REVSTAT – Statistical Journal
Volume 19, Number 4, October 2021, 531–551

AN EFFICIENT MIXED RANDOMIZED RESPONSE MODEL FOR SENSITIVE CHARACTERISTIC IN SAMPLE SURVEY

Authors: Amod Kumar
– Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India
amod.ism01@gmail.com

Gajendra K. Vishwakarma
– Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India
vishwagk@rediffmail.com

G.N. Singh
– Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India
gnsingh_ism@yahoo.com

Received: April 2019 Revised: September 2019 Accepted: September 2019

Abstract:

- This paper proposes an efficient mixed randomized response (RR) model for estimating the proportion of individuals who possess to the sensitive attribute in the given population under both the conditions completely truthful reporting as well as less than completely truthful reporting and examined its properties. The proposed models are found to be dominant over Kim and Warde [13] model. It has also been extended for stratified random sampling. Numerical illustrations are presented to support the theoretical results.

Keywords:

- *randomized response technique; dichotomous; estimation of proportion; privacy; innocuous variable; sensitive characteristic.*

AMS Subject Classification:

- 62D05.

1. INTRODUCTION

In situations where potentially embarrassing or incriminating responses are sought, the randomized response technique (RRT) is effective in reducing non-sampling errors in sample surveys. In survey methodology, refusal to respond and lying are two major sources of non-sampling errors, as the stigma attached to certain practices (e.g. abortion and the use of illegal drugs) often leads to discrimination. Warner [28] did the pioneering work by suggesting a randomized response technique (RRT), which minimizes under reporting in survey data related to a socially undesirable or incriminating behaviour questions such as illegal earning or homosexuality among others. Warner [28] model requires the interviewee to give a "Yes" or "No" answers either to the sensitive question or to its negative depending on the outcome of a random device not reported to the interviewer. Further by introducing a choice of an unrelated question Greenberg *et al.* [7] modifying the Warner [28] randomized response model (RRM), the randomized response technique was further modified for different practical situation by Moors [17], Cochran [5], Fox and Tracy [6], Chaudhuri and Mukherjee [4], Hedayat and Sinha [8], Ryu *et al.* [19], Singh and Mangat [22], Tracy and Mangat [26], Tracy and Osahan [27], Singh [21], Singh and Tarray [23, 24, 25] and Kim and Warde [13] among others.

Kim and Warde [13] suggested a mixed randomized response model using simple random sampling with replacement which rectifies the privacy problem. Following the work of Kim and Warde [13], Amitava [1] and Hussain and Shabbir [10] suggested mixed randomized response technique (RRT) for complex survey designs and illustrated the superiority of their models over Kim and Warde [13] model.

Motivated with the above works, we have suggested a modified version of Kim and Warde [13] model and studied its properties in detail. We also present the less than completely truthful reporting counterpart of suggested model. It has been demonstrated that the suggested models perform better than the mixed randomized response model (RRM) of Kim and Warde [13]. We have also introduced the suggested model for stratified random sampling. The empirical studies are carried out; which showed dominance of proposed mixed randomized response models and stratified random sampling as well.

2. SUGGESTED MODEL

Let a sample of size n be selected from a finite population of size N using simple random sampling with replacement (SRSWR) scheme. Each respondent from the sample is instructed to answer the direct question "whether he/she is a member of the innocuous group?" If the answer to the initial direct question is "Yes" then he/she is instructed to go to the random device R_1 consisting of two statements:

(i) "I am a member of the sensitive trait group",

(ii) "I am a member of the innocuous trait group",

with probabilities P_1 and $(1 - P_1)$ respectively. If a respondent answers "No" to the direct question, then the respondent is instructed to use the random device R_2 consisting of the statements on the first stage which is same as Mangat and Singh [16]:

- **(i)** "Do you possess the sensitive attribute A", with probability T,
- **(ii)** "Go to the random device R_3 in the second stage", with probability $(1-T)$.

The respondents at the second stage are instructed to use the random device R_3 using three statements:

- **(i)** "I possess the sensitive attribute A",
- **(ii)** "Yes",
- **(iii)** "No",

with probabilities P, $(1-P)/2$ and $(1-P)/2$ respectively. When the outcome of random device R_3 is either (ii) or (iii), all the respondents, irrespective of whether they possess attribute A or not, are supposed to say "Yes" or "No" respectively. It is to be mentioned that the random device R_3 is due to Tracy and Osahan [27]. The survey procedures are performed under the assumption that both the sensitive and innocuous questions are unrelated and independent in a random device R_1. To protect the respondents' privacy, the respondents should not disclose to the interviewer the question they answered from either R_1 or R_2 or R_3. 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. Since all respondents using a random device R_1 already responded "Yes" from the initial direct question.

The probability 'Y' of getting "Yes" answers from the respondents using random device R_1 is given by

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

where π_s is the proportion of "Yes" answer from the sensitive trait group and π_1 is the proportion of "Yes" so that $(\pi_1 = 1)$ answer from the innocuous question

$$Y = P_1\pi_s + (1-P_1)\,. \tag{2.2}$$

The probability 'Y^*' of getting "No" answers from the respondents using random device R_1 is given as

$$Y^* = 1 - \big[P_1\pi_s + (1-P_1)\big]\,. \tag{2.3}$$

Thus the maximum likelihood function is given by

$$L = \binom{n}{n_1}\big[P_1\pi_s + (1-P_1)\big]^{n_1}\big[P_1(1-\pi_s)\big]^{(n-n_1)}\,. \tag{2.4}$$

Taking log on the both sides of equation (2.4):

$$\log L = \log\binom{n}{n_1} + n_1\log\big[P_1\pi_s + (1-P_1)\big] + (n-n_1)\log\big[P_1(1-\pi_s)\big]\,. \tag{2.5}$$

Differentiating on both sides of equation (2.5) with respect to π_s and equating to zero, we have

$$P_1\pi_s + (1-P_1) = \frac{n_1}{n}\,. \tag{2.6}$$

This is maximum likelihood estimator of Y.

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

$$\hat{\pi}_a = \frac{\hat{Y}-(1-P_1)}{P_1}\,, \tag{2.7}$$

where $\hat{Y}$ is the sample proportion of "Yes" response, thus expected value of $\hat{\pi}_a$ is

$$E(\hat{\pi}_a) = \frac{E(\hat{Y})-(1-P_1)}{P_1} = \pi_s\,. \tag{2.8}$$

The variance of $\hat{\pi}_a$ is obtained as

$$V(\hat{\pi}_a) = \frac{1}{n_1}\left[\pi_s(1-\pi_s) + \frac{(1-\pi_s)\,(1-P_1)}{P_1}\right]. \tag{2.9}$$

The probability X of "Yes" answers from the respondents using random devices R_2 and R_3 is given as

$$X = T\pi_s + (1-T)\left[P\pi_s + \frac{(1-P)}{2}\right]. \tag{2.10}$$

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

$$\hat{\pi}_b = \frac{\hat{X}-(1-T)\,\frac{(1-P)}{2}}{T+P(1-T)}\,. \tag{2.11}$$

The variance of unbiased estimator $\hat{\pi}_b$ is obtained as

$$V(\hat{\pi}_b) = \frac{1}{n_2}\left[\pi_s(1-\pi_s) + \frac{(1-T)\,(1-P)\,\big[2-(1-T)\,(1-P)\big]}{4\big[T+P(1-T)\big]^2}\right]. \tag{2.12}$$

The estimator of π_s, in the terms of the sample proportion of "Yes" response $\hat{\pi}_a$ and $\hat{\pi}_b$, is

$$\hat{\pi}_{A1} = \left(\frac{n_1}{n}\right)\hat{\pi}_a + \left(\frac{n_2}{n}\right)\hat{\pi}_b\,, \qquad \text{for} \quad 0 < \frac{n_1}{n} < 1\,. \tag{2.13}$$

Since $\hat{\pi}_a$ and $\hat{\pi}_b$ are unbiased estimators, therefore the expected value of $\hat{\pi}_{A1}$ is

$$E(\hat{\pi}_{A1}) = \frac{n_1}{n}\,E(\hat{\pi}_a) + \frac{(n-n_1)}{n}\,E(\hat{\pi}_b) = \frac{n_1}{n}\,\pi_s + \frac{(n-n_1)}{n}\,\pi_s = \pi_s\,. \tag{2.14}$$

Thus, the proposed estimator $\hat{\pi}_{A1}$ is an unbiased estimator of π_s.

Since the random device R_1 and Tracy and Osahan [27] randomized response technique (consists of two random devices R_2 and R_3) used are independent. We derive the expression of variance of $\hat{\pi}_{A1}$ as

$$
\begin{aligned}
V(\hat{\pi}_{A1}) &= \frac{n_1^2}{n^2} V(\hat{\pi}_a) + \frac{n_2^2}{n^2} V(\hat{\pi}_b) \\
&= \frac{n_1}{n^2}\left[\frac{(1-\pi_s)\left[P_1\pi_s + (1-P_1)\right]}{P_1}\right] \\
&\quad + \frac{n_2}{n^2}\left[\pi_s(1-\pi_s) + \frac{(1-T)(1-P)\left[2-(1-T)(1-P)\right]}{4\left[T+P(1-T)\right]^2}\right].
\end{aligned} \tag{2.15}
$$

Under the circumstances that the Warner [28] and Simmons *et al.* [20] method (known π_1) are equally confidential to respondents, Lanke [14] obtain a unique value of P as $P = 1/2 + P_1/\left[2P_1 + 4(1-P_1)\pi_1\right]$, for every P_1 and every π_1.

Since our proposed mixed model also use Simmons *et al.* [20] method when $\pi_1 = 1$, we may apply Lanke [14] technique in our proposed model. Thus we get

$$
P = \frac{1}{(2-P_1)}\,. \tag{2.16}
$$

Putting $P = 1/(2-P_1)$ in equation (2.12), we get

$$
\begin{aligned}
V(\hat{\pi}_b) &= \frac{1}{n_2}\left[\pi_s(1-\pi_s) + \frac{(1-T)\left(1-\frac{1}{(1-2P_1)}\right)\left[2-(1-T)\left(1-\frac{1}{(2-P_1)}\right)\right]}{4\left[T+\frac{1}{(2-P_1)}(1-T)\right]^2}\right] \\
&= \left[\frac{\pi_s(1-\pi_s)}{n_2} + \frac{(1-T)(1-P_1)\left[2(2-P_1)-(1-T)(1-P_1)\right]}{4n_2\left[1+T(1-P_1)\right]^2}\right].
\end{aligned} \tag{2.17}
$$

Thus, we have the following theorem.

Theorem 2.1. *The variance of $\hat{\pi}_{A1}$ is given by*

$$
\begin{aligned}
V(\hat{\pi}_{A1}) &= \frac{\pi_s(1-\pi_s)}{n} \\
&\quad + \frac{1}{n}\left[\frac{\lambda(1-\pi_s)(1-P_1)}{P_1} + \frac{(1-\lambda)(1-T)(1-P_1)\left[2(2-P_1)-(1-T)(1-P_1)\right]}{4\left[1+T(1-P_1)\right]^2}\right],
\end{aligned} \tag{2.18}
$$

for $n = n_1 + n_2$ and $\lambda = n_1/n$.

2.1. Efficiency comparison

In this section, the comparison of the proposed model under completely truthful reporting case has been made with Kim and Warde [13] model.

From Kim and Warde [13] model, we have

$$V(\hat{\pi}_{kw}) = \frac{\pi_s(1-\pi_s)}{n} + \frac{(1-P_1)\left[\lambda P_1(1-\pi_s)+(1-\lambda)\right]}{n P_1^2}\,. \tag{2.19}$$

The estimator $\hat{\pi}_{A1}$ is always more efficient than that of Kim and Warde [13] estimator $\hat{\pi}_{kw}$ if

$$V(\hat{\pi}_{kw}) > V(\hat{\pi}_{A1})\,,$$

which gives the conditions, when

$$\left[4\big(1+T(1-P_1)\big)^2 - P_1^2(1-T)\big(3+T(1-P_1)-P_1\big)\right] > 0\,.$$

To have a tangible idea about the performance of the proposed estimator $\hat{\pi}_{A1}$ over Kim and Warde [13] estimator $\hat{\pi}_{kw}$, we compute the percent relative efficiency $\mathrm{PRE}(\hat{\pi}_{A1},\hat{\pi}_{kw})$ for $\lambda=(0.7,0.5,0.3)$, $n=1000$ and for different values of T, π_s, n_1, n_2 and P_1, and presented in Table 1:

$$\mathrm{PRE}(\hat{\pi}_{A1},\hat{\pi}_{kw}) = \frac{V(\hat{\pi}_{kw})}{V(\hat{\pi}_{A1})}\times 100\,. \tag{2.20}$$

Table 1: Percent relative efficiency of the proposed estimator $\hat{\pi}_{A1}$ with respect to Kim and Warde [13] 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	554.04	313.85	232.53	191.00	165.40	147.71	134.41	123.50	113.26
	500	500	0.5	0.5	1161.80	603.05	414.50	318.32	258.82	217.29	185.45	158.55	132.55
	300	700	0.3	0.9	2581.70	1278.40	838.29	613.08	472.76	373.61	296.28	230.14	167.51
0.2	700	300	0.7	0.1	601.54	331.57	240.35	193.90	165.40	145.84	131.31	119.75	109.77
	500	500	0.5	0.5	1266.60	638.56	427.09	319.64	253.65	208.18	174.11	146.69	122.86
	300	700	0.3	0.9	2794.00	1330.20	838.29	589.05	436.59	332.21	254.93	194.15	143.81
0.3	700	300	0.7	0.1	662.49	354.70	251.19	198.87	167.08	145.58	129.96	117.97	108.28
	500	500	0.5	0.5	1401.90	686.31	446.91	326.51	253.65	204.49	168.75	141.24	118.97
	300	700	0.3	0.9	3073.90	1410.90	858.93	584.52	421.02	312.88	236.26	179.24	135.16
0.4	700	300	0.7	0.1	743.37	385.71	266.30	206.54	170.71	146.89	129.96	117.37	107.67
	500	500	0.5	0.5	1567.20	739.28	466.33	331.77	252.32	200.28	163.76	136.86	116.27
	300	700	0.3	0.9	3454.00	1531.80	903.88	598.39	421.02	307.13	229.09	173.12	131.60
0.5	700	300	0.7	0.1	855.64	428.97	287.89	218.17	176.97	150.05	131.31	117.72	107.61
	500	500	0.5	0.5	1834.20	844.82	520.90	362.66	270.27	210.52	169.25	139.40	117.08
	300	700	0.3	0.9	3993.00	1713.50	982.27	634.07	436.59	312.88	230.23	172.49	130.82

It is observed from Table 1 and Figure 1 that:

(**a**) For all the parametric combinations, the values of percent relative efficiencies are substantially exceeding 100, which indicate that the proposed estimator $\hat{\pi}_{A1}$ is uniformly better than Kim and Warde [13] estimator $\hat{\pi}_{kw}$.

(**b**) It may also be seen that the values of the percent relative efficiencies decrease with the increasing values of P_1. However, the values of the percent relative efficiencies are showing increasing trend with the decreasing values of λ when the values of P_1 are fixed.

(**c**) From Figure 1 it may be observed that there is a large gain in efficiency by using the proposed estimator $\hat{\pi}_{A1}$ over Kim and Warde [13] estimator $\hat{\pi}_{kw}$, when the proportion of stigmatizing attribute is moderately large.

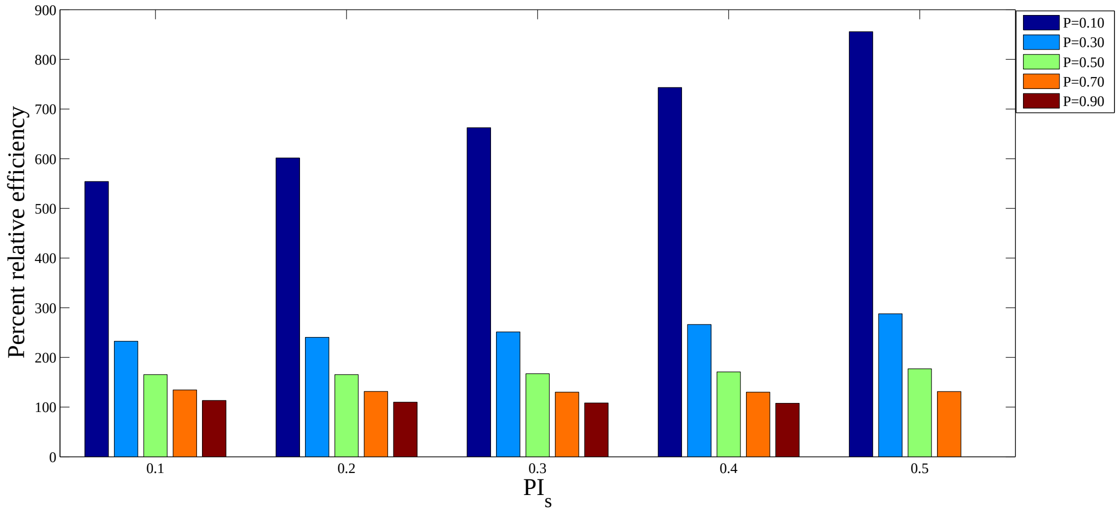

Figure 1: Percent relative efficiency of the proposed estimator $\hat{\pi}_{A1}$ with respect to Kim and Warde [13] estimator $\hat{\pi}_{kw}$ when $T = 0.1$ and $\lambda = 0.7$.

3. LESS THAN COMPLETELY TRUTHFUL REPORTING

Various authors including Mangat [15], Tracy and Osahan [27], Chang and Huang [2], Chang *et al.* [3], Kim and Warde [12], Kim and Elam [11], Nazuk and Shabbir [18] and cited therein has been consider the problem of "Less than completely truthful reporting" in RR technique. It is reasonably assumed that the persons who belong to sensitive trait group state truthful answers with probabilities T_1, T_2 and T_3 in random devices R_1, R_2 and R_3 respectively. The respondents in the non-sensitive group have no reason to tell a lie, they may lie for the sensitive group.

Since all respondents using a random device R_1 already responded "Yes" from the initial direct question, therefore $\pi_1 = 1$ in R_1. Thus, the probability Y' of "Yes" answer for the random device R_1 is given by

$$Y' = P_1 \pi_s T_1 + (1 - P_1)\,. \tag{3.1}$$

An estimator of π_s, in term of the sample proportion of "Yes" responses is given as

$$\hat{\pi}_{a(1)} = \frac{\hat{Y'} - (1-P_1)}{P_1}\,. \tag{3.2}$$

Since each $\hat{Y'}$ follows Binomial distribution $B(n_1, Y')$, therefore the estimator $\hat{\pi}_{a(1)}$ has the following bias and mean square error (MSE):

$$B\big(\hat{\pi}_{a(1)}\big) = \pi_s (T_1 - 1) \tag{3.3}$$

and

$$V\big(\hat{\pi}_{a(1)}\big) = \frac{Y'(1-Y')}{n_1 P_1^2} = \frac{(1-\pi_s T_1)\left[1 - P_1(1-\pi_s T_1)\right]}{n_1 P_1}\,. \tag{3.4}$$

Thus, the MSE of $\hat{\pi}_{a(1)}$ is given by

$$\begin{aligned} \mathrm{MSE}\big(\hat{\pi}_{a(1)}\big) &= V\big(\hat{\pi}_{a(1)}\big) + \left[B\big(\hat{\pi}_{a(1)}\big)\right]^2 \\ &= \frac{(1-\pi_s T_1)\left[1 - P_1(1-\pi_s T_1)\right]}{n_1 P_1} + \pi_s^2 (T_1 - 1)^2\,. \end{aligned} \tag{3.5}$$

On the basis of the proposed procedure, the probability for the respondents who response "Yes" answer using random devices R_2 and R_3 is given by

$$X' = T\pi_s T_2 + (1-T)\left[P\,\pi_s T_3 + \frac{(1-P)}{2}\right]. \tag{3.6}$$

By the method of moments, an estimator of population proportion π_s is obtained as

$$\hat{\pi}_{b(1)} = \frac{\hat{X'} - (1-T)\,\frac{(1-P)}{2}}{T + P(1-T)}\,. \tag{3.7}$$

In random devices R_2 and R_3, the same sensitive question is asked from the respondents who belong to rare sensitive group in the sample, so we take $T_2 = T_3$ in our case which is unlike as in case of Kim and Elam [11].

Since each $\hat{X'}$ follows Binomial distribution $B(n_1, X')$, therefore the estimator $\hat{\pi}_{b(1)}$ has the following bias and MSE:

$$B\big(\hat{\pi}_{b(1)}\big) = \pi_s (T_2 - 1) \tag{3.8}$$

and

$$\begin{aligned} V\big(\hat{\pi}_{b(1)}\big) &= \frac{X'(1-X')}{n_2\left[T + P(1-T)\right]^2} \\ &= \frac{\pi_s T_2\,(1-\pi_s T_2)}{n_2} + \frac{(1-T)\,(1-P)\left[2 - (1-T)\,(1-P)\right]}{4\,n_2\left[T + P(1-T)\right]^2}\,. \end{aligned} \tag{3.9}$$

Therefore, the MSE of $\hat{\pi}_{b(1)}$ is given by

$$\begin{aligned} \mathrm{MSE}\big(\hat{\pi}_{b(1)}\big) &= V\big(\hat{\pi}_{b(1)}\big) + \left[B\big(\hat{\pi}_{b(1)}\big)\right]^2 \\ &= \frac{\pi_s T_2\,(1-\pi_s T_2)}{n_2} + \frac{(1-T)\,(1-P)\left[2 - (1-T)\,(1-P)\right]}{4\,n_2\left[T + P(1-T)\right]^2} + \pi_s^2 (T_2 - 1)^2\,. \end{aligned} \tag{3.10}$$

Now, we propose the estimator for population proportion π_s in the terms of the sample proportion of "Yes" response $\hat{\pi}_{a(1)}$ and $\hat{\pi}_{b(1)}$ as

$$\hat{\pi}_A = \left(\frac{n_1}{n}\right)\hat{\pi}_{a(1)} + \left(\frac{n_2}{n}\right)\hat{\pi}_{b(1)} \qquad \text{for} \quad 0 < \frac{n_1}{n} < 1\,, \tag{3.11}$$

where $n_1 + n_2 = 1$.

Since both the estimators $\hat{\pi}_{a(1)}$ and $\hat{\pi}_{b(1)}$ are bias estimator of π_s, therefore the bias of $\hat{\pi}_A$ is given by

$$B(\hat{\pi}_A) = \pi_s\left[\left(\frac{n_1}{n}\right)(T_1-1) + \left(\frac{n_2}{n}\right)(T_2-1)\right], \tag{3.12}$$

and

$$\begin{aligned} \mathrm{MSE}(\hat{\pi}_A) &= \frac{\lambda(1-\pi_s T_1)\left[1-P_1(1-\pi_s T_1)\right]}{n P_1} \\ &+ \frac{(1-\lambda)}{n}\left[\pi_s T_2\,(1-\pi_s T_2) + \frac{(1-T)\,(1-P)\left[2-(1-T)\,(1-P)\right]}{4\left[T+P(1-T)\right]^2}\right] \\ &+ \pi_s^2\left[\lambda^2\,(T_1-1)^2 + (1-\lambda)^2\,(T_2-1)^2\right]. \end{aligned} \tag{3.13}$$

Inserting Lanke [14] a unique value $P = 1/(2-P_1)$ in equation (3.10), we get

$$\begin{aligned} \mathrm{MSE}\big(\hat{\pi}_b(1)\big) &= \frac{\pi_s T_2\,(1-\pi_s T_2)}{n_2} \\ &+ \frac{(1-T)\,(1-P_1)\left[2\,(2-P_1)-(1-T)\,(1-P_1)\right]}{4\,n_2\left[1+T(1-P_1)\right]^2} + \pi_s^2\,(T_2-1)^2\,. \end{aligned} \tag{3.14}$$

Thus, we have the following theorem.

Theorem 3.1. *The MSE of $\hat{\pi}_A$ is given by*

$$\begin{aligned} \mathrm{MSE}(\hat{\pi}_A) &= \frac{\pi_s\left[\lambda T_1\,(1-\pi_s T_1) + (1-\lambda)\,T_2\,(1-\pi_s T_2)\right]}{n} \\ &+ \frac{(1-P_1)}{n}\left[\frac{\lambda(1-\pi_s T_1)}{P_1} + \frac{(1-\lambda)\,(1-T)\left[2\,(2-P_1)-(1-T)\,(1-P_1)\right]}{4\left[1+T(1-P_1)\right]^2}\right] \\ &+ \pi_s^2\left[\lambda^2\,(T_1-1)^2 + (1-\lambda)^2\,(T_2-1)^2\right], \end{aligned} \tag{3.15}$$

for $n = n_1 + n_2$ and $\lambda = n_1/n$.

3.1. Efficiency comparison

We compare the proposed model with Kim and Warde [13] model, under "Less than completely truthful reporting" situation.

The MSE of Kim and Warde [13] estimator $\hat{\pi}_{kw}$ under less than completely truthful reporting is given as

$$\begin{aligned} \text{MSE}(\hat{\pi}_{kw}) &= \frac{\pi_s\left[\lambda T_1(1-\pi_s T_1)+(1-\lambda)\,T_2(1-\pi_s T_2)\right]}{n} \\ &\quad + \frac{(1-P_1)\left[\lambda P_1(1-\pi_s T_1)+(1-\lambda)\right]}{n P_1^2} \\ &\quad + \pi_s^2\left[\lambda^2(T_1-1)^2+(1-\lambda)^2(T_2-1)^2\right]. \end{aligned} \tag{3.16}$$

The estimator $\hat{\pi}_A$ is always more efficient than that of Kim and Warde [13] estimator $\hat{\pi}_{kw}$ if

$$\text{MSE}(\hat{\pi}_{kw}) > \text{MSE}(\hat{\pi}_A)\,,$$

which is true if

$$\left[\frac{(1-T)\left[2(2-P_1)-(1-T)(1-P_1)\right]}{4\left[1+T(1-P_1)\right]^2}-\frac{1}{P_1^2}\right] > 0\,. \tag{3.17}$$

To have an idea about the magnitude of the percent relative efficiency of the proposed model in relation to Kim and Warde [13] model, we resort to an empirical investigation for $\lambda = (0.7, 0.5, 0.3)$, $n = 1000$, $T_1\,(T_2) = 0.7, 0.8, 0.9\ (0.6, 0.7, 0.8)$ and for different values of T, π_s, n_1, n_2 and P_1. The percent relative efficiency of the proposed estimator $\hat{\pi}_A$ with respect to Kim and Warde [13] estimator $\hat{\pi}_{kw}$ is defined as

$$\text{PRE}(\hat{\pi}_A, \hat{\pi}_{kw}) = \frac{\text{MSE}(\hat{\pi}_{kw})}{\text{MSE}(\hat{\pi}_A)} \times 100\,. \tag{3.18}$$

The following interpretations may be read out from Table 2 and Figure 2:

(**a**) For all the parametric combinations, the values of percent relative efficiencies are substantially exceeding 100, which indicate that the proposed estimator $\hat{\pi}_A$ is uniformly better than Kim and Warde [13] estimator $\hat{\pi}_{kw}$.

(**b**) Table 2 makes it visible that the values of percent relative efficiencies decrease with the increasing values of P_1. Further, we observe that the percent relative efficiencies increase with the decreasing values of λ (and increasing values of T_1, T_2) when the values of P_1 are fixed.

(**c**) It may also be seen that with the increase in the values of π_s there is the decreasing pattern in values of the percent relative efficiencies for fix values of P_1.

(**d**) From Figure 2 it is clear that there is less gain in the efficiency by using the proposed estimator $\hat{\pi}_A$ over Kim and Warde [13] estimator $\hat{\pi}_{kw}$, when the proportion of sensitive attribute is moderately large.

Table 2: Percent relative efficiency of the proposed estimator $\hat{\pi}_A$ with respect to Kim and Warde [13] estimator $\hat{\pi}_{kw}$ under the situation of "Less than completely truthful reporting".

π_s	$n=1000$		λ	T	T_1	T_2	P_1						
	n_1	n_2					0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	700	300	0.7	0.1	0.7	0.6	197.29	159.83	137.91	123.84	114.33	107.74	103.14
	500	500	0.5	0.5	0.8	0.7	347.62	257.62	203.92	168.40	143.33	124.85	110.82
	300	700	0.3	0.9	0.9	0.8	681.61	470.63	345.02	262.04	203.38	159.88	126.44
0.2	700	300	0.7	0.1	0.7	0.6	155.82	129.78	116.64	109.33	105.04	102.46	100.90
	500	500	0.5	0.5	0.8	0.7	253.81	185.33	149.67	129.12	116.52	108.51	103.33
	300	700	0.3	0.9	0.9	0.8	450.78	294.67	214.06	167.73	139.19	120.74	108.41
0.3	700	300	0.7	0.1	0.7	0.6	132.48	116.20	108.61	104.65	102.43	101.16	100.41
	500	500	0.5	0.5	0.8	0.7	194.00	148.49	126.73	115.02	108.23	104.11	101.57
	300	700	0.3	0.9	0.9	0.8	312.22	209.97	161.31	135.06	119.69	110.16	104.03
0.4	700	300	0.7	0.1	0.7	0.6	120.45	109.88	105.14	102.73	101.41	100.66	100.23
	500	500	0.5	0.5	0.8	0.7	160.79	130.25	116.28	108.98	104.85	102.40	100.90
	300	700	0.3	0.9	0.9	0.8	236.93	168.65	137.44	121.08	111.69	105.98	102.35
0.5	700	300	0.7	0.1	0.7	0.6	113.84	106.58	103.39	101.79	100.92	100.43	100.15
	500	500	0.5	0.5	0.8	0.7	141.78	120.40	110.84	105.93	103.18	101.56	100.58
	300	700	0.3	0.9	0.9	0.8	194.13	146.38	125.02	113.98	107.71	103.92	101.54

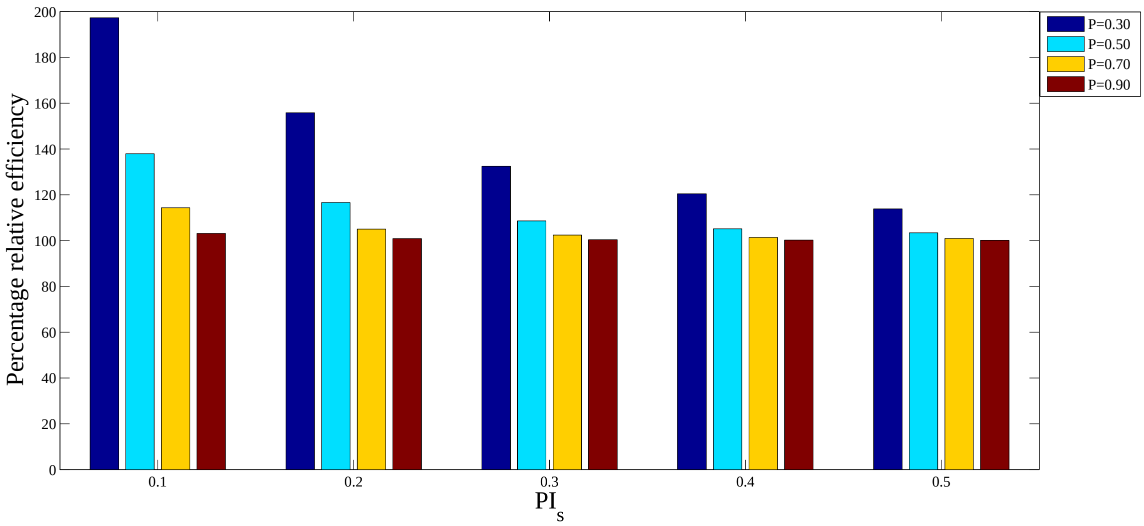

Figure 2: Percent relative efficiency of the proposed estimator $\hat{\pi}_A$ with respect to Kim and Warde [13] estimator $\hat{\pi}_{kw}$ under the condition of "Less than completely truthful reporting", when $T=0.1$ and $\lambda=0.7$.

4. A MIX RANDOMIZED RESPONSE MODEL USING STRATIFICATION

4.1. A mixed Stratified randomized response (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. The main advantage of the stratified random sampling is that the technique overcomes the limitation of the loss of individual characteristics of the respondents. A randomized response (RR) technique using stratified random sampling yields the group characteristics associated to each stratum estimator. Also, stratified random sampling protects a researcher from the possibility of obtaining a poor sample. Hong *et al.* [9] suggested a stratified RR technique using a proportional allocation. Kim and Warde [12] proposed a stratified randomized response model using an optimum allocation which is more efficient than that using a proportional allocation. Kim and Elam [11] suggested a two stage stratified Warner's RR model using optimal allocation. Further Kim and Warde [13] suggested a mixed stratified RR model.

In the proposed models, we assume that the population is partitioned into strata, and a sample is selected by using simple random sampling with replacement (SRSWR) scheme from 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 the respondent is instructed to go to the random device R_{k1} consisting of statements:

(i) "I am a member of the sensitive trait group",

(ii) "I am a member of the innocuous trait group",

with probabilities Q_k and $(1-Q_k)$ respectively. If a respondent answers "No", then the respondent is instructed to use the random device R_{k2} consisting of two statements (see Mangat and Singh [16]):

(i) "Do you possess the sensitive attribute A", with probability T_k,

(ii) "Go to the third random device R_{k3} in the second stage", with probability $(1-T_k)$.

The random device R_{k3} at the second stage consists of three statements:

(i) "I possess the sensitive attribute A",

(ii) "Yes",

(iii) "No",

with probabilities P_k, $(1-P_k)/2$ and $(1-P_k)/2$. When the outcome of random device R_{k3} is either (ii) or (iii), all the respondents, irrespective of whether they possess attribute A or not, are supposed to say "Yes" or "No" respectively. 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}. Let m_k denote 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 are set by researcher. Thus, the probability Y_k of "Yes" answer from the respondents using the random device R_{k1} is given by

$$Y_k = Q_k\,\pi_{sk} + (1 - Q_k)\,\pi_{1k} \qquad \text{for} \quad k = 1, 2, ..., r\,, \tag{4.1}$$

where π_{sk} is the proportion of respondents with the sensitive trait in stratum k, π_{1k} is the proportion of respondents with the innocuous trait group in stratum k.

Since the respondent performing a random device R_{k1} answered "Yes" to the direct question of the innocuous trait, if the respondent selects the same innocuous question from R_{k1}, then $\pi_{1k} = 1$ (see Kim and Warde [13]). Therefore, equation (4.1) becomes

$$Y_k = Q_k\,\pi_{sk} + (1 - Q_k) \qquad \text{for} \quad k = 1, 2, ..., r\,. \tag{4.2}$$

An unbiased estimator of π_{sk} is given as

$$\hat{\pi}_{a1k} = \frac{\hat{Y}_k - (1 - Q_k)}{Q_k} \qquad \text{for} \quad k = 1, 2, ..., r\,, \tag{4.3}$$

where $\hat{Y}_k$ is the proportion of "Yes" answer in a sample in stratum k. Since each $\hat{Y}_k$ follows Binomial distribution i.e. $\hat{Y}_k \sim B(m_{k1}, Y_k)$.

The variance of unbiased estimator $\hat{\pi}_{a1k}$ is given by

$$V(\hat{\pi}_{a1k}) = \frac{(1 - \pi_{sk})\left[Q_k\,\pi_{sk} + (1 - Q_k)\right]}{m_{k1}\,Q_k}\,. \tag{4.4}$$

The probability X_k of "Yes" answer from the respondents using random devices R_{k2} and R_{k3} will be

$$X_k = T_k\,\pi_{sk} + (1 - T_k)\left[P_k\,\pi_{sk} + \frac{(1 - P_k)}{2}\right], \tag{4.5}$$

where π_{sk} is the proportion of respondents with the sensitive treat in stratum k.

An unbiased estimator of π_{sk} is given by

$$\hat{\pi}_{b1k} = \frac{\hat{X}_k - (1 - T_1)\,\frac{(1-P_k)}{2}}{T_k + P_k(1 - T_k)}\,, \tag{4.6}$$

where $\hat{X}_k$ is the proportion of "Yes" responses in a sample from a stratum k. Since each $\hat{X}_k$ follows Binomial distribution i.e. $\hat{X}_k \sim B(m_{k1}, X_k)$. By using $m_k = m_{k1} + m_{k2}$ and $P_k = (2 - Q_k)^{-1}$ (see Lanke [14]), the variance of estimator $\hat{\pi}_{b1k}$ is given by

$$V(\hat{\pi}_{b1k}) = \left[\frac{\pi_{sk}(1 - \pi_{sk})}{m_{k2}} + \frac{(1 - T_k)\,(1 - Q_k)\left[2\,(2 - Q_k) - (1 - T_k)\,(1 - Q_k)\right]}{4\,m_2\left[1 + T_k(1 - Q_k)\right]^2}\right]. \tag{4.7}$$

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

$$\hat{\pi}_{msk} = \left(\frac{m_{k1}}{m_k}\right)\hat{\pi}_{a1k} + \left(\frac{m_{k2}}{m_k}\right)\hat{\pi}_{b1k} \qquad \text{for} \quad 0 < \frac{m_{k1}}{m_k} < 1\,. \tag{4.8}$$

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

$$\begin{aligned} V(\hat{\pi}_{msk}) = \Bigg[& \frac{\pi_{sk}(1-\pi_{sk})}{m_k} + \frac{\lambda_k(1-\pi_{sk})(1-Q_k)}{m_k Q_k} \\ & + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)\left[2(2-Q_k)-(1-T_k)(1-Q_k)\right]}{4\,m_k\left[1+T_k(1-Q_k)\right]^2}\Bigg], \end{aligned} \tag{4.9}$$

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

Thus, the unbiased estimator of $\pi_s = \sum_{k=1}^{r} w_k \pi_{sk}$ is obtained as

$$\hat{\pi}_{Ak} = \sum_{k=1}^{r} w_k\,\hat{\pi}_{msk} = \sum_{k=1}^{r} w_k\left[\frac{m_{k1}}{m_k}\,\hat{\pi}_{a1k} + \frac{m_{k2}}{m_k}\,\hat{\pi}_{b1k}\right], \tag{4.10}$$

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

$$\begin{aligned} V(\hat{\pi}_{Ak}) = \sum_{k=1}^{r} \frac{w_k^2}{m_k}\Bigg[& \pi_{sk}(1-\pi_{sk}) + \frac{\lambda_k(1-\pi_{sk})(1-Q_k)}{Q_k} \\ & + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)\left[2(2-Q_k)-(1-T_k)(1-Q_k)\right]}{4\left[1+T_k(1-Q_k)\right]^2}\Bigg]. \end{aligned} \tag{4.11}$$

Here, the requirement of doing the optimal allocation of a sample size n, we need to know $\lambda_k = m_{k1}/m_k$ and π_{sk}. In practice the information on $\lambda_k = m_{k1}/m_k$ and π_{sk} is usually unavailable. But if prior information about $\lambda_k = m_{k1}/m_k$ and π_{sk} is available from past experience, it will help to derive the following optimal allocation formula.

Theorem 4.1. *The optimum allocation of m to $m_1, m_2, ..., m_{r-1}$ and m_r to derive the minimum variance of the $\hat{\pi}_{Ak}$ subject to $n = \sum_{k=1}^{r} m_k$ is approximately given by*

$$\frac{m_k}{n} = \frac{A}{B}\,, \tag{4.12}$$

where

$$\begin{aligned} A = w_k\Bigg[& \pi_{sk}(1-\pi_{sk}) + \frac{\lambda_k(1-\pi_{sk})(1-Q_k)}{Q_k} \\ & + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)\left[2(2-Q_k)-(1-T_k)(1-Q_k)\right]}{4\left[1+T_k(1-Q_k)\right]^2}\Bigg]^{\frac{1}{2}}, \end{aligned}$$

$$\begin{aligned} B = \sum_{k=1}^{r} w_k\Bigg[& \pi_{sk}(1-\pi_{sk}) + \frac{\lambda_k(1-\pi_{sk})(1-Q_k)}{Q_k} \\ & + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)\left[2(2-Q_k)-(1-T_k)(1-Q_k)\right]}{4\left[1+T_k(1-Q_k)\right]^2}\Bigg]^{\frac{1}{2}}. \end{aligned}$$

Thus, the minimal variance of the estimator $\hat{\pi}_{Ak}$ is given by

$$V(\hat{\pi}_{Ak}) = \frac{1}{n}\left[\sum_{k=1}^{r} w_k \left[\pi_{sk}(1-\pi_{sk}) + \frac{\lambda_k(1-\pi_{sk})(1-Q_k)}{Q_k} + \frac{(1-\lambda_k)(1-T_k)(1-Q_k)\left[2(2-Q_k)-(1-T_k)(1-Q_k)\right]}{4\left[1+T_k(1-Q_k)\right]^2}\right]^{\frac{1}{2}}\right]^2, \tag{4.13}$$

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

4.2. Efficiency comparison

To show the efficacious performance of the proposed stratified mixed randomized response model, we examine the efficiency comparison of the proposed estimator $\hat{\pi}_{Ak}$ over the proposed mixed randomized estimator $\hat{\pi}_{A1}$ and Kim and Warde [13] estimator $\hat{\pi}_{kw}$ respectively. The comparisons are given in the form of following theorems.

Theorem 4.2. *Suppose there are two strata (i.e. $k=2$) in the population and $\lambda = m_{k1}/m_k$. The proposed stratified estimator $\hat{\pi}_{Ak}$ is always more efficient than that of usual proposed estimator $\hat{\pi}_{A1}$ where $P_1 = Q_1 = Q_2,\ \lambda=\lambda_1=\lambda_2$ and $T=T_1=T_2$.*

Proof: Under the assumption $k=2,\ P_1=Q_1=Q_2,\ \lambda=\lambda_1=\lambda_2$ and $T=T_1=T_2$, the equation (4.13) can be rewritten as

$$V(\hat{\pi}_{Ak}) = \frac{1}{n}\left[w_1\left[\pi_{s1}(1-\pi_{s1}) + \frac{\lambda(1-\pi_{s1})(1-P_1)}{P_1} + \frac{(1-\lambda)(1-T)(1-P_1)\left[2(2-P_1)-(1-T)(1-P_1)\right]}{4\left[1+T(1-P_1)\right]^2}\right]^{\frac{1}{2}} + w_2\left[\pi_{s2}(1-\pi_{s2}) + \frac{\lambda(1-\pi_{s2})(1-P_1)}{P_1} + \frac{(1-\lambda)(1-T)(1-P_1)\left[2(2-P_1)-(1-T)(1-P_1)\right]}{4\left[1+T(1-P_1)\right]^2}\right]^{\frac{1}{2}}\right]^2. \tag{4.14}$$

If we denote

$$a_1 = \frac{(1-\pi_{s1})(1-P_1)}{P_1},$$

$$a_2 = \frac{(1-\pi_{s2})(1-P_1)}{P_1},$$

$$b = \frac{(1-\lambda)(1-T)(1-P_1)\left[2(2-P_1)-(1-T)(1-P_1)\right]}{4\left[1+T(1-P_1)\right]^2},$$

we can write equation (4.14) as

$$V(\hat{\pi}_{Ak}) = \frac{1}{n}\left[w_1\Big[\pi_{s1}(1-\pi_{s1}) + \lambda a_1 + b\Big]^{\frac{1}{2}} + w_2\Big[\pi_{s2}(1-\pi_{s2}) + \lambda a_2 + b\Big]^{\frac{1}{2}}\right]^2. \tag{4.15}$$

From equation (2.18), we have

$$V(\hat{\pi}_{A1}) = \frac{1}{n}\Big[(w_1\pi_{s1} + w_2\pi_{s2})\,(1 - w_1\pi_{s1} - w_2\pi_{s2}) + \lambda(w_1 a_1 + w_2 a_2) + b\Big]. \tag{4.16}$$

Now, subtracting equation (4.15) from equation (4.16), we have

$$\begin{aligned}
n\Big[V(\hat{\pi}_{A1}) - V(\hat{\pi}_{Ak})\Big] &= \Big[(w_1\pi_{s1} + w_2\pi_{s2})\,(1 - w_1\pi_{s1} - w_2\pi_{s2}) + \lambda(w_1 a_1 + w_2 a_2) + b\Big] \\
&\quad - \left[w_1\Big[\pi_{s1}(1-\pi_{s1}) + \lambda a_1 + b\Big]^{\frac{1}{2}} + w_2\Big[\pi_{s2}(1-\pi_{s2}) + \lambda a_2 + b\Big]^{\frac{1}{2}}\right]^2 \\
&= w_1\pi_{s1} + w_2\pi_{s2} - 2\,w_1 w_2 \pi_{s1}\pi_{s2} - w_1^2\pi_{s1} - w_2^2\pi_{s2} \\
&\quad - w_1^2(\lambda a_1 + b) - w_2^2(\lambda a_2 + b) + \lambda(w_1 a_1 + w_2 a_2) + b \\
&\quad - 2\,w_1 w_2\Big[\pi_{s1}(1-\pi_{s1}) + \lambda a_1 + b\Big]^{\frac{1}{2}}\Big[\pi_{s2}(1-\pi_{s2}) + \lambda a_2 + b\Big]^{\frac{1}{2}} \\
&= w_1(\pi_{s1} + \lambda a_1) + w_2(\pi_{s2} + \lambda a_2) \\
&\quad - w_1^2(\pi_{s1} + \lambda a_1 + b) - w_2^2(\pi_{s2} + \lambda a_2 + b) - 2\,w_1 w_2 \pi_{s1}\pi_{s2} + b \\
&\quad - 2\,w_1 w_2\Big[\pi_{s1}(1-\pi_{s1}) + \lambda a_1 + b\Big]^{\frac{1}{2}}\Big[\pi_{s2}(1-\pi_{s2}) + \lambda a_2 + b\Big]^{\frac{1}{2}} \\
&> 0\,,
\end{aligned}$$

which proves the theorem. □

Theorem 4.3. *Suppose there are two strata (i.e.* $k = 2$*) in the population and* $\lambda = m_{k1}/m_k$*. The proposed stratified estimator* $\hat{\pi}_{Ak}$ *is always more efficient than that of Kim and Warde [13] estimator* $\hat{\pi}_{kw}$ *where* $P_1 = Q_1 = Q_2,\ \lambda = \lambda_1 = \lambda_2$ *and* $T = T_1 = T_2$*.*

Proof: Under the assumption $P_1 = Q_1 = Q_2$, $\lambda = \lambda_1 = \lambda_2$ and $T = T_1 = T_2$, the minimal variance of the Kim and Warde [13] estimator $\hat{\pi}_{kw}$ is given by

$$V(\hat{\pi}_{kw}) = \frac{1}{n}\Big[w_1(A_1 + b_1)^{\frac{1}{2}} + w_2(A_2 + b_1)^{\frac{1}{2}}\Big]^2, \tag{4.17}$$

where

$$\begin{aligned}
A_1 &= \pi_{s1}(1-\pi_{s1}) + \frac{\lambda(1-P_1)\,(1-\pi_{s1})}{P_1}\,, \\
A_2 &= \pi_{s2}(1-\pi_{s2}) + \frac{\lambda(1-P_1)\,(1-\pi_{s2})}{P_1}\,, \\
b_1 &= \frac{(1-\lambda)\,(1-P_1)}{P_1^2}\,.
\end{aligned}$$

Equation (4.15) can be rewritten as

$$V(\hat{\pi}_{Ak}) = \frac{1}{n}\left[w_1(A_1+b)^{\frac{1}{2}} + w_2(A_2+b)^{\frac{1}{2}}\right]^2. \tag{4.18}$$

From equations (4.17) and (4.18), we have

$$\begin{aligned}
n\Big[V(\hat{\pi}_{kw}) - V(\hat{\pi}_{Ak})\Big] &= \\
&= \left[w_1(A_1+b_1)^{\frac{1}{2}} + w_2(A_2+b_1)^{\frac{1}{2}}\right]^2 - \left[w_1(A_1+b)^{\frac{1}{2}} + w_2(A_2+b)^{\frac{1}{2}}\right]^2 \\
&= \left[w_1^2 b_1 + w_2^2 b_1 - w_1^2 b - w_2^2 b + 2\,w_1 w_2\left[(A_1+b_1)^{\frac{1}{2}}(A_2+b_1)^{\frac{1}{2}} - (A_1+b)^{\frac{1}{2}}(A_2+b)^{\frac{1}{2}}\right]\right] \\
&= \left[(b_1-b)\,(w_1^2+w_2^2) + 2\,w_1 w_2\left[(A_1+b_1)^{\frac{1}{2}}(A_2+b_1)^{\frac{1}{2}} - (A_1+b)^{\frac{1}{2}}(A_2+b)^{\frac{1}{2}}\right]\right] \\
&= (b_1-b)\left[w_1^2 w_2^2 + 2\,w_1 w_2 \frac{(A_1+A_2+b_1+b)}{\left[(A_1+b_1)^{\frac{1}{2}}(A_2+b_1)^{\frac{1}{2}} - (A_1+b)^{\frac{1}{2}}(A_2+b)^{\frac{1}{2}}\right]}\right] \\
&> 0\,,
\end{aligned}$$

since $(b_1-b) > 0$.

Therefore, $n\big[V(\hat{\pi}_{kw}) - V(\hat{\pi}_{Ak})\big]$ is always positive. Thus the theorem is proved. □

We have shown the performance of proposed stratified estimator $\hat{\pi}_{Ak}$ over suggested mixed estimator $\hat{\pi}_{A1}$ and Kim and Warde [13] estimator $\hat{\pi}_{kw}$ in case of two strata (i.e. $k=2$). Now, we calculate the percent relative efficiencies $\mathrm{PRE}(\hat{\pi}_{Ak},\hat{\pi}_{A1})$ and $\mathrm{PRE}(\hat{\pi}_{Ak},\hat{\pi}_{kw})$ for different values of T, π_s, n_1, n_2 and P_1, by using the following formulas:

$$\mathrm{PRE}(\hat{\pi}_{Ak},\hat{\pi}_{A1}) = \frac{V(\hat{\pi}_{A1})}{V(\hat{\pi}_{Ak})} \times 100 \tag{4.19}$$

and

$$\mathrm{PRE}(\hat{\pi}_{Ak},\hat{\pi}_{kw}) = \frac{V(\hat{\pi}_{kw})}{V(\hat{\pi}_{Ak})} \times 100\,, \tag{4.20}$$

where

$$V(\hat{\pi}_{kw}) = \left[\sum_{k=1}^{2} w_k\left[\frac{\pi_{sk}(1-\pi_{sk})}{n} + \frac{(1-Q_k)\left[\lambda_k Q_k(1-\pi_{sk}) + (1-\lambda_k)\right]}{n\,Q_k^2}\right]^{\frac{1}{2}}\right]^2.$$

We may observe from Tables 3–4:

(a) For all the parametric combinations the values of percent relative efficiencies are substantially exceeding 100, which indicate that the proposed stratified estimator $\hat{\pi}_{Ak}$ is uniformly better than the proposed mixed estimator $\hat{\pi}_{A1}$ and Kim and Warde [13] estimator $\hat{\pi}_{kw}$ under optimum allocation condition.

(b) It is also noted, from Table 3, the percent relative efficiencies increasing as the values of P_1 increases. Also the percent relative efficiencies almost increasing as the values of π_s increases for fixed values of λ and T.

(**c**) From Table 4, we observe that with the increase in the values of P_1 there is a decreasing pattern in the values of percent relative efficiencies.

(**d**) Figures 3–4 also show that there is a large gain in efficiencies by using the proposed stratified estimator $\hat{\pi}_{Ak}$ over the mixed estimator $\hat{\pi}_{A1}$ and Kim and Warde [13] stratified estimator, when the proportion of stigmatizing attribute is moderately large.

Table 3: Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{Ak}$ with respect to mixed estimator $\hat{\pi}_{A1}$.

π_{s1}	π_{s2}	π_s	w_1	w_2	λ	T	$P_1 = Q_1 = Q_2$						
							0.10	0.15	0.20	0.25	0.30	0.35	0.40
0.08	0.13	0.1	0.6	0.4	0.2	0.1	100.031	100.041	100.051	100.063	100.076	100.090	100.107
					0.4	0.3	100.026	100.032	100.038	100.046	100.056	100.067	100.080
					0.6	0.1	100.023	100.027	100.031	100.036	100.042	100.049	100.058
					0.8	0.3	100.023	100.026	100.029	100.033	100.038	100.044	100.051
0.18	0.23	0.2	0.6	0.4	0.2	0.1	100.035	100.043	100.052	100.061	100.072	100.082	100.094
					0.4	0.3	100.031	100.037	100.043	100.050	100.058	100.067	100.078
					0.6	0.1	100.028	100.032	100.036	100.041	100.046	100.053	100.061
					0.8	0.3	100.028	100.031	100.035	100.039	100.043	100.049	100.055
0.28	0.33	0.3	0.6	0.4	0.2	0.1	100.040	100.047	100.055	100.063	100.071	100.080	100.089
					0.4	0.3	100.038	100.044	100.050	100.056	100.063	100.071	100.080
					0.6	0.1	100.035	100.039	100.043	100.047	100.053	100.059	100.066
					0.8	0.3	100.036	100.039	100.043	100.047	100.051	100.057	100.063
0.38	0.43	0.4	0.6	0.4	0.2	0.1	100.047	100.054	100.060	100.067	100.074	100.081	100.089
					0.4	0.3	100.049	100.054	100.060	100.066	100.072	100.079	100.087
					0.6	0.1	100.046	100.049	100.053	100.058	100.063	100.069	100.075
					0.8	0.3	100.047	100.051	100.055	100.059	100.064	100.069	100.075

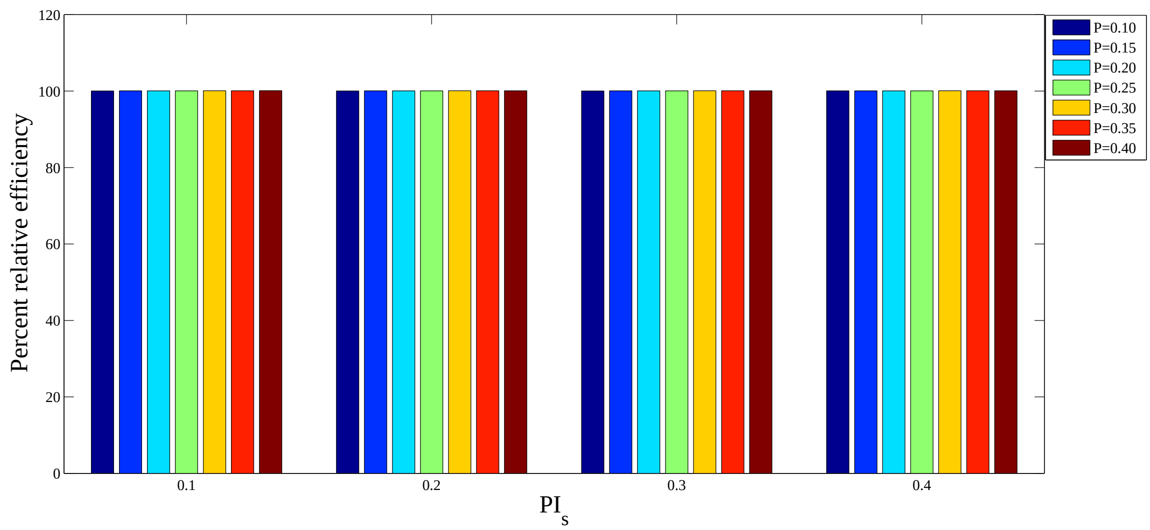

Figure 3: Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{Ak}$ with respect to mixed estimator $\hat{\pi}_{A1}$ when $T = 0.1$ and $\lambda = 0.2$.

Table 4: Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{Ak}$ with respect to Kim and Warde [13] stratified estimator $\hat{\pi}_{kw}$.

π_{s1}	π_{s2}	π_s	w_1	w_2	λ	T	$P_1 = Q_1 = Q_2$						
							0.10	0.15	0.20	0.25	0.30	0.35	0.40
0.08	0.13	0.1	0.6	0.4	0.2	0.1	3481.8	2101.6	1442.9	1066.2	826.74	663.42	546.22
					0.4	0.3	1631.1	1075.7	797.99	631.41	520.33	440.92	381.28
					0.6	0.1	794.72	546.74	422.30	347.29	297.01	260.87	233.57
					0.8	0.3	370.55	277.34	230.42	202.03	182.88	169.02	158.45
0.18	0.23	0.2	0.6	0.4	0.2	0.1	3667.2	2161.6	1454.6	1056.2	806.48	638.26	518.94
					0.4	0.3	1768.8	1146.8	837.14	652.25	529.65	442.57	377.60
					0.6	0.1	864.48	586.05	446.58	362.71	306.65	266.47	236.23
					0.8	0.3	399.97	295.03	242.18	210.17	188.57	172.92	160.97
0.28	0.33	0.3	0.6	0.4	0.2	0.1	3914.3	2255.9	1490.7	1066.2	803.68	629.02	506.52
					0.4	0.3	1946.4	1240.7	891.40	684.34	548.12	452.19	381.28
					0.6	0.1	953.86	636.79	478.46	383.62	320.52	275.53	241.87
					0.8	0.3	437.83	317.90	257.52	220.96	196.31	178.47	164.89
0.38	0.43	0.4	0.6	0.4	0.2	0.1	4245.8	2394.1	1555.4	1097.4	817.92	634.11	506.52
					0.4	0.3	2182.3	1367.1	966.80	731.53	578.17	471.21	392.93
					0.6	0.1	1072.2	704.20	521.29	412.31	340.23	289.19	251.26
					0.8	0.3	488.32	348.50	278.17	235.65	207.04	186.39	170.73

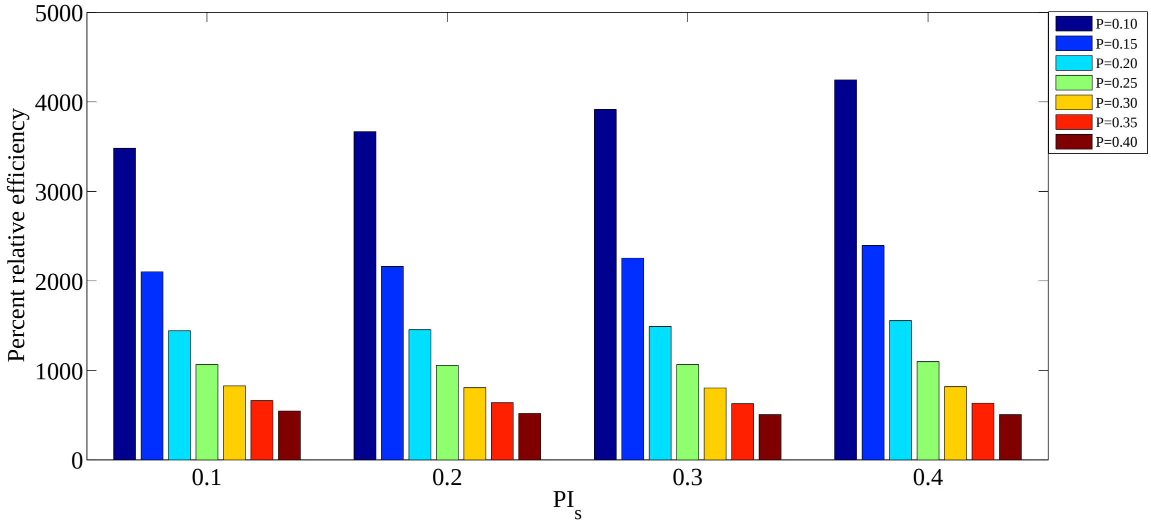

Figure 4: Percent relative efficiency of the proposed stratified estimator $\hat{\pi}_{Ak}$ with respect to Kim and Warde [13] stratified estimator $\hat{\pi}_{kw}$ when $T = 0.1$ and $\lambda = 0.2$.

5. CONCLUSIONS

In this paper, we have estimated the population proportion who possess to the sensitive attribute in the given population under both the situations of completely truthful reporting and less than completely truthful reporting as well as its stratified randomized response model. It has been shown that the proposed mixed randomized response models are better than the

Kim and Warde [13] mixed randomized response model with larger gain in efficiencies.

ACKNOWLEDGMENTS

Authors are heartily thankful to Editor-in-Chief and learned referee for their valuable comments which have made substantial improvement to bring the original manuscript to its present form. Authors are also grateful to the IIT(ISM), Dhanbad for providing the financial assistance and necessary infrastructure to carry out the present work.

REFERENCES

[1] Amitava, S. (2005). Kim and Warde's mixed randomized response technique for complex surveys, *Journal of Modern Applied Statistical Methods*, **4**(2), 538–544.

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

[3] Chang, H.J.; Wang, C.L. and Huang, K.C. (2004). Using randomized response to estimate the proportion and truthful reporting probability in a dichotomous finite population, *Journal of Applied Statistics*, **31**(5), 565–573.

[4] Chaudhuri, A. and Mukerjee, R. (1998). *Randomized Response: Theory and Techniques*, Marcel-Dekker, New York, USA.

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

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

[7] Greenberg, B.; Abdul-Ela, A.; Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question randomized response: theoretical framework, *Journal of American Statistical Association*, **64**, 529–539.

[8] Hedayat, A.S. and Sinha, B.K. (1991). *Design and Inference in Finite Population Sampling*, New York, Wiley.

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

[10] Hussain, Z. and Shabbir, J. (2007). Improvement of Kim and Warde's mixed randomized response technique for complex surveys, *InterStat*, July 003.

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

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

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

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

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

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

[17] Moors, J.A. (1971). Optimization of the unrelated question randomized response model, *Journal of the American Statistical Association*, **66**, 627–929.

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

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

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

[21] Singh, S. (2003). *Advance Sampling Theory with Applications*, Kluwer Academic Publishers, Dordrecht, the Netherlands.

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

[23] Singh, H.P. and Tarray, T.A. (2013a). An alternative to Kim and Warde's mixed randomized response model, *Statistics and Operation Research Transactions*, **37**(2), 189–210.

[24] Singh, H.P. and Tarray, T.A. (2013b). An improve mixed randomized response model, *Model Assisted Statistics and Applications*, **9**, 73–87.

[25] Singh, H.P. and Tarray, T.A. (2013c). An alternative to Kim and Warde's mixed randomized response technique, *Statistica*, **LXXIII**(3), 379–402.

[26] 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.

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

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