Hacettepe Journal of Mathematics and Statistics
Volume 47 (2) (2018), 403 – 422

Two mixed randomized response models under simple and stratified random sampling with replacement schemes

Fatima Batool*† Kaneez Fatima‡ and Zawar Hussain§

Abstract

We have proposed two mixed randomized response models for surveying sensitive issues. The properties of proposed estimators are derived under the simple and stratified random sampling schemes while considering completely and less than completely truthful reporting cases. We proved that the proposed models are unconditionally efficient than [13, 15, 16, 20]. In order to get the idea of gain in efficiency and model stability numerical and graphical efficiency comparisons are done for the two models under two mentioned cases.

Keywords: Categorical responses, Randomization, Less than completely truthful reporting, Stratification, Optimum allocation.

2000 AMS Classification: 62D05

Received : 15.02.2016 *Accepted* : 31.05.2016 *Doi* : 10.15672/HJMS.201612818614

1. Introduction

Surveys studies related to sensitive issues have wide applicability in social sciences. One may need to conduct surveys on topics such as alcohol consumption patterns, cocaine use, crime victims, illegitimacy, sexual orientation, weight/leprosy stigma, experience of a mental illness associated with a particular race, religion, belief etc and much more. Several researcher have considered these problems for instance [2, 3, 4, 5, 7, 9, 14, 23, 24, 26]. A major challenge confronting researchers in such studies is to gain respondent cooperation without offending them and gather truthful response. Phillips in [18] has discussed that the problem of response bias is likely to arise, even in the survey of innocuous nature

*Department of Statistics, COMSATS Institute of Information Technology, Lahore, Pakistan
Email: gbatoolfatima@gmail.com

†Corresponding Author.

‡Government Girls Higher Secondary School Hyderabad Town, Sargodha, Pakistan,
Email: kaneez_fatima79@hotmail.com

§Department of Statistics, Quaid - i - Azam University, Islamabad, Pakistan
Email: Zhlangah@yahoo.com

and this bias can raise potentially in surveys that aim at gathering sensitive information. Perhaps, survey sampling will become worthless unless we succeed in controlling response and non-response bias. Long run advantages can only be achieved by carefully designing survey procedures for sensitive issues. Otherwise, the results would not reflect true picture of the population and would be misleading due to presence response and non-response bias. These two kind of bias are mainly raised due to two major factors refusal to response and evasive answer bias.

The solution to these problems was originally suggested by Warner in [27] by securing the privacy of the interviewee that eventually increases confidence of respondent not only to respondent but also to provide true response. From here a fully new technique was emerged. A technique which aims at collecting the information on sensitive question without disclosing the identity of the respondent is known as randomized response technique (RRT). Warner successfully presented estimate of the proportion of the qualitative sensitive characteristic of population to improve respondent cooperation and to get reliable data. Greenberg et al. in [6] extended the idea of Warner to unrelated question model. In [16], Moors tried to improve the precision of Greenberg et al. model. Unfortunately, in an attempt to increase precision he neglects privacy of the model as he sets probability $P_2 = 0$ which means only one sample is required instead of two. This endangers not only the randomization of the device but also this implies that the estimate for innocuous question will be obtained through direct questioning. A detail discussion on this can be found in [15]. Some authors for instance Mangat et al. and Singh et al. have tried to improve the privacy problem of Moors model but with a drawback of high cost in data collection.

Another attempt in this direction is the introduction of mixed randomized response technique. Dating back to the history of these models Kim and Warde's [10] model is the most famous one. Kim and Warde proposed the mixed randomized response strategy by involving direct questioning to the procedure. Therefore, two devices R_1 and R_2 are required for two answers 'Yes' and 'No' to the direct question respectively. R_1 of this procedure can be viewed as Greenberg et al. randomization device when proportion of non-sensitive attribute is known such that only one sample is required to estimate one unknown which is proportion of sensitive attribute. Whereas, Kim and Ward utilizes Warner's device as R_2. From this discussion it is clear that Kim and Warde's model is better than Moors, Mangat et al. and Singh et al. models.

Nazuk and Shabbir [17] follows the same procedure with the difference that they keep similar structure in R_2 as that in R_1 but with different probabilities in both devices. Singh and Tarray [20] provided the modification of Kim and Warde's [13] by using Tracy and Osahan [25] model as R_2 instead of Warner's [27] device. Also, Singh and Tarray [21] proposed mixed model by using Singh et al. [19] device as R_2. Moving in the same direction we plan to further improve the efficiency of mixed randomized response models thus models presented in this paper are cost effective. In addition, it is shown intuitively that these models are superior than Moors [16], Mangat et al. [15] and kim and Warde [13] under efficiency as a performance criterion.

In Section 2, we aim at presenting two designs for sensitive surveys and derive the estimators. A researcher must be prepared for incomplete and/or untruthful answers while dealing with highly sensitive issues. Therefore, in Section 3, we have considered less than completely truthful reporting problem for both models and have derived the bias and MSE associated with them.

As Singh and Tarray [20] have proved that their model out performs Kim and Warde's [13] model unconditionally. Therefore, this model seems to be a fairly good competitor of the proposed models. In Section 4, we have compared the efficiency of both models algebraically and numerically under completely and less than completely truthful reporting cases.

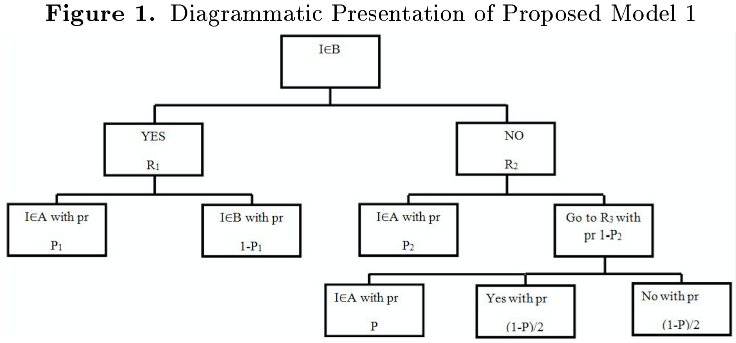
In Section 5, we have extended the proposed models to stratified random sampling scheme. Since population is divided into L strata therefore, it is necessary to provide minimum variance by optimally distributing the complete sample size among these strata. In Section 6, we conclude this study by giving important comments.

2. Proposed models

In this section we have proposed two modified mixed randomized response models. The estimation of sensitive parameter is done by simple random sampling with replacement (SRSWR) scheme by assuming that respondent will provide the truthful response.

2.1. Proposed model 1. Let a random sample of size n be selected using SRSWR. Each respondent from the sample is instructed to answer the direct question: "Do you belong to an innocuous group?" If the respondent answers "Yes" then he or she is instructed to go to randomization device R_1 consisting of the statements (i) "I belong to sensitive Group A" (ii) "I belong to an innocuous Group B" with respective probabilities P_1 and $(1 - P_1)$. If a respondent answers "No" to the direct question then he/she is instructed to go to randomization device R_2 consisting of the statements (i) "I belong to sensitive Group A". (ii) "Go to randomization device R_3" with respective probabilities P_2 and $(1 - P_2)$. Randomization device R_3 consists of statements (i) "I belong to sensitive Group A" (ii) "Say Yes" (iii) "Say No" with respective known probabilities P, $(1 - P)/2$ and $(1 - P)/2$. For the second and third statements, the respondent is asked to report "Yes" or "No" with no relevance to his/her actual status. The underlying assumption of the survey procedure is that the sensitive and innocuous questions are unrelated and independent. The privacy of respondent is protected in this way that respondents will not disclose to the interviewer the question they answered from either R_1 or R_2. The diagrammatic presentation of Proposed Model 1 is given in Figure 1, below.

Figure 1. Diagrammatic Presentation of Proposed Model 1



Let n be the sample size of respondents confronted with the direct question such that $n_1 + n_2 = n$, where n_1 and n_2 denote the number of "Yes" and "No" responses from the sample, respectively.

2.1. Theorem. *An unbiased estimator of population proportion π_s of sensitive attribute is given by*

$$\hat{\pi}_1 = \frac{n_1}{n}\left[\frac{\hat{Y}-(1-P_1)}{P_1}\right] + \frac{n_2}{n}\left[\frac{\hat{X}-\frac{(1-P)(1-P_2)}{2}}{P_2+(1-P_2)P}\right].$$

Proof. Let Y be the probability of Yes responses from the respondents using R_1. Then,

$$Y = P_1\pi_s + (1-P_1)\pi_b,$$

where π_s and π_b be the population proportions of sensitive and innocuous groups, respectively. Note that the respondent coming to R_1 have reported "Yes" to the initial direct question, therefore $\pi_b = 1$ in R_1. The unbiased estimator for population proportion π_s from the respondents who say 'Yes' to the direct question is given as:

$$\text{(2.1)} \qquad \hat{\pi}_{sa} = \frac{\hat{Y}-(1-P_1)}{P_1}.$$

Let "X" be the probability of a "Yes" answer from the respondent using R_2. Then,

$$X = P_2\pi_s + (1-P_2)\left[P\pi_s + \frac{1-P}{2}\right].$$

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

$$\text{(2.2)} \qquad \hat{\pi}_{sb} = \frac{\hat{X}-\frac{(1-P)(1-P_2)}{2}}{P_2+(1-P_2)P}.$$

The mix overall estimator for population proportion having sensitive behavior is given by:

$$\text{(2.3)} \qquad \hat{\pi}_1 = \frac{n_1}{n}\hat{\pi}_{sa} + \frac{n_2}{n}\hat{\pi}_{sb}, \quad \text{for } 0 < \frac{n_1}{n} < 1.$$

using (2.1) and (2.2) in (2.3) will given the required result which completes the proof. □

2.2. Theorem. *The variance of the proposed estimator $\hat{\pi}_1$ is given by*

$$\text{(2.4)} \qquad V(\hat{\pi}_1) = \frac{\pi_s(1-\pi_s)}{n} + \frac{\lambda(1-\pi_s)(1-P_1)}{nP_1} + \frac{(1-\lambda)(1-\alpha^2)}{4n\alpha^2},$$

where $\lambda = \frac{n_1}{n}$, $1-\lambda = \frac{n_2}{n}$ and $\alpha = P_2 + P(1-P_2)$.

Proof. The variance of $\hat{\pi}_{sa}$ is given by:

$$\text{(2.5)} \qquad V(\hat{\pi}_{sa}) = \frac{V(\hat{Y})}{P_1^2} = \frac{Y(1-Y)}{nP_1^2} = \frac{\pi_s(1-\pi_s)}{n_1} + \frac{(1-\pi_s)(1-P_1)}{P_1 n_1}.$$

Similarly, the variance of $\hat{\pi}_{sb}$ is given by:

$$\text{(2.6)} \qquad V(\hat{\pi}_{sb}) = \frac{\pi_s(1-\pi_s)}{n_2} + \frac{1-\alpha^2}{4n_2\alpha^2}.$$

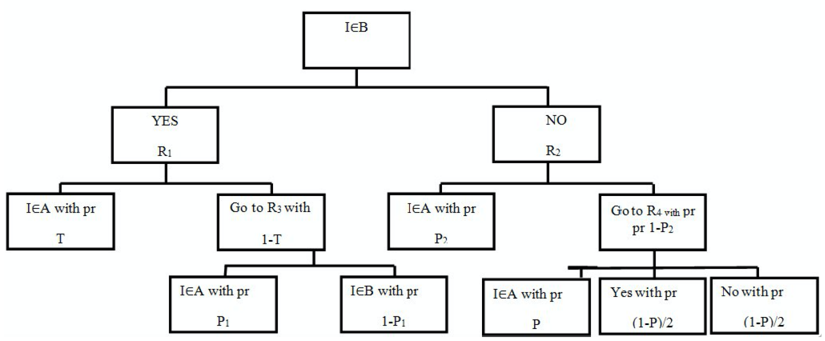
The variance of estimator $\hat{\pi}_1$ is defined as:

$$\text{(2.7)} \qquad V(\hat{\pi}_1) = \frac{n_1^2}{n^2}V(\hat{\pi}_{sa}) + \frac{n_2^2}{n^2}V(\hat{\pi}_{sb}).$$

using (2.5) and (2.6) in (2.7) and after some simplifications we get the required result, hence the theorem. □

2.2. Proposed model 2. Consider a random sample of size n selected by using SRSWR scheme. Each respondent from the sample is instructed to answer the direct question: "Do you belong to an innocuous group?" If the respondent answers "Yes" then he/she is instructed to go to randomization device R_1 consisting of the statements (i) "I belong to sensitive Group A" (ii) "Go to randomization device R_3" with respective probabilities T and $(1-T)$. The randomization device R_3 consists of the statements (i) "I belong to sensitive Group A" (ii) "I belong to an innocuous Group B" with respective probabilities P_1 and $(1-P_1)$. If a respondent responds "No" to direct question then he/she is instructed to go to randomization device R_2 consisting of the statements (i) "I belong to sensitive Group A" (ii) "Go to randomization device R_4" with respective probabilities P_2 and $(1-P_2)$. Randomization device R_4 consists of the statements (i) "I belong to sensitive Group A" (ii) "Say Yes" (iii) "Say No" with respective known probabilities P, $(1-P)/2$ and $(1-P)/2$. For the second and third statements, the respondent is to report "Yes" or "No" with no relevance to his/her actual status. The diagrammatic presentation of Proposed Model 2 is given in Figure 2, below.

Figure 2. Diagrammatic Presentation of Proposed Model 2



Let π_s and π_b be the population proportions of sensitive and innocuous group respectively and n be the sample size confronted with the direct question and n_1 and n_2 $(n_1+n_2=n)$ denote the number of "Yes" and "No" responses from the sample. Note that the respondents coming to R_1 have reported "Yes" to the initial direct question, therefore $\pi_b = 1$ in R_1.

2.3. Theorem. *An unbiased estimator of population proportion π_s of sensitive attribute is given by*

$$\hat{\pi}_2 = \frac{n_1}{n}\left[\frac{\hat{Y}-(1-\alpha_1)}{\alpha_1}\right] + \frac{n_2}{n}\left[\frac{\hat{X}-(1-\alpha_2)/2}{\alpha_2}\right],$$

where $\alpha_1 = T + P_1(1-T)$ and $\alpha_2 = P + P_2(1-P)$.

Proof. Let "Y" and "X" be the probability of "Yes" responses from the respondent coming from R_1 and R_2 respectively. Then $Y = T\pi_s + (1-T)[P_1\pi_s + (1-P_1)]$ and $X = P_2\pi_s + (1-P_2)[P\pi_s + (1-P)/2]$. Taking expectation of Y and X we get

$$\hat{\pi}_{sa} = \frac{\hat{Y}-(1-\alpha_1)}{\alpha_1}, \tag{2.8}$$

and

$$\hat{\pi}_{sb} = \frac{\hat{X} - (1-\alpha_2)/2}{\alpha_2}. \tag{2.9}$$

Using (2.8) and (2.9) in (2.3), we get the required result which complete the proof. □

2.4. Theorem. *The variance of* $\hat{\pi}_2$ *is given by*

$$V(\hat{\pi}_2) = \frac{\pi_s(1-\pi_s)}{n} + \frac{\lambda(1-\pi_s)(1-\alpha_1)}{n\alpha_1} + \frac{(1-\lambda)(1-\alpha_2^2)}{4n\alpha_2^2}, \tag{2.10}$$

where $\lambda = \frac{n_1}{n}$ *and* $1-\lambda = \frac{n_2}{n}$.

Proof. The variances of $\hat{\pi}_{sa}$ and $\hat{\pi}_{sb}$ given in (2.8) and (2.9) are respectively given as:

$$V(\hat{\pi}_{sa}) = \frac{\pi_s(1-\pi_s)}{n_1} + \frac{(1-\pi_s)(1-\alpha_1)}{n_1\alpha_1}, \tag{2.11}$$

and

$$V(\hat{\pi}_{sb}) = \frac{\pi_s(1-\pi_s)}{n_2} + \frac{1-\alpha_2^2}{4n_2\alpha_2^2}. \tag{2.12}$$

using (2.11) and (2.12) in (2.7) we get the required result which completes the proof. □

3. Proposed models under less than completely truthful reporting

While surveying sensitive issues, there is high chance of the problem of less than completely truthful reporting. It is assumed that respondents do not lie about the innocuous question but they may lie for the sensitive issue. In the Proposed Models, we assign the probabilities of truthful responses to the proposed models. Let W_i, for $i = 1, 2$ be the probability of less than completely truthful reporting for first and second randomization devices, where $0 \leq W_i \leq 1$. Here we will only consider the case of known W_i's. However, if W_i are not know and their estimation is essential one can extend this whole set up for two samples from each respondent such that two equations can be obtained to estimate two unknowns.

3.1. Proposed Model 1. Let Y^* and X^* be the probability of "Yes" from the respondents using R_1 and R_2. Then $Y^* = P_1\pi_s W_1 + (1-P_1)$ and $X^* = \alpha\pi_s W_2 + (1-\alpha)/2$, where $\alpha = P_2 + (1-P_2)P$. Thus $\hat{\pi}^*_{sa}$ and $\hat{\pi}^*_{sb}$ are given by:

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

and

$$\hat{\pi}^*_{sb} = \frac{\hat{X}^* - (1-\alpha)/2}{\alpha}.$$

By definition, the overall mixed estimator $\hat{\pi}^*_1$ is now given as:

$$\begin{aligned}\hat{\pi}^*_1 &= \frac{n_1}{n}\hat{\pi}^*_{sa} + \frac{n_2}{n}\hat{\pi}^*_{sb}, \\ &= \frac{n_1}{n}\left[\frac{\hat{Y}^* - (1-P_1)}{P_1}\right] + \frac{n_2}{n}\left[\frac{\hat{X}^* - (1-\alpha)/2}{\alpha}\right].\end{aligned}$$

3.1. Theorem. *The bias and MSE in proposed estimator $\hat{\pi}_1$ under less than truthful reporting case is given as:*

$$Bias(\hat{\pi}_1^*) = \frac{n_1}{n}\pi_s(W_1 - 1) + \frac{n_2}{n}\pi_s(W_2 - 1). \tag{3.1}$$

and

$$\begin{aligned} MSE(\hat{\pi}_1^*) =& \frac{\lambda\pi_s W_1(1 - W_1\pi_s)}{n} + \frac{\lambda(1 - P_1)(1 - W_1\pi_s)}{nP_1} + \frac{(1 - \lambda)(1 - \alpha^2)}{4n\alpha^2} \\ &+ \frac{\pi_s W_2(1 - W_2\pi_s)(1 - \lambda)}{n} \\ &+ \pi_s^2[\lambda(W_1 - 1) + (1 - \lambda)(W_2 - 1)]^2. \end{aligned} \tag{3.2}$$

Proof. We have

$$\begin{aligned} Bias(\hat{\pi}_{sa}^*) &= E[\hat{\pi}_{sa}^* - \pi_s] = E\Big[\frac{\hat{Y}^* - Y}{P_1}\Big] \\ &= E\Big[\frac{P_1\hat{\pi}_{sa}W_1 + (1 - P_1) - P_1\pi_s - (1 - P_1)}{P_1}\Big]. \end{aligned}$$

Since $\hat{\pi}_{sa}$ is unbiased estimator of π_s, therefore,

$$Bias(\hat{\pi}_{sa}^*) = \pi_s(W_1 - 1)$$

Similarly,

$$\begin{aligned} Bias(\hat{\pi}_{sb}^*) =& E[\hat{\pi}_{sb}^* - \pi_s] \\ =& E\Big[\frac{\hat{X}^* - (1 - \alpha)/2}{\alpha} - \frac{X - (1 - \alpha)/2}{\alpha}\Big] \\ =& E\Big[\frac{\alpha\hat{\pi}_{sb}W_2 + (1 - \alpha)/2 - \alpha\pi_s - (1 - \alpha)/2}{\alpha}\Big]. \end{aligned}$$

Since $E(\hat{\pi}_{sb}) = \pi_s$, therefore,

$$Bias(\hat{\pi}_{sb}^*) = \pi_s(W_2 - 1)$$

The bias in overall mix estimator $\hat{\pi}_1^*$ is defined as

$$Bias(\hat{\pi}_1^*) = \frac{n_1}{n}Bias(\hat{\pi}_{sa}^*) + \frac{n_2}{n}Bias(\hat{\pi}_{sb}^*). \tag{3.3}$$

Substitution of results of $Bias(\hat{\pi}_{sa}^*)$ and $Bias(\hat{\pi}_{sb}^*)$ in (3.3) will give the required expression.

The variance of the estimator $\hat{\pi}_{sa}^*$ is given by:

$$V(\hat{\pi}_{sa}^*) = \frac{Y^*(1 - Y^*)}{n_1 P_1^2} = \frac{\pi_s W_1(1 - W_1\pi_s)}{n_1} + \frac{(1 - P_1)(1 - W_1\pi_s)}{n_1 P_1}. \tag{3.4}$$

Similarly, the variance and MSE of the estimator $\hat{\pi}_{sb}^*$ can be derived as:

$$V(\hat{\pi}_{sb}^*) = \frac{X^*(1 - X^*)}{n_2\alpha^2} = \frac{1 - \alpha^2}{4n_2\alpha^2} + \frac{\pi_s W_2(1 - W_2\pi_s)}{n_2}. \tag{3.5}$$

The MSE for the estimator $\hat{\pi}_1^*$ under less than completely truthful reporting case is defined as:

$$MSE(\hat{\pi}_1^*) = V(\hat{\pi}_1^*) + [Bias(\hat{\pi}_1^*)]^2 = \lambda^2 V(\hat{\pi}_{sa}^*) + (1 - \lambda)^2 V(\hat{\pi}_{sb}^*) + [Bias(\hat{\pi}_1^*)]^2. \tag{3.6}$$

Using (3.1), (3.4) and (3.5) in (3.6) will give the required result which completes the proof. □

3.2. Proposed Model 2. Let we denote the probability of "Yes" responses from the respondents coming from R_1 and R_2 by Y^* and X^*. Then $Y^* = P_1\pi_s W_1 + (1-P_1)T\pi_s W_1 + (1-P_1)(1-T)$ and $X^* = P_2\pi_s W_2 + (1-P_2)\left[P\pi_s W_2 + (1-P)/2\right]$, such that

$$\hat{\pi}^*_{sa} = \frac{\hat{Y}^* - (1-\alpha_1)}{\alpha_1}, \tag{3.7}$$

and

$$\hat{\pi}^*_{sb} = \frac{\hat{X}^* - (1-\alpha_2)/2}{\alpha_2}. \tag{3.8}$$

3.2. Theorem. *The bias and MSE in proposed estimator $\hat{\pi}_2$ under less than truthful reporting case are given by:*

$$B(\hat{\pi}^*_2) = \frac{n_1}{n}\pi_s(W_1-1) + \frac{n_2}{n}\pi_s(W_2-1),$$

and

$$\begin{aligned} MSE(\hat{\pi}^*_2) =& \frac{\lambda\pi_s W_1(1-W_1\pi_s)}{n} + \frac{\lambda(1-\alpha_1)(1-W_1\pi_s)}{n\alpha_1} + \frac{(1-\lambda)(1-\alpha_2^2)}{4n\alpha_2^2} \\ &+ \frac{\pi_s W_2(1-\lambda)(1-W_2\pi_s)}{n} \\ &+ \pi_s^2[\lambda(W_1-1) + (1-\lambda)(W_2-1)]^2. \end{aligned} \tag{3.9}$$

The proof is similar to Theorem 3.1.

4. Algebraic and numeric efficiency comparisons

Algebraic efficiency comparisons of Proposed Models, under completely truthful reporting case, has been done with Singh and Tarray (2013) model. The $V(\hat{\pi}_t)$ for Singh and Tarray (2013) model under completely truthful reporting case is

$$V(\hat{\pi}_t) = \frac{\pi_s(1-\pi_s)}{n} + \frac{\lambda(1-\pi_s)(1-P_1)}{nP_1} + \frac{(1-\lambda)(1-P^2)}{4nP^2}. \tag{4.1}$$

The Proposed Models will be efficient than Singh and Tarray (2013) model if

$$V(\hat{\pi}_i) < V(\hat{\pi}_t), \quad \text{for i=1, 2.} \tag{4.2}$$

Similarly, algebraic efficiency comparisons of Proposed Models, under less than completely truthful reporting case, has been done with Singh and Tarray (2013) model. For Singh and Tarray (2013) model, we have

$$\begin{aligned} MSE(\hat{\pi}^*_t) =& \frac{\lambda\pi_s W_1(1-W_1\pi_s)}{n} + \frac{\lambda(1-P_1)(1-W_1\pi_s)}{nP_1} + \frac{(1-\lambda)(1-P^2)}{4nP^2} \\ &+ \frac{\pi_s W_2(1-W_2\pi_s)(1-\lambda)}{n} \\ &+ \pi_s^2[\lambda(W_1-1) + (1-\lambda)(W_2-1)]^2 \end{aligned} \tag{4.3}$$

The Proposed Models will be efficient than Singh and Tarray (2013) model if

$$MSE(\hat{\pi}^*_i) < MSE(\hat{\pi}^*_t), \quad \text{for i= 1, 2.} \tag{4.4}$$

4.1. Proposed model 1 under completely truthful reporting case. Using (2.4) and (4.1) in (4.2), we get

$$\begin{aligned}&\frac{(1-\lambda)(1-\alpha^2)}{4n\alpha^2} < \frac{(1-\lambda)(1-P^2)}{4nP^2}\\ \Longleftrightarrow &\frac{1-\alpha^2}{\alpha^2} < \frac{1-P^2}{P^2}\\ \Longleftrightarrow &P^2 < \alpha^2,\end{aligned}$$

since $\alpha^2 = P_2^2 + (1-P_2)^2P^2 + 2P_2P(1-P_2)$. Therefore,

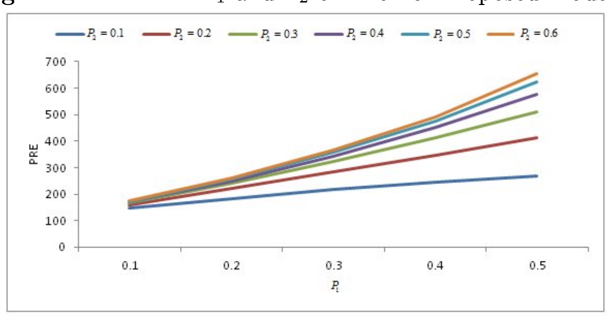
$$\Longleftrightarrow (1-P_2)^2P + 2P_2(1-P_2) > 0, 0 < P, P_2 < 1.$$

which is always true for all values of P, and P_2. Thus, the Proposed Model 1 is always more efficient than the Singh and Tarray (2013) model.
Next, we plan to check the magnitude of gain in efficiency in order to investigate the model's stability. The percent relative efficiency of Proposed Model 1 relative to Singh and Tarray (2013) model is calculated by using the following formula:

$$PRE(\hat{\pi}_1, \hat{\pi}_t) = \frac{V(\hat{\pi}_t)}{V(\hat{\pi}_1)} * 100,$$

for different values of P, P_1 and P_2.
The values of P_1 vary from $0.1 - 0.9$ with a hump of 0.1, $P_2 = 0.3$, $n = 1000$, $\lambda = 0.8$ and $\pi_s = 0.1$. Some of the results are reported in Table 1. Figure 3, shows upward trend in PRE of Proposed Model 1 with increase in P_1.

Figure 3. Effect of P_1 and P_2 on PRE of Proposed Model 1



4.2. Proposed model 2 under completely truthful reporting case. By using (2.10) and (4.1) in (4.2) we get

$$\begin{aligned}&\frac{\lambda(1-\pi_s)(1-\alpha_1)}{n\alpha_1} + \frac{(1-\lambda)(1-\alpha_2^2)}{4n\alpha_2^2} < \frac{\lambda(1-\pi_s)(1-P_1)}{nP_1} + \frac{(1-\lambda)(1-P^2)}{4nP^2}\\ \Longleftrightarrow &\frac{\lambda(1-\pi_s)}{n}\left\{\frac{1-\alpha_1}{\alpha_1} - \frac{1-P_1}{P_1}\right\} + \frac{1-\lambda}{4n}\left\{\frac{1-\alpha_2^2}{\alpha_2^2} - \frac{1-P^2}{P^2}\right\} < 0\\ \Longleftrightarrow &\frac{\lambda(1-\pi_s)}{n}\left\{\frac{P_1-\alpha_1}{P_1\alpha_1}\right\} + \frac{1-\lambda}{4n}\left\{\frac{P^2-\alpha_2^2}{P^2\alpha_2^2}\right\} < 0,\end{aligned}$$

Table 1. Percent Relative Efficiency of Model-I relative to Singh and Tarray (2013) model for $n = 1000$ and $\lambda = \frac{n_1}{n} = 0.8$

PRE			π_1				
P	P_1	P_2	0.1	0.2	0.3	0.4	0.5
0.3	0.1	0.1	102.76	103.05	103.42	103.90	104.56
		0.3	105.41	105.99	106.73	107.72	109.10
		0.5	106.58	107.30	108.22	109.44	111.15
		0.7	107.20	107.99	109.00	110.35	112.25
		0.9	107.56	108.39	109.46	110.89	112.90
	0.3	0.1	109.12	109.66	110.39	111.36	112.70
		0.3	118.99	120.23	121.90	124.18	127.40
		0.5	123.78	125.39	127.58	130.61	134.93
		0.7	126.41	128.24	130.74	134.20	139.18
		0.9	127.99	129.96	132.65	136.38	141.78
	0.5	0.1	116.91	117.06	117.53	118.38	119.72
		0.3	138.15	138.56	139.82	142.13	145.86
		0.5	149.77	150.35	152.15	155.45	160.86
		0.7	156.59	157.27	159.42	163.38	169.90
		0.9	160.86	161.61	163.99	168.38	175.66
	0.7	0.1	126.66	125.39	124.87	125.01	125.83
		0.3	167.18	163.01	161.34	161.79	164.45
		0.5	193.64	186.98	184.34	185.05	189.26
		0.7	210.90	202.37	199.02	199.91	205.29
		0.9	222.48	212.58	208.72	209.75	215.96
	0.9	0.1	139.22	134.83	132.39	131.27	131.21
		0.3	216.36	197.31	187.66	183.41	183.20
		0.5	283.43	246.00	228.37	220.88	220.51
		0.7	337.58	282.00	257.25	246.99	246.48
		0.9	379.93	308.33	277.74	265.28	264.68

since $P_1 - \alpha_1 = -T(1-P_1)$ and $P^2 - \alpha_2^2 = -P[P(1-P_2)^2 + 2P_2(1-P_2)]$,

$$\iff \frac{\lambda T(1-\pi_s)(1-P_1)}{\alpha_1 P_1} + \frac{(1-\lambda)\left[P(1-P_2)^2 + 2P_2(1-P_2)\right]}{4\alpha_2^2 P} > 0,$$

the left hand side of the above inequality is always greater than 0 irrespective of the parametric values involved. Thus the Proposed Model 2 is always more efficient than the Singh and Tarray (2013) model. The percent relative efficiency of Proposed Model 2 relative to Singh and Tarray (2013) model is calculated to check the magnitude of gain in efficiency by using the following formula:

$$PRE(\hat{\pi}_2, \hat{\pi}_t) = \frac{V(\hat{\pi}_t)}{V(\hat{\pi}_2)} * 100,$$

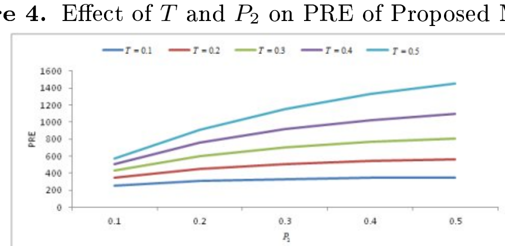
for different values of P, P_1, P_2, and T. The parameter are fixed at 0.1, 0.3, 0.5, 0.7 and 0.9. Some of the PRE results are reported in Table 2. It is evident from Figure 4, that PRE increases by increasing P_2 and T.

4.3. Proposed model 1 vs proposed model 2 under completely truthful reporting case. Proposed model 1 will be efficient than Proposed model 2 if $V(\hat{\pi}_1) < V(\hat{\pi}_2)$.

Table 2. Percent Relative Efficiency of Model-II relative to Singh and Tarray (2013) model for $T = 0.3$, $n = 1000$ and $\lambda = \frac{n_1}{n} = 0.8$

PRE			π_2				
P	P_1	P_2	0.1	0.2	0.3	0.4	0.5
		0.1	433.77	410.58	389.22	369.06	349.47
		0.3	485.23	461.61	440.78	422.29	405.80
	0.1	0.5	511.13	487.49	467.22	450.01	435.80
		0.7	525.64	502.06	482.19	465.83	453.14
		0.9	534.59	510.96	491.37	475.59	463.90
		0.1	207.43	198.04	190.18	183.54	177.86
		0.3	246.27	235.39	227.14	221.19	217.48
	0.3	0.5	267.67	256.02	247.71	242.43	240.36
		0.7	280.28	268.18	259.87	255.10	254.18
		0.9	288.20	275.82	267.54	263.13	263.01
		0.1	165.91	159.24	154.28	150.61	148.03
		0.3	212.21	201.83	195.11	191.29	190.18
0.3	0.5	0.5	240.93	227.86	219.96	216.22	216.50
		0.7	259.06	244.15	235.49	231.86	233.21
		0.9	270.97	254.78	245.60	242.07	244.20
		0.1	151.54	145.63	141.75	139.31	138.03
		0.3	213.42	198.96	190.69	186.58	185.93
	0.7	0.5	258.51	235.86	223.67	218.21	218.28
		0.7	290.24	260.89	245.66	239.18	239.87
		0.9	312.63	278.11	260.60	253.39	254.57
		0.1	147.04	140.74	137.08	135.09	134.37
		0.3	235.86	210.23	197.21	190.95	189.42
	0.9	0.5	317.85	266.41	242.68	231.92	229.58
		0.7	387.56	309.15	275.55	260.86	257.88
		0.9	444.43	341.07	299.19	281.36	277.86

Figure 4. Effect of T and P_2 on PRE of Proposed Model 2



Using (2.4) and (2.10) in this inequality, we get

$$\frac{\lambda(1-\pi_s)(1-P_1)}{nP_1} + \frac{(1-\lambda)(1-\alpha^2)}{4n\alpha^2} < \frac{\lambda(1-\pi_s)(1-\alpha_1)}{n\alpha_1} + \frac{(1-\lambda)(1-\alpha_2^2)}{4n\alpha_2^2},$$

since $\alpha_2 = \alpha$,

$$\Longleftrightarrow \frac{(1-P_1)}{P_1} < \frac{(1-\alpha_1)}{\alpha_1}$$
$$\Longleftrightarrow \alpha_1 < P_1$$
$$\Longleftrightarrow T(1-P_1) < 0,$$

since $0 < T, P_1 < 1$ thus above inequality will never exist. Which further implies that converse which is $V(\hat{\pi}_2) < V(\hat{\pi}_1)$ will always be true thus model 2 outperforms model 1 in terms of efficiency.

4.4. Proposed model 1 under less than completely truthful reporting case. By using (3.2) and (4.3) in (4.4), we get

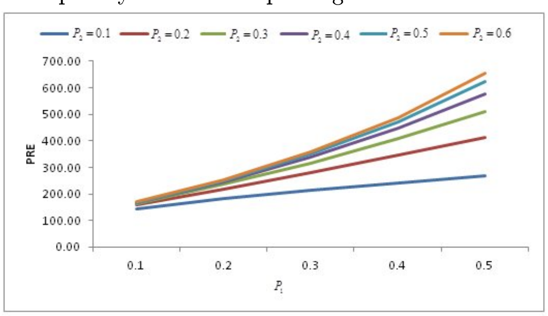
$$\frac{(1-\lambda)(1-\alpha^2)}{4n\alpha^2} < \frac{(1-\lambda)(1-P^2)}{4nP^2},$$
$$\Longleftrightarrow \alpha^2 - P^2 > 0$$
$$\Longleftrightarrow (1-P_2)^2 P + 2P_2(1-P_2) > 0$$

which is always true for all values of P and P_2. Thus $\hat{\pi}_1^*$ is always efficient than $\hat{\pi}_t^*$. The PRE of Proposed Model 1 relative to Singh and Tarray (2013) model is calculated by using the following formula:

$$PRE(\hat{\pi}_1^*, \hat{\pi}_t^*) = \frac{MSE(\hat{\pi}_t^*)}{MSE(\hat{\pi_1}^*)} * 100,$$

for different values of P, P_1, W_1 and W_2. Some findings are reported in Table 3 in Appendix B and graphical representation in Figure 5. Figure 5, shows that there is increase in PRE with the increase in P_1 and P_2.

Figure 5. Effect of P_1 and P_2 on PRE of Proposed Model 1 under Less than Completely Truthful Reporting case



4.5. Proposed model 2 under less than completely truthful reporting case. By using (3.9) and (4.3) in (4.4) we get

$$\frac{\lambda(1-\alpha_1)(1-W_1\pi_s)}{n\alpha_1} + \frac{(1-\lambda)(1-\alpha_2^2)}{4n\alpha_2^2}$$
$$< \frac{\lambda(1-P_1)(1-W_1\pi_s)}{nP_1} + \frac{(1-\lambda)(1-P^2)}{4nP^2},$$

Table 3. Percent Relative Efficiency of Model-I relative to Singh and Tarray (2013) model under Less Than Completely Truthful Reporting for $P_1 = 0.7$, $P_2 = 0.3$, $n = 1000$ and $\lambda = \frac{n_1}{n} = 0.8$

PRE			π_1^*				
P	W_1	W_2	0.1	0.2	0.3	0.4	0.5
0.3	0.5	0.5	183.55	163.01	176.59	179.52	183.55
		0.6	174.48	175.23	177.00	179.74	183.45
		0.7	174.79	175.65	177.36	179.89	183.24
		0.8	175.09	176.04	177.67	179.95	182.90
		0.9	175.38	176.42	177.94	179.93	182.44
	0.6	0.5	173.49	173.45	174.6	176.89	180.31
		0.6	173.81	173.91	175.07	177.22	180.38
		0.7	174.12	174.36	175.48	177.47	180.32
		0.8	174.42	174.77	175.85	177.63	180.13
		0.9	174.72	175.17	176.16	177.70	179.81
	0.7	0.5	172.87	172.32	173.06	175.00	178.19
		0.6	173.19	172.81	173.57	175.42	178.41
		0.7	173.50	173.27	174.04	175.77	178.50
		0.8	173.81	173.71	174.46	176.02	178.46
		0.9	174.12	174.12	174.83	176.19	178.27
	0.8	0.5	172.30	171.39	171.91	173.80	177.15
		0.6	172.63	171.90	172.48	174.33	177.53
		0.7	172.94	172.38	173.00	174.77	177.77
		0.8	173.26	172.84	173.47	175.12	177.87
	0.9	0.5	171.79	170.65	171.17	173.30	177.22
		0.6	172.12	171.18	171.79	173.92	177.76
		0.7	172.44	171.68	172.36	174.46	178.16
		0.8	172.76	172.16	172.88	174.90	178.40
		0.9	173.07	172.62	173.35	175.25	178.49

$$\Longleftrightarrow \lambda(1 - W_1\pi_s)\left[\frac{P_1 - \alpha_1}{\alpha_1 P_1}\right] + \frac{(1-\lambda)}{4}\left[\frac{P^2 - \alpha_2^2}{\alpha_2^2 P_2}\right] < 0$$

$$\Longleftrightarrow \frac{T\lambda(1 - W_1\pi_s)(1 - P_1)}{\alpha_1 P_1} + \frac{(1-\lambda)\left[P(1-P_2)^2 + 2P_2(1-P_2)\right]}{4\alpha_2^2 P} > 0,$$

which is always true for all values of T, W_1, P, P_1, P_2 and π_s. Thus $\hat{\pi}_2^*$ is always better than $\hat{\pi}_t^*$. The percent relative efficiency of Proposed Model 2 relative to Singh and Tarray (2013) model is calculated by using the following formula:

$$PRE(\hat{\pi}_2^*, \hat{\pi}_t^*) = \frac{MSE(\hat{\pi}_t^*)}{MSE(\hat{\pi_2}^*)} * 100,$$

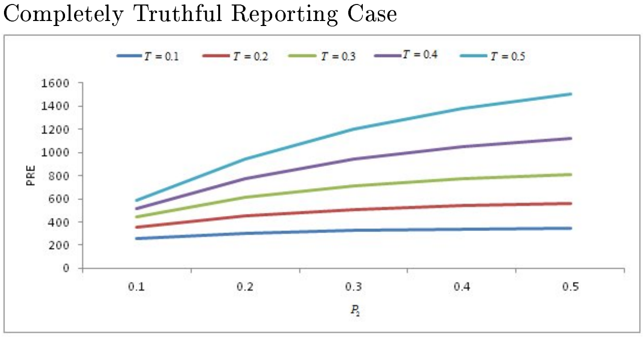
for different values of T, W_1, P, P_1, P_2 and π_s and W_2. Some of the PRE results are reported in Table 4 and graphical representation in Figure 6. Figure 6 depicts increase in PRE with an increase in T and P_2.

4.6. Proposed model 1 vs proposed model 2 under less than completely truthful reporting case. Proposed model 1 will be efficient than Proposed model

Table 4. Percent Relative Efficiency of Model-II relative to Singh and Tarray (2013) model under Less Than Completely Truthful Reporting case for $T = 0.3$, $P_1 = 0.7$, $P_2 = 0.3$, $n = 1000$ and $\lambda = \frac{n_1}{n} = 0.8$

PRE			π_2^*				
P	W_1	W_2	0.1	0.2	0.3	0.4	0.5
0.3	0.5	0.5	223.52	210.81	200.21	191.24	183.55
		0.6	223.03	210.21	199.73	191.00	183.61
		0.7	222.55	209.66	199.32	190.85	183.79
		0.8	222.08	209.14	198.97	190.79	184.09
		0.9	221.63	208.66	198.68	190.83	184.53
	0.6	0.5	221.87	208.88	198.74	190.64	184.05
		0.6	221.38	208.27	198.21	190.31	183.98
		0.7	220.90	207.70	197.75	190.07	184.05
		0.8	220.44	207.17	197.35	189.93	184.23
		0.9	219.99	206.67	197.02	189.87	184.55
	0.7	0.5	220.27	207.03	197.35	190.12	184.70
		0.6	219.78	206.41	196.77	189.71	184.50
		0.7	219.31	205.82	196.27	196.27	189.38
		0.8	218.84	205.28	195.82	189.15	184.50
		0.9	218.39	204.76	195.43	189.00	184.70
	0.8	0.5	218.71	205.25	196.03	189.70	185.51
		0.6	218.22	204.62	195.41	189.19	185.18
		0.7	217.75	204.02	195.41	188.77	184.98
		0.8	217.29	203.46	194.36	188.45	184.92
		0.9	216.84	202.93	193.93	188.22	184.98
	0.9	0.5	217.19	203.54	194.78	189.35	186.50
		0.6	216.71	202.89	194.12	188.75	186.02
		0.7	216.24	202.29	193.52	188.25	185.68
		0.8	215.78	201.71	192.98	187.84	185.48
		0.9	215.33	201.17	192.51	187.52	185.41

Figure 6. Effect of T and P_2 on PRE of Proposed Model 2 under Less than Completely Truthful Reporting Case



2 if $MSE(\hat{\pi}_1^*) < MSE(\hat{\pi}_2^*)$. Using (3.2) and (3.9) in this inequality, we get

$$\frac{\lambda(1-P_1)(1-W_1\pi_s)}{nP_1} + \frac{(1-\lambda)(1-\alpha^2)}{4n\alpha^2}$$
$$< \frac{\lambda(1-\alpha_1)(1-W_1\pi_s)}{n\alpha_1} + \frac{(1-\lambda)(1-\alpha_2^2)}{4n\alpha_2^2},$$

since $\alpha = \alpha_2$, therefore,

$$\Longleftrightarrow \frac{1-P_1}{P_1} < \frac{1-\alpha_1}{\alpha_1}$$
$$\Longleftrightarrow \alpha_1 - P_1 < 0$$
$$\Longleftrightarrow (1-P_1)T < 0$$

since $0 \leq P_1, T \leq 1$ therefore, above inequality never holds, which implies that $MSE(\hat{\pi}_2^*) < MSE(\hat{\pi}_1^*)$. Thus model 2 is always efficient than model 1 under less than completely truthful reporting case.

5. Proposed models under stratified random sampling

Stratification gives the best representative sample of the population being studied, that reduces sample selection bias by ensuring certain part of the population are not under or over represented. Thus, the benefits are in terms of accuracy, greater precision and less cost. It is applied by partitioned the population into non-overlapping groups called strata and sample is selected by using SRSWR. Then randomized response (RR) technique is applied to each stratum. Many researchers have extended RR models to stratification such as [1, 8, 10, 11, 12, 13, 20, 22]. In the following sub section we also extend proposed models to stratified sampling framework.

5.1. Proposed model 1. In this sampling design, the population of N units is subdivided into L strata, such that the j^{th} stratum consists of N_j units, where $j = 1, 2, ..., L$ and $\sum_{j=1}^{L} N_j = N$. Let we want to draw a sample of size m. From j^{th} population stratum, a sample of size m_j is drawn using SRSWR is drawn such that $\sum_{j=1}^{L} m_j = m$. Each selected unit in the sample is instructed to answer the direct question: "Do you belong to an innocuous group?" If the respondent answers "Yes" then he/she is instructed to go to randomization device R_{1j} consisting of the statements (i) "I belong to sensitive Group A" (ii) "I belong to an innocuous Group B" with respective probabilities P_{1j} and $(1 - P_{1j})$. If a respondent answers "No" to the direct question then he/she is instructed to go to randomization device R_{2j} consisting of the statements: (i) "I belong to sensitive Group A" (ii) "Go to randomization device R_{3j}" with respective probabilities P_{2j} and $(1 - P_{2j})$. Randomization device R_{3j} consists of statements (i) "I belong to sensitive Group A" (ii) "Say Yes" (iii) "Say No" with respective known probabilities P_j, $(1 - P_j)/2$ and $(1 - P_j)/2$. For the second and third statements, the respondent is asked to report "Yes" or "No" with no relevance to his/her actual status.

Note that in each stratum there will be some respondents that will respond 'Yes' to the direct question and some who will respond 'No' to the direct question therefore, sample size in each stratum is further subdivided as m_{1j} and m_{2j}, respectively, such that $m_{1j} + m_{2j} = m_j$.

Assume the response of i^{th} unit of the study variable selected from the j^{th} stratum using randomization device R_{1j} is denoted by Y_{ji}, where $i = 1, 2, ..., m_j$ and $G_j = N_j/N$ is the known stratum weights in the population. Then,

$$Y_j = P_{1j}\pi_{sj} + (1 - P_{1j})\pi_{bj}.$$

An unbiased estimator of π_{sj} in terms of sample proportion $\hat{Y}_j$ of "Yes" responses is given by:

$$(5.1) \qquad \hat{\pi}_{saj} = \frac{\hat{Y}_j - (1 - P_{1j})}{P_{1j}}.$$

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

$$(5.2)\qquad V(\hat{\pi}_{saj}) = \frac{\pi_{sj}(1-\pi_{sj})}{m_{1j}} + \frac{(1-\pi_{sj})(1-P_{1j})}{P_{1j}m_{1j}}.$$

Assume "X_j" be the probability of "Yes" responses from the respondents in the j^{th} stratum using randomization device using R_{2j}. Then

$$X_j = P_{2j}\pi_{sj} + (1-P_{2j})\left[P_j\pi_{sj} + \frac{1-P_j}{2}\right].$$

An unbiased estimator of π_{sj} in terms of sample proportion $\hat{X}_j$ of "Yes" responses is given by:

$$(5.3)\qquad \hat{\pi}_{sbj} = \frac{\hat{X}_j - \frac{(1-P_j)(1-P_{2j})}{2}}{P_{2j} + (1-P_{2j})P_j}.$$

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

$$(5.4)\qquad V(\hat{\pi}_{sbj}) = \frac{\pi_{sj}(1-\pi_{sj})}{m_{2j}} + \frac{1-\alpha_j^2}{4m_{2j}\alpha_j^2},$$

where $\alpha_j = P_{2j} + (1-P_{2j})P_j$.

5.1. Theorem. *An unbiased estimator of population proportion of sensitive attribute under stratified random sampling with replacement is given by*

$$\hat{\pi}_1 = \sum_{j=1}^{L}\frac{G_j}{m_j}\left[m_{1j}\left(\frac{\hat{Y}_j-(1-P_{1j})}{P_{1j}}\right) + (m_j - m_{1j})\left(\frac{\hat{X}_j-(1-P_j)(1-P_{2j})/2}{P_{2j}+(1-P_{2j})P_j}\right)\right].$$

and

$$V(\hat{\pi}_1) = \sum_{j=1}^{L}\frac{G_j^2}{m_j}\left[\pi_{sj}(1-\pi_{sj})\right.$$
$$(5.5)\qquad \left. + \frac{4\alpha_j^2\lambda_j(1-\pi_{sj})(1-P_{1j}) + P_{1j}(1-\lambda_j)(1-\alpha_j^2)}{4P_{1j}\alpha_j^2}\right],$$

where $\lambda_j = m_{1j}/m_j$.

Proof. $\hat{\pi}_{1j}$ is defined as

$$(5.6)\qquad \hat{\pi}_{1j} = \frac{m_{1j}}{m_j}\hat{\pi}_{saj} + \frac{m_j - m_{1j}}{m_j}\hat{\pi}_{sbj}, \quad \text{for } 0 < \frac{m_{1j}}{m_j} < 1.$$

We have

$$(5.7)\qquad \hat{\pi}_1 = \sum_{j=1}^{L} G_j\hat{\pi}_{1j} = \sum_{j=1}^{L} G_j\left[\frac{m_{1j}}{m_j}\hat{\pi}_{saj} + \frac{m_j - m_{1j}}{m_j}\hat{\pi}_{sbj}\right].$$

Using (5.1) and (5.3) in (5.7) and some simplification will give the required result.
Let the sample are drawn independently in different strata then, the variance of $\hat{\pi}_1$ is defined as:

$$(5.8)\qquad V(\hat{\pi}_1) = \sum_{j=1}^{L} G_j^2 V(\hat{\pi}_{1j})$$

where

$$(5.9)\qquad V(\hat{\pi}_{1j}) = \frac{m_{1j}^2}{m_j^2}V(\hat{\pi}_{saj}) + \frac{m_{2j}^2}{m_j^2}V(\hat{\pi}_{sbj})$$

Using (5.2) and (5.4) in (5.9) we get

$$V(\hat{\pi}_{1j}) = \frac{\pi_{sj}(1-\pi_{sj})}{m_j} + \frac{4\alpha_j^2\lambda_j(1-\pi_{sj})(1-P_{1j}) + P_{1j}(1-\lambda_j)(1-\alpha_j^2)}{4m_jP_{1j}\alpha_j^2}, \quad (5.10)$$

where $m_j = m_{1j} + m_{2j}$ and $\lambda_j = \frac{m_{1j}}{m_j}$.
Using (5.10) in (5.8) will give the required result hence the theorem. □

If prior information about $\lambda_j = \frac{m_{1j}}{m_j}$ and π_{sj} is known from previous some surveys, then optimal allocation of sample size can be derived by using following theorem.

5.2. Theorem. *The optimal allocation of sample size m to strata sizes $m_1, m_2, ..., m_L$ subject to $m = \sum_{j=1}^{L} m_j$ is given by:*

$$\frac{m_j}{m} = \frac{G_j\left[\pi_{sj}(1-\pi_{sj}) + \frac{4\alpha_j^2\lambda_j(1-\pi_{sj})(1-P_{1j})+P_{1j}(1-\lambda_j)(1-\alpha_j^2)}{4P_{1j}\alpha_j^2}\right]^{\frac{1}{2}}}{\sum_{j=1}^{L} G_j\left[\pi_{sj}(1-\pi_{sj}) + \frac{4\alpha_j^2\lambda_j(1-\pi_{sj})(1-P_{1j})+P_{1j}(1-\lambda_j)(1-\alpha_j^2)}{4P_{1j}\alpha_j^2}\right]^{\frac{1}{2}}},$$

where $m_j = m_{1j} + m_{2j}$ and $\lambda_j = \frac{m_{1j}}{m_j}$.
Thus the minimum variance of the estimator $\hat{\pi}_1$ is given by:

$$V_{min}(\hat{\pi}_1) = \frac{1}{m}\left[\sum_{j=1}^{L} G_j\left\{\pi_{sj}(1-\pi_{sj}) + \frac{4\alpha_j^2\lambda_j(1-\pi_{sj})(1-P_{1j}) + P_{1j}(1-\lambda_j)(1-\alpha_j^2)}{4\alpha_j^2P_{1j}}\right\}^{\frac{1}{2}}\right]^2.$$

5.2. Proposed model 2. In the Proposed Model 2 the whole setup is same as that in Proposed model 1 except each respondent selected in the sample is instructed to answer the direct question: "Do you belong to an innocuous group"? If the respondent answers "Yes" then he/she is instructed to go to randomization device R_{1j} consisting of the statements (i) "I belong to sensitive Group A" (ii) "Go to randomization device R_{3j}" with respective probabilities T_j and $(1-T_j)$. The randomization device R_{3j} consists of the statements (i) "I belong to sensitive Group A" (ii) "I belong to an innocuous Group B" with respective probabilities P_{1j} and $(1-P_{1j})$. If a respondent responds "No" to direct question then he/she is instructed to go to randomization device R_{2j} consisting of the statements (i) "I belong to sensitive Group A" (ii) "Go to randomization device R_{4j}" with respective probabilities P_{2j} and $(1-P_{2j})$. Randomization device R_{4j} consists of the statements (i) "I belong to sensitive Group A" (ii) "Say Yes" (iii) "Say No" with respective known probabilities P_j, $(1-P_j)/2$ and $(1-P_j)/2$. For the second and third statements, the respondent is to report "Yes" or "No" with no relevance to his/her actual status.
Let m_j be the number of units in the sample from the j^{th} stratum and m be the total number of units from all strata. Let m_{1j} be the respondents which respond "Yes" to the direct question and m_{2j} be the respondents which respond "No" to the direct question such that $m = \sum_{j=1}^{L}(m_{1j} + m_{2j})$. Let Y_j be the probability of "Yes" response using randomization device R_{1j} in the j^{th} stratum. Assuming π_{sj} and π_{bj} be the population proportions of sensitive and innocuous group in stratum j, respectively. Let "Y_j" be the probability of "Yes" from the respondent using R_{1j}. Then

$$Y_j = T_j\pi_{sj} + (1-T_j)\Big[P_{1j}\pi_{sj} + (1-P_{1j})\pi_{bj}\Big],$$

An unbiased estimator of π_{sj} in terms of sample proportion of "Yes" responses $\hat{Y}_j$is given by:

$$\hat{\pi}_{saj} = \frac{\hat{Y}_j - (1-T_j)(1-P_{1j})}{T_j + P_{1j}(1-T_j)}. \tag{5.11}$$

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

$$V(\hat{\pi}_{saj}) = \frac{\pi_{sj}(1-\pi_{sj})}{m_{1j}} + \frac{(1-\pi_{sj})(1-\alpha_{1j})}{m_{1j}\alpha_{1j}}, \tag{5.12}$$

where $\alpha_{1j} = T_j + (1-T_j)P_{1j}$.
Let "X_j" be the probability of "Yes" from the respondent using R_{2j}, then

$$X_j = P_{2j}\pi_{sj} + (1-P_{2j})\left[P_j\pi_{sj} + \frac{1-P_j}{2}\right],$$

An unbiased estimator of π_{sj} in terms of sample proportion of "Yes" responses $\hat{X}j$ is given by:

$$\hat{\pi}_{sbj} = \frac{\hat{X} - \frac{(1-P_2)(1-P)}{2}}{P_2 + (1-P_2)P}. \tag{5.13}$$

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

$$V(\hat{\pi}_{sbj}) = \frac{\pi_{sj}(1-\pi_{sj})}{m_{2j}} + \frac{1-\alpha_{2j}^2}{4m_{2j}\alpha_{2j}^2}. \tag{5.14}$$

where $\alpha_{2j} = P_{2j} + (1-P_{2j})P_j$.

5.3. Theorem. *An unbiased estimator of population proportion of sensitive attribute under stratified random sampling with replacement is given by*

$$\hat{\pi}_2 = \sum_{j=1}^{L} \frac{G_j}{m_j}\left[m_{1j}\left(\frac{\hat{Y}_j - (1-T_j)(1-P_{1j})}{T_j + P_{1j}(1-T_j)}\right) + (m_j - m_{1j})\left(\frac{\hat{X} - \frac{(1-P_2)(1-P)}{2}}{P_2 + (1-P_2)P}\right)\right].$$

and

$$\begin{aligned} V(\hat{\pi}_2) = \sum_{j=1}^{L} \frac{G_j^2}{m_j}\Bigg[&\pi_{sj}(1-\pi_{sj}) \\ &+ \frac{4\alpha_{2j}^2\lambda_j(1-\pi_{sj})(1-\alpha_{1j}) + \alpha_{1j}(1-\lambda_j)(1-\alpha_{2j}^2)}{4\alpha_{1j}\alpha_{2j}^2}\Bigg], \end{aligned} \tag{5.15}$$

where $\lambda_j = m_{1j}/m_j$.

Proof. The unbiased estimator of $\hat{\pi}_2$ is given by:

$$\hat{\pi}_2 = \sum_{j=1}^{L}(G_j\pi_{2j}) = \sum_{j=1}^{L} G_j\left[\frac{m_{1j}}{m_j}\hat{\pi}_{saj} + \frac{m_j - m_{1j}}{m_j}\hat{\pi}_{sbj}\right].$$

Using (5.11) and (5.13) in above equation will prove the first part of the theorem.

The variance of $\hat{\pi}_{2j}$ for the j^{th} stratum is given by:

$$V(\hat{\pi}_{2j}) = \frac{\pi_{sj}(1-\pi_{sj})}{m_j} + \frac{4\alpha_{2j}^2\lambda_j(1-\pi_{sj})(1-\alpha_{1j}) + \alpha_{1j}(1-\lambda_j)(1-\alpha_{2j}^2)}{4m_j\alpha_{1j}\alpha_{2j}^2}, \tag{5.16}$$

where $m_j = m_{1j} + m_{2j}$ and $\lambda_j = \frac{m_{1j}}{m_j}$.
Thus the variance of estimator $\hat{\pi}_2$ is given by:

$$V(\hat{\pi}_2) = \sum_{j=1}^{L} G_j^2 V(\hat{\pi}_{2j}).$$

Using 5.16 in above equation will prove the theorem. □

5.4. Theorem. *The optimal allocation of total sample size m to strata sizes $m_1, m_2, ..., m_L$ subject to $m = \sum_{j=1}^{L} m_j$ is given by:*

$$\frac{m_j}{m} = \frac{G_j\left[\pi_{sj}(1-\pi_{sj}) + \frac{4\alpha_{2j}^2\lambda_j(1-\pi_{sj})(1-\alpha_{1j})+\alpha_{1j}(1-\lambda_j)(1-\alpha_{2j}^2)}{4\alpha_{1j}\alpha_{2j}^2}\right]^{\frac{1}{2}}}{\sum_{j=1}^{L} G_j\left[\pi_{sj}(1-\pi_{sj}) + \frac{4\alpha_{2j}^2\lambda_j(1-\pi_{sj})(1-\alpha_{1j})+\alpha_{1j}(1-\lambda_j)(1-\alpha_{2j}^2)}{4\alpha_{1j}\alpha_{2j}^2}\right]^{\frac{1}{2}}},$$

where $\alpha_{1j} = T_j + (1-T_j)P_{1j}$, $\alpha_{2j} = P_{2j} + (1-P_{2j})P_j$, $m_j = m_{1j} + m_{2j}$ and $\lambda_j = \frac{m_{1j}}{m_j}$.
Thus the minimum variance for $\hat{\pi}_2$ is given by:

$$V_{min}(\hat{\pi}_2) = \frac{1}{m}\left[\sum_{j=1}^{L} G_j\left\{\pi_{sj}(1-\pi_{sj}) + \frac{4\alpha_{2j}^2\lambda_j(1-\pi_{sj})(1-\alpha_{1j})+\alpha_{1j}(1-\lambda_j)(1-\alpha_{2j}^2)}{4\alpha_{1j}\alpha_{2j}^2}\right\}^{\frac{1}{2}}\right]^2.$$

6. Discussion

The main idea of this paper is to propose such mixed randomized response models which yield efficient results and secure the privacy of respondent while asking question about sensitive issue. As argument developed in Section 1, Kim and Warde's [13] model is better than Moors [16] and Mangat et al. [15]. Also, Singh and Tarray [20] have proved that their model is efficient than Kim and Warde's [13] model. In Section 4, of this study we have proved that the proposed models are unconditionally efficient than Singh and Tarray's (2013) model for completely and less than completely truthful reporting case. Thus it also follows that proposed models are also more efficient than Moors [16], Mangat et al. [15] and Kim and Warde's [13] models. Further more for less than completely truthful reporting case proposed model 2 is always more efficient than proposed model 1. Moreover, it is interesting to note that the efficiency conditions of $\hat{\pi}_1$ relative to $\hat{\pi}_t$ turns out to be same irrespective of completely truthful or less than completely truthful reporting.

References

[1] Abdelfatah, S., and Mazloum, R. *An improved stratified randomized response model using two decks of cards*, Model Assisted Statistics and Applications, **10** (4), 309-320, 2015.

[2] Al-Menayes, J. *Measuring the Spiritual and Behavioral Dimensions of Religiosity in a Muslim Sample*, Asian Social Science, **12** (2), 158, 2016.

[3] Blair, Graeme, Kosuke Imai, and Yang-Yang Zhou. *Design and Analysis of the Randomized Response Technique*, Journal of the American Statistical Association, **110** (511), 1304 - 1319, 2015.

[4] Clark, S. J., and Desharnais, R. A. *Honest answers to embarrassing questions: Detecting cheating in the randomized response model*, Psychological Methods, **3** (2), 160, 1998.

[5] Coutts, E., and Jann, B. *Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT)*, Sociological Methods and Research, **40** (1), 169-193, 2011.

[6] Greenberg, B. G., Abul-Ela, A. L. A., Simmons, W. R. and Horvitz, D. G. *The unrelated question randomized response model: Theoretical framework*, Journal of the American Statistical Association, **64** (326), 520-539, 1969.

[7] Holbrook, A. L., and Krosnick, J. A. *Measuring Voter Turnout By Using The Randomized Response Technique Evidence Calling Into Question The Method's Validity* Public Opinion Quarterly, **74** (2), 328-343, 2010.

[8] Hong, K., Yum, J. and Lee, H. *A stratified randomized response technique* Korean Journal of Applied Statistics, **7**, 141-147, 1994.

[9] Jin, Z., Zhu, Z., Yu, B. and Gao, Ge. *A Monte Carlo Simulation Investigating the Validity and Reliability of Two Stage Cluster Sampling Survey with Sensitive Topics* Computer Modelling and New Technologies, **17** (3), 69 - 75, 2013.

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

[11] Kim, J. M. and Elam, M. E. *A stratified unrelated question randomized response model* Statistical Papers, **48** (2), 215-233, 2007.

[12] Kim, J. M. and Warde, W. D. *A stratified Warner's randomized response model* Journal of Statistical Planning and Inference, **120** (1), 155-165, 2004.

[13] Kim, J. M. and Warde, W. D. *A mixed randomized response model* Journal of Statistical Planning and Inference, **133** (1), 211-221, 2005.

[14] Locander, W., Sudman, S., and Bradburn, N. *An investigation of interview method, threat and response distortion* Journal of the American Statistical Association, **71** (354), 269-275, 1976.

[15] Mangat, N. S., Singh, R., and Singh, S. *Violation of respondent's privacy in moors' model-its rectification through a random group strategy* Communications in Statistics-Theory and Methods, **26** (3), 743-754, 1997.

[16] Moors, J. A. *Optimization of the unrelated question randomized response model*, Journal of the American Statistical Association, **66** (335), 627-629, 1971.

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

[18] Phillips, D. L. *Knowledge from what?*, Chicago: Rand-McNally, 1971.

[19] Singh, S., Singh, R., and Mangat, N. S. *Some alternative strategies to Moors' model in randomized response sampling*, Journal of Statistical Planning and inference, **83** (1), 243-255, 2000.

[20] Singh, H. P. and Tarray, T. A. *An alternative to Kim and Warde's mixed randomized response model*, SORT-Statistics and Operations Research Transactions, **1** (2), 189-210, 2013.

[21] Singh, H. P. and Tarray, T. A. *An alternative to stratified Kim and Warde's randomized response model using optimal (Neyman) allocation*, Model Assisted Statistics and Applications, **9** (1), 37-62, 2014.

[22] Son, C. K., and Kim, J. M. *Bayes linear estimator for two-stage and stratified randomized response models*, Model Assisted Statistics and Applications, **10** (4), 321-333, 2015.

[23] Soudarssanane, M., Naik, B., Sahai, A., and Bazroy, J. *Randomised Response Technique-An Innovative Method To Measure Culturally Sensitive Variables: Results From A Pilot Study*, Indian Journal of Community Medicine, **28** (3), 138, 2003.

[24] Stem, D. E., and Bozman, C. S. *Respondant Anxiety Reduction with Randomized Response Technique*, Advanced in consumer Research, **15**, 595-599, 1988.

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

[26] Usman, A., and Oshungade, I. O. *A Two-way Randomized Response Technique in Stratification for Tracking HIV Seroprevalence*, Mathematical Theory and Modeling, **2** (7), 86-97, 2012.

[27] Warner, S. L. *Randomized response: A Survey Technique for Eliminating Evasive Answer Bias*, Journal of the American Statistical Association, **60** (309), 63-69, 1965.