South African Statistical Journal
Vol. 54, No. 1, 45–63
https://doi.org/10.37920/sasj.2020.54.1.4

AN IMPROVED UNRELATED QUESTION RANDOMIZED RESPONSE MODEL

Khadiga Sayed
Department of Statistics, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt

Reda Mazloum[1]
Department of Statistics, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt
e-mail: *rmazloum@feps.edu.eg*

In this paper we restrict the design probabilities of Mahmood, Singh and Horn (1998) unrelated question randomized response model. Besides its simplicity, the resulted restricted model has two advantages over Mahmood et al. (1998) model with other design probabilities. First, the restricted model requires selecting only one simple random sample and not two which reduces the cost of survey. Second, the efficiency of the estimator of the proportion π_s of the population bearing a sensitive characteristic is increased. In addition, efficiency comparisons showed that this estimator can be easily adjusted to be more efficient than other competitors that were developed after 1998. A simulation study is performed to determine the minimum sample size required for the estimator to lie inside the unit interval. Moreover, the restricted model is extended to stratified random sampling and the resulting estimator is shown to be more efficient than the Kim and Elam (2007) and Singh and Tarray (2016) estimators.

Key words: Efficiency, Estimation of proportion, Randomized response technique, Simple random sampling, Simulation, Stratified random sampling, Unrelated characteristic.

1. Introduction

Warner (1965) introduced an indirect questioning technique called "Randomized Response Technique" to reduce evasive answer bias occurring when the participants in a sample survey are faced with questions of sensitive matters such as illegal use of drugs, drunken driving, tax evasion and bribery.

According to a model by Warner (1965), every respondent in a simple random sample with replacement (SRSWR) is provided with an identical random device. This random device can be a deck of cards consisting of two different types of cards bearing the statements: (I) "I belong to the sensitive group" and (II) "I do not belong to the sensitive group". Unobserved by the interviewer, each respondent is asked to choose one card randomly and answer "Yes" or "No" according to the statement selected and his /her actual status with respect to the sensitive characteristic A .

Several modifications on Warner's model have been proposed by various authors in order to improve its efficiency and/or increase the respondent's cooperation. For a literature overview on the topic, see Chaudhuri and Christofides (2013).

To increase the respondent's cooperation, Horvitz, Shah and Simmons (1967) suggested a modification on Warner's (1965) model. This modification involved using two unrelated questions, one

[1] Corresponding author.
MSC2010 subject classifications. 62D05.

about the sensitive characteristic A and the other about a non-sensitive characteristic Y. The non-sensitive characteristic Y could be, for example, whether the respondent was born in a certain area or whether the respondent is a left-handed. The model in this case is called "Unrelated Question Randomized Response Model". The theoretical framework for the unrelated question model was developed by Greenberg, Abul-Ela, Simmons and Horvitz (1969).

Mahmood et al. (1998) proposed an unrelated question randomized response. The structure of that model is as follows: two independent SRSWR of sizes n_1 and n_2 are selected from the population. Each respondent in the first sample is provided with a random device consisting of three different types of cards bearing the three statements: (I) "I possess the sensitive characteristic A ", (II) "I do not possess the non-sensitive characteristic Y" and (III) "I possess the non-sensitive characteristic Y". The statements are represented with probabilities q_1, q_2 and q_3 respectively, where $q_1+q_2+q_3=1$. The respondents in the second sample are asked a direct question regarding the non-sensitive characteristic Y. Assuming the proportion π_y of the population bearing a non-sensitive characteristic Y is unknown, they obtained an unbiased estimator $\hat{\pi}_M$ for the proportion π_s of the population bearing a sensitive characteristic A, given by

$$\hat{\pi}_M = \frac{\hat{\beta} + (q_2 - q_3)\,\hat{\pi}_y - q_2}{q_1}, \quad q_1 \neq 0,$$

where $\hat{\beta}$ is the observed proportion of "Yes" answers obtained from the respondents in the first sample and $\hat{\pi}_y$ is the observed proportion of "Yes" answers on the direct question obtained from the respondents in the second sample.

The minimum variance of $\hat{\pi}_M$ is given by

$$V(\hat{\pi}_M) = \frac{\left[\sqrt{\beta(1-\beta)} + |q_2 - q_3|\sqrt{\pi_y\left(1-\pi_y\right)}\right]^2}{nq_1^2}, \quad q_1 \neq 0, \tag{1}$$

where $\beta = q_1\pi_s + q_2\left(1-\pi_y\right) + q_3\pi_y$.

Kim and Warde (2005) proposed an alternative randomized response model. According to this model, each respondent in a SRSWR of size n has to answer the direct question "I possess the non-sensitive characteristic Y". If a respondent answers "Yes", then he/she has to use a random device consisting of two statements: (I) "I possess the sensitive characteristic A" and (II) "I possess the non-sensitive characteristic Y" represented with probabilities Q and $(1-Q)$ respectively. If a respondent answers "No" to the direct question, then the respondent is instructed to use Warner's random device which consists of the statements: (I) "I possess the sensitive characteristic A" and (II) "I do not possess the sensitive characteristic A" represented with probabilities P and $(1-P)$, respectively. They obtained an unbiased estimator $\hat{\pi}_{KW}$ for π_s with variance given by

$$V(\hat{\pi}_{KW}) = \frac{\pi_s(1-\pi_s)}{n} + \frac{(1-Q)(\lambda Q(1-\pi_s) + (1-\lambda))}{nQ^2}, \quad Q \neq 0, \tag{2}$$

where λ is the proportion of people who answer "Yes" to the direct question.

Kim and Elam (2007) extended the unrelated question randomized response model of Greenberg et al. (1969) to stratified random sampling. Each respondent in a SRSWR of size n_h from stratum h; $h = 1, 2, \ldots, k$ is provided with the random device R_h which consists of the statements: (I) "I possess the sensitive characteristic A" and (II) "I possess the non-sensitive characteristic Y" represented with probabilities P_h and $(1 - P_h)$, respectively. Under the assumption that the proportion π_{yh} of the population bearing the non-sensitive characteristic Y in stratum h is known, they obtained an unbiased estimator $\hat{\pi}_{KE}$ of π_s with minimum variance, under Neyman allocation, given by

$$V_{Ney}(\hat{\pi}_{KE}) = \frac{1}{n}\left[\sum_{h=1}^{k} \frac{W_h\sqrt{\lambda_h(1-\lambda_h)}}{P_h}\right]^2, \quad P_h \neq 0, \tag{3}$$

where $\lambda_h = P_h\pi_{sh} + (1 - P_h)\pi_{yh}$, π_{sh} is the population proportion of people bearing the sensitive characteristic A in stratum h, $W_h = N_h/N$ (N is the number of units in the whole population and N_h is the number of units in stratum h) so that $\sum_{h=1}^{k} W_h = 1$ and $n = \sum_{h=1}^{k} n_h$ is the total sample size from all strata.

In case π_{yh} is unknown, two independent non-overlapping SRSWR of sizes n_{h1} and n_{h2} are selected from each stratum such that $n_{h1} + n_{h2} = n_h$, total sample size from stratum h. The respondent in the ith sample, $i = 1, 2$, from stratum h is provided with the random device R_{hi} which consists of the statements: (I) "I possess the sensitive characteristic A" and (II) "I possess the non-sensitive characteristic Y" represented with probabilities P_{hi} and $(1 - P_{hi})$, $i = 1, 2$, respectively. They obtained an unbiased estimator $\hat{\hat{\pi}}_{KE}$ of π_s with minimum variance, under Neyman allocation, given by

$$V_{Ney}\left(\hat{\hat{\pi}}_{KE}\right) = \frac{1}{n}\left[\sum_{h=1}^{k} W_h \frac{(1-P_{h2})\sqrt{\lambda_{h1}(1-\lambda_{h1})} + (1-P_{h1})\sqrt{\lambda_{h2}(1-\lambda_{h2})}}{(P_{h1}-P_{h2})}\right]^2, \quad P_{h1} \neq P_{h2}, \tag{4}$$

where $n = \sum_{h=1}^{k} n_h$ is the total sample size from all strata and $\lambda_{hi} = P_{hi}\pi_{sh} + (1 - P_{hi})\pi_{yh}$, $i = 1, 2$.

Singh and Tarray (2016) extended the Singh, Horn, Singh and Mangat (2003) unrelated question model to stratified random sampling. Each respondent in a SRSWR of size n_h from the hth stratum, $h = 1, 2, \ldots, k$, is provided with a random device consisting of three types of cards bearing the three statements: (I) "I possess the sensitive characteristic A", (II) "I possess the non-sensitive characteristic Y" and (III) "Blank card". The statements are represented with probabilities P_{h1}, P_{h2} and P_{h3} respectively, where $P_{h1} + P_{h2} + P_{h3} = 1$, $h = 1, 2, \ldots, k$. In case of a blank card being chosen, the respondent is instructed to report "No" irrespective of his/her actual status. Assuming that the proportion π_{yh} is known, they obtained an unbiased estimator $\hat{\pi}_{ST}$ of π_s with minimum variance, under Neyman allocation, given by

$$V_{Ney}(\hat{\pi}_{ST}) = \frac{1}{n}\left[\sum_{h=1}^{k} \frac{W_h\sqrt{\theta_h(1-\theta_h)}}{P_{h1}}\right]^2, \quad P_{h1} \neq 0, \tag{5}$$

where $\theta_h = P_{h1}\pi_{sh} + P_{h2}\pi_{yh}$, $h = 1, 2, \ldots, k$.

In the following section, an improved restricted version of the Mahmood et al. (1998) unrelated question randomized response model is presented where only one simple random sample and not two as in the general set up of the Mahmood et al. (1998) model, is required which reduces the cost of survey. It is also shown that the resulting estimator of π_s is more efficient than the other versions of the Mahmood et al. (1998) estimator. In addition, it is shown that this estimator can be easily adjusted to be more efficient than other estimators that are suggested after 1998, for example, Kim and Warde (2005) when π_y is unknown and both Singh et al. (2003) and Perri (2008) estimators when π_y is known. In Section 3, following Lee, Sedory and Singh (2013), a simulation study is performed to determine the minimum sample sizes required for the proposed estimator to take values within the unit interval. In Section 4, the model is extended to stratified random sampling and it is shown that the resulting stratified estimator is more efficient than its counterpart in simple random sampling, namely the Kim and Elam (2007) and Singh and Tarray (2016) estimators. Finally, a conclusion of this work is presented in Section 5.

2. The model

This model is a restricted version of the Mahmood et al. (1998) model where both the probabilities q_2 and q_3 are restricted to $0.5\,(1-q_1)$. The motivation behind this restriction is not only to increase the efficiency, but also to obtain both an estimator of the proportion π_s and its variance that are free from π_y. Consequently, this will allow one to select a single simple random sample and not two as in the Mahmood et al. (1998) model. This in turn reduces the cost of survey.

According to this model, each respondent in a SRSWR of size n is provided with a random device R as shown in Figure 1. The random device R consists of three different types of cards bearing the three statements: (I) "I possess the sensitive characteristic A ", (II) "I do not possess the non-sensitive characteristic Y" and (III) "I possess the non-sensitive characteristic Y", with probabilities q_1, $0.5\,(1-q_1)$ and $0.5\,(1-q_1)$, respectively. The respondent is requested to answer "Yes" or "No" according to the statement selected and his/her actual status. The whole procedure is completed by the respondent, unobserved by the interviewer.

In the following subsection an unbiased estimator of π_s, along with its variance, is obtained. In Subsection 2.2 the efficiency comparisons of the proposed estimator relative to the Mahmood et al. (1998) (with other design probabilities) and Kim and Warde (2005) estimators in the case where the proportion π_y is unknown are examined, while in Subsection 2.3 we investigate the relative efficiency of the proposed estimator with respect to the Singh et al. (2003) and Perri (2008) estimators in the case where π_y is known.

2.1 Estimation of the population proportion π_s

The probability, α, of getting a "Yes" answer is

$$\alpha = \Pr(Yes) = q_1\pi_s + 0.5\,(1-q_1)\,. \tag{6}$$

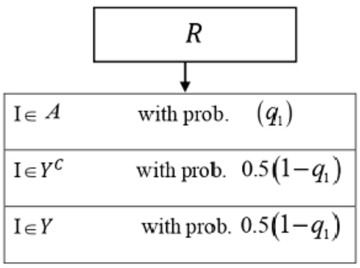

Figure 1. The model.

The probability, $(1-\alpha)$, of getting a "No" answer is

$$1-\alpha = \Pr(No) = q_1(1-\pi_s) + 0.5(1-q_1). \tag{7}$$

Remark 1. It is obvious from (6) and (7) that the probabilities α and $(1-\alpha)$ do not depend on π_y.

This is an interesting remark especially if π_y is unknown since the estimator of π_s and its variance, as will be shown, are free from π_y. This implies that the proposed model utilizes the use of the unrelated question Y and at the same time does not require knowledge or estimation of π_y. Hence, unlike the Mahmood et al. (1998) model in the general set up, the estimation process here requires only one sample and not two samples. This in turn saves time and/or money and avoids the problems of the optimal allocation of the sample sizes. Also, it is worth mentioning that the restricted model is simple and depends only on one design probability, namely q_1.

The following theorem gives an unbiased estimator for π_s along with its variance.

Theorem 1. *An unbiased estimator of the population proportion π_s is given by*

$$\hat{\pi} = \frac{\hat{\alpha} - 0.5(1-q_1)}{q_1}, \quad q_1 \neq 0, \tag{8}$$

where $\hat{\alpha} = \grave{n}/n$ is the observed proportion of "Yes" answers.

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

$$V(\hat{\pi}) = \frac{\pi_s(1-\pi_s)}{n} + \frac{0.5(1-q_1)[1-0.5(1-q_1)]}{nq_1^2}, \quad q_1 \neq 0. \tag{9}$$

Proof. The proof of the unbiasedness is immediate by taking the expected values on both sides of (8).

$$V(\hat{\pi}) = \frac{V(\hat{\alpha})}{q_1^2} = \frac{\alpha(1-\alpha)}{nq_1^2}, \quad q_1 \neq 0. \tag{10}$$

Substituting (6) and (7) into (10) and after some algebraic manipulations, we get $V(\hat{\pi})$ as given by (9). ∎

Remark 2. The variance of $\hat{\pi}$ is symmetric around $\pi_s = 0.5$.

Remark 3. It is clear from (8) and (9) that the estimator $\hat{\pi}$ and its variance do not depend on π_y.

Theorem 2. *An unbiased estimator of the variance of $\hat{\pi}$ is given by*

$$\hat{V}(\hat{\pi}) = \frac{1}{(n-1)}\left[\hat{\pi}(1-\hat{\pi}) + \frac{0.5(1-q_1)[1-0.5(1-q_1)]}{q_1^2}\right], \quad q_1 \neq 0. \tag{11}$$

Proof. The proof is immediate by taking the expected values on both sides of (11). ∎

2.2 Efficiency comparisons when π_y is unknown

We examine the relative efficiency of the proposed estimator $\hat{\pi}$ given by (8) with respect to the Mahmood et al. (1998) (with $q_2 \neq q_3$) and Kim and Warde (2005) estimators in the case where π_y is unknown.

Table 1. Summary statistics of percent relative efficiency (RE_1) for different levels of π_s.

π_s	**Frequency**	**Mean**	**Standard Deviation**	**Minimum**	**1^{st} Quartile**	**Median**	**3^{rd} Quartile**	**Maximum**
0.1	286	159.9	42.7	102.5	124.1	148.4	181.5	289.8
0.2	288	158.6	42.0	104.1	122.7	146.8	178.4	289.4
0.3	288	158.0	41.6	107.0	121.8	144.9	178.5	289.2
0.4	288	157.7	41.3	109.4	121.1	144.3	180.4	289.0
0.5	288	157.6	41.2	111.7	122.2	144.0	183.3	289.0
0.6	288	157.7	41.3	109.4	121.1	144.3	180.4	289.0
0.7	288	158.0	41.6	107.0	121.8	144.9	178.5	289.2
0.8	288	158.6	42.0	104.1	122.7	146.8	178.4	289.4
0.9	286	159.9	42.7	102.5	124.1	148.4	181.5	289.8

2.2.1 *Comparing $\hat{\pi}$ with the Mahmood et al. (1998) estimator*

The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to the Mahmood et al. (1998) estimator $\hat{\pi}_M$, where $q_2 \neq q_3$, is given by

$$RE_1 = \frac{V(\hat{\pi}_M)}{V(\hat{\pi})} \times 100,$$

where $V(\hat{\pi}_M)$ and $V(\hat{\pi})$ are as given in (1) and (9), respectively.

For each value of π_s where π_s takes values from 0.1 to 0.9 with a step of 0.1, the relative efficiencies are calculated for all possible combinations (288 combinations) from the values of π_y, q_1 and q_2 where the parameter π_y takes values from 0.1 to 0.9 with a step of 0.1, while the values of q_1 range from 0.1 to 0.7 with a step of 0.1, and q_2 from 0.1 to 0.8 with a step of 0.1 such that $q_2 \neq q_3$ and $q_3 = 1 - q_1 - q_2 > 0$. It is found that the proposed estimator $\hat{\pi}$ is more efficient than the Mahmood et al. (1998) estimator $\hat{\pi}_M$ in about 99.8% of the cases. Table 1 presents summary statistics of RE_1 for each value of π_s. For example, for $\pi_s = 0.1$, there are 286 different combinations of the parameters where $RE_1 > 100\%$. Among these, the values of RE_1 range from a minimum of 102.5% to a maximum of 289.8% with a median of 148.4%, a mean of 159.9%, a standard deviation of 42.7% and an interquartile range (IQR) of 57.4%. It can be easily observed from Table 1 that, for π_s=0.1 or 0.9, there is only two cases out of all the cases where $RE_1 < 100\%$ namely RE_1=99.97%. Otherwise, the proposed estimator $\hat{\pi}$ is always more efficient than the Mahmood et al. (1998) estimator with $q_2 \neq q_3$. From the numerical study it is observed that the RE_1 for the combination $(\pi_s, \pi_y) = (a, b)$ is the same for the combination $(\pi_s, \pi_y) = (1 - a, 1 - b)$.

2.2.2 *Comparing $\hat{\pi}$ with the Kim and Warde (2005) estimator*

The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to the Kim and Warde (2005) estimator $\hat{\pi}_{KW}$ is given by

$$RE_2 = \frac{V(\hat{\pi}_{KW})}{V(\hat{\pi})} \times 100,$$

where $V(\hat{\pi}_{KW})$ and $V(\hat{\pi})$ are as given in (2) and (9), respectively.

For each value of π_s where π_s takes values from 0.1 to 0.9 with a step of 0.1, we compute the relative efficiency of the proposed estimator $\hat{\pi}$ with respect to $\hat{\pi}_{KW}$ of Kim and Warde (2005) for all possible combinations (729 combinations) from the values of Q, λ and q_1 where each of the parameters Q, λ and q_1 takes values from 0.1 to 0.9 with a step of 0.1. It is found that the proposed

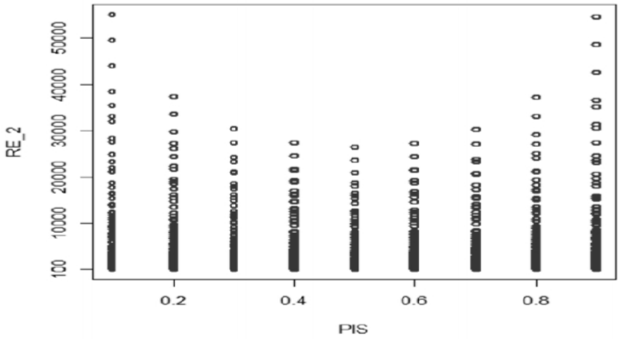

Figure 2. The relationship between RE_2 and π_s.

Table 2. Summary statistics of percent relative efficiency (RE_2) for different levels of π_s.

π_s	f	**Mean**	**StDev**	**Minimum**	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**	**Maximum**
0.1	435	2713.1	6480.6	100.4	227.5	538.9	1920.1	55098.8
0.2	429	2133.1	4726.7	100.1	205.8	479.0	1599.2	37449.3
0.3	425	1865.3	3992.5	100.5	193.8	436.4	1373.6	30464.3
0.4	420	1740.2	3662.7	100.6	187.0	409.1	1316.3	27384.0
0.5	415	1700.5	3565.9	100.8	186.0	403.2	1309.7	26470.8
0.6	407	1744.9	3673.8	100.5	187.3	395.9	1308.7	27323.7
0.7	404	1853.7	3996.9	100.5	186.5	412.2	1354.6	30330.3
0.8	392	2142.0	4761.6	101.5	195.2	462.8	1586.0	37202.4
0.9	385	2733.1	6549.4	103.0	213.3	516.4	1856.8	54614.4

estimator $\hat{\pi}$ is more efficient than the Kim and Warde (2005) estimator $\hat{\pi}_{KW}$ in about 57% of the cases. For these cases, Figure 2 shows that RE_2 is symmetric around π_s=0.5 and it increases as the value of π_s gets close to zero or one.

Table 2 presents summary statistics of RE_2 for each value of π_s. For example, for $\pi_s = 0.1$, there are 435 different combinations of the parameters where RE_2>100%. Among these, the values of RE_2 range from a minimum of 100.4% to a maximum of 55098.8% with a median of 538.9%, a mean of 2713.1%, a standard deviation of 6480.6% and an IQR of 1692.6%.

2.3 Efficiency comparisons when π_y is known

In the following Subsections the efficiency comparisons of the estimator $\hat{\pi}$ relative to the Singh et al. (2003) and Perri (2008) estimators in the case where π_y is known are examined.

2.3.1 Comparing $\hat{\pi}$ with the Singh et al. (2003) estimator

The variance of the Singh et al. (2003) estimator $\hat{\pi}_S$ of π_s is given by

$$V(\hat{\pi}_S) = \frac{\pi_s(1-\pi_s)}{n} + \frac{\pi_s\left(1-P_1-2P_2\pi_y\right)}{nP_1} + \frac{P_2\pi_y\left(1-P_2\pi_y\right)}{nP_1^2}, \quad P_1 \neq 0, \tag{12}$$

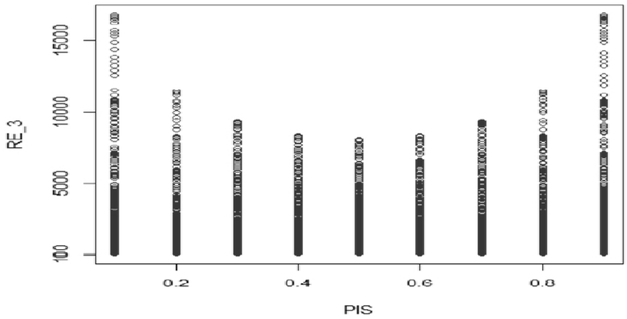

Figure 3. The relationship between RE_3 and π_s.

where P_1, P_2 and P_3 are the corresponding probabilities of selecting the statements: (I) "I possess the sensitive characteristic A", (II) "I possess the non-sensitive characteristic Y" and (III) "Blank card", respectively. In case of a blank card being chosen, the respondent is instructed to report "no", irrespective of his/her actual status with respect to the sensitive characteristic.

The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to the Singh et al. (2003) estimator $\hat{\pi}_S$ is given by

$$RE_3 = \frac{V(\hat{\pi}_S)}{V(\hat{\pi})} \times 100,$$

where $V(\hat{\pi}_S)$ and $V(\hat{\pi})$ are as given in (12) and (9), respectively.

For each value of π_s, where π_s takes values from 0.1 to 0.9 with a step of 0.1, the relative efficiencies are calculated for all possible combinations (2916 combinations) from the values of π_y, P_1, P_2 and q_1 where each of the parameters π_y and q_1 takes values from 0.1 to 0.9 with a step of 0.1 while the values of P_1 and P_2 range from 0.1 to 0.8 with a step of 0.1 such that $P_3 = 1 - P_1 - P_2 > 0$. It is found that the proposed estimator $\hat{\pi}$ is more efficient than the Singh et al. (2003) estimator $\hat{\pi}_S$ in about 64% of the cases. For such cases, Figure 3 shows that RE_3 is symmetric around π_s=0.5 and it increases as the value of π_s gets close to zero or one.

Table 3 presents summary statistics of RE_3 for each value of π_s. For example, for $\pi_s = 0.1$, there are 1702 different combinations of the parameters where RE_3>100%. Among these the values of RE_3 range from a minimum of 100.1% to a maximum of 16818.9% with a median of 505.4%, a mean of 1557.8%, a standard deviation of 2695.7% and an IQR of 1311.3%.

2.3.2 Comparing $\hat{\pi}$ with the Perri (2008) estimator

The variance of the Perri (2008) estimator $\hat{\pi}_P$ is given by

$$V(\hat{\pi}_P) = \frac{\pi_s(1-\pi_s)}{n} + \frac{\pi_s(1-Q_1-\theta Q_3)\left(1-2\pi_y\right)}{n(Q_1+\theta Q_3)} + \frac{\pi_y(1-Q_1-\theta Q_3)\left(1-\pi_y(1-Q_1-\theta Q_3)\right)}{n(Q_1+\theta Q_3)^2}, \tag{13}$$

where Q_1, Q_2 and Q_3 are the corresponding probabilities of selecting the statements: (I) "I possess the sensitive characteristic A", (II) "I possess the non-sensitive characteristic Y" and (III) "Blank card", respectively, of the first random device, while θ denotes the probability of selecting the sensitive question of the second random device which consists of two statements:(I) "I possess the sensitive

Table 3. Summary statistics of percent relative efficiency (RE_3) for different levels of π_s.

π_s	f	**Mean**	**StDev**	**Minimum**	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**	**Maximum**
0.1	1702	1557.8	2695.7	100.1	220.4	505.4	1531.7	16818.9
0.2	1755	1240.3	1948.2	100.2	209.8	444.7	1279.8	11434.2
0.3	1798	1106.9	1644.9	100.1	207.9	418.4	1178.1	9302.3
0.4	1827	1057.1	1515.5	100.2	210.4	409.3	1142.4	8367.9
0.5	1831	1068.4	1491.6	100.3	219.9	433.4	1179.4	8100.0
0.6	1886	1096.3	1540.7	100.2	223.5	447.8	1212.0	8367.9
0.7	1937	1179.1	1695.0	100.1	225.0	465.8	1295.2	9302.3
0.8	1969	1360.6	2036.9	100.2	237.6	518.9	1449.3	11434.2
0.9	1995	1756.8	2860.2	100.6	280.0	659.6	1764.7	16812.2

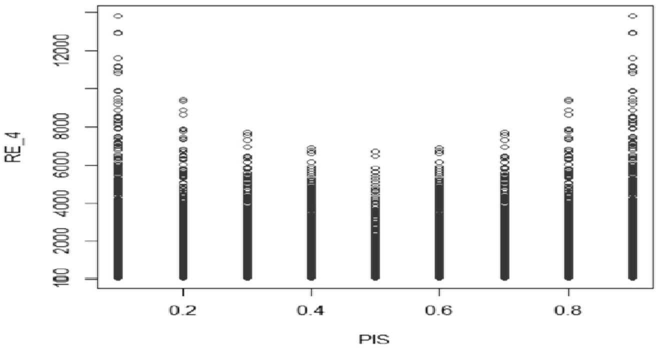

Figure 4. The relationship between RE_4 and π_s.

characteristic A" and (II) "I possess the non-sensitive characteristic Y" represented with probabilities θ and $(1-\theta)$.

The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to the Perri (2008) estimator $\hat{\pi}_P$ is given by

$$RE_4 = \frac{V(\hat{\pi}_P)}{V(\hat{\pi})} \times 100,$$

where $V(\hat{\pi}_P)$ and $V(\hat{\pi})$ are as given in (13) and (9) respectively.

For each value of π_s where π_s takes values from 0.1 to 0.9 with a step of 0.1, the relative efficiencies are calculated for all possible combinations (26244 combinations) from the values of π_y, θ, Q_1, Q_2 and q_1 where each of the parameters π_y, θ and q_1 takes values from 0.1 to 0.9 with a step of 0.1 while the values of Q_1 and Q_2 range from 0.1 to 0.8 with a step of 0.1 where $Q_3 = 1 - Q_1 - Q_2 > 0$. It is found that the proposed estimator $\hat{\pi}$ is more efficient than the Perri (2008) estimator $\hat{\pi}_P$ in about 48% of the cases. For such cases, Figure 4 shows that RE_4 is symmetric around π_s=0.5 and it increases as the value of π_s gets close to zero or one.

Table 4 presents summary statistics of RE_4 for each value of π_s. For example, for $\pi_s = 0.1$, there are 12353 different combinations of the parameters where RE_4>100%. Among these, the values of RE_4 range from a a minimum of 100.051% to a maximum of 13792.3 % with a median of 308.1%, a mean of 663.0%, a standard deviation of 1057.9% and an IQR of 500.9%.

Table 4. Summary statistics of percent relative efficiency (RE_4) for different levels of π_s.

π_s	f	**Mean**	**StDev**	**Minimum**	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**	**Maximum**
0.1	12353	663.0	1057.9	100.051	171.0	308.1	671.9	13792.3
0.2	12490	525.9	757.4	100.004	158.8	267.4	546.9	9408.6
0.3	12566	465.7	633.7	100.085	153.3	247.6	483.0	7676.1
0.4	12646	436.6	577.4	100.007	150.0	236.8	456.0	6915.0
0.5	12636	428.9	561.4	100.174	149.0	234.5	449.7	6694.2
0.6	12646	436.6	577.4	100.007	150.0	236.8	456.0	6915.0
0.7	12566	465.7	633.7	100.085	153.3	247.6	483.0	7676.1
0.8	12490	525.9	757.4	100.004	158.8	267.4	546.9	9408.6
0.9	12353	663.0	1057.9	100.051	171.0	308.1	671.9	13792.3

Table 5. Minimum sample sizes recommended for $\hat{\pi}$.

	π_s								
q_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	10000	3110	1340	730	480	690	1430	3110	10000
0.4	2770	690	240	150	100	150	240	690	2770
0.6	920	300	90	60	40	60	100	300	920
0.8	370	130	50	40	20	40	50	130	370

3. Simulation Study

Like other estimators of the proportion π_s, the estimator $\hat{\pi}$ given by (8) can take values outside the unit interval [0, 1]. For example, when $q_1 = 0.6$, $n = 40$ and $\grave{n} = 5$, $\hat{\pi}$ takes the value -0.125. A simulation study is performed to determine the minimum sample sizes required for $\hat{\pi}$ to take values inside the unit interval. For each combination of π_s and q_1 where π_s takes values from 0.1 to 0.9 with a step of 0.1 and q_1 ranges from 0.2 to 0.8 with a step of 0.2, the probability of a "Yes" answer, α, given by (6) is computed. Then for each sample size n, where n ranges from 10 to 10,000 with a step of 10, and each α, 10 000 samples are simulated from the binomial distribution. For each combination of (n, α), 10 000 estimates of π_s are obtained and the number of times where the estimator $\hat{\pi}$ takes values outside the unit interval [0, 1] are counted. The sample sizes beyond which the proportions of estimates that lie outside [0,1] become less than 0.0001 are deleted.

Figure 5 presents the proportions of the estimated values lying outside [0,1] versus the sample sizes for all chosen values of π_s at $q_1 = 0.4$. From Figure 5, the proportions of the estimates lying outside [0,1] decrease as the sample size increases.

Table 5 presents the minimal sample sizes recommended for the estimator $\hat{\pi}$ given by (8) at each value of π_s and different values of the design parameter q_1. For example, when $q_1 = 0.4$ and $\pi_s = 0.2$, the minimum sample size of $n = 690$ respondents is required to keep $\hat{\pi}$ inside [0, 1]. It can be easily observed that the minimum sample sizes increase as the value of π_s gets close to zero or one.

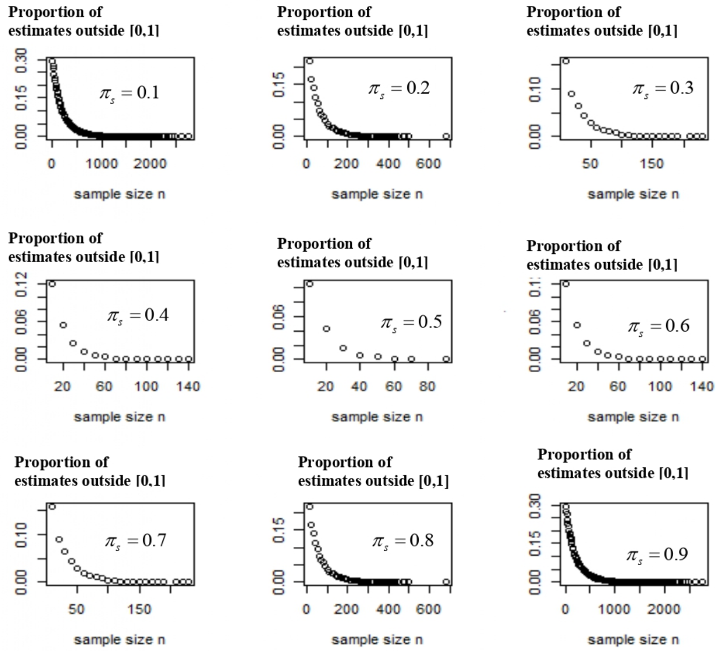

Figure 5. Proportion of estimates outside [0, 1] versus sample size n for all values of π_s at $q_1 = 0.4$.

4. The proposed stratified randomized response model

Suppose a population of size N is divided into k strata and a SRSWR of size n_h is selected from stratum h, $h = 1, 2, \dots, k$. It is assumed that the number of units in stratum h, N_h, is known and the selections in different strata are made independently. Each respondent in the sample from stratum h is provided with a random device R_h as shown in Figure 6. The random device R_h consists of three different types of cards bearing the three statements: (I) "I possess the sensitive characteristic A ", (II) "I do not possess the non-sensitive characteristic Y" and (III) "I possess the non-sensitive characteristic Y", with probabilities q_h, $0.5\,(1 - q_h)$ and $0.5\,(1 - q_h)$, $h = 1, 2, \dots, k$, respectively. The respondent is asked to answer "Yes" or "No" according to the statement selected and his/her actual status.

The probability, α_h, of getting a "Yes" answer is

$$\alpha_h = \Pr(Yes) = q_h \pi_{sh} + 0.5\,(1 - q_h)\,, h = 1, 2, \dots, k. \tag{14}$$

The probability, $(1 - \alpha_h)$, of getting a "No" answer is

$$1 - \alpha_h = \Pr(No) = q_h\,(1 - \pi_{sh}) + 0.5\,(1 - q_h)\,, h = 1, 2, \dots, k, \tag{15}$$

where π_{sh} is the population proportion of people having the sensitive characteristic A in stratum h.

Remark 4. It is obvious from (14) and (15) that the probabilities α_h and $(1 - \alpha_h)$ do not depend on π_{yh}; the population proportion of people having the non-sensitive characteristic Y in stratum h.

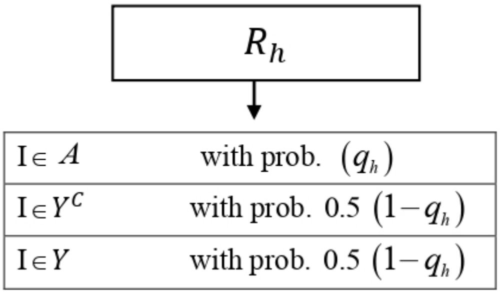

Figure 6. The stratified model.

This is an interesting remark especially if π_{yh} is unknown since the estimator of π_s and its variance, as will be shown in Subsection 4.1, are free from π_{yh}. This implies that the estimation process requires selecting only one simple random sample from each stratum and not two as most of the existing stratified unrelated question models which in turn reduces the cost of survey in each stratum.

In the following subsection an unbiased estimator of π_s along with its variance is obtained. In Subsection 4.2 the efficiencies of the proposed estimator relative to the Kim and Elam (2007) and Singh and Tarray (2016) estimators are examined.

4.1 Estimation of the population proportion π_s

Following the same procedure of Section 2.1 for stratum h, we get the following estimator $\hat{\pi}_{sh}$ of the proportion π_{sh}:

$$\hat{\pi}_{sh} = \frac{\hat{\alpha_h} - 0.5\,(1 - q_h)}{q_h}, \quad q_h \neq 0, h = 1, 2, \ldots, k, \tag{16}$$

where $\hat{\alpha_h} = \grave{n}_h / n_h$ is the observed proportion of "Yes" answers in the sample from stratum h.

The estimator $\hat{\pi}_{sh}$ given by (16) is unbiased with variance given by

$$V\left(\hat{\pi}_{sh}\right) = \frac{\pi_{sh}\,(1 - \pi_{sh})}{n_h} + \frac{0.5\,(1 - q_h)\,[1 - 0.5\,(1 - q_h)]}{n_h q_h^2}, \quad q_h \neq 0, h = 1, 2, \ldots, k. \tag{17}$$

Hence, for estimating the population proportion π_s, we have the following theorem.

Theorem 3. *An unbiased estimator of the population proportion π_s is*

$$\hat{\pi}_s = \sum_{h=1}^{k} W_h \hat{\pi}_{sh} = \sum_{h=1}^{k} W_h \frac{\hat{\alpha_h} - 0.5\,(1 - q_h)}{q_h}, \quad q_h \neq 0, \tag{18}$$

where $W_h = N_h/N$ for $h = 1, 2, \ldots\ k$ (N is the number of units in the whole population) so that $\sum_{h=1}^{k} W_h = 1$ and $\hat{\pi}_{sh}$ is given by (16).

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

$$V\left(\hat{\pi}_s\right) = \sum_{h=1}^{k} W_h^2 \left[\frac{\pi_{sh}\,(1 - \pi_{sh})}{n_h} + \frac{0.5\,(1 - q_h)\,[1 - 0.5\,(1 - q_h)]}{n_h q_h^2}\right], \quad q_h \neq 0, \tag{19}$$

Proof. The proof of the unbiasedness is immediate by taking the expected values on both sides of (18).

Since the selections in different strata are made independently, the variance of $\hat{\pi}_s$ is

$$V(\hat{\pi}_s) = \sum_{h=1}^{k} W_h^2 V(\hat{\pi}_{sh}). \tag{20}$$

Substituting $V(\hat{\pi}_{sh})$ of(17) into (20) gives (19) ■

Remark 5. It is obvious from (19) that the variance of $\hat{\pi}_s$ is symmetric around $\pi_{sh} = 0.5$; $h = 1, 2, \ldots, k$.

Theorem 4. *An unbiased estimator of the variance of π_s is given by*

$$\hat{V}(\hat{\pi}_s) = \sum_{h=1}^{k} \frac{W_h^2}{(n_h - 1)} \left[\hat{\pi}_{sh}(1 - \hat{\pi}_{sh}) + \frac{0.5(1 - q_h)[1 - 0.5(1 - q_h)]}{q_h^2} \right], \quad q_h \neq 0. \tag{21}$$

Proof. The proof is immediate by taking the expected values on both sides of (21). ■

Theorem 5. *Under Neyman allocation, the optimal allocation of the total sample size n is given by*

$$\frac{n_h}{n} = \frac{W_h \left[\pi_{sh}(1 - \pi_{sh}) + 0.5(1 - q_h)[1 - 0.5(1 - q_h)]/q_h^2\right]^{1/2}}{\sum_{h=1}^{k} W_h \left[\pi_{sh}(1 - \pi_{sh}) + 0.5(1 - q_h)[1 - 0.5(1 - q_h)]/q_h^2\right]^{1/2}} \tag{22}$$

and the minimum variance of $\hat{\pi}_s$ is

$$V_{Ney}(\hat{\pi}_s) = \frac{1}{n} \left[\sum_{h=1}^{k} W_h \left\{ \pi_{sh}(1 - \pi_{sh}) + \frac{0.5(1 - q_h)[1 - 0.5(1 - q_h)]}{q_h^2} \right\}^{1/2} \right]^2, \quad q_h \neq 0. \tag{23}$$

Proof. The proof of (22) is immediate following Section 5.5 of Cochran (1977). Substituting n_h of (22) into (19), we get the minimum variance of $\hat{\pi}_s$ given by (23). ■

The Neyman allocation given by (22), requires prior information on π_{sh} which is usually unavailable. In practice, these can be obtained from a previous study or a good guess.

4.2 Efficiency comparisons

In what follows, we examine the relative efficiency of the estimator $\hat{\pi}_s$ of (18) with respect to the Kim and Elam (2007) and Singh and Tarray (2016) estimators and the estimator of Section 2.

4.2.1 Comparing the estimator $\hat{\pi}_s$ with the Kim and Elam (2007) estimators

Case 1: π_{yh} is known

The relative efficiency of the proposed estimator $\hat{\pi}_s$ with respect to the estimator $\hat{\pi}_{KE}$ under Neyman allocation is given by

$$RE_5 = \frac{V_{Ney}(\hat{\pi}_{KE})}{V_{Ney}(\hat{\pi}_s)} \times 100,$$

where $V_{Ney}(\hat{\pi}_{KE})$ and $V_{Ney}(\hat{\pi}_s)$ are as given in (3) and (23), respectively.

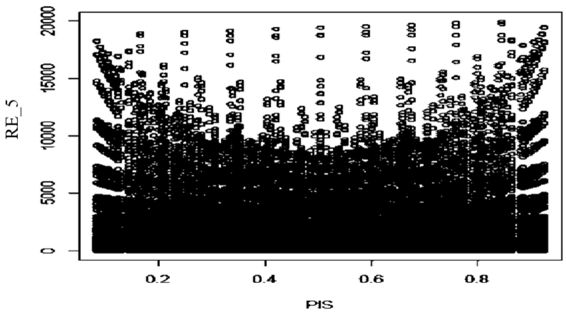

Figure 7. The relationship between RE_5 and π_s.

We assume that there are two strata in the population, i.e. $k = 2$, $\pi_{s1} \neq \pi_{s2}$, $\pi_{y1} = \pi_{y2} = \pi_y$, $P_1 = P_2 = P$ and $q_1 = q_2 = q$. The values of RE_5 are calculated for different combinations from the values of $\pi_{s1}, \pi_{s2}, \pi_y, W_1, W_2 = 1 - W_1, P$ and q where each of the parameters π_y, W_1, P and q takes values from 0.1 to 0.9 with a step of 0.1 while the values of π_{s1} range from 0.08 to 0.88 with a step of 0.2, and π_{s2} ranges from 0.13 to 0.93 with a step of 0.2. It is found that the proposed estimator $\hat{\pi}_s$ is more efficient than $\hat{\pi}_{KE}$ in about 47% of the cases. Figure 7 shows that the value of RE_5 increases as the value of π_s gets close to zero or one.

Table 6 presents summary statistics of RE_5 for each (π_{s1}, π_{s2}) pair. At $\pi_s = 0.105$ there are 343 different combinations of the parameters where $RE_5 > 100\%$. Among these the values of RE_5 range from a minimum of 100.5% to a maximum of 16414.8% with a median of 413.1%, a mean of 1384.7%, a standard deviation of 2607.8% and an IQR of 945.9%. It is observed that the values of the descriptive statistics of RE_5 increase as the value of π_s gets close to zero or one.

Case 2: π_{yh} is unknown

The relative efficiency of the proposed estimator $\hat{\pi}_s$ with respect to the estimator $\hat{\hat{\pi}}_{KE}$ under Neyman allocation is given by

$$RE_6 = \frac{V_{Ney}\left(\hat{\hat{\pi}}_{KE}\right)}{V_{Ney}\left(\hat{\pi}_s\right)} \times 100,$$

where $V_{Ney}\left(\hat{\hat{\pi}}_{KE}\right)$ and $V_{Ney}\left(\hat{\pi}_s\right)$ are as given in (4) and (23), respectively.

We assume that there are two strata in the population, i.e. $k = 2$, $\pi_{s1} \neq \pi_{s2}$, $\pi_{y1} = \pi_{y2} = \pi_y$, $P_{11} = P_{21} = P_1, P_{12} = P_{22} = P_2$ with $P_1 + P_2 = 1$ and $q_1 = q_2 = q$. The values of RE_6 are calculated for different combinations from the values of π_{s1}, π_{s2}, π_y, W_1, $W_2 = 1 - W_1$, P_1, $P_2 = 1 - P_1$ ($P_2 \neq 0.5$) and q where each of the parameters π_y, W_1, P_1 and q takes values from 0.1 to 0.9 with a step of 0.1 while the values of π_{s1} range from 0.08 to 0.88 with a step of 0.2 and π_{s2} from 0.13 to 0.93 with a step of 0.2. It is found that the proposed estimator $\hat{\pi}_s$ is more efficient than $\hat{\hat{\pi}}_{KE}$ in about 44% of the cases. Figure8 shows that the value of RE_6 increases as the value of π_s gets close to zero or one.

Table 7 presents summary statistics of RE_6 for each (π_{s1}, π_{s2}) pair. At $\pi_s = 0.105$ there are 276 different combinations of the parameters where $RE_6 > 100\%$. Among these the values of RE_6 range from a minimum of 106.3% to a maximum of 4023.4% with a median of 359.4%, a mean of 718.1%,

Table 6. Summary statistics of percent relative efficiency (RE_5) for different levels of π_s.

π_{s1}	π_{s2}	W_1	π_s	f	**Mean**	**StDev**	**Min.**	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**	**Max.**
0.08	0.13	0.9	0.085	341	1486.7	2862.6	100.4	191.9	437.4	1195.0	18260.9
0.08	0.13	0.7	0.095	342	1434.0	2730.3	100.1	188.7	428.1	1155.8	17300.8
0.08	0.13	0.5	0.105	343	1384.7	2607.8	100.5	186.8	413.1	1132.7	16414.8
0.08	0.13	0.3	0.115	343	1342.0	2497.1	101.5	184.3	407.3	1111.4	15595.3
0.08	0.13	0.1	0.125	343	1301.8	2394.2	101.3	182.9	398.3	1094.9	14836.0
0.28	0.33	0.9	0.285	344	987.5	1644.0	100.2	172.1	340.0	907.5	9518.4
0.28	0.33	0.7	0.295	345	976.1	1622.3	100.1	170.1	338.0	902.5	9381.9
0.28	0.33	0.5	0.305	343	972.5	1606.0	101.3	174.7	338.4	898.1	9248.2
0.28	0.33	0.3	0.315	343	964.0	1586.8	101.2	174.6	336.1	893.0	9117.5
0.28	0.33	0.1	0.325	345	950.6	1564.7	100.8	172.6	331.9	883.0	8989.5
0.48	0.53	0.9	0.485	327	935.3	1460.8	100.0	177.5	349.0	880.2	8111.7
0.48	0.53	0.7	0.495	324	943.2	1465.7	119.7	178.7	354.1	883.4	8114.3
0.48	0.53	0.5	0.505	324	943.4	1466.1	119.6	178.8	354.2	883.6	8116.9
0.48	0.53	0.3	0.515	327	935.8	1461.9	100.0	177.6	349.1	880.8	8119.5
0.48	0.53	0.1	0.525	330	928.4	1457.8	100.1	176.4	344.8	879.3	8122.1
0.68	0.73	0.9	0.685	343	963.2	1585.2	101.2	174.6	336.0	892.7	9106.3
0.68	0.73	0.7	0.695	343	972.3	1605.5	101.3	174.7	338.4	898.0	9244.7
0.68	0.73	0.5	0.705	345	976.4	1622.9	100.1	170.1	338.0	902.7	9386.2
0.68	0.73	0.3	0.715	344	988.3	1645.8	100.2	172.1	340.2	907.9	9531.1
0.68	0.73	0.1	0.725	344	997.9	1667.4	100.3	171.5	340.4	911.5	9679.4
0.88	0.93	0.9	0.885	343	1339.3	2489.8	101.6	184.1	407.3	1110.0	15538.6
0.88	0.93	0.7	0.895	343	1383.9	2605.8	100.5	186.8	412.8	1132.4	16399.3
0.88	0.93	0.5	0.905	342	1435.5	2734.4	100.1	188.7	428.6	1156.7	17333.9
0.88	0.93	0.3	0.915	341	1490.8	2874.0	100.4	192.1	438.3	1197.7	18351.0
0.88	0.93	0.1	0.925	340	1550.2	3026.0	100.7	195.7	442.9	1228.1	19460.8

a standard deviation of 866.6% and an IQR of 676.5%. It is observed that the values of the descriptive statistics of RE_6 increase as the value of π_s gets close to zero or one.

4.2.2 Comparing the estimator $\hat{\pi}_s$ with the Singh and Tarray (2016) estimator $\hat{\pi}_{ST}$

The relative efficiency of the proposed estimator $\hat{\pi}_s$ with respect to the estimator $\hat{\pi}_{ST}$ under Neyman allocation is given by

$$RE_7 = \frac{V_{Ney}\left(\hat{\pi}_{ST}\right)}{V_{Ney}\left(\hat{\pi}_s\right)} \times 100,$$

where $V_{Ney}\left(\hat{\pi}_{ST}\right)$ and $V_{Ney}\left(\hat{\pi}_s\right)$ are as given in (5) and (23), respectively.

We assume that there are two strata in the population, i.e. $k = 2$, $\pi_{s1} \neq \pi_{s2}$, $\pi_{y1} = \pi_{y2} = \pi_y$, $P_{11} = P_{21} = P_1$, $P_{12} = P_{22} = P_2$ with $P_3 = 1 - P_1 - P_2 > 0$ and $q_1 = q_2 = q$. The values of RE_7 are calculated for different combinations from the values of π_{s1}, π_{s2}, π_y, W_1, $W_2 = 1 - W_1$, P_1, P_2 and q where each of the parameters π_y, W_1, P_1, P_2 and q takes values from 0.1 to 0.9 with a step of 0.1 such that $P_3 = 1 - P_1 - P_2 > 0$ while the values of π_{s1} range from 0.08 to 0.88 with a step of 0.2 and π_{s2} from 0.13 to 0.93 with a step of 0.2. It is found that the proposed estimator $\hat{\pi}_s$ is more efficient than $\hat{\pi}_{ST}$ in about 64% of the cases. Figure 9 shows that the value of RE_7 increases as the value of

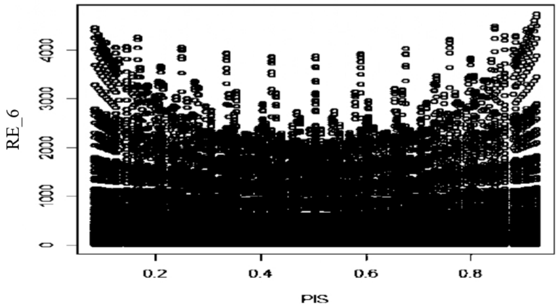

Figure 8. The relationship between RE_6 and π_s.

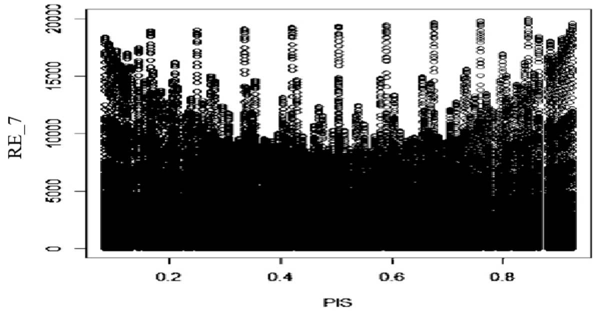

Figure 9. The relationship between RE_7 and π_s.

π_s gets close to zero or one.

Table 8 presents summary statistics of RE_7 for each (π_{s1}, π_{s2}) pair. At $\pi_s = 0.105$ there are 1707 different combinations of the parameters where RE_7>100%. Among these the values of RE_7 range from a minimum of 100.358% to a maximum of 16517.8% with a median of 494.3%, a mean of 1538.5%, a standard deviation of 2652.6% and an IQR of 1302.5%.

4.2.3 Comparing the stratified estimator $\hat{\pi}_s$ with the estimator $\hat{\pi}$

The following theorem shows that the stratified estimator $\hat{\pi}_s$ given by (18) is always more efficient than its counterpart $\hat{\pi}$ given by (8) in simple random sampling.

Theorem 6. *Under Neyman allocation, the stratified estimator $\hat{\pi}_s$ given by* (18)*, where $q_h = q$, $h = 1, 2, \ldots, k$, is always more efficient than the estimator $\hat{\pi}$ given by* (8) *in simple random sampling where $q_1 = q$.*

Proof. The proof is obtained by showing that

$$n\left[V(\hat{\pi}) - V_{Ney}(\hat{\pi}_s)\right] = \sum_{h=1}^{l-1}\sum_{l>h}^{k} W_h W_l \left\{ \left[b_h^{\frac{1}{2}} - b_l^{\frac{1}{2}} \right]^2 + (\pi_{sh} - \pi_{sl})^2 \right\} \geq 0,$$

where $b_h = \pi_{sh}(1 - \pi_{sh}) + 0.5\,(1 - q)\,[1 - 0.5\,(1 - q)]/q^2$, $q \neq 0$, for $h = 1, 2, \ldots, k$, $V(\hat{\pi})$ is given by (9) where $q_1 = q$, and $V_{Ney}(\hat{\pi}_s)$ is given in (23) where $q_h = q$ for $h = 1, 2, \ldots, k$. ■

Table 7. Summary statistics of percent relative efficiency (RE_6) for different levels of π_s.

π_{s1}	π_{s2}	W_1	π_s	f	**Mean**	**StDev**	**Min.**	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**	**Max.**
0.08	0.13	0.9	0.085	274	754.7	942.7	103.3	193.2	371.0	889.6	4468.1
0.08	0.13	0.7	0.095	276	733.5	902.3	102.8	181.7	366.0	884.0	4236.9
0.08	0.13	0.5	0.105	276	718.1	866.6	106.3	181.1	359.4	857.6	4023.4
0.08	0.13	0.3	0.115	276	703.7	833.5	107.9	180.5	351.2	856.7	3825.9
0.08	0.13	0.1	0.125	276	690.2	802.8	109.3	180.1	345.4	822.8	3642.8
0.28	0.33	0.9	0.285	288	570.2	579.4	102.7	182.0	319.8	647.0	2365.3
0.28	0.33	0.7	0.295	288	567.6	573.7	103.8	181.7	317.6	643.7	2332.2
0.28	0.33	0.5	0.305	288	565.1	568.1	104.8	182.6	316.1	641.5	2299.9
0.28	0.33	0.3	0.315	288	562.6	562.7	105.8	182.9	315.8	640.2	2268.2
0.28	0.33	0.1	0.325	288	560.2	557.5	106.9	181.9	316.3	638.9	2237.2
0.48	0.53	0.9	0.485	288	543.3	521.0	117.7	185.1	319.8	656.3	2026.9
0.48	0.53	0.7	0.495	288	543.4	521.0	118.1	186.3	321.4	658.0	2027.4
0.48	0.53	0.5	0.505	288	543.4	521.1	118.1	186.3	321.4	658.0	2027.8
0.48	0.53	0.3	0.515	288	543.5	521.2	117.7	185.0	319.8	656.3	2028.2
0.48	0.53	0.1	0.525	288	543.5	521.4	117.3	183.7	318.3	654.6	2028.7
0.68	0.73	0.9	0.685	288	562.4	562.3	105.9	182.8	315.8	640.2	2266.2
0.68	0.73	0.7	0.695	288	565.0	568.0	104.8	182.6	316.1	641.5	2299.2
0.68	0.73	0.5	0.705	288	567.7	573.9	103.7	181.7	317.7	643.7	2333.0
0.68	0.73	0.3	0.715	288	570.4	579.9	102.7	182.1	319.9	646.9	2367.6
0.68	0.73	0.1	0.725	288	573.2	586.0	101.6	182.2	322.1	650.2	2403.0
0.88	0.93	0.9	0.885	276	702.7	831.3	108.0	180.5	350.7	854.1	3813.3
0.88	0.93	0.7	0.895	276	717.9	866.0	106.4	181.1	359.2	857.6	4020.1
0.88	0.93	0.5	0.905	276	734.1	903.7	102.7	181.7	366.3	885.4	4244.5
0.88	0.93	0.3	0.915	274	756.3	946.3	103.1	193.1	371.3	889.5	4488.6
0.88	0.93	0.1	0.925	274	775.2	990.8	100.3	191.5	381.7	890.2	4754.7

5. Conclusion

In an attempt to increase the efficiency when estimating π_s, the population proportion bearing a sensitive characteristic A, and at the same time reduce the cost of survey, an improved unrelated question randomized response model is proposed. This model is a restricted version of the model proposed by Mahmood et al. (1998) where the design probabilities are chosen so that the resulting estimator of π_s along with its variance do not depend on π_y, the population proportion bearing the non-sensitive characteristic Y. As a result, the estimation process of π_s requires selecting only a single simple random sample and not two as in Mahmood et al.'s model in its general set up. In addition, it is shown that the resulting estimator of π_s is more efficient than the estimator of Mahmood et al. in 99.8% of the cases. Moreover, it is also shown that this estimator can be easily adjusted to be more efficient than the Singh et al. (2003), Kim and Warde (2005) and Perri (2008) estimators. The minimum sample sizes required for the proposed estimator of π_s to lie inside the unit interval [0, 1] are determined through a simulation study. Moreover, the restricted model is extended to stratified random sampling and it is shown that the stratified estimator under Neyman allocation is more efficient than its counterpart in simple random sampling, as well as being more efficient than the Kim and Elam (2007) and Singh and Tarray (2016) estimators in stratified random sampling.

Table 8. Summary statistics of percent relative efficiency (RE_7) for different levels of π_s.

π_{s1}	π_{s2}	W_1	π_s	f	**Mean**	**StDev**	**Min.**	1^{st} **Quartile**	**Median**	3^{rd} **Quartile**	**Max.**
0.08	0.13	0.9	0.085	1686	1650.0	2912.8	100.395	224.3	520.4	1608.4	18387.4
0.08	0.13	0.7	0.095	1695	1593.7	2778.5	100.131	221.5	513.5	1554.9	17415.1
0.08	0.13	0.5	0.105	1707	1538.5	2652.6	100.358	220.4	494.3	1522.9	16517.8
0.08	0.13	0.3	0.115	1712	1492.5	2539.7	100.567	219.8	487.4	1504.7	15688.2
0.08	0.13	0.1	0.125	1714	1451.6	2436.2	100.543	219.2	476.6	1455.7	14919.5
0.28	0.33	0.9	0.285	1794	1120.2	1677.2	100.048	207.5	421.1	1187.2	9533.3
0.28	0.33	0.7	0.295	1799	1111.1	1657.2	100.016	207.9	418.7	1178.7	9394.3
0.28	0.33	0.5	0.305	1799	1105.0	1639.1	100.108	208.0	417.7	1180.2	9258.2
0.28	0.33	0.3	0.315	1798	1099.5	1621.6	100.084	208.5	417.6	1177.6	9125.1
0.28	0.33	0.1	0.325	1802	1091.5	1603.0	100.060	208.2	416.2	1170.8	8994.9
0.48	0.53	0.9	0.485	1829	1065.2	1491.0	100.032	219.7	429.6	1174.0	8111.7
0.48	0.53	0.7	0.495	1829	1068.8	1493.3	101.102	221.6	432.2	1176.2	8114.3
0.48	0.53	0.5	0.505	1831	1071.5	1495.2	101.012	220.3	436.3	1186.0	8116.9
0.48	0.53	0.3	0.515	1836	1072.5	1496.3	100.020	221.4	437.4	1181.5	8119.5
0.48	0.53	0.1	0.525	1840	1074.1	1497.7	100.060	222.3	437.3	1183.6	8122.1
0.68	0.73	0.9	0.685	1932	1161.6	1663.9	100.045	224.3	467.2	1274.8	9114.0
0.68	0.73	0.7	0.695	1937	1173.5	1686.5	100.108	224.0	464.1	1291.5	9254.7
0.68	0.73	0.5	0.705	1940	1186.8	1710.1	100.064	224.0	469.9	1299.6	9398.6
0.68	0.73	0.3	0.715	1943	1200.2	1734.3	100.156	225.1	477.4	1306.3	9546.0
0.68	0.73	0.1	0.725	1945	1214.5	1759.4	100.18	226.3	473.9	1330.0	9696.8
0.88	0.93	0.9	0.885	1995	1671.7	2679.9	100.097	272.1	624.4	1696.8	15626.8
0.88	0.93	0.7	0.895	1995	1733.0	2810.9	100.492	277.7	650.4	1746.3	16496.3
0.88	0.93	0.5	0.905	1998	1796.1	2951.5	100.131	282.6	672.5	1789.1	17442.0
0.88	0.93	0.3	0.915	1998	1866.2	3105.8	100.395	287.1	697.8	1845.2	18473.0
0.88	0.93	0.1	0.925	2000	1939.6	3272.7	100.535	291.8	717.0	1899.4	19598.0

Acknowledgements. The authors are thankful to a learned referee for valuable contributions that brought the original manuscript to its present form.

References

CHAUDHURI, A. AND CHRISTOFIDES, T. (2013). *Indirect Questioning in Sample Surveys*. Springer-Verlag, Berlin.

COCHRAN, W. (1977). *Sampling Techniques*. John Wiley & Sons, New York.

GREENBERG, B. G., ABUL-ELA, A. L. A., SIMMONS, W. R., AND HORVITZ, D. G. (1969). The unrelated question randomized response model-theoretical framework. *Journal of the American Statistical Association*, **64**, 520–539.

HORVITZ, D. G., SHAH, B. V., AND SIMMONS, W. R. (1967). The unrelated question randomized response model. *Social Statistics Section Proceedings of the American Statistical Association*, 65–72.

KIM, J. M. AND ELAM, M. E. (2007). A stratified unrelated question randomized response model. *Statistical Papers*, **48**, 215–233.

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

Lee, C., Sedory, S. A., and Singh, S. (2013). Simulated minimum sample sizes for various randomized response models. *Communications in Statistics – Simulation and Computation*, **42**, 771–789.

Mahmood, M., Singh, S., and Horn, S. (1998). On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. *Biometrical Journal*, **40**, 237–242.

Perri, P. F. (2008). Modified randomized devices for Simmons' model. *Model Assisted Statistics and Applications*, **3**, 233–239.

Singh, H. P. and Tarray, T. A. (2016). A sinuous stratified unrelated question randomized response model. *Communications in Statistics – Theory and Methods*, **45**, 6510–6520.

Singh, S., Horn, S., Singh, R., and Mangat, N. S. (2003). On the use of modified randomization device for estimating the prevalence of a sensitive attribute. *Statistics in Transition*, **6**, 515–522.

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

Manuscript received 2018-11-17, revised 2019-06-30, accepted 2019-08-25.