde's [11] to propose a mixed randomized response model using simple random sampling which rectifies the privacy problem. In this paper we have suggested a mixed randomized response model using Mangat and Singh [13] two – stage randomized response technique in place of second randomization device in Kim and Warde mixed randomized response model where they have used exactly the Warner's randomized response model. It has been shown that the proposed mixed randomized response model is better than Kim and Warde's [11] mixed randomized response model. Numerical illustration is given in support of the present study. The proposed model has been also extended in stratified sampling.

Keywords: Randomized response technique, dichotomous population, estimation of proportion, privacy of respondents, sensitive characteristics

1. Introduction

In survey methodology wherever the variable under investigation is sensitive in nature either because it pertains to something that is too personal or stigmatizing or illegal such as induced abortions, crimes, trade in contraband goods, susceptibility to intoxication, expenditures on addictions of various forms, humus-sexuality, and similar issues which are customarily disapproved of by society, randomized response techniques (RRTs) are used to collect the data. Warner [31] was first to develop a model for estimating the proportion of individuals possessing a sensitive attribute without requiring the individual respondent to report the interviewer whether or not he possesses the sensitive attribute. Further modifications in the model and in choice of unrelated questions were suggested by Greenberg et al. [4], Moors [18], Horvitz et al. [6], Singh [24], Mangat and Singh [13], Chaudhuri and Mukerjee [1], Singh and Singh [25–27], Mangat et al. [14–16], Singh et al. [28], Singh and Tracy [29], Kim and Warde [10], Singh and Tarray [20] and others.

To implement the privacy problem with the Moors [18] model, Mangat et al. [17] and Singh et al. [28] presented several strategies as an alternative to Moors model, but their models used simple random sampling without replacement (SRSWOR) which led to a high – cost survey compared with the Moors model using simple random sampling with replacement (SRSWR). Keeping the drawbacks with the previous alternative models for the Moors models Kim and Warde [11] proposed a mixed randomized response model using simple random sampling which modifies the privacy problem. We have also extended the proposed model to stratified sampling.

*Corresponding author: Housila P. Singh, School of Studies in Statistics, Vikram University, Ujjain 456010, India. E-mail: hpsujn@gmail.com.

ISSN 1574-1699/14/$27.50

2. The suggested model

Let a random sample of size n be selected using simple random sampling with replacement (SRSWR) from the population. Each respondent from the sample is instructed to answer an innocuous question "I posses the innocuous characteristic "Z". If the answer to the initial direct question is "Yes" then the respondent 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 pre-assigned 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 which is the same randomization device used in Mangat and Singh's [13] two – stage randomized response (RR) model. 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. Mangat and Singh [13] two – stage RR model is described as follows:
In the first stage of the surveys interview an individual respondent in the sample of size n_2 is instructed to use the randomization device R_2 which consists of a sensitive question (S) card with probability T and a "Go to the randomization device R_3 in the second stage" direction card with probability $(1 - T)$. The respondents in the second stage are instructed to use the randomization device R_3 which consists of a sensitive question (S) card with probability P and its negative question $(\bar{S})$ card with probability $(1 - P)$".

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 or R_3. Since all the respondents using a randomization device R_1 already responded "Yes" from the initial direct innocuous question, the proportion 'Y' of getting "Yes" answers from the respondents using randomization device R_1 should be

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

where π_S is the proportion of "Yes" answers from the sensitive trait and π_1 is the proportion of "Yes" answers from the innocuous question.

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

$$\hat{\pi}_{t1} = \frac{\hat{Y} - (1 - P_1)}{P_1} \tag{2}$$

whose variance is given by

$$\begin{aligned} V(\hat{\pi}_{t1}) &= \frac{Y(1-Y)}{n_1 P_1^2} = \frac{(1-\pi_S)[P_1\pi_S + (1-P_1)]}{n_1 P_1} \\ &= \frac{1}{n_1}\left\{\pi_S(1-\pi_S) + \frac{(1-\pi_S)(1-P_1)}{P_1}\right\} \end{aligned} \tag{3}$$

See Kim and Warde's [11].

The proportion of "Yes" answers from the respondents using Mangat and Singh [13] two – stage RR technique follows:

$$X = T\pi_S + (1 - T)[P\pi_S + (1 - P)(1 - \pi_S)] \tag{4}$$

where X is the proportion of "Yes" responses.

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

$$\hat{\pi}_{t2} = \frac{\hat{X} - (1 - T)(1 - P)}{[2P - 1 + 2\,T(1 - P)]} \tag{5}$$

The variance of the estimator $\hat{\pi}_{t2}$ is given by

$$V(\hat{\pi}_{t2}) = \frac{X(1 - X)}{n_2[2P - 1 + 2\,T(1 - P)]^2}$$

$$V(\hat{\pi}_{t2}) = \left\{ \frac{\pi_S(1-\pi_S)}{(n-n_1)} + \frac{(1-T)(1-P)[1-(1-T)(1-P)}{(n-n_1)[2P-1+2\,T(1-P)]^2} \right\} \tag{6}$$

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

$$\hat{\pi}_t = \frac{n_1}{n}\hat{\pi}_{t1} + \frac{n_2}{n}\hat{\pi}_{t2} = \frac{n_1}{n}\hat{\pi}_{t1} + \frac{(n-n_1)}{n}\hat{\pi}_{t2}, \text{ for } 0 < \frac{n_1}{n} < 1 \tag{7}$$

Since $\hat{\pi}_{t1}$ and $\hat{\pi}_{t2}$ are unbiased estimators of π_S, therefore the expected value of $\hat{\pi}_t$ is

$$E(\hat{\pi}_t) = E\left\{ \frac{n_1}{n}\hat{\pi}_{t1} + \frac{n_2}{n}\hat{\pi}_{t2} \right\} = \frac{n_1}{n}\pi_S + \frac{(n-n_1)}{n}\pi_S = \pi_S$$

Since the randomization device R_1 and the Mangat and Singh [13] two-stage RR technique (consisting of two randomization device R_2 and R_3) used are independent, we can obtain the following variance of $\hat{\pi}_t$:

$$V(\hat{\pi}_t) = \frac{n_1}{n^2}\left\{ \frac{(1-\pi_S)[P_1\pi_S + (1-P_1)]}{P_1} \right\} + \frac{n-n_1}{n^2}\left\{ \pi_S(1-\pi_S) + \frac{(1-T)(1-P)[1-(1-T)(1-P)]}{[2P-1+2\,T(1-P)]^2} \right\} \tag{8}$$

Under the circumstance that the Warner [31] model and Simmon's [5] method (known π_1) are equally confidential to respondents, Lanke [12] obtained a unique value of P as

$$P = \frac{1}{2} + \frac{P_1}{2P_1 + 4(1-P_1)\pi_1} \textit{ for every } P_1 \textit{ and every } \pi_1.$$

Since our mixed model also use Simmon's [5] method when $\pi_1 = 1$, we can apply Lanke's [12] idea to our proposed model. Thus we get

$$P = \frac{1}{2-P_1} \tag{9}$$

Putting $P = (2-P_1)^{-1}$ in Eq. (6), we get

$$\begin{aligned} V(\hat{\pi}_{t2}) &= \frac{\pi_S(1-\pi_S)}{(n-n_1)} + \frac{(1-T)(1-P_1)[2-P_1-(1-T)(1-P_1)]}{(n-n_1)[P_1+2\,T(1-P_1)]^2} \\ &= \frac{\pi_S(1-\pi_S)}{n} + \frac{(1-T)(1-P_1)[1+T(1-P_1)]}{(n-n_1)[P_1+2\,T(1-P_1)]^2} \end{aligned} \tag{10}$$

Thus we established the following theorem.

Theorem 1. The variance of $\hat{\pi}_t$ is given by

$$V(\hat{\pi}_t) = \frac{\pi_S(1-\pi_S)}{n} + \frac{1}{n}\left\{ \frac{\lambda(1-\pi_S)(1-P_1)}{P_1} + \frac{(1-\lambda)(1-T)(1-P_1)[1+T(1-P_1)]}{[P_1+2\,T(1-P_1)]^2} \right\} \tag{11}$$

where $n = n_1 + n_2$ and $\lambda = \frac{n_1}{n}$.

3. Efficiency comparisons

Under completely truthful reporting case we have made efficiency comparison of our proposed RR model with that of Kim and Warde's [11] model.

From Kim and Warde's ([11, p. 213], Eq. (10)), we have

$$V(\hat{\pi}_{kw}) = \frac{\pi_S(1-\pi_S)}{n} + \frac{(1-P_1)[\lambda P_1(1-\pi_S) + (1-\lambda)]}{nP_1^2} \tag{12}$$

Table 1
Percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's [11] estimator $\hat{\pi}_{kw}$

π_S	$n = 1000$		λ	T	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	0.1	358.48	195.23	149.48	129.33	118.51	112.01	107.77	104.81	102.50
0.1	500	500	0.5	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	0.9	2560.83	1258.07	818.84	594.81	456.08	359.00	284.35	221.61	163.12
0.2	700	300	0.7	0.1	375.35	200.00	151.31	129.96	118.51	111.64	107.20	104.14	101.89
0.2	500	500	0.5	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	0.9	2769.50	1308.17	818.84	572.30	422.64	321.12	246.76	188.89	141.34
0.3	700	300	0.7	0.1	395.32	205.79	153.74	131.00	118.85	111.59	106.95	103.82	101.62
0.3	500	500	0.5	0.5	1311.90	614.28	389.22	280.64	217.80	177.31	149.27	128.72	112.89
0.3	300	700	0.3	0.9	3044.12	1386.05	838.46	568.04	408.16	303.28	229.58	175.14	133.30
0.4	700	300	0.7	0.1	419.27	212.90	156.90	132.55	119.57	111.85	106.95	103.70	101.51
0.4	500	500	0.5	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	0.9	3416.30	1502.29	881.10	581.06	408.16	297.96	222.94	169.45	129.96
0.5	700	300	0.7	0.1	448.48	221.73	161.04	134.74	120.75	112.45	107.20	103.77	101.50
0.5	500	500	0.5	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	0.9	3942.52	1676.42	955.16	614.43	422.64	303.28	224.00	168.87	129.23
0.6	700	300	0.7	0.1	484.81	232.91	166.50	137.83	122.57	113.50	107.77	104.04	101.58
0.6	500	500	0.5	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	0.9	4734.47	1947.57	1078.91	677.37	456.08	321.12	233.18	173.14	130.73

From Eqs (11) and (12) we have

$$\{V(\hat{\pi}_{kw}) - V(\hat{\pi}_t)\} = \frac{(1-\lambda)(1-P_1)}{nP_1^2} - \frac{(1-\lambda)(1-T)(1-P_1)[1+T(1-P_1)]}{n[P_1+2\,T(1-P_1)]^2}$$
$$= \frac{(1-\lambda)(1-P_1)}{n}\left\{\frac{1}{P_1^2} - \frac{(1-T)[1+T(1-P_1)]}{[P_1+2\,T(1-P_1)]^2}\right\} > 0 \text{ if}$$
$$\frac{1}{P_1^2} - \frac{(1-T)[1+T(1-P_1)]}{[P_1+2\,T(1-P_1)]^2} > 0$$

i.e. if $T[4(1-P_1)^2 + P_1^2(1-P_1)] + 4P_1(1-P_1) + P_1^3 > 0$ which is always true.

Thus the proposed model is always better than Kim and Warde's [11] model. To have tangible idea about the performance of the proposed estimator $\hat{\pi}_t$ over Kim and Warde's [11] estimator $\hat{\pi}_{kw}$, we have computed the percent relative efficiency (*PRE*) of the proposed estimator $\hat{\pi}_t$ with respect to the Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ by using the formula:

$$PRE(\hat{\pi}_t, \hat{\pi}_{kw}) = \frac{V(\hat{\pi}_{kw})}{V(\hat{\pi}_t)} \times 100$$
$$= \frac{\left\{\pi_S(1-\pi_S) + \frac{\lambda(1-\pi_S)(1-P_1)}{P_1} + \frac{(1-\lambda)(1-P_1)}{P_1^2}\right\}}{\left\{\pi_S(1-\pi_S) + \frac{\lambda(1-\pi_S)(1-P_1)}{P_1} + \frac{(1-\lambda)(1-T)(1-P_1)[1+T(1-P_1)]}{[P_1+2\,T(1-P_1)]^2}\right\}} \times 100$$

for $n = 1000$ and different values of n_1, n_2, P_1 and T.

We have obtained the values of the percent relative efficiencies *PRE*$(\hat{\pi}_t, \hat{\pi}_{kw})$ for $\lambda = (0.3, 0.5, 0.7)$, $n = 1000$ and for different cases of π_S, T, n_2, n_1 *and* P_1. Findings are shown in Tables 1 and 2. Diagrammatic representations are also given in Figs 1 and 2.

It is observed from Tables 1 and 2 that:

The values of percent relative efficiencies P$RE(\hat{\pi}_t, \hat{\pi}_{kw})$ are more than 100. We can say that the envisaged estimator $\hat{\pi}_t$ is more efficient than the Kim and Warde's [11] estimator $\hat{\pi}_{kw}$.

Table 2
Percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's [11] estimator $\hat{\pi}_{kw}$

π_S	$n = 1000$		λ	T	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	0.9	567.12	327.95	247.34	206.17	180.55	162.37	147.96	135.00	121.02
0.1	500	500	0.3	0.5	2201.05	971.19	586.67	406.25	304.10	239.21	194.28	160.44	131.68
0.1	300	700	0.5	0.1	502.93	243.14	174.55	144.78	128.73	118.96	112.48	107.86	104.14
0.2	700	300	0.7	0.9	617.26	347.77	256.62	209.84	180.55	159.69	143.25	128.98	115.21
0.2	500	500	0.3	0.5	2352.39	1000.00	586.67	396.59	290.90	224.95	180.24	147.72	122.23
0.2	300	700	0.5	0.1	519.23	247.05	175.78	144.90	128.26	118.12	111.42	106.70	103.11
0.3	700	300	0.7	0.9	681.97	373.85	269.61	216.15	182.81	159.33	141.23	126.19	112.80
0.3	500	500	0.3	0.5	2546.20	1043.68	596.11	394.73	284.93	217.71	173.16	141.75	118.39
0.3	300	700	0.5	0.1	537.88	251.84	177.60	145.53	128.26	117.76	110.88	106.11	102.65
0.4	700	300	0.7	0.9	768.48	409.22	287.93	226.02	187.75	161.19	141.23	125.26	111.82
0.4	500	500	0.3	0.5	2799.61	1106.50	616.14	400.38	284.93	215.48	170.29	139.15	116.73
0.4	300	700	0.5	0.1	559.34	257.65	180.11	146.70	128.73	117.83	110.73	105.87	102.45
0.5	700	300	0.7	0.9	889.79	459.36	314.60	241.22	196.36	165.75	143.25	125.81	111.73
0.5	500	500	0.3	0.5	3140.87	1195.65	649.41	414.43	290.90	217.71	170.75	138.88	116.37
0.5	300	700	0.5	0.1	584.20	264.70	183.43	148.50	129.72	118.34	110.93	105.90	102.42
0.6	700	300	0.7	0.9	1071.84	535.13	355.59	265.36	210.88	174.29	147.96	128.06	112.49
0.6	500	500	0.3	0.5	3620.09	1323.74	701.08	439.39	304.10	224.95	174.68	140.85	117.12
0.6	300	700	0.5	0.1	613.23	273.26	187.75	151.08	131.35	119.36	111.54	106.22	102.54

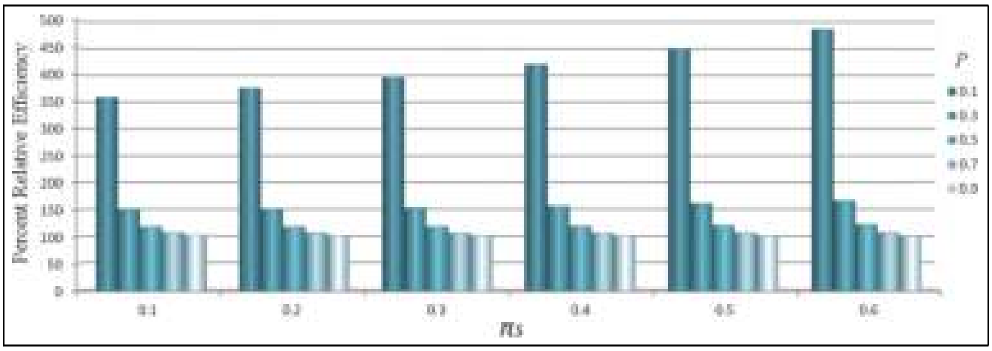

Fig. 1. Percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ when $T = 0.1$, $\lambda = 0.7$ and $n = 1000$.

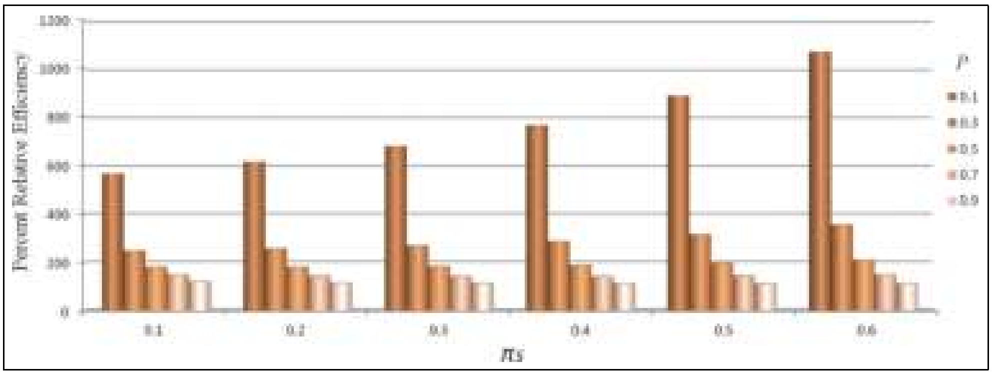

Fig. 2. Percent relative efficiency of the proposed estimator $\hat{\pi}_t$ with respect to Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ when $T = 0.9$, $\lambda = 0.7$ and $n = 1000$.

We note from Tables 1 and 2 that the values of the percent relative efficiencies $PRE(\hat{\pi}_t, \hat{\pi}_{kw})$ decrease as the value of P_1 increases. Also the values of the percent relative efficiencies $PRE(\hat{\pi}_t, \hat{\pi}_{kw})$ increase as the value of λ decreases for fixed value of P_1.

We further note from the results of Figs 1 and 2 that there is large gain in efficiency by using the suggested estimator $\hat{\pi}_t$ over the Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ when the proportion of stigmatizing attribute is moderately large.

4. A mixed randomized response model using stratification

4.1. A mixed stratified RR model

Stratified random sampling is generally obtained by dividing the population into non-overlapping groups called strata and selecting a simple random sample from each stratum. An RR technique using a stratified random sampling gives the group characteristics related to each stratum estimator. Also, stratified sampling protects a researcher from the possibility of obtaining a poor sample. Hong et al. [7] suggested a stratified RR technique using a proportional allocation. Kim and Warde [11] presented a stratified RR model based on Warner [31] model that has an optimal allocation and large gain in precision. Kim and Elam [8] suggested a two – stage stratified Warner's randomized response model using optimal allocation. Further Kim and Warde [11] suggested a mixed stratified randomized response model.

In the proposed models, the assumptions for a stratified mixed RR model are similar to Kim and Warde [11] and Kim and Elam [8] model. 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 answer 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_{k1} consisting of the statements: (i) "I am a member of the sensitive trait group" and (ii) "I am a member of the innocuous trait group" with pre-assigned probabilities Q_k and $(1 - Q_k)$, respectively. If a respondent answers "No", then the respondent is instructed to use the randomization device R_{k2} based on Mangat and Singh's [13] two-stage randomized RR model which consists of a sensitive question (S) card with probability T_k and a "Go to the randomization device R_{k3} in the second stage" direction card with probability $(1 - T_k)$. The respondent in the second stage are instructed to use the randomization device R_{k3} which consists of a sensitive question (S) card with probability P_k and its negative question $(\bar{S})$ card with probability $(1-P_k)$". To protect the respondent's privacy, the respondents should not disclose to the interviewer the question they answered from either R_{k1} or R_{k2} or R_{k3}. Suppose we denote m_k as the number of units in the sample from stratum k and n as the total number of units in samples from all strata. Let m_{k1} be the number of people responding "Yes" when respondents in a sample m_k were asked the direct question and m_{k2} be the number of people responding "No' when respondents in a sample m_k were asked the direct question so that $n = \sum_{k=1}^{r} m_k = \sum_{k=1}^{r} (m_{k1} + m_{k2})$. Under the assumption that these "Yes" or "No" reports are made truthfully, and Q_k and P_k $(\neq 0.5)$ are set by the researcher, then the proportion of "Yes" answers from the respondents using the randomization device R_{k1} will be

$$Y_k = Q_k \pi_{S_k} + (1 - Q_k)\pi_{1_k} \text{ for } k = 1, 2, \ldots r, \tag{13}$$

where Y_k is the proportion of "Yes" answers in stratum k, π_{S_k} is the proportion of respondents with the sensitive traits in stratum k, π_{1_k} is the proportion of respondents with the innocuous trait in stratum k, and Q_k is the probability that a respondent in the sample stratum k is asked a sensitive question.

Since the respondent performing a randomization device R_{k1} respond "Yes" to the direct question of the innocuous trait, if he or she chooses the same innocuous question from R_{k1}, then π_{1_k} is equal to one. Therefore, Eq. (13) becomes $Y_k = Q_k \pi_{S_k} + (1 - Q_k)$. The estimator of π_{S_k} is

$$\hat{\pi}_{t1_k} = \frac{\hat{Y}_k - (1 - Q_k)}{Q_k} for k = 1, 2, \ldots, r, \tag{14}$$

where $\hat{Y}_k$ is the proportion of "Yes" answers in a sample in stratum k and $\hat{\pi}_{t1_k}$ is the proportion of respondents with the sensitive trait in a sample from stratum k. Since each $\hat{Y}_k$ is a binomial distribution $B(m_{k1}, Y_k)$, the estimator $\hat{\pi}_{t1_k}$ is an unbiased for π_{S_k} with

$$V(\hat{\pi}_{t1_k}) = \frac{Q_k(1 - \pi_{S_k})[Q_k \pi_{S_k} + (1 - Q_k)]}{m_{k1} Q_k^2} = \frac{(1 - \pi_{S_k})[Q_k \pi_{S_k} + (1 - Q_k)]}{m_{k1} Q_k} \tag{15}$$

See Kim and Warde [11].

The proportion of "Yes" answers from the respondents using Mangat and Singh [13] two – stage RR technique follows:

$$X_k = T_k\pi_{S_k} + (1 - T_k)[P_k\pi_{S_k} + (1 - P_k)(1 - \pi_{S_k})] \tag{16}$$

where X_k is the proportion of "Yes" responses in stratum k, π_{S_k} is the proportion of respondents with the sensitive trait in stratum k, and P_k is the probability that a respondent in the sample stratum k has a sensitive question card. The maximum likelihood estimator in this case is

$$\hat{\pi}_{t2_k} = \frac{\hat{X}_k - (1 - T_k)(1 - P_k)}{[2P_k - 1 + 2T_k(1 - P_k)]} \tag{17}$$

where $\hat{X}_k$ is the proportion of "Yes" responses in a sample from a stratum k and $\hat{\pi}_{t2_k}$ is the proportion of respondents with the sensitive trait in a sample from stratum k. Since each $\hat{X}_k$ is a binomial distribution $B(m_k, X_k)$, the estimator $\hat{\pi}_{t2_k}$ is an unbiased for π_{S_k}. By using $m_k = m_{k1} + m_{k2}$ and Eq. (10), its variance is

$$V(\hat{\pi}_{t2_k}) = \frac{\pi_{S_k}(1 - \pi_{S_k})}{m_k} + \frac{(1 - T_k)(1 - Q_k)[1 + T_k(1 - Q_k)]}{(m_k - m_{k1})[Q_k + 2\,T_k(1 - Q_k)]^2} \tag{18}$$

Thus the unbiased estimator of π_{S_k}, in terms of sample proportion of "Yes" responses $\hat{Y}_k$ and $\hat{X}_k$, is

$$\hat{\pi}_{mS_k} = \frac{m_{k_1}}{m_k}\hat{\pi}_{t1_k} + \frac{m_k - m_{k1}}{m_k}\hat{\pi}_{t2_k} \quad \textit{for}\; 0 < \frac{m_{k1}}{m_k} < 1 \tag{19}$$

Its variance is

$$V(\hat{\pi}_{mS_k}) = \frac{\pi_{S_k}(1-\pi_{S_k})}{m_k} + \frac{1}{m_k}\left\{\frac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k} + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)[1+T_k(1-Q_k)]}{[Q_k+2T_k(1-Q_k)]^2}\right\} \tag{20}$$

where $m_k = m_{k1} + m_{k2}$ and $\lambda_k = m_{k1}/m_k$.

The unbiased estimator of π_{S_k} is shown to be

$$\hat{\pi}_{mS} = \sum_{k=1}^{r} w_k\hat{\pi}_m S_k = \sum_{k=1}^{r} w_k\left\{\frac{m_{k_1}}{m_k}\hat{\pi}_t 1_k + \frac{m_k - m_{k1}}{m_k}\hat{\pi}_{t2_k}\right\} \tag{21}$$

where N is the number of units in the whole population, N_k is the total number of units in stratum k, and $w_k = N_k/N$ for $k = 1, 2, \ldots, r$, so that $w = \sum_{k=1}^{r} w_k = 1$. It can be shown that the proposed estimator $\hat{\pi}_{mS}$ is unbiased for the population proportion π_S. The variance of an estimator $\hat{\pi}_{mS}$ is given by

$$V(\hat{\pi}_{mS}) = \sum_{k=1}^{r}\frac{w_k^2}{m_k} \times \left\{\pi_{S_k}(1-\pi_{S_k}) + \frac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k} + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)[1+T_k(1-Q_k)]}{[Q_k+2\,T_k(1-Q_k)]^2}\right\} \tag{22}$$

In order to do the optimal (Neyman) allocation of a sample size n, we need to know $\lambda_k = m_{k1}/m_k$ and π_{S_k}. Information on $\lambda_k = m_{k1}/m_k$ and π_{S_k} is usually unavailable. But if prior information about them is available from past experience it will help to derive the following optimal allocation formula.

Theorem 2. The optimal (Neyman) allocation of m to $m_1, m_2 \ldots m_{r-1}$ and m_r to derive the minimum variance of the $\hat{\pi}_{mS}$ subject to $n = \sum_{k=1}^{r} m_k$ is approximately given by

$$\frac{m_k}{n} = \frac{w_k\left\{\pi_{S_k}(1-\pi_{S_k}) + \frac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k} + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)[1+T_k(1-Q_k)]}{[Q_k+2\,T_k(1-Q_k)]^2}\right\}^{1/2}}{\sum_{k=1}^{r} w_k\left\{\pi_{S_k}(1-\pi_{S_k}) + \frac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k} + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)[1+T_k(1-Q_k)]}{[Q_k+2\,T_k(1-Q_k)]^2}\right\}^{1/2}} \tag{23}$$

where $m_k = m_{k1} + m_{k2}$ and $\lambda_k = m_{k1}/m_k$.

The minimal variance of the estimator $\widehat{\pi}_{mS}$ is given by

$$V(\hat{\pi}_{mS}) = \frac{1}{n}\left\{\sum_{k=1}^{r} w_k\left[\pi_{S_k}(1-\pi_{S_k}) + \frac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k} + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)[1+T_k(1-Q_k)]}{[Q_k+2\,T_k(1-Q_k)]^2}\right]^{1/2}\right\}^2 \tag{24}$$

where $n = \sum_{k=1}^{r} m_k, m_k = m_{k1} + m_{k2}$ and $\lambda_k = m_{k1}/m_k$.

4.2. An efficiency comparison of a stratified RR model

We do an efficiency comparison of a stratified mixed randomized response technique and the mixed randomized response model by comparing $V(\hat{\pi}_{mS})$ and $V(\hat{\pi}_t)$ in the following theorem.

Theorem 3. Suppose there are two strata in the population and $\lambda_k = m_{k1}/m_k$. The proposed estimator $\hat{\pi}_{mS}$ of a stratified mixed RR is more efficient than the estimator $\hat{\pi}_t$ of a mixed model where $P_1 = Q_1 = Q_2, \lambda = \lambda_1 = \lambda_2$ and $T = T_1 = T_2$.

Proof Under the assumptions $k = 2, P_1 = Q_1 = Q_2, \lambda = \lambda_1 = \lambda_2$ and $T = T_1 = T_2$ we write Eq. (24) as

$$V(\hat{\pi}_{mS}) = \frac{1}{n}\left\{w_1[\pi_{S1}(1-\pi_{S1}) + \lambda a_1 + b]^{1/2} + w_2[\pi_{S2}(1-\pi_{S2}) + \lambda a_2 + b]^{1/2}\right\}^2 \tag{25}$$

where

$$a_1 = \frac{(1-\pi_{S1})(1-P_1)}{P_1}, a_2 = \frac{(1-\pi_{S2})(1-P_1)}{P_1}, b = \frac{(1-\lambda)(1-T)(1-P_1)[1+T(1-P_1)]}{[P_1+2T(1-P_1)]^2}$$

From Eqs (11) and (25) we have

$$\begin{aligned} n[V(\hat{\pi}_t) - V(\hat{\pi}_{mS}) = &\{\pi_S(1-\pi_S) + \lambda(w_1 a_1 + w_2 a_2) + b \\ &-w_1^2[\pi_{S1}(1-\pi_{S1}) + \lambda a_1 + b] - w_2^2[\pi_{S2}(1-\pi_{S2}) + \lambda a_2 + b] \\ &-2w_1 w_2[\pi_{S1}(1-\pi_{S1}) + \lambda a_1 + b]^{1/2}[\pi_{S2}(1-\pi_{S2}) + \lambda a_2 + b]^{1/2}\} \\ =& w_1\pi_{S1} + w_2\pi_{S2} - 2w_1 w_2 \pi_{S1}\pi_{S2} - w_1^2\pi_{S1} - w_2^2\pi_{S2} - w_1^2(\lambda a_1 + b) - w_2^2(\lambda a_2 + b) \\ &+\lambda(w_1 a_1 + w_2 a_2) + b - 2w_1 w_2[\pi_{S1}(1-\pi_{S1}) + \lambda a_1 + b]^{1/2}[\pi_{S2}(1-\pi_{S2}) + \lambda a_2 + b]^{1/2} \\ =& w_1 w_2\{\pi_{S1} + \pi_{S2} - 2\pi_{S1}\pi_{S2} + 2b + \lambda(a_1 + a_2) \\ &-[\pi_{S1}(1-\pi_{S1}) + \lambda a_1 + b]^{1/2}[\pi_{S2}(1-\pi_{S2}) + \lambda a_2 + b]^{1/2}\} \\ =& w_1 w_2\{[(\pi_{S1}(1-\pi_{S1}) + \lambda a_1 + b)^{1/2} - (\pi_{S2}(1-\pi_{S2}) + \lambda a_2 + b)^{1/2}]^2 + (\pi_{S1} - \pi_{S2})^2\} \end{aligned}$$

which is always positive. Thus the theorem is proved.

Theorem 4. Suppose there are two strata in the population and $\lambda_k = m_{k1}/m_k$. The proposed estimator $\hat{\pi}_{mS}$ of a stratified mixed RR is more efficient than the Kim and Warde's [11] estimator $\hat{\pi}_{kw}$of a mixed model, where $P_1 = Q_1 = Q_2, \lambda = \lambda_1 = \lambda_2$ and $T = T_1 = T_2$.

Proof Under the assumptions $k = 2, P_1 = Q_1 = Q_2, \lambda = \lambda_1 = \lambda_2$ and $T = T_1 = T_2$ we write the minimal variance of the estimator $\hat{\pi}_{kw}$ (from Kim and Warde [11, p. 218], Eq. (24)) as

$$V(\hat{\pi}_{kw}) = \frac{1}{n}\{w_1(A_1 + b_1)^{1/2} + w_2(A_2 + b_1)^{1/2}\}^2 \tag{26}$$

where

$$A_1 = \pi_{S1}(1-\pi_{S1}) + \frac{\lambda(1-P_1)(1-\pi_{S1})}{P_1}; A_2 = \pi_{S2}(1-\pi_{S2}) + \frac{\lambda(1-P_1)(1-\pi_{S2})}{P_1},$$
$$b_1 = \frac{(1-P_1)(1-\lambda)}{P_1^2};$$

Expression Eq. (25) can be re - written as

$$V(\hat{\pi}_{mS}) = \frac{1}{n}\{w_1(A_1 + b_2)^{1/2} + w_2(A_2 + b_2)^{1/2}\}^2 \tag{27}$$

From Eqs (26) and (27) we have

$$n[V(\hat{\pi}_{kw})-V(\hat{\pi}_{mS})=\Big\{w_1^2(A_1+b_1)+w_2^2(A_2+b_1)+2w_1w_2(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}$$
$$-\ [w_1^2(A_1+b_2)+w_2^2(A_2+b_2)+2w_1w_2(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]\Big\}$$
$$=\Big\{w_1^2b_1+w_2^2b_1-w_1^2b_2-w_2^2b_2+2w_1w_2[(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}-(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]\Big\}$$
$$=\Big\{w_1^2(b_1-b_2)+w_2^2(b_1-b_2)+2w_1w_2[(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}-(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]\Big\}$$
$$=(b_1-b_2)(w_1^2+w_2^2)+2w_1w_2[(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}-(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]$$
$$=(b_1-b_2)(w_1^2+w_2^2)+2w_1w_2\frac{[(A_1+b_1)(A_2+b_1)-(A_1+b_2)(A_2+b_2)]}{[(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}+(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]}$$
$$=(b_1-b_2)\left\{w_1^2+w_2^2+\frac{2w_1w_2[A_1+A_2+b_1+b_2]}{[(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}+(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]}\right\}$$

Since

$$(b_1-b_2)=\frac{(1-\lambda)(2-P_1)^2(1-P_1)T[P_1+T(1-P_1)]}{P_1^2[P_1+2\,T(1-P_1)]^2}>0$$

Therefore

$$n[V(\hat{\pi}_{kw})-V(\hat{\pi}_{mS}]=(b_1-b_2)\left\{w_1^2+w_2^2+\frac{2w_1w_2[A_1+A_2+b_1+b_2]}{[(A_1+b_1)^{1/2}(A_2+b_1)^{1/2}+(A_1+b_2)^{1/2}(A_2+b_2)^{1/2}]}\right\}$$

is always positive. It follows that the proposed mixed randomized response estimator $\hat{\pi}_{mS}$ is always superior to that of Kim and Warde [11] estimator $\hat{\pi}_{kw}$. Thus the theorem is proved.

To judge the performance of the proposed stratified estimator $\hat{\pi}_{mS}$ over Kim and Warde [11] stratified estimator $\hat{\pi}_{kw}$ and suggested mixed estimator $\hat{\pi}_t$, we have computed the percent relative efficiency by using the formulae:

$$PRE(\hat{\pi}_{mS},\hat{\pi}_t)=\frac{V(\hat{\pi}_t)}{V(\hat{\pi}_{mS})}\times 100$$
$$=\frac{\left\{\pi_S(1-\pi_S)+\dfrac{\lambda(1-\pi_S)(1-P_1)}{P_1}+\dfrac{(1-\lambda)(1-T)(1-P_1)[1+T(1-P_1)]}{[P_1+2\,T(1-P_1)]^2}\right\}}{\left\{\sum_{k=1}^{2}w_k\left[\begin{array}{l}\pi_{S_k}(1-\pi_{S_k})+\dfrac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k}\\+\dfrac{(1-\lambda_k)(1-T_k)(1-Q_k)[1+T_k(1-Q_k)]}{[Q_k+2\,T_k(1-Q_k)]^2}\end{array}\right]^{1/2}\right\}^2}\times 100$$

and

$$PRE(\hat{\pi}_{mS},\hat{\pi}_{kw})=\frac{V(\hat{\pi}_{kw})}{V(\hat{\pi}_{mS})}\times 100$$
$$=\frac{\left\{\sum_{k=1}^{2}w_k\left[\pi_{S_k}(1-\pi_{S_k})+\dfrac{(1-Q_k)(\lambda_kQ_k(1-\pi_{S_k})+(1-\lambda_k))}{Q_k^2}\right]^{1/2}\right\}^2}{\left\{\sum_{k=1}^{2}w_k\left[\begin{array}{l}\pi_{S_k}(1-\pi_{S_k})+\dfrac{\lambda_k(1-\pi_{S_k})(1-Q_k)}{Q_k}\\+\dfrac{(1-\lambda_k)(1-T_k)(1-Q_k)\{1+T_k(1-Q_k)]}{[Q_k+2\,T_k(1-Q_k)]^2}\end{array}\right]^{1/2}\right\}^2}\times 100$$

Table 3
Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to suggested mixed model estimator $\hat{\pi}_t$

π_{S1}	π_{S2}	π_S	W_1	W_2	λ	T	$P_1 = Q_1 = Q_2$						
							$P_1 \to 0.1$	0.15	0.2	0.25	0.3	0.35	0.4
0.08	0.13	0.1	0.6	0.4	0.2	0.1	192.23	192.28	192.33	191.89	191.97	192.05	192.15
0.08	0.13	0.1	0.6	0.4	0.4	0.3	191.78	191.93	192.06	191.96	192.07	192.19	192.32
0.08	0.13	0.1	0.6	0.4	0.6	0.1	191.89	191.96	192.03	192.24	192.30	192.37	192.46
0.08	0.13	0.1	0.6	0.4	0.8	0.3	191.51	191.60	191.69	192.48	192.59	192.71	192.84
0.18	0.23	0.2	0.6	0.4	0.2	0.1	192.22	192.26	192.30	191.81	191.88	191.95	192.03
0.18	0.23	0.2	0.6	0.4	0.4	0.3	191.70	191.85	191.97	191.86	191.96	192.07	192.18
0.18	0.23	0.2	0.6	0.4	0.6	0.1	191.85	191.92	191.98	192.20	192.24	192.30	192.36
0.18	0.23	0.2	0.6	0.4	0.8	0.3	191.40	191.49	191.58	192.37	192.46	192.56	192.65
0.28	0.33	0.3	0.6	0.4	0.2	0.1	192.20	192.24	192.26	191.71	191.77	191.83	191.90
0.28	0.33	0.3	0.6	0.4	0.4	0.3	191.61	191.76	191.88	191.74	191.83	191.93	192.02
0.28	0.33	0.3	0.6	0.4	0.6	0.1	191.80	191.87	191.92	192.15	192.18	192.22	192.27
0.28	0.33	0.3	0.6	0.4	0.8	0.3	191.26	191.36	191.45	192.27	192.34	192.41	192.48
0.38	0.43	0.4	0.6	0.4	0.2	0.1	192.18	192.21	192.23	191.60	191.64	191.69	191.74
0.38	0.43	0.4	0.6	0.4	0.4	0.3	191.50	191.65	191.77	191.60	191.69	191.77	191.85
0.38	0.43	0.4	0.6	0.4	0.6	0.1	191.74	191.81	191.86	192.10	192.12	192.15	192.17
0.38	0.43	0.4	0.6	0.4	0.8	0.3	191.08	191.18	191.27	192.16	192.22	192.27	192.32

Table 4
Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to suggested mixed model estimator $\hat{\pi}_t$

π_{S1}	π_{S2}	π_S	W_1	W_2	λ	T	$P_1 = Q_1 = Q_2$						
							$P_1 \to 0.1$	0.15	0.2	0.25	0.3	0.35	0.4
0.08	0.13	0.1	0.6	0.4	0.8	0.1	191.68	191.76	191.82	191.89	191.97	192.05	192.15
0.08	0.13	0.1	0.6	0.4	0.6	0.3	191.62	191.74	191.85	191.96	192.07	192.19	192.32
0.08	0.13	0.1	0.6	0.4	0.4	0.1	192.07	192.14	192.19	192.24	192.30	192.37	192.46
0.08	0.13	0.1	0.6	0.4	0.2	0.3	192.07	192.23	192.36	192.48	192.59	192.71	192.84
0.18	0.23	0.2	0.6	0.4	0.8	0.1	191.61	191.68	191.75	191.81	191.88	191.95	192.03
0.18	0.23	0.2	0.6	0.4	0.6	0.3	191.52	191.64	191.75	191.86	191.96	192.07	192.18
0.18	0.23	0.2	0.6	0.4	0.4	0.1	192.05	192.11	192.15	192.20	192.24	192.30	192.36
0.18	0.23	0.2	0.6	0.4	0.2	0.3	192.01	192.16	192.27	192.37	192.46	192.56	192.65
0.28	0.33	0.3	0.6	0.4	0.8	0.1	191.52	191.60	191.66	191.71	191.77	191.83	191.90
0.28	0.33	0.3	0.6	0.4	0.6	0.3	191.40	191.53	191.64	191.74	191.83	191.93	192.02
0.28	0.33	0.3	0.6	0.4	0.4	0.1	192.02	192.08	192.11	192.15	192.18	192.22	192.27
0.28	0.33	0.3	0.6	0.4	0.2	0.3	191.93	192.08	192.18	192.27	192.34	192.41	192.48
0.38	0.43	0.4	0.6	0.4	0.8	0.1	191.42	191.49	191.55	191.60	191.64	191.69	191.74
0.38	0.43	0.4	0.6	0.4	0.6	0.3	191.26	191.39	191.50	191.60	191.69	191.77	191.85
0.38	0.43	0.4	0.6	0.4	0.4	0.1	191.99	192.04	192.07	192.10	192.12	192.15	192.17
0.38	0.43	0.4	0.6	0.4	0.2	0.3	191.85	191.99	192.09	192.16	192.22	192.27	192.32

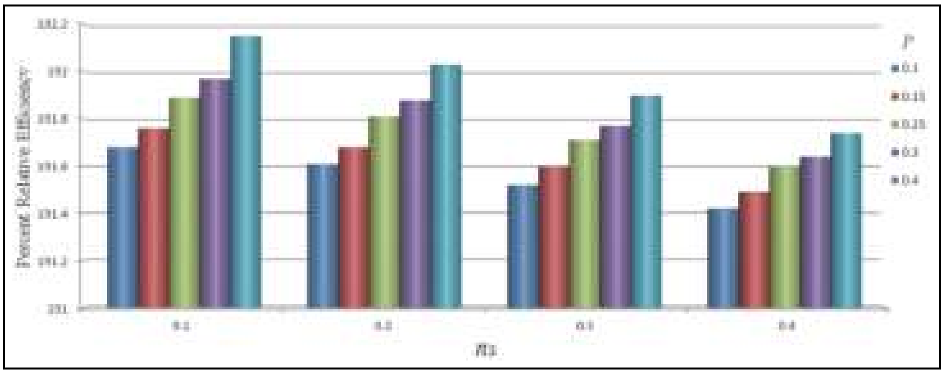

Fig. 3. Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to suggested mixed model estimator $\hat{\pi}_t$ when $T = 0.1$, $\lambda = 0.2$ and $n = 1000$.

Table 5

Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$

π_{S1}	π_{S2}	π_S	W_1	W_2	λ	T	$P_1 = Q_1$						
							$P_1 \to 0.1$	0.15	0.2	0.25	0.3	0.35	0.4
							$Q_2 \to 0.2$	0.25	0.3	0.35	0.4	0.45	0.5
0.08	0.13	0.1	0.6	0.4	0.2	0.1	543.42	349.80	265.22	218.74	189.69	169.95	155.76
0.08	0.13	0.1	0.6	0.4	0.4	0.3	1000.81	647.71	475.63	375.13	309.98	264.75	231.74
0.08	0.13	0.1	0.6	0.4	0.6	0.1	356.65	249.37	199.62	171.42	153.52	141.28	132.47
0.08	0.13	0.1	0.6	0.4	0.8	0.3	303.99	232.86	196.00	173.24	157.73	146.48	137.96
0.18	0.23	0.2	0.6	0.4	0.2	0.1	546.85	350.63	265.07	218.10	188.72	168.76	154.39
0.18	0.23	0.2	0.6	0.4	0.4	0.3	1053.60	671.41	486.99	380.17	311.39	263.91	229.44
0.18	0.23	0.2	0.6	0.4	0.6	0.1	368.85	255.37	203.02	173.43	154.67	141.86	132.64
0.18	0.23	0.2	0.6	0.4	0.8	0.3	323.55	244.26	203.26	178.00	160.85	148.44	139.07
0.28	0.33	0.3	0.6	0.4	0.2	0.1	551.42	352.28	265.64	218.11	188.41	168.22	153.67
0.28	0.33	0.3	0.6	0.4	0.4	0.3	1119.44	702.24	503.18	388.91	315.90	265.82	229.67
0.28	0.33	0.3	0.6	0.4	0.6	0.1	383.13	262.56	207.25	176.10	156.38	142.92	133.26
0.28	0.33	0.3	0.6	0.4	0.8	0.3	348.42	258.82	212.66	184.33	165.17	151.37	140.99
0.38	0.43	0.4	0.6	0.4	0.2	0.1	557.20	354.79	266.94	218.79	188.72	168.28	153.56
0.38	0.43	0.4	0.6	0.4	0.4	0.3	1202.61	742.26	525.33	402.04	323.90	270.66	232.48
0.38	0.43	0.4	0.6	0.4	0.6	0.1	399.95	271.19	212.51	179.56	158.74	144.55	134.37
0.38	0.43	0.4	0.6	0.4	0.8	0.3	380.98	277.97	225.12	192.86	171.14	155.59	143.95

Table 6

Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$

π_{S1}	π_{S2}	π_S	W_1	W_2	λ	T	$P_1 = Q_1$						
							$P_1 \to 0.1$	0.15	0.2	0.25	0.3	0.35	0.4
							$Q_2 \to 0.2$	0.25	0.3	0.35	0.4	0.45	0.5
0.08	0.13	0.1	0.6	0.4	0.8	0.1	239.66	182.91	155.58	139.77	129.64	122.69	117.70
0.08	0.13	0.1	0.6	0.4	0.6	0.3	585.02	407.26	317.19	262.80	226.50	200.67	181.42
0.08	0.13	0.1	0.6	0.4	0.4	0.1	456.72	304.02	235.46	197.25	173.19	156.80	145.00
0.08	0.13	0.1	0.6	0.4	0.2	0.3	1682.45	1001.70	692.08	520.43	413.59	341.76	290.72
0.18	0.23	0.2	0.6	0.4	0.8	0.1	250.00	188.39	158.86	141.84	130.94	123.49	118.14
0.18	0.23	0.2	0.6	0.4	0.6	0.3	624.07	427.98	329.23	269.93	230.57	202.68	182.00
0.18	0.23	0.2	0.6	0.4	0.4	0.1	465.95	308.06	237.45	198.18	173.46	156.63	144.50
0.18	0.23	0.2	0.6	0.4	0.2	0.3	1721.42	1010.39	690.75	515.03	406.34	333.63	282.18
0.28	0.33	0.3	0.6	0.4	0.8	0.1	262.55	195.11	162.96	144.49	132.71	124.66	118.89
0.28	0.33	0.3	0.6	0.4	0.6	0.3	673.05	454.30	344.96	279.75	236.75	206.47	184.14
0.28	0.33	0.3	0.6	0.4	0.4	0.1	476.66	313.04	240.18	199.75	174.32	157.01	144.55
0.28	0.33	0.3	0.6	0.4	0.2	0.3	1775.37	1028.08	696.15	515.19	403.98	329.97	277.83
0.38	0.43	0.4	0.6	0.4	0.8	0.1	278.07	203.49	168.15	147.94	135.07	126.30	120.03
0.38	0.43	0.4	0.6	0.4	0.6	0.3	735.92	488.32	365.68	293.13	245.64	212.43	188.09
0.38	0.43	0.4	0.6	0.4	0.4	0.1	489.10	319.09	243.74	202.01	175.81	157.98	145.15
0.38	0.43	0.4	0.6	0.4	0.2	0.3	1847.15	1055.84	708.63	520.92	406.29	330.38	277.13

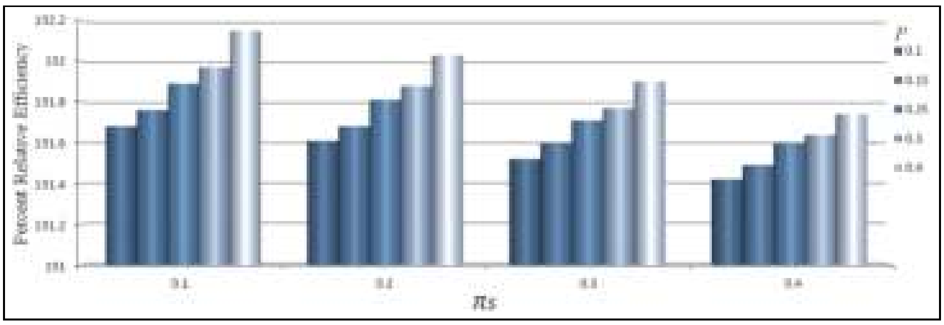

Fig. 4. Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to suggested mixed model estimator $\hat{\pi}_t$. when $T = 0.1$, $\lambda = 0.8$ and $n = 1000$.

Table 7
Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$

π_{S1}	π_{S2}	π_S	W_1	W_2	λ	T	$P_1 = Q_1$						
							$P_1 \to$ 0.6 $Q_2 \to$ 0.65	0.65 0.7	0.7 0.75	0.75 0.8	0.8 0.85	0.85 0.9	0.9 0.95
0.08	0.13	0.1	0.6	0.4	0.2	0.1	126.24	121.56	117.61	114.19	111.15	108.35	105.58
0.08	0.13	0.1	0.6	0.4	0.4	0.3	162.44	151.26	141.80	133.61	126.32	119.59	112.96
0.08	0.13	0.1	0.6	0.4	0.6	0.1	114.45	111.69	109.40	107.46	105.78	104.27	102.82
0.08	0.13	0.1	0.6	0.4	0.8	0.3	118.90	115.65	112.86	110.41	108.22	106.17	104.13
0.18	0.23	0.2	0.6	0.4	0.2	0.1	124.45	119.70	115.68	112.23	109.21	106.50	103.99
0.18	0.23	0.2	0.6	0.4	0.4	0.3	157.76	146.33	136.72	128.50	121.34	114.97	109.12
0.18	0.23	0.2	0.6	0.4	0.6	0.1	113.81	110.94	108.56	106.57	104.86	103.38	102.04
0.18	0.23	0.2	0.6	0.4	0.8	0.3	118.28	114.78	111.78	109.20	106.92	104.88	102.99
0.28	0.33	0.3	0.6	0.4	0.2	0.1	123.41	118.62	114.60	111.17	108.21	105.62	103.32
0.28	0.33	0.3	0.6	0.4	0.4	0.3	155.43	143.79	134.09	125.92	118.93	112.88	107.55
0.28	0.33	0.3	0.6	0.4	0.6	0.1	113.58	110.59	108.15	106.11	104.41	102.96	101.71
0.28	0.33	0.3	0.6	0.4	0.8	0.3	118.26	114.50	111.33	108.63	106.31	104.29	102.52
0.38	0.43	0.4	0.6	0.4	0.2	0.1	122.95	118.12	114.09	110.67	107.74	105.22	103.03
0.38	0.43	0.4	0.6	0.4	0.4	0.3	154.96	142.99	133.10	124.85	117.89	111.96	106.87
0.38	0.43	0.4	0.6	0.4	0.6	0.1	113.70	110.59	108.05	105.96	104.22	102.78	101.57
0.38	0.43	0.4	0.6	0.4	0.8	0.3	118.83	114.75	111.36	108.51	106.10	104.06	102.32

Table 8
Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$

π_{S1}	π_{S2}	π_S	W_1	W_2	λ	T	$P_1 = Q_1$						
							$P_1 \to$ 0.6 $Q_2 \to$ 0.65	0.65 0.7	0.7 0.75	0.75 0.8	0.8 0.85	0.85 0.9	0.9 0.95
0.08	0.13	0.1	0.6	0.4	0.8	0.1	107.61	106.10	104.86	103.83	102.94	102.16	101.41
0.08	0.13	0.1	0.6	0.4	0.6	0.3	139.62	132.68	126.75	121.59	116.97	112.69	108.45
0.08	0.13	0.1	0.6	0.4	0.4	0.1	120.63	116.83	113.64	110.91	108.52	106.33	104.20
0.08	0.13	0.1	0.6	0.4	0.2	0.3	187.70	171.64	158.16	146.58	136.33	126.91	117.69
0.18	0.23	0.2	0.6	0.4	0.8	0.1	107.39	105.79	104.49	103.41	102.50	101.72	101.03
0.18	0.23	0.2	0.6	0.4	0.6	0.3	137.51	130.20	124.02	118.69	114.04	109.87	106.03
0.18	0.23	0.2	0.6	0.4	0.4	0.1	119.46	115.55	112.28	109.50	107.09	104.97	103.03
0.18	0.23	0.2	0.6	0.4	0.2	0.3	179.11	163.19	149.92	138.64	128.86	120.18	112.26
0.28	0.33	0.3	0.6	0.4	0.8	0.1	107.38	105.69	104.32	103.21	102.28	101.52	100.87
0.28	0.33	0.3	0.6	0.4	0.6	0.3	136.74	129.09	122.69	117.27	112.62	108.58	105.04
0.28	0.33	0.3	0.6	0.4	0.4	0.1	118.86	114.87	111.55	108.76	106.38	104.33	102.53
0.28	0.33	0.3	0.6	0.4	0.2	0.3	174.35	158.58	145.53	134.58	125.24	117.17	110.07
0.38	0.43	0.4	0.6	0.4	0.8	0.1	107.59	105.78	104.33	103.16	102.21	101.44	100.80
0.38	0.43	0.4	0.6	0.4	0.6	0.3	137.14	129.07	122.39	116.79	112.06	108.04	104.61
0.38	0.43	0.4	0.6	0.4	0.4	0.1	118.74	114.65	111.27	108.44	106.06	104.04	102.31
0.38	0.43	0.4	0.6	0.4	0.2	0.3	172.31	156.50	143.52	132.70	123.58	115.80	109.11

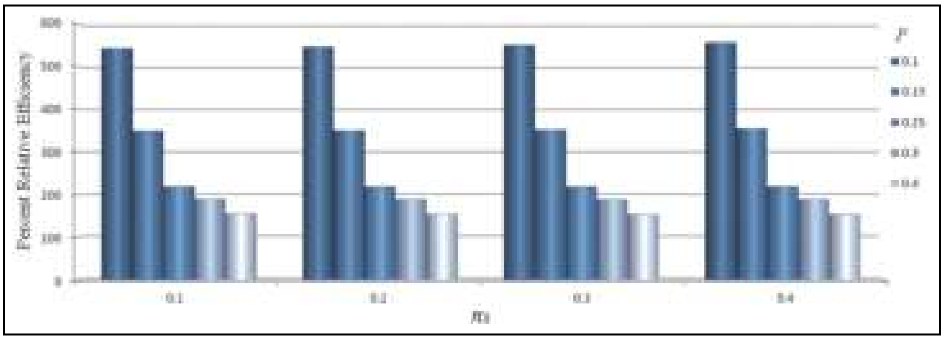

Fig. 5. Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$ when $T = 0.1$, $\lambda = 0.2$ and $n = 1000$.

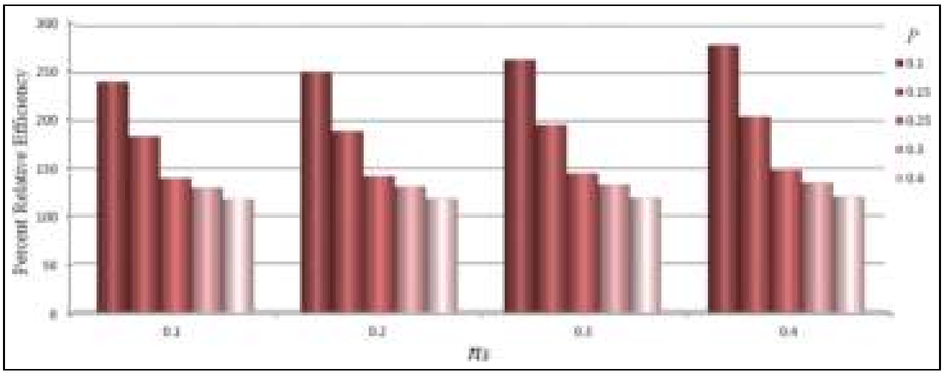

Fig. 6. Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$ when $T = 0.1$, $\lambda = 0.8$ and $n = 1000$.

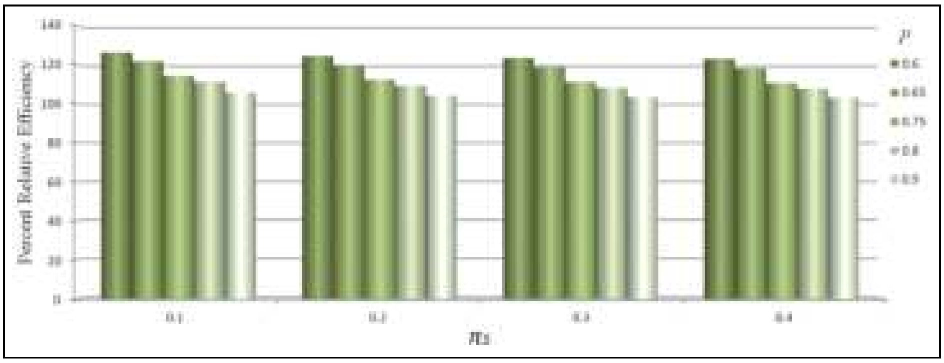

Fig. 7. Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$ when $T = 0.1$, $\lambda = 0.2$ and $n = 1000$.

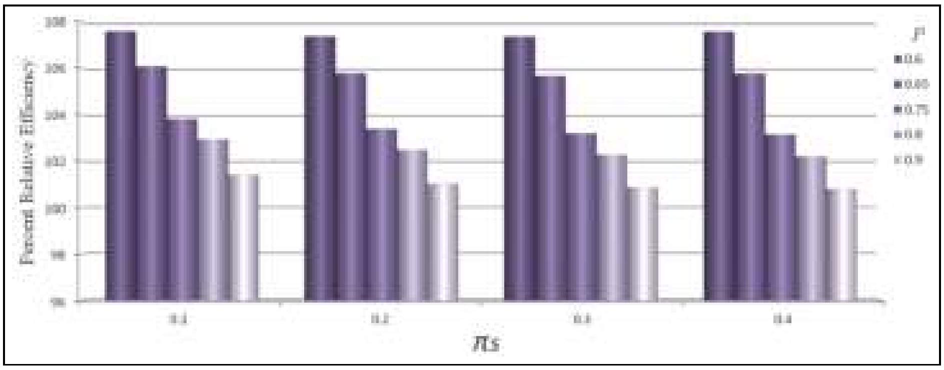

Fig. 8. Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] stratified estimator $\hat{\pi}_{kw}$ when $T = 0.1$, $\lambda = 0.8$ and $n = 1000$.

To judge the merits of the proposed estimators $(\hat{\pi}_t, \hat{\pi}_{mS})$ over Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ for $\lambda =$ (0.2, 0.4, 0.6, 0.8), $n = 1000$ and for different levels of π_S, T, n_2, n_1 and P_1. Findings are shown in Tables 3 to 8. Diagrammatic representations are also given in Figs 3 to 8.

It is observed from Tables 3 and 4 and Figs 3 and 4 that:

The values of percent relative efficiencies $PRE(\hat{\pi}_{mS}, \hat{\pi}_t)$ are more than 100. We can say that the envisaged estimator $\hat{\pi}_{ms}$ is more efficient than the suggested mixed estimator $\hat{\pi}_t$.

It is observed from Tables 3 and 4 that the zig – zag trend in the values of the percent relative efficiencies $PRE(\hat{\pi}_{mS}, \hat{\pi}_t)$ is observed as the value of P_1 increases. Also the values of the percent relative efficiencies $PRE(\hat{\pi}_{mS}, \hat{\pi}_t)$ decrease as the value of π_S increases for fixed values of (T, λ).

We further note from the results of Figs 3 and 4 that there is large gain in efficiency by using the suggested estimator $\hat{\pi}_{mS}$ over the proposed midxed estimator $\hat{\pi}_t$.

Tables 5–8 and Figs 5–8 – exhibit that the percent relative efficiencies of the proposed estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ decrease as the value of P_1 increases. The values of percent relative efficiency $PRE(\hat{\pi}_{mS}, \hat{\pi}_{kw})$ are more than 100 for all parametric values considered here, therefore the proposed estimator $\hat{\pi}_{mS}$ is better than Kim and Warde's [11] estimator $\hat{\pi}_{kw}$.

Finally from the above discussion we conclude that the proposed estimator $\hat{\pi}_{mS}$ is better than the Kim and Warde's [11] estimator $\hat{\pi}_{kw}$ as well as proposed mixed estimator $\hat{\pi}_t$.

Remark The results of this paper can be also extended in PPS sampling.

5. Discussion

In this paper, we have envisaged a mixed randomized response model as well as its stratified randomized response model to estimate the proportion of qualitative sensitive character. It has been shown that the proposed mixed randomized response model and its stratified randomized response model is better than the Kim and Warde's [11] mixed randomized response model with larger gain in efficiency.

Acknowledgments

The authors are thankful to the Editor – in – Chief – Dr. Sarjinder Singh and to the learned referees for their valuable suggestions regarding improvement of the paper.

References

[1] A. Chaudhuri and R. Mukerjee, Randomized response technique: A review, *Statist Neerl* **41** (1987), 27–44.
[2] W.G. Cochran, Sampling technique, 3rd Edition, New York: John Wiley and Sons, USA, 1977.
[3] J.A. Fox and P.E. Tracy, Randomized response: A method of sensitive surveys, Newbury Park, CA: SEGE Publications, 1986.
[4] B. Greenberg, A.A.-E. Kuebler, R. Roy, J.R. Abernathy and D.G. Horvitz, Applications of the randomized response technique in obtaining quantative data, *Jour Amer Statist Assoc* **66** (1971), 243–250.
[5] D.G. Horvitz, B.V. Shah and W.R. Simmons, The unrelated question randomized response model, *Proc Soc Statist Sec Amer Statistical Assoc* (1967), 65–72.
[6] D.G. Horvitz, G. Daniel, B.G. Greenberg and J.R. Abernathy, Recent developments in randomized response designs: A survey of statistical design and linear model, edited by Srivastava JN, *North Holland Publishing Company*, Amsterdam, (1975), 271–285.
[7] K. Hong, J. Yum and H. Lee, A stratified randomized response technique, *Korean Jour Appl Statist* **7** (1994), 141–147.
[8] J.M. Kim and M.E. Elam, A two – stage stratified Warner's randomized response model using optimal allocation, *Metrika* **61** (2005), 1–7.
[9] J.M. Kim, J.M. Tebbs and S.W. An, Extensions of Mangat's randomized response model, *Jour Statist Plan Inference* **136** (2006), 1554–1567.
[10] J. Kim and W. Warde, A stratified warner randomized response model, *Jour Statist Plan Inference* **120** (2004), 155–165.
[11] J.M. Kim and W.D. Warde, A mixed randomized response model, *Jour Statist Plan Inference* **133** (2005), 211–221.
[12] J. Lank, On the degree of protection in randomized interview, *Int Statist Rev* **44** 80–83.
[13] N.S. Mangat and R. Singh, An alternative randomized procedure, *Biometrika* **77** (1990), 439–442.
[14] N.S. Mangat, R. Singh and S. Singh, An alternative estimator in randomized response technique, *Aligarh Jour Statist* **11** (1991), 75–80.
[15] N.S. Mangat, R. Singh and S. Singh, An improved unrelated question randomized response strategy, *Calcutta Statist Assoc Bull* (1992), 277–281.
[16] N.S. Mangat, R. Singh, S. Singh and B. Singh, On the Moors' randomized response model, *Biom Jour* **35**(6) (1993), 727–755.
[17] N.S. Mangat, R. Singh and S. Singh, Violation of respondent's privacy in Moors model – its rectification through a random group strategy response model, *Comm Statist Theo – Metho* **3** (1997), 243–255.
[18] J.A. Moors, Optimization of the unrelated question randomized response model, *Jour Amer Statist Assoc* **66** (1971), 627–629.
[19] J.B. Ryu, K.H. Hong and G.S. Lee, Randomized response model, *Freedom Academy*, Seoul, Korea, (1993).
[20] H.P. Singh and T.A. Tarray, A stratified unknown repeated trials in randomized response sampling, *Commun Kor Statist Soc* **19**(6) (2012), 751–759.

[21] R. Singh and N.S. Mangat, A mail survey design for sensitive character without using randomization device, *Comm Statist Theo – Metho* **22**(9) (1993), 2661–2668.
[22] R. Singh and N.S. Mangat, Elements of survey sampling, kluwer academic publishers, Dordrecht, The Netherlands, 1996.
[23] S. Singh, R. Singh, N.S. Mangat and D.S. Tracy, An improved two – stage randomized response strategy, *Statist* **36** (1995), 265–271.
[24] S. Singh, Advanced sampling theory with applications, Kluwer Academic Publishers, Dordrecht, 2003.
[25] S. Singh and R. Singh, Improved Franklin's model for randomized response sampling, *Jour Ind Statist Assoc* **30** (1992), 109–122.
[26] S. Singh and R Singh, An alternative estimator for randomized response technique, *Jour Ind Soc Agril Statist* **44**(2) (1992), 146–154.
[27] S. Singh and R. Singh, Generalized Franklin's model for randomized response sampling, *Comm Statist Theo Metho* **22**(2) (1993), 741–755.
[28] S. Sing, R. Singh and N.S. Mangat, Some alternative strategies to Moor's model in randomized response model, *Jour Statist Plan Inference* **83** (2000), 243–255.
[29] S. Singh and D.S. Tracy, Ridge regression using scrambled responses, *Metron* **57** (1999), 47–68.
[30] D.S. Tracy and N.S. Mangat, Some developments in randomized response sampling during the last decade – A follow up of review by Chaudhuri and Mukherjee, *Jour Applied Statist Sci* **4**(2/3) (1996), 147–158.
[31] S.L. Warner, Randomized response: A survey technique for eliminating evasive answer bias, *Jour Amer Statist Assoc* **60** (1965), 63–69.