Sankhyā : The Indian Journal of Statistics
2015, Volume 77-B, Part 1, pp. 91-107

An Improvement Over Kim and Elam Stratified Unrelated Question Randomized Response Model Using Neyman Allocation

Housila P. Singh and Tanveer A. Tarray
Vikram University, Ujjain, India

Abstract

The present study considers the use of stratified random sampling with Neyman allocation to Mangat (Jour. Ind. Soc. Agril. Statist. **44**, 82–87, 1992) unrelated question randomized response strategy for completely truthful reporting. It has been shown that, for the prior information given, our new model is more efficient in terms of variance (in the case of completely truthful reporting) than Kim and Elam's (Statist. Papers **48**, 215–233, 2007) model. Numerical illustrations and graphs are also given in support of the present study.

AMS (2000) *subject classification.* Primary 62D05; Secondary 94A20.
Keywords and phrases. Randomized response technique, stratified random sampling, dichotomous population, estimation of proportion, mean square error, optimal allocation.

1 Introduction

Sample surveys on human populations have established the fact that refusal to respond and intentional giving of incorrect answers are two main sources of non - sampling error. The bias produced by these two sources of error is sometimes large enough to make the sample estimate seriously misleading. This problem becomes more serious when respondents are questioned about sensitive matters, especially those on which truthful answers may place them in an unfavorable light. Warner (1965) introducing an ingenious interviewing procedure known as randomized response (RR) technique that requests information to the questions randomized on a probability basis rather than from a direct reply to the given question. Feeling that the confidence of the respondent provided by RR technique might be further enhanced if one of the two questions is referred to non - stigmatized attribute. Horvitz et al. (1967) developed an unrelated question RR model. While developing theory for this model, Greenberg et al. (1969) dealt with both the situations when , the proportion innocuous character (say) Y in population is known and when it is unknown. Some modifications in the randomized response (RR) model has been suggested by Chaudhuri and Mukerjee (1988, 2011), Mangat et al. (1992), Mangat and Singh (1990), Mangat (1994), Grewal et al. (2005–2006), Singh and Mathur (2004) and Singh and Tarray (2012, 2013a, 2013b, 2013c, 2013d). Hong et al. (1994) envisaged a stratified RR technique under the proportional sampling assumption. Under Hong et al. (1994) proportional sampling assumption, it may be easy to derive the variance of the proposed estimator. However, it may come at a high cost in terms of time, effort and money. For example, obtaining a fixed number of samples from a rural country in India through a proportional sampling method may be very difficult compared to the researcher's time, effort and money. To overcome this problem, Kim and Warde (2004) and Kim and Elam (2005, 2007) suggested stratified RR techniques using an optimal allocation which are more efficient than a stratified RR technique using a proportional allocation. The extension of the randomized response technique to stratified random sampling may be useful if the investigator is interested in estimating the proportion of HIV/AIDS positively affected persons at different levels such as by rural areas or urban areas, age group or income group, for instance, see Kim and Elam (2005, p. 216). A primary focus of this paper is the implementation of unrelated Stratified RR technique using Mangat (1992) unrelated question RR Strategy. In Section 2 we present our suggested model in the case where the proportion of respondents with the non sensitive trait in a stratum is known and unknown. In Section 2.1 we demonstrate the findings of four empirical studies, in the case of completely truthful reporting. Table 1 demonstrates that, for the given prior information, the proposed model is more efficient in terms of variance than Kim and Elam (2007) stratified unrelated question RR model. In Section 2.2 we present the less than completely truthful reporting counterparts to Sections 2 and 2.1. The empirical studies in Section 3 show that, for the prior information given, the proposed model is more efficient in terms of mean square error than Kim and Elam (2007) model. In Section 4 we offer some conclusion remarks.

2 Suggested Model: For Completely Truthful Reporting

2.1. The Proportion When the Non-Sensitive Trait π_{yi} is Known. In the suggested model, the population is divided into strata and a sample is drawn by simple random sampling with replacement (SRSWR) from each stratum. To get the full benefit from stratification, we suppose that the number of units in each stratum is known. The randomization model requires two randomization devices R_{1i} and R_{2i}. The randomization device R_{2i} is same as used by Greenberg et al. (1969) model. In the first stage of the survey interview, an individual respondent in the sample from stratum i is instructed to use the randomization device R_{1i} which consists of a sensitive question (S) cards with probability T_i and a 'Go to the random device R_{2i} in the second stage' direction card with probability $(1-T_i)$. The respondents in the second stage of stratum i are instructed to use the randomization device R_{2i} which consists of a sensitive question (S) card with probability P_i and a non - sensitive question (Y) card with probability $(1-P_i)$. The respondents selects randomly one of these statements unobserved by the interviewer and reports 'Yes' if he / she possesses statement and 'No' otherwise. Let n_i denote the number of units in the sample from stratum i and n denote the total number of units in all strata so that $\sum_{i=1}^{k} n_i = n$. Under the assumption that these 'Yes' and 'No' reports are made truthfully and P_i and T_i are set by the researcher, the probability X_i of a 'Yes' answer in stratum i for this procedure is:

$$X_i = \pi_{Si}T_i + (1-T_i)[\pi_{Si}P_i + (1-P_i)\pi_{yi}] \quad \text{for} \quad i = 1, 2, \cdots, k \tag{2.1}$$

where π_{Si} is the proportion of people with sensitive traits in i and π_{yi} is the proportion of people with the non-sensitive traits in i.
Under the condition that π_{yi} is known, the unbiased estimator $\hat{\pi}_{Si}$ of π_{Si} is:

$$\hat{\pi}_{Si} = \frac{\hat{X}_i - (1-T_i)(1-P_i)\pi_{yi}}{T_i + P_i(1-T_i)} \quad \text{for} \quad i = 1, 2, \cdots, k \tag{2.2}$$

where $\hat{X}_i$ is the proportion of 'Yes' answer in the sample from stratum for i. Since each $\hat{X}_i$ is a binomial distribution B(n_i, $\hat{X}_i$), the variance of the estimator $\hat{\pi}_{Si}$ is

$$V(\hat{\pi}_{Si}|\pi_{yi}) = \frac{X_i(1-X_i)}{n_i\{T_i + P_i(1-T_i)\}^2} \tag{2.3}$$

Since the selections in different strata are made independently, the estimators for individual strata can be added together to obtain an estimator for the entire population. Thus the unbiased estimator of π_S is

$$\hat{\pi}_S = \sum_{i=1}^{k} w_i\hat{\pi}_{Si} = \sum_{i=1}^{k} w_i \frac{\hat{X}_i - (1-T_i)(1-P_i)\pi_{yi}}{T_i + P_i(1-T_i)} \tag{2.4}$$

The variance of the unbiased estimator $\hat{\pi}_S$ given π_{yi} is:

$$V(\hat{\pi}_S|\pi_{yi}) = \sum_{i=1}^{k} w_i^2 \frac{X_i(1-X_i)}{n_i\{T_i + P_i(1-T_i)\}^2} \tag{2.5}$$

Information on π_{Si} and π_{yi} are usually unavailable. But if prior information on π_{Si} and π_{yi} are available from the past experience then it helps to derive the following optimal allocation formula.

THEOREM 1. *The Neyman allocation* n *to* n_1, n_2, ..., n_{k-1} *and* n_k *to derive the minimum variance of* $\hat{\pi}_S$ *subject* $n = \sum_{i=1}^{k} n_i$ *is approximately given by*

$$\frac{n_i}{n} = \frac{\dfrac{w_i\sqrt{X_i(1-X_i)}}{\{T_i + P_i(1-T_i)\}}}{\sum_{i=1}^{k} \dfrac{w_i\sqrt{X_i(1-X_i)}}{\{T_i + P_i(1-T_i)\}}} \tag{2.6}$$

PROOF. Follows, for example, from section 5.5 of Cochran (1977). Putting Eq. 2.6 in Eq. 2.5 we get the minimum variance of the estimator $\hat{\pi}_S$ given π_{yi} is given by:

$$V(\hat{\pi}_S|\pi_{yi}) = \frac{1}{n}\left[\sum_{i=1}^{k} \frac{w_i\sqrt{X_i(1-X_i)}}{T_i + P_i(1-T_i)}\right]^2 \tag{2.7}$$

The unbiased estimator of the minimum variance of the estimator $\hat{\pi}_S$ given by π_{yi} is obtained upon replacing X_i by $\hat{X}_i$ and $n_i$by $(n_i - 1)$ in Eq. 2.5

REMARK 2.1. If we put $T_i = 0$, the proposed model reduced to the Kim and Elam (2007) model.

2.2. The proportion when the non-sensitive trait π_{yi} *is unknown.* In practice, π_{yi} is rarely known and may be thorny to obtain. In the suggested model the population is partitioned into strata, and two independent non-overlapping simple random samples are drawn from each stratum. To obtain the full benefit from stratification, we assume that the number of units in each stratum is known. In this procedure, two sets of the randomization devices [such as $\{(R_{i1}, R_{i2})$ & $(R^*_{i1}, R^*_{i2})\}$] in each stratum need to be employed (as stated in the case of known $\hat{\pi}_{yi}$). The first set is employed for respondents in the first sample, and the second set is used for respondents in the second sampled. In the first sample at the first stage of the survey interview, an individual respondent of the first sample from stratum i is instructed to use the randomization device R_{i1}which consists of a sensitive question (S) card with probability T_{i1} and a "Go to the random device R_{i2} in the second stage" direction card with probability $(1 - T_{i1})$. The respondents in the second stage of stratum i are instructed to use the randomization device R_{2i} which consists of a sensitive question (S) card with probability P_{i1} and a non sensitive question (Y) card with probability $(1 - P_{i1})$. In the second sample at the first stage of the survey interview, an individual this sample from stratum i is instructed to use R^*_{i1} which consist of a sensitive question (S) card with probability T_{i2} and "Go to the randomization device in the second stage" direction card with probability $(1-T_{i2})$. The respondents in the second stage of stratum i are instructed to use R^*_{i1} the randomization device R^*_{i2} which consists of a sensitive question (S) card with probability P_{i2} and a non sensitive question (Y) card with probability $(1-P_{i2})$. The respondents selects randomly one of these statements unobserved by the interviewer and reports 'Yes' if he / she possesses statement and 'No' otherwise. So a respondent in different strata will perform different set of randomization devices, each having different pre-assigned probabilities. Suppose n_{i1} is the number of units in the first sample from stratum i, n_{i2} is the number of units in the second sample from stratum i, and n_i is the total number of units in two samples from each stratum. So $n = \sum_{i=1}^{k} n_i$ is the total number of units in the samples from every strata. Under the supposition that these 'Yes' and 'No' reports are made truthfully, the probability of a 'Yes' answer in stratum i for our proposed procedure is:

$$X_{i1} = \pi_{Si}T_{i1} + (1-T_{i1})[\pi_{Si}P_{i1} + (1-P_{i1})\pi_{yi}] \tag{2.8}$$
$$X_{i2} = \pi_{Si}T_{i2} + (1-T_{i2})[\pi_{Si}P_{i2} + (1-P_{i2})\pi_{yi}], \quad \text{for} \quad i = 1, 2, \ldots, k \tag{2.9}$$

where X_{i1} and X_{i2} are the proportions of 'Yes' answers in the first and second samples, respectively, from stratum i. Solving (2.8) and (2.9) for π_{Si}, we get

$$\pi_{Si} = \frac{\{(1-P_{i2})(1-T_{i2})X_{i1} - (1-T_{i1})(1-P_{i1})X_{i2}\}}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})}, P_{i1} \neq P_{i2}, T_{i1} \neq T_{i2} \tag{2.10}$$

Suppose the observed proportion of 'Yes' answers reported in the first and second samples be $\hat{X}_{i1} = n'_{i1}/n_{i1}$ and $\hat{X}_{i2} = n'_{i2}/n_{i2}$ respectively, from stratum i, where n'_{i1} and n'_{i2} are numbers of 'Yes' answers in the two corresponding samples from stratum i. Then, the sample estimate, $\hat{\pi}^*_{Si}$ is obtained by replacing (X_{i1}, X_{i2}) by $(\hat{X}_{i1}, \hat{X}_{i2})$ in Eq. 2.10 and it follows that

$$\hat{\pi}^*_{Si} = \frac{\hat{X}_{i1}(1-P_{i2})(1-T_{i2}) - \hat{X}_{i2}(1-P_{i1})(1-T_{i1})}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})} \tag{2.11}$$

The observed proportions, $\hat{X}_{i1}$ and $\hat{X}_{i2}$ are binomially distributed with parameters $(n_{i1,X_{i1}})$ and $(n_{i2,X_{i2}})$ respectively. It therefore follows that the expression in Eq. 2.11 is unbiased and its variance is given by:

$$\begin{aligned} V(\hat{\pi}^*_{Si}) &= \left(\frac{1}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})} \right)^2 \Big\{ (1-P_{i2})^2(1-T_{i2})^2 V(\hat{X}_{i1}) \\ &\qquad + (1-P_{i1})^2(1-T_{i1})^2 V(\hat{X}_{i2}) \Big\} \\ &= \left(\frac{1}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})} \right)^2 \left\{ \frac{X_{i1}(1-X_{i1})}{n_{i1}T^{*2}_{i1}}(1-P_{i2})^2(1-T_{i2})^2 \right. \\ &\qquad \left. + \frac{X_{i2}(1-X_{i2})}{n_{i2}T^{*2}_{i2}}(1-P_{i1})^2(1-T_{i1})^2 \right\} \end{aligned} \tag{2.12}$$

By using Cauchy-Schwarz inequality; we can derive the following:

$$\begin{aligned} &= \left(\frac{1}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})} \right)^2 \left\{ \frac{X_{i1}(1-X_{i1})}{n_{i1}T^{*2}_{i1}}(1-P_{i2})^2(1-T_{i2})^2 \right. \\ &\qquad \left. + \frac{X_{i2}(1-X_{i2})}{n_{i2}T^{*2}_{i2}}(1-P_{i1})^2(1-T_{i1})^2 \right\} (n_{i1}+n_{i2}) \\ &\geq \left(\frac{(1-P_{i2})(1-T_{i2})}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})} \frac{\sqrt{X_{i1}(1-X_{i1})}}{T^*_{i1}} + \frac{(1-P_{i1})(1-T_{i1})}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})} \frac{\sqrt{X_{i2}(1-X_{i2})}}{T^*_{i2}} \right)^2 \end{aligned}$$

By using the inequality, we can derive the minimum variance of the estimator $\hat{\pi}^*_{Si}$ as follows

$$\begin{aligned} V(\hat{\pi}^*_{Si}) &= \frac{1}{n_i(P_{i1}-P_{i2})^2(T_{i1}-T_{i2})^2} \left\{ (1-P_{i2})(1-T_{i2}) \frac{\sqrt{X_{i1}(1-X_{il})}}{T^*_{i1}} \right. \\ &\qquad \left. + (1-P_{i1})(1-T_{i1}) \frac{\sqrt{X_{i2}(1-X_{i2})}}{T^*_{i2}} \right\} \end{aligned} \tag{2.13}$$

where

$$T^*_{i1} = T_{i1} + P_{i1}(1-T_{i1}) \ and \ T^*_{i2} = T_{i2} + P_{i2}(1-T_{i2})$$

Since the selection in different strata are made independently, the estimators for individual strata can be added together to obtain an estimator for the whole population. Thus the estimator for π_S is given by

$$\hat{\pi}_S^* = \sum_{i=1}^{k} w_i \hat{\pi}_{Si}^* = \sum_{i=1}^{k} \frac{w_i}{(P_{i1} - P_{i2})(T_{i1} - T_{i2})} \times \left[\hat{X}_{i1}(1 - P_{i2})(1 - T_{i2}) - \hat{X}_{i2}(1 - P_{i1})(1 - T_{i1}) \right] \tag{2.14}$$

THEOREM 1. *The proposed estimator $\hat{\pi}_S^*$ is an unbiased estimator for the population proportion π_S.*

PROOF. This follows from taking the expected value of Eq. 2.14.

THEOREM 2. *The variance of the estimator $\hat{\pi}_S^*$ is:*

$$V(\hat{\pi}_S^*) = \frac{w_i^2}{n_i (P_{i1} - P_{i2})^2 (T_{i1} - T_{i2})^2} \times \left\{ (1 - P_{i2})(1 - T_{i2}) \frac{\sqrt{X_{i1}(1 - X_{il})}}{T_{i1}^*} + (1 - P_{i1})(1 - T_{i1}) \frac{\sqrt{X_{i2}(1 - X_{i2})}}{T_{i2}^*} \right\} \tag{2.15}$$

Information on π_{Si} and π_{yi} are generally not available. But if prior information on π_{Si} and π_{yi} are available from the past experience then it assists to obtain the following Neyman allocation formula.

THEOREM 3. *The Neyman allocation n to n_1, n_2, $\cdots$, n_{k-1} and n_k to derive the minimum variance of $\hat{\pi}_S^*$ subject $n = \sum_{i=1}^{k} n_i$ is approximately given by*

$$\frac{n_i}{n} = \frac{\dfrac{w_i \left\{ (1 - P_{i2})(1 - T_{i2}) \frac{\sqrt{X_{i1}(1 - X_{i1})}}{T_{i1}^*} + (1 - P_{i1})(1 - T_{i1}) \frac{\sqrt{X_{i2}(1 - X_{i2})}}{T_{i2}^*} \right\}}{(P_{i1} - P_{i2})(T_{i1} - T_{i2})}}{\sum_{i=1}^{k} \dfrac{w_i \left\{ (1-P_{i2})(1-T_{i2}) \frac{\sqrt{X_{i1}(1 - X_{i1})}}{T_{i1}^*} + (1-P_{i1})(1-T_{i1}) \frac{\sqrt{X_{i2}(1 - X_{i2})}}{T_{i2}^*} \right\}}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})}} \tag{2.16}$$

Proof. Follows, for example, from section 5.5 of Cochran 1977 . The minimum variance of $\hat{\pi}_S^*$ is given by

$$V(\hat{\pi}_S^*) = \frac{1}{n}\left[\sum_{i=1}^{k} \frac{w_i(1-P_{i2})(1-T_{i2})\frac{\sqrt{X_{i1}(1-X_{i1})}}{T_{i1}^*} + (1-P_{i1})(1-T_{i1})\frac{\sqrt{X_{i2}(1-X_{i2})}}{T_{i2}^*}}{(P_{i1}-P_{i2})(T_{i1}-T_{i2})}\right]^2 \quad (2.17)$$

The unbiased minimum variance estimator of the estimator $\hat{\pi}_S$ is obtained upon replacing (X_{i1}, X_{i2}) by $(\hat{X}_{i1}, \hat{X}_{i2})$ and n_i by $(n_1 - 1)$ in Eq. 2.15.

3 Efficiency Comparison in the Case of Completely Truthful Reporting

To have tangible idea about the performance of the proposed model relative to Kim and Elam (2007) model, we have computed the relative efficiency of the proposed estimators $\hat{\pi}_S$ and $\hat{\pi}_S^*$ with respect to Kim and Elam (2007) estimators $\hat{\pi}_{ke}$ and $\hat{\pi}_{ke}^*$ respectively in both the cases when π_{yi} is known and unknown.

Case 1. *When is π_{yi} known*
The percent relative efficiency of the proposed estimator $\hat{\pi}_S$ with respect to Kim and Elam (2007) estimator $\hat{\pi}_{ke}$ is given by

$$PRE(\hat{\pi}_S, \hat{\pi}_{ke}) = \frac{V(\hat{\pi}_{ke})}{V(\hat{\pi}_S)} \times 100 \quad (3.1)$$

The value of $PRE(\hat{\pi}_S, \hat{\pi}_{ke})$ have been computed for $n = 1000$, $k = 2$, $\pi_{yi} = \pi_{y1} = \pi_{y2}$, $P_1 = P_{11} = P_{21}$, $P_2 = P_{12} = P_{22}$, $P_1 + P_2 = 1$, $P_1 \neq P_2$, $T_1 \neq T_2$ and findings are shown in Table 1, *where $V(\hat{\pi}_{ke})$ and $V(\hat{\pi}_S)$ are respectively given by* Eq. 2.7 *and* Kim and Elam *(2007, equation (3.7), p.223).*

Case 2. *When π_{yi} is unknown*
The percent relative efficiency of the proposed estimator $\hat{\pi}_S^*$ with respect to Kim and Elam (2007) *estimator $\hat{\pi}_{ke}^*$ is given by*

$$PRE(\hat{\pi}_S^*, \hat{\pi}_{ke}^*) = \frac{V(\hat{\pi}_{ke}^*)}{V(\hat{\pi}_S^*)} \times 100 \quad (3.2)$$

We have computed the values of $PRE(\hat{\pi}_S^*, \hat{\pi}_{ke}^*)$ for $n = 1000$, $k = 2$, $\pi_{yi} = \pi_{y_1} = \pi_{y_2}$, $P_1 = P_{11} = P_{21}$, $P_2 = P_{12} = P_{22}$, $P_1 + P_2 = 1$, $P_1 \neq P_2$, $T_1 \neq T_2$ and results are depicted in Table 2, where $V(\hat{\pi}_{ke}^*)$ and $V(\hat{\pi}_S^*)$ are respectively given by (2.2) and Kim and Elam (2007, equation (3.14), p.226).

Table 1: The percent relative efficiency of $\hat{\pi}_{ke}$ with respect to $\hat{\pi}_S$ when π_{yi} known, $n = 1000$, $P_2 = P_{12} = P_{22}$ and $P_1 = P_{11} = P_{21}$.

π_{S1}	π_{S2}	w_1	w_2	π_y	T_1	T_2	$P_1 = 0.6$	$P_1 = 0.7$	$P_1 = 0.8$	$P_1 = 0.9$
							$P2 = 0.4$	$P2 = 0.3$	$P2 = 0.2$	$P2 = 0.1$
0.48	0.53	0.7	0.3	0.95	0.495	0.95	176.84	178.77	219.98	455.74
0.48	0.53	0.7	0.3	0.93	0.495	0.93	179.49	182.96	231.78	517.59
0.48	0.53	0.7	0.3	0.91	0.495	0.91	182	186.94	242.95	575.88
0.48	0.53	0.6	0.4	0.95	0.5	0.95	187.07	208.5	290.67	684.63
0.48	0.53	0.6	0.4	0.93	0.5	0.93	190.62	215.44	310.56	784.65
0.48	0.53	0.6	0.4	0.91	0.5	0.91	193.99	221.99	329.27	878.09
0.48	0.53	0.4	0.6	0.95	0.51	0.95	209.68	266.53	412.1	1030.02
0.48	0.53	0.4	0.6	0.93	0.51	0.93	215.15	278.44	444.72	1183.21
0.48	0.53	0.4	0.6	0.91	0.51	0.91	220.31	289.6	475.08	1324.51
0.48	0.53	0.3	0.7	0.95	0.515	0.95	217.58	284.66	446.18	1117.6
0.48	0.53	0.3	0.7	0.93	0.515	0.93	223.69	298.01	482.11	1283.45
0.48	0.53	0.3	0.7	0.91	0.515	0.91	229.44	310.5	515.45	1435.93
0.58	0.63	0.7	0.3	0.95	0.595	0.95	185.59	185.84	227.84	479.89
0.58	0.63	0.7	0.3	0.93	0.595	0.93	189.54	191.58	242.71	554.51
0.58	0.63	0.7	0.3	0.91	0.595	0.91	193.3	197.03	256.82	625.04
0.58	0.63	0.6	0.4	0.95	0.6	0.95	198.14	220.93	311.85	765.82
0.58	0.63	0.6	0.4	0.93	0.6	0.93	203.41	230.53	338.12	895.35
0.58	0.63	0.6	0.4	0.91	0.6	0.91	208.41	239.62	362.89	1016.82
0.58	0.63	0.4	0.6	0.95	0.61	0.95	227.48	296.16	476.7	1278.33
0.58	0.63	0.4	0.6	0.93	0.61	0.93	235.69	313.54	523.79	1499.83
0.58	0.63	0.4	0.6	0.91	0.61	0.91	243.44	329.85	567.68	1704.69
0.58	0.63	0.3	0.7	0.95	0.615	0.95	238.28	321.7	528.42	1427.25
0.58	0.63	0.3	0.7	0.93	0.615	0.93	247.53	341.58	581.63	1673.94

Table 1: (continued)

π_{S1}	π_{S2}	w_1	w_2	π_y	T_1	T_2	$P_1 = 0.6$ $P2 = 0.4$	$P_1 = 0.7$ $P2 = 0.3$	$P_1 = 0.8$ $P2 = 0.2$	$P_1 = 0.9$ $P2 = 0.1$
0.58	0.63	0.3	0.7	0.91	0.615	0.91	256.25	360.17	631.05	1901.19
0.68	0.73	0.7	0.3	0.95	0.695	0.95	194.56	193.34	237.75	517.77
0.68	0.73	0.7	0.3	0.93	0.695	0.93	200.71	201.59	257.45	611.81
0.68	0.73	0.7	0.3	0.91	0.695	0.91	206.59	209.47	276.17	700.87
0.68	0.73	0.6	0.4	0.95	0.7	0.95	209.65	234.2	336.42	869.96
0.68	0.73	0.6	0.4	0.93	0.7	0.93	217.78	248.08	372.44	1042.22
0.68	0.73	0.6	0.4	0.91	0.7	0.91	225.52	261.24	406.45	1204.14
0.68	0.73	0.4	0.6	0.95	0.71	0.95	246.93	329.59	553.85	1601.58
0.68	0.73	0.4	0.6	0.93	0.71	0.93	259.78	355.92	623.71	1926.8
0.68	0.73	0.4	0.6	0.91	0.71	0.91	271.93	380.64	688.86	2227.92
0.68	0.73	0.3	0.7	0.95	0.715	0.95	261.41	364.66	629.58	1843.01
0.68	0.73	0.3	0.7	0.93	0.715	0.93	276.03	395.34	710.54	2215.93
0.68	0.73	0.3	0.7	0.91	0.715	0.91	289.82	424.05	785.72	2559.51
0.78	0.83	0.7	0.3	0.95	0.795	0.95	207.24	205	256.78	599.33
0.78	0.83	0.7	0.3	0.93	0.795	0.93	217.72	218.15	285.77	730.67
0.78	0.83	0.7	0.3	0.91	0.795	0.91	227.78	230.73	313.37	855.24
0.78	0.83	0.6	0.4	0.95	0.8	0.95	226.18	254.73	378.97	1062.06
0.78	0.83	0.6	0.4	0.93	0.8	0.93	240.01	276.94	433.55	1313.78
0.78	0.83	0.6	0.4	0.91	0.8	0.91	253.23	298.06	485.19	1550.8
0.78	0.83	0.4	0.6	0.95	0.81	0.95	276.36	382.15	681.5	2168.4
0.78	0.83	0.4	0.6	0.93	0.81	0.93	298.8	426.53	796.13	2692.13
0.78	0.83	0.4	0.6	0.91	0.81	0.91	320.03	468.22	903.03	3176.88
0.78	0.83	0.3	0.7	0.95	0.815	0.95	297.29	433.49	799.31	2584.69
0.78	0.83	0.3	0.7	0.93	0.815	0.93	323.19	486.39	936.05	3205.07

Table 1: (continued)

π_{S1}	π_{S2}	w_1	w_2	π_y	T_1	T_2	$P_1 = 0.6$	$P_1 = 0.7$	$P_1 = 0.8$	$P_1 = 0.9$
							$P2 = 0.4$	$P2 = 0.3$	$P2 = 0.2$	$P2 = 0.1$
0.78	0.83	0.3	0.7	0.91	0.815	0.91	347.62	535.86	1062.87	3775.63
0.88	0.93	0.7	0.3	0.95	0.895	0.95	238.08	235.82	313.35	848.36
0.88	0.93	0.7	0.3	0.93	0.895	0.93	260.77	262.66	368.28	1084.4
0.88	0.93	0.7	0.3	0.91	0.895	0.91	282.66	288.47	420.75	1308.66
0.88	0.93	0.6	0.4	0.95	0.9	0.95	268.12	310.45	502.73	1629.6
0.88	0.93	0.6	0.4	0.93	0.9	0.93	298.69	356.86	610.76	2108.4
0.88	0.93	0.6	0.4	0.91	0.9	0.91	328.01	401.14	713.18	2559.73
0.88	0.93	0.4	0.6	0.95	0.91	0.95	361.1	537.59	1069.45	3927.74
0.88	0.93	0.4	0.6	0.93	0.91	0.93	415.16	640.78	1327.78	5078.15
0.88	0.93	0.4	0.6	0.91	0.91	0.91	466.29	737.54	1567.91	6138.17
0.88	0.93	0.3	0.7	0.95	0.915	0.95	406.91	647.46	1338.49	4998.37
0.88	0.93	0.3	0.7	0.93	0.915	0.93	472.07	776.65	1663.74	6440.39
0.88	0.93	0.3	0.7	0.91	0.915	0.91	533.31	896.85	1963.32	7755.27

Table 2: The percent relative efficiency of $\hat{\pi}^*_{ke}$ with respect to with respect to $\hat{\pi}^*_S$ when π_{yi} unknown, $n = 1000$, $P_2 = P_{12} = P_{22}$ and $P_1 = P_{11} = P_{21}$.

π_{S1}	π_{S2}	w_1	w_2	T_1	T_2	π_y	$P_1 = 0.6$	$P_1 = 0.7$	$P_1 = 0.8$	$P_1 = 0.9$
							$P2 = 0.4$	$P2 = 0.3$	$P2 = 0.2$	$P2 = 0.1$
0.48	0.53	0.7	0.3	0.3	0.9	0.95	114.26	139.18	175.49	231.27
0.48	0.53	0.7	0.3	0.3	0.9	0.93	116.41	141.64	178.14	233.6
0.48	0.53	0.7	0.3	0.3	0.9	0.91	118.44	143.95	180.58	235.68
0.48	0.53	0.6	0.4	0.3	0.9	0.95	113.9	138.84	175.23	231.14
0.48	0.53	0.6	0.4	0.3	0.9	0.93	116.07	141.32	177.89	233.48
0.48	0.53	0.6	0.4	0.3	0.9	0.91	118.12	143.66	180.36	235.57
0.48	0.53	0.4	0.6	0.3	0.9	0.95	113.16	138.16	174.7	230.88
0.48	0.53	0.4	0.6	0.3	0.9	0.93	115.38	140.69	177.4	233.24
0.48	0.53	0.4	0.6	0.3	0.9	0.91	117.48	143.06	179.9	235.35
0.48	0.53	0.3	0.7	0.3	0.9	0.95	112.8	137.82	174.43	230.75
0.48	0.53	0.3	0.7	0.3	0.9	0.93	115.04	140.37	177.16	233.1
0.48	0.53	0.3	0.7	0.3	0.9	0.91	117.16	142.76	179.68	235.24
0.58	0.63	0.7	0.3	0.33	0.9	0.95	110.38	137.93	179.48	244.79
0.58	0.63	0.7	0.3	0.33	0.9	0.93	113.16	141.08	182.8	247.6
0.58	0.63	0.7	0.3	0.33	0.9	0.91	115.8	144.04	185.85	250.1
0.58	0.63	0.6	0.4	0.33	0.9	0.95	110.14	137.71	179.32	244.72
0.58	0.63	0.6	0.4	0.33	0.9	0.93	112.96	140.9	182.67	247.55
0.58	0.63	0.6	0.4	0.33	0.9	0.91	115.64	143.89	185.76	250.08
0.58	0.63	0.4	0.6	0.33	0.9	0.95	109.68	137.28	179	244.59
0.58	0.63	0.4	0.6	0.33	0.9	0.93	112.57	140.54	182.41	247.46
0.58	0.63	0.4	0.6	0.33	0.9	0.91	115.31	143.6	185.56	250.02
0.58	0.63	0.3	0.7	0.33	0.9	0.95	109.44	137.06	178.84	244.52
0.58	0.63	0.3	0.7	0.33	0.9	0.93	112.37	140.36	182.29	247.41

Table 2: (continued)

π_{S1}	π_{S2}	w_1	w_2	T_1	T_2	π_y	$P_1 = 0.6$	$P_1 = 0.7$	$P_1 = 0.8$	$P_1 = 0.9$
							$P2 = 0.4$	$P2 = 0.3$	$P2 = 0.2$	$P2 = 0.1$
0.58	0.63	0.3	0.7	0.33	0.9	0.91	115.15	143.46	185.47	249.99
0.68	0.73	0.7	0.3	0.36	0.9	0.95	107.93	137.44	183.02	255.84
0.68	0.73	0.7	0.3	0.36	0.9	0.93	111.71	141.66	187.35	259.37
0.68	0.73	0.7	0.3	0.36	0.9	0.91	115.3	145.62	191.34	262.5
0.68	0.73	0.6	0.4	0.36	0.9	0.95	107.85	137.37	182.99	255.87
0.68	0.73	0.6	0.4	0.36	0.9	0.93	111.69	141.66	187.39	259.43
0.68	0.73	0.6	0.4	0.36	0.9	0.91	115.34	145.68	191.43	262.59
0.68	0.73	0.4	0.6	0.36	0.9	0.95	107.68	137.24	182.94	255.92
0.68	0.73	0.4	0.6	0.36	0.9	0.93	111.65	141.66	187.46	259.55
0.68	0.73	0.4	0.6	0.36	0.9	0.91	115.42	145.81	191.61	262.78
0.68	0.73	0.3	0.7	0.36	0.9	0.95	107.59	137.17	182.91	255.94
0.68	0.73	0.3	0.7	0.36	0.9	0.93	111.63	141.66	187.49	259.62
0.68	0.73	0.3	0.7	0.36	0.9	0.91	115.46	145.87	191.7	262.88
0.78	0.83	0.7	0.3	0.39	0.9	0.95	108.21	139.25	187.67	265.47
0.78	0.83	0.7	0.3	0.39	0.9	0.93	113.82	145.42	193.82	270.21
0.78	0.83	0.7	0.3	0.39	0.9	0.91	119.13	151.17	199.43	274.4
0.78	0.83	0.6	0.4	0.39	0.9	0.95	108.4	139.47	187.91	265.69
0.78	0.83	0.6	0.4	0.39	0.9	0.93	114.15	145.78	194.19	270.5
0.78	0.83	0.6	0.4	0.39	0.9	0.91	119.58	151.67	199.91	274.76
0.78	0.83	0.4	0.6	0.39	0.9	0.95	108.79	139.93	188.42	266.13
0.78	0.83	0.4	0.6	0.39	0.9	0.93	114.82	146.53	194.95	271.11
0.78	0.83	0.4	0.6	0.39	0.9	0.91	120.51	152.68	200.91	275.51
0.78	0.83	0.3	0.7	0.39	0.9	0.95	108.99	140.17	188.68	266.36
0.78	0.83	0.3	0.7	0.39	0.9	0.93	115.16	146.92	195.35	271.42

Table 2: (continued)

π_{S1}	π_{S2}	w_1	w_2	T_1	T_2	π_y	$P_1 = 0.6$	$P_1 = 0.7$	$P_1 = 0.8$	$P_1 = 0.9$
							$P2 = 0.4$	$P2 = 0.3$	$P2 = 0.2$	$P2 = 0.1$
0.78	0.83	0.3	0.7	0.39	0.9	0.91	120.99	153.2	201.42	275.89
0.88	0.93	0.7	0.3	0.42	0.9	0.95	117.99	150.71	200.45	278.56
0.88	0.93	0.7	0.3	0.42	0.9	0.93	128.38	161.8	211	286.13
0.88	0.93	0.7	0.3	0.42	0.9	0.91	137.99	171.95	220.47	292.77
0.88	0.93	0.6	0.4	0.42	0.9	0.95	119.36	152.18	201.86	279.58
0.88	0.93	0.6	0.4	0.42	0.9	0.93	130.22	163.75	212.82	287.41
0.88	0.93	0.6	0.4	0.42	0.9	0.91	140.25	174.3	222.65	294.27
0.88	0.93	0.4	0.6	0.42	0.9	0.95	122.32	155.36	204.9	281.78
0.88	0.93	0.4	0.6	0.42	0.9	0.93	134.2	167.95	216.74	290.16
0.88	0.93	0.4	0.6	0.42	0.9	0.91	145.11	179.38	227.32	297.5
0.88	0.93	0.3	0.7	0.42	0.9	0.95	123.92	157.07	206.54	282.96
0.88	0.93	0.3	0.7	0.42	0.9	0.93	136.34	170.21	218.86	291.64
0.88	0.93	0.3	0.7	0.42	0.9	0.91	147.74	182.11	229.83	299.23

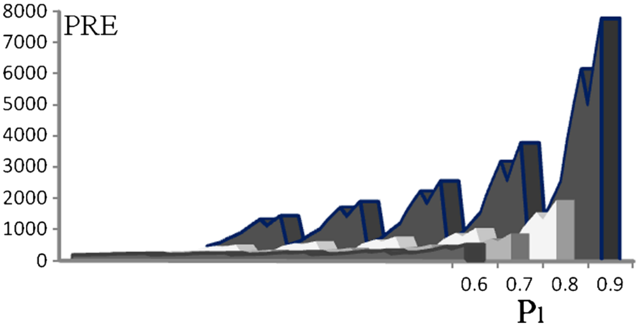

Figure 1: The percent relative efficiency of $\hat{\pi}_{ke}$ with respect to $\hat{\pi}_S$ when π_y is known.

It is observed from Tables 1 and 2, Figures 1 and 2 that:

(i) for fixed values of (π_{S1}, π_{S2}), π_y and (T_1, T_2), the values of $PRE(\hat{\pi}_{S1}, \hat{\pi}_{ke})$ and $PRE(\hat{\pi}^*_S, \hat{\pi}^*_{ke})$ increases as $P_1(P_2)$ increases (decreases).

(ii) for fixed values of (π_{S1}, π_{S2}), (P_1, P_2) and (T_1, T_2) the values of $PRE(\hat{\pi}_S, \hat{\pi}_{ke})$ and $PRE(\hat{\pi}^*_S, \hat{\pi}^*_{ke})$ decreases as π_y decreases.

(iii) for fixed values of π_y, (P_1, P_2) and (T_1, T_2), the values of $PRE(\hat{\pi}_S, \hat{\pi}_{ke})$ increases in a speedy manner as (π_{S1}, π_{S2}) increases while the values of $PRE(\hat{\pi}^*_S, \hat{\pi}^*_{ke})$ decreases.

(iv) for fixed values of (π_{S1}, π_{S2}), (P_1, P_2), T_1 and π_y, the values of $PRE(\hat{\pi}_S, \hat{\pi}_{ke})$ increases as T_2 increases.

(v) for fixed values of (π_{S1}, π_{S2}), π_y, (P_1, P_2) and T_2 the values of $PRE(\hat{\pi}_S, \hat{\pi}_{ke})$ increases as T_1 increases while the values of $PRE(\hat{\pi}^*_S, \hat{\pi}^*_{ke})$ decreases.

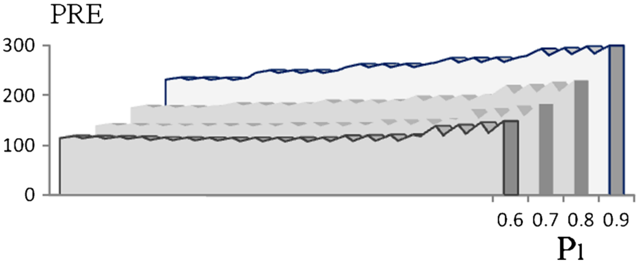

Figure 2: The percent relative efficiency of $\hat{\pi}^*_{ke}$ with respect to $\hat{\pi}^*_S$ when π_y is unknown.

Since all the PRE values in Table 1 are greater than 100, our stratified unrelated question RR model using Neyman allocation is more efficient in terms of variance than Kim and Elam (2007) stratified RR model under the assumptions given and the prior information used.

4 Discussion

This paper addresses the problem of estimating the proportion of the population belonging to a sensitive group using randomized response technique in stratified unrelated question randomized response sampling. A stratified unrelated question randomized response model using Mangat (1992) improved unrelated question RR model for completely truthful reporting has been proposed. It has been shown that for the prior information given, the proposed stratified unrelated question randomized response model using Neyman allocation is more efficient in terms of Variance than Kim and Elam (2007) unrelated question stratified RR model. In addition to the gain in efficiency, the proposed method is more beneficial than the previous method as stratified randomize response method assists to solve the limitations of randomized response that is the loss of the individual characteristics of the respondents. A notable point in this study is that the proposed model is more precious (with considerable gain in efficiency) than the one earlier considered by Kim and Elam (2007).

Acknowledgement. The authors are thankful to the Editor - in- Chief Dr. B.L.S Prakasa Rao and to the learned referees for their valuable suggestions regarding improvement of the paper.

References

CHAUDHURI, A. (2011). *Randomized response and indirect questioning techniques in surveys.* CRC Press, London.

CHAUDHURI, A. and MUKERJEE, R. (1988). *Randomized Response: Theory and Techniques.* Marcel- Dekker, New York.

COCHRAN, W.G. (1977). *Sampling techniques 3rd edition.* John Wiley and Sons, NY.

GREENBERG, B., ABUL- ELA, A., SIMMONS, W.R. and HORVITZ, D.G. (1969). The unreleased question randomized response: Theoretical framework. *Jour. Amer. Statist. Assoc.,* **64**, 529–539.

GREWAL, I.S., BANSAL, M.L. and SIDHU, S.S. (2005–2006). Population mean corresponding to HorvitzThompson estimator for multi-characteristics using randomized response technique. *Model Assist. Stat. Appl.,* **1**, 215–220.

HORVITZ, D.G., SHAH, B.V. and SIMMONS, W.R. (1967) The unrelated question randomized response model. Proceedings of the ASA social statistics section, pp. 62–72.

HONG, K., YUM, J. and LEE, H. (1994). A stratified randomized response technique. *Korean Jour. Appl. Statist.,* **7**, 141–147.

KIM, J.M. and ELAM, M.E. (2005). A two stage stratified Warner randomized response model using optimal allocation. *Metrika,* **61**, 1–7.

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

KIM, J.M. and WARDE, W.D. (2004). A stratified Warner randomized response sampling procedure using unrelated question. *Jour. Statist. Plan. Inference,* **120**, 155–165.

MANGAT, N.S. (1992). Two stage randomized response sampling procedure using unrelated question. *Jour. Ind. Soc. Agril. Statist.,* **44**, 82–87.

MANGAT, N.S., SINGH, R. and SINGH, S. (1992). An improved unrelated question randomized response strategy. *Calcuta Statist. Assoc. Bull.,* **42**, 277–281.

MANGAT, N.S. (1994). An improved randomized response strategy. *Jour. Roy. Statist. Soc., B,* **56**, 93–95.

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

SINGH, H.P. and MATHUR, N. (2004). Unknown repeated trails in the unrelated question randomized response model. *Biometrica Jour.,* **46**, 375–378.

SINGH, H.P. and TARRAY, T.A. (2012). A Stratified Unknown repeated trials in randomized response sampling. *Commun. Korean Statist. Soc.,* **19**, 751–759.

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

SINGH, H.P. and TARRAY, T.A. (2013b). An alternative to stratified Kim and Wardes randomized response model using optimal (Neyman) allocation. *Model Assist. Stat. Appl.,* **9**, 37–62.

SINGH, H.P. and TARRAY, T.A. (2013c). An improved mixed randomized response model. *Model Assist. Stat. Appl.,* **9**, 73–87.

SINGH, H.P. and TARRAY, T.A. (2013d). An alternative to Kim and Wardes mixed randomized response model. *Statist. Oper. Res. Trans.,* **37**, 189–210.

WARNER, S.L. (1965).Randomized response: A survey technique for eliminating evasive answer bias. *Jour. Amer. Statist. Assoc.,* **60**, 63–69.

HOUSILA P. SINGH
AND TANVEER A. TARRAY
SCHOOL OF STUDIES IN STATISTICS
VIKRAM UNIVERSITY
UJJAIN 456010, INDIA
E-mail: tanveerstat@gmail.com

Paper received: 9 December 2013.