Original Article

Sociological Methods & Research
1-17

DOI: 10.1177/0049124114553309
smr.sagepub.com

An Efficient Alternative Mixed Randomized Response Procedure

Housila P. Singh[1] and Tanveer A. Tarray[1]

Abstract

In this article, we have suggested a new modified mixed randomized response (RR) model and studied its properties. It is shown that the proposed mixed RR model is always more efficient than the Kim and Warde's mixed RR model. The proposed mixed RR model has also been extended to stratified sampling. Numerical illustrations and graphical representations are also given in support of this study.

Keywords

randomized response technique, dichotomous population, estimation of proportion, privacy of respondents, innocuous variable, sensitive characteristics

Introduction

In survey of human population, questions requiring personal or controversial assertions often run into trouble in terms of resistance. It is difficult to collect reliable data from interviewees and hard to raise the quality of responses when the survey topic is sensitive. The purpose is to formulate an effective random device so as to induce each respondent to give truthful answers to

[1] School of Studies in Statistics, Vikram University, Ujjain, India

Corresponding Author:
Tanveer A. Tarray, School of Studies in Statistics, Vikram University, Ujjain, Madhya Pradesh 456010, India.
Email: tanveerstat@gmail.com

sensitive questions without exposing his or her true identity to the interviewer. A large number of rectifications on Warner's (1965) pioneering technique have been cited in the literature, for instance, see Fox and Tracy (1986), Chaudhuri and Mukherjee (1987, 1988), Hedayat and Singh (1991), Ryu, Hong, and Lee (1993), Tracy and Mangat (1996), and Singh (2003), for the reviews. Some other developments on randomized response (RR) sampling in recent years include Mahmood, Singh, and Horn (1998), Ryu et al. (2005–2006), Hong (2005–2006), Grewal, Bansal, and Sidhu (2005–2006), Perri (2008), Mahajan, Sharma, and Gupta (2007), Chua and Tsui (2000), Singh, Singh, and Mangat (2000), Chang and Huang (2001), Chang, Wang, and Huang (2004), Huang (2004), Land, Singh, and Sedory (2011), Kim and Warde (2004), Kim and Elam (2005), and so on.

To implement the privacy problem with the Moors (1971) model, Mangat, Singh, and Singh (1997) and Singh et al. (2000) have given several strategies as an alternative to Moors (1971) model, but their models may lose a large portion of data information and require a high cost to obtain confidentiality of the respondents. These drawbacks with the previous alternative models for the Moors model motivated Kim and Warde (2005) to envisage a mixed RR model using simple random sampling with replacement that modifies the privacy problem.

In this article, we have suggested a modification of Kim and Warde's (2005) model to estimate the proportion of a qualitative sensitive variable. It has been demonstrated that the suggested model performs better than the mixed RR model of Kim and Warde (2005). In the fourth section, we introduce the stratified version of the proposed mixed RR model along with its properties. Comparison of the proposed stratified mixed RR model with that of Kim and Warde's (2005) model has been given in the fifth section. The empirical studies performed show that, the proposed mixed RR model is more efficient than Kim and Warde's (2005) model in stratified sampling too.

The Suggested Model

A single sample with size n is selected by simple random sampling with replacement (SRSWR) from the population. Each respondent from the sample is instructed to answer the direct question, "I am a member of the innocuous trait group." If a respondent answers "Yes" to direct question, then he or she is instructed to go to randomization device R_1 consisting of the statements (i) "I am a member of the sensitive trait group" and (ii) "I am a member of the innocuous trait group" with probabilities of selection P_1 and $(1 - P_1)$, respectively. If a respondent answers "No" to the direct question, then the respondent is instructed to use a randomization procedure due to Mangat (1994). In the Mangat's (1994) RR procedure, each respondent is instructed to say "Yes" if he or she is a member of the sensitive trait group. If he or she is not a member of the sensitive trait group, then the respondent is required to use the Warner's (1965) randomization device R_2 consisting of the statements: (i) "I belong to the sensitive trait group" and (b) "I do not belong to the sensitive trait group" with preassigned probabilities P and $(1 - P)$, respectively. Then he or she is to report "Yes" or "No" according to the outcome of the randomization device R_2 and the actual status that he or she has with respect to the sensitive trait group. The survey procedures are performed under the assumption that both the sensitive and the innocuous questions are unrelated and independent in a randomization device R_1. To protect the respondent's privacy, the respondents should not disclose to the interviewer the question they answered from either R_1 or R_2.

Let n be the sample size confronted with a direct question, and n_1 and n_2 $(= n - n_1)$ denote the number of "Yes" and "No" answers from the sample. Since all the respondents using a randomization device R_1 already responded "Yes" from the initial direct innocuous question, the proportion "Y" of getting "Yes" answers from the respondents using randomization device R_1 should be

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

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

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

$$\hat{\pi}_{h1} = \frac{\hat{Y} - (1 - P_1)}{P_1}, \tag{2}$$

whose variance is given by:

$$V(\hat{\pi}_{h1}) = \frac{Y(1 - Y)}{n_1 P_1^2} = \left[\frac{[(1 - \pi_S)\pi_S}{n_1} + \frac{(1 - \pi_S)\ (1 - P_1)]}{n_1 P_1}\right]. \tag{3}$$

The proportion of "Yes" answers from the respondents using Mangat's (1994) randomization procedure is given by:

$$X = \pi_S + (1 - \pi_S)(1 - P), \tag{4}$$

where X is the proportion of "Yes" responses.

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

$$\hat{\pi}_{h2} = \frac{\hat{X} - (1 - P)}{P}. \tag{5}$$

Its variance is given by:

$$\begin{aligned} V(\hat{\pi}_{h2}) &= \frac{X(1-X)}{n_2 P^2} \\ &= \frac{1}{n_2}\left[\pi_S(1-\pi_S) + \frac{(1-P)(1-\pi_S)}{P}\right]. \end{aligned} \tag{6}$$

Now we shall pool the two unbiased estimators using weights, to formulate an estimator for π_S as

$$\begin{aligned} \hat{\pi}_h &= \frac{n_1}{n}\hat{\pi}_{h1} + \frac{n_2}{n}\hat{\pi}_{h2} \\ &= \frac{n_1}{n}\hat{\pi}_{h1} + \frac{(n-n_1)}{n}\hat{\pi}_{h2}, \ for\ 0 < \frac{n_1}{n} < 1. \end{aligned} \tag{7}$$

Since the randomization device R_1 and the used randomization procedure due to Mangat (1994) are independent, therefore,

$$\begin{aligned} V(\hat{\pi}_h) = &\frac{n_1}{n^2}\left[\pi_S(1-\pi_S) + \frac{(1-\pi_S)(1-P_1)}{P_1}\right] \\ &+ \frac{(n-n_1)}{n^2}\left[\pi_S(1-\pi_S) + \frac{(1-\pi_S)(1-P)}{P}\right]. \end{aligned} \tag{8}$$

Horvitz, Shah, and Simmons (1967) presented Simmon's method, when π_1 is known and unknown. Under the situation that the Warner (1965) model and Simmon's method (known π_1) are equally confidential to respondents, Lanke (1976) obtained a unique value of P as

$$P = \frac{1}{2} + \frac{P_1}{2P_1 + 4(1-P_1)\pi_1} \quad \text{for every } P_1 \text{ and every } \pi_1. \tag{9}$$

Since the proposed mixed RR model also use Simmon's method $\pi_1 = 1$, we can apply Lanke's (1976) arguments to the suggested model. Putting $\pi_1 = 1$ in (9), we get

$$P = (2 - P_1)^{-1}. \tag{10}$$

Setting $P = (2 - P_1)^{-1}$ in equation (8), we get the variance of $\hat{\pi}_h$ as

$$V(\hat{\pi}_h) = \frac{1}{n}\left[\pi_S(1-\pi_S) + (1-\pi_S)(1-P_1)\left\{\frac{\lambda}{P_1} + (1-\lambda)\right\}\right], \quad (11)$$

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

Thus, we established the following theorem.

Theorem 2.1: The variance of $\hat{\pi}_h$ is given by:

$$V(\hat{\pi}_h) = \frac{1}{n}\left[\pi_S(1-\pi_S) + (1-\pi_S)(1-P_1)\left\{\frac{\lambda}{P_1} + (1-\lambda)\right\}\right]. \quad (12)$$

Efficiency Comparisons

An efficiency comparison of the proposed mixed RR model, under completely truthful reporting case, has been done with Kim and Warde's (2005) model.

From Kim and Warde's (2005, eq. 2.10:213), the variance of the Kim and Warde (2005) estimator $\hat{\pi}_{kw}$ based on mixed RR model is given by:

$$V(\hat{\pi}_{kw}) = \frac{1}{n}\left[\pi_S(1-\pi_S) + (1-P_1)\left\{\frac{\lambda(1-\pi_S)}{P_1} + \frac{(1-\lambda)}{P_1^2}\right\}\right]. \quad (13)$$

From equations (12) and (13), we have

$$n\left[V(\hat{\pi}_{kw}) - V(\hat{\pi}_h)\right] = (1-\lambda)(1-P_1)\left[\frac{1}{P_1^2} - 1 + \pi_S\right], \quad (14)$$

which is always positive.

Thus, the proposed estimator $\hat{\pi}_h$ is more efficient than the Kim and Warde (2005) estimator $\hat{\pi}_{kw}$.

To have the insight over the performance of the proposed mixed RR model, we have computed the percentage relative efficiency (PRE) of the proposed estimator $\hat{\pi}_h$ with respect to the Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$ using the formula:

$$\begin{aligned} \text{PRE}\ (\hat{\pi}_h, \hat{\pi}_{kw}) &= \frac{V(\hat{\pi}_{kw})}{V(\hat{\pi}_h)} \times 100 \\ &= \frac{\left[\pi_S(1-\pi_S) + (1-P_1)\left\{\frac{\lambda(1-\pi_S)}{P_1} + \frac{(1-\lambda)}{P_1^2}\right\}\right]}{\left[\pi_S(1-\pi_S) + (1-P_1)(1-\pi_S)\left\{\frac{\lambda}{P_1} + (1-\lambda)\right\}\right]} \times 100, \end{aligned} \quad (15)$$

Table 1. Percentage Relative Efficiency of the Proposed Estimator $\hat{\pi}_h$ with Respect to Kim and Warde's (2005) Estimator $\hat{\pi}_{kw}$.

	$n = 1{,}000$		$\lambda=$	P_1						
π_S	n_1	n_2	n_1/n	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	200	800	0.2	663.92	424.24	296.82	220.66	171.46	137.99	114.70
	400	600	0.4	427.87	301.88	229.16	182.53	150.54	127.60	110.90
	600	400	0.6	278.47	215.05	176.54	150.57	131.87	117.84	107.19
	800	200	0.8	175.39	150.23	134.44	123.39	115.12	108.66	103.55
0.2	200	800	0.2	688.40	433.67	300.00	221.08	170.80	137.19	114.37
	400	600	0.4	448.49	311.42	233.33	183.96	150.62	127.23	110.70
	600	400	0.6	291.96	222.01	180.00	152.04	132.24	117.73	107.08
	800	200	0.8	181.77	153.78	136.36	124.31	115.43	108.66	103.51
0.3	200	800	0.2	727.80	452.38	309.52	226.08	173.46	138.65	115.18
	400	600	0.4	477.82	326.53	241.42	188.31	152.92	128.43	111.32
	600	400	0.6	310.32	232.14	185.71	155.22	133.94	118.59	107.51
	800	200	0.8	190.26	158.73	139.28	125.99	116.35	109.12	103.73
0.4	200	800	0.2	787.64	483.05	326.66	236.11	179.41	142.07	116.84
	400	600	0.4	519.64	349.26	254.54	196.07	157.52	131.04	112.57
	600	400	0.6	335.82	246.75	194.44	160.49	137.07	120.37	108.34
	800	200	0.8	201.90	165.69	143.58	128.65	117.94	110.02	104.15

for different cases of $\pi_S,\ n,\ n_1,\ P_1$, and $\lambda = (0.2, 0.4, 0.6, 0.8)$. Findings are shown in Table 1 and diagrammatic representation is given in Figure 1.

It is observed from Table 1 and Figure 1 that the values of $\mathrm{PRE}(\hat{\pi}_h, \hat{\pi}_{kw})$ is more than 100. We can say that the envisaged estimator $\hat{\pi}_h$ is more efficient than the Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$. Figure 1 show results for $\pi_S = (0.1 \text{ and } 0.4)$, $\lambda = (0.2, 0.4, 0.6, 0.8)$, and different values of $P_1,\ n,$ and n_1, which also justify the previous fact.

We note from Table 1 that the values of the $\mathrm{PRE}(\hat{\pi}_h, \hat{\pi}_{kw})$ decrease as the value of P_1 increases. Also the values of $\mathrm{PRE}(\hat{\pi}_h, \hat{\pi}_{kw})$ increase as the value of λ decreases for fixed values of P_1. We further note from the results of Figure 1 that there is large gain in efficiency by using the suggested estimator $\hat{\pi}_h$ over the Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$ when the proportion of stigmatizing attribute is moderately large.

A Mixed RR Model Using Stratification

Stratified random sampling is generally obtained by dividing the population into nonoverlapping groups called strata and selecting a simple random

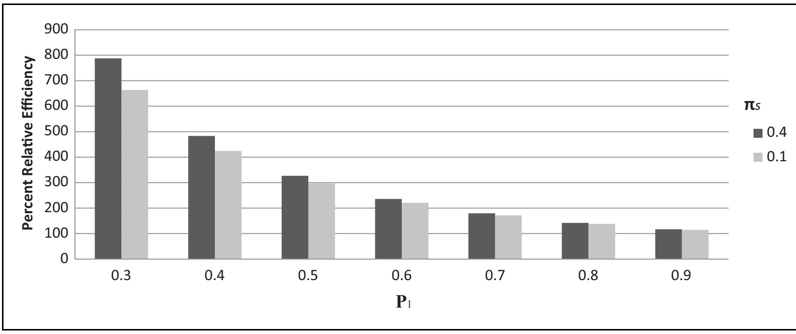

Figure 1. Percentage relative efficiency of the proposed estimator $\hat{\pi}_h$ with respect to Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$.

sample from each stratum. The main advantage of the stratified approach is that the technique overcomes the limitation of the loss of individual characteristics of the respondents. An RR technique using a stratified random sampling gives the group characteristics related to each stratum estimator. Also, stratified sampling protects a researcher from the possibility of obtaining a poor sample. Hong, Yum, and Lee (1994) suggested a stratified RR technique using a proportional allocation. Kim and Warde (2004) presented a stratified RR model based on Warner (1965) model that has an optimal allocation and large gain in precision. Kim and Elam (2005) suggested a two-stage stratified Warner's RR model using optimal allocation. Further Kim and Warde (2005) suggested a mixed stratified RR model.

In the proposed model, the assumptions for a stratified mixed RR model are similar to Kim and Warde (2004) and Kim and Elam (2005) model. We assume that the population is partitioned into "r" nonoverlapping strata, and a sample is selected by simple random sampling with replacement from each stratum. To get the full benefit from stratification, we assumed that the number of units in each stratum is known. In this proposed model, an individual respondent in a sample from each stratum is instructed to answer a direct question "I am a member of the innocuous trait group." Respondents answer the direct question by "Yes" or "No." If a respondent answers "Yes," then he or she is instructed to go to the randomization device R_{k1} consisting of statements: (i) "I am the member of the sensitive trait group" and (ii) "I am a member of the innocuous trait group" with preassigned probabilities Q_k and $(1 - Q_k)$, respectively. If a respondent answers "No," then the respondent is instructed to use a randomization procedure due to Mangat (1994). In the Mangat's (1994) RR procedure, each respondent is instructed to say "Yes" if he or she is a member of the sensitive trait group. If he or she is not a member of the sensitive trait group, then the respondent is required to use the Warner's (1965) randomization device R_{k2} consisting of the statement: (i) "I belong to the sensitive trait group" and (b) "I do not belong to the sensitive trait group" with preassigned probabilities P_k and $(1 - P_k)$, respectively. Then he or she is to report "Yes" or "No" according to the outcome of the randomization device R_{k2} and the actual status that he or she has with respect to the sensitive trait group. The survey procedures are performed under the assumption that both the sensitive and the innocuous questions are unrelated and independent in a randomization device R_{k1}. To protect the respondent's privacy, the respondents should not disclose to the interviewer the question they answered from either R_{k1} or R_{k2}. Suppose we denote m_k as the number of units in the sample from stratum k and n as the total number of units in samples from all strata. Let m_{k1} be the number of people responding "Yes" when respondents in a sample m_k were asked the direct question and m_{k2} be the number of people responding "No" when respondents in a sample m_k were asked the direct question so that $n = \sum_{k=1}^{r} m_k = \sum_{k=1}^{r} (m_{k1} + m_{k2})$. Under the assumption that these "Yes" or "No" reports are made truthfully, and Q_k and P_k ($\neq 0.5$) are set by the researcher, then the proportion of "Yes" answer from the respondents using the randomization device R_{k1} will be

$$Y_k = Q_k \pi_{S_k} + (1 - Q_k)\pi_{1_k} \quad \text{for } k = 1, 2, \ldots, r, \tag{16}$$

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

Since the respondent performing a randomization device R_{k1} respond "Yes" to the direct question of the innocuous trait, if he or she chooses the same innocuous question from R_{k1}, then π_{1_k} is equal to one (i.e., $\pi_{1_k} = 1$). Therefore, equation (16) becomes $Y_k = Q_k \pi_{S_k} + (1 - Q_k)$. The estimator of π_{S_k} is given by:

$$\hat{\pi}_{h1_k} = \frac{\hat{Y}_k - (1 - Q_k)}{Q_k} \quad \text{for } k = 1, 2, \ldots, r, \tag{17}$$

where $\hat{Y}_k$ is the proportion of "Yes" answers in a sample in stratum k and $\hat{\pi}_{h1_k}$ is the proportion of respondents with the sensitive trait in a sample from stratum k. Since each $\hat{Y}_k$ is a binomial distribution B (m_{k1}, Y_k), the estimator $\hat{\pi}_{h1_k}$ is an unbiased estimator for π_{S_k} with the variance,

$$V(\hat{\pi}_{h1_k}) = \frac{Q_k(1-\pi_{S_k})\,[Q_k\pi_{S_k} + (1-Q_k)]}{m_{k1}Q_k^2} = \frac{(1-\pi_{S_k})\,[Q_k\pi_{S_k} + (1-Q_k)]}{m_{k1}Q_k}. \tag{18}$$

The proportion of "Yes" answers from the respondents using Mangat (1994) RR technique will be

$$X_k = \pi_{S_k} + (1-\pi_{S_k})(1-P_k), \tag{19}$$

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

$$\hat{\pi}_{h2_k} = \frac{\hat{X}_k - (1-P_k)}{P_k}, \tag{20}$$

where $\hat{X}_k$ is the proportion of "Yes" responses in a sample from a stratum k and $\hat{\pi}_{h2_k}$ is the proportion of respondents with the sensitive trait in a sample from stratum k. By using $m_k = m_{k1} + m_{k2}$ and $P_k = (2-Q_k)^{-1}$, the variance of $\hat{\pi}_{h2_k}$ is given by:

$$V(\hat{\pi}_{h2_k}) = \frac{1}{(m_k - m_{k1})}\left[\pi_{S_k}(1-\pi_{S_k}) + \frac{(1-Q_k)(1-\pi_{S_k})}{Q_k}\right]. \tag{21}$$

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

$$\hat{\pi}_{mS_k} = \frac{m_{k1}}{m_k}\hat{\pi}_{h1_k} + \frac{m_k - m_{k1}}{m_k}\hat{\pi}_{h2_k}, \ \text{for } 0 < \frac{m_{k1}}{m_k} < 1. \tag{22}$$

Its variance is

$$V(\hat{\pi}_{mS_k}) = \frac{\pi_{S_k}(1-\pi_{S_k})}{m_k} + \frac{1}{m_k}\left[(1-\pi_{S_k})(1-Q_k)\left\{\frac{\lambda_k}{Q_k} + (1-\lambda_k)\right\}\right], \tag{23}$$

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

Thus, the unbiased estimator of $\pi_S = \sum_{k=1}^{r} w_k\pi_S k$ is given by:

$$\hat{\pi}_{mS} = \sum_{k=1}^{r} w_k \hat{\pi}_{mS_k} = \sum_{k=1}^{r} w_k \left[\frac{m_{k1}}{m_k}\hat{\pi}_{h1_k} + \frac{m_k - m_{k1}}{m_k}\hat{\pi}_{h2_k}\right], \tag{24}$$

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

$$V(\hat{\pi}_{mS}) = \sum_{k=1}^{r} \frac{w_k^2}{m_k} \times \left[\pi_{S_k}(1-\pi_{S_k}) + \left((1-\pi_{S_k})(1-Q_k)\left\{\frac{\lambda_k}{Q_k} + (1-\lambda_k)\right\}\right)\right], \tag{25}$$

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

Theorem 4.1: The optimal (Neyman) allocation of m to $m_1, m_2 \ldots m_{r-1}$ and m_r to derive the minimum variance of the $\hat{\pi}_{mS}$ subject to $n = \sum_{k=1}^{r} m_k$ is approximately given by:

$$\frac{m_k}{n} = \frac{w_k\left[\pi_{S_k}(1-\pi_{S_k}) + \left((1-\pi_{S_k})(1-Q_k)\left\{\frac{\lambda_k}{Q_k} + (1-\lambda_k)\right\}\right)\right]^{1/2}}{\sum_{k=1}^{r} w_k\left[\pi_{S_k}(1-\pi_{S_k}) + \left((1-\pi_{S_k})(1-Q_k)\left\{\frac{\lambda_k}{Q_k} + (1-\lambda_k)\right\}\right)\right]^{1/2}}, \tag{26}$$

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

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

$$V(\hat{\pi}_{mS}) = \frac{1}{n}\left[\sum_{k=1}^{r} w_k \left\{\pi_{S_k}(1-\pi_{S_k}) + \left((1-\pi_{S_k})(1-Q_k)\left\{\frac{\lambda_k}{Q_k} + (1-\lambda_k)\right\}\right)\right\}^{1/2}\right]^2, \tag{27}$$

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

Efficiency Comparison

In this section, we do an efficiency comparison of Kim and Warde's (2005) stratified mixed RR model with our proposed stratified mixed RR model by way of variance comparison.

The minimal variance of the Kim and Warde's (2005) estimator $\hat{\pi}_{kw}$ based on stratified mixed RR model is given by:

$$\begin{aligned} V(\hat{\pi}_{kw}) &= \frac{1}{n}\left[\sum_{k=1}^{r} w_k\left\{\pi_{S_k}(1-\pi_{S_k}) + \left(\frac{(1-Q_k)\{\lambda_k Q_k(1-\pi_{S_k}) + (1-\lambda_k)\}}{Q_k^2}\right)\right\}^{1/2}\right]^2 \\ &= \frac{1}{n}\left[\sum_{k=1}^{2} w_k\{a_k + b_k + c_k\}^{1/2}\right]^2, \end{aligned} \tag{28}$$

(see Kim and Warde's 2005, eq. (4.12):218), where

$$a_k = \pi_{S_k}(1-\pi_{S_k}),\ b_k = \frac{\{\lambda_k(1-\pi_{S_k})(1-Q_k)\}}{Q_k}$$
$$\text{and } c_k = \frac{\{(1-Q_k)(1-\lambda_k)\}}{Q_k^2}.$$

The minimal variance of the proposed estimator $\hat{\pi}_{mS}$ in equation (27) is rewritten as

$$V(\hat{\pi}_{mS}) = \frac{1}{n}\left[\sum_{k=1}^{r} w_k\{a_k + b_k + d_k\}^{1/2}\right]^2, \tag{29}$$

where a_k and b_k are same as defined in equation (28) and $d_k = \frac{\{(1-Q_k)(1-\lambda_k)(1-\pi_{S_k}\}}{Q_k}$.

From equations (28) and (29), we have

$$
\begin{aligned}
&V(\hat{\pi}_{kw}) - V(\hat{\pi}_{mS}) \\
&= \frac{1}{n}\left\{\left[\sum_{k=1}^{r} w_k\{a_k+b_k+c_k\}^{1/2}\right]^2 - \left[\sum_{k=1}^{r} w_k\{a_k+b_k+d_k\}^{1/2}\right]^2\right\} \\
&= \frac{1}{n}\left\{\sum_{k=1}^{r} w_k\{a_k+b_k+c_k\}^{1/2} - \sum_{k=1}^{r} w_k\{a_k+b_k+d_k\}^{1/2}\right\} \\
&\quad \times \left\{\sum_{k=1}^{r} w_k\{a_k+b_k+c_k\}^{1/2} + \sum_{k=1}^{r} w_k\{a_k+b_k+d_k\}^{1/2}\right\} \\
&= \frac{1}{n}\left[\sum_{k=1}^{r}\left\{w_k\{a_k+b_k+c_k\}^{1/2} - \{a_k+b_k+d_k\}^{1/2}\right\}\right] \\
&\quad \times \left[\sum_{k=1}^{r}\left\{w_k\{a_k+b_k+c_k\}^{1/2} + \{a_k+b_k+d_k\}^{1/2}\right\}\right],
\end{aligned}
$$

which is positive if

$$
\begin{aligned}
&\left[\{a_k+b_k+c_k\}^{1/2} - \{a_k+b_k+d_k\}^{1/2}\right] > 0 \quad \forall k = 1, 2, \ldots, r \\
&\text{that is, if } \{a_k+b_k+c_k\} > \{a_k+b_k+d_k\} \quad \forall k = 1, 2, \ldots, r \\
&\text{that is, if } \{a_k+b_k+c_k\} > \{a_k+b_k+d_k\} \quad \forall k = 1, 2, \ldots, r \\
&\text{that is, if } c_k > d_k \quad \forall k = 1, 2, \ldots, r \\
&\text{that is, if } \frac{(1-\lambda_k)(1-Q_k)}{Q_k^2} > \frac{(1-Q_k)(1-\pi_{S_k})(1-\lambda_k)}{Q_k}, \qquad (30) \\
&\quad \forall k = 1, 2, \ldots, r \\
&\text{that is, if } \frac{1}{Q_k} > (1-\pi_{S_k}), \quad \forall k = 1, 2, \ldots, r \\
&\text{that is, if } \left(\frac{1}{Q_k} - 1 + \pi_{S_k}\right) > 0, \quad \forall k = 1, 2, \ldots, r,
\end{aligned}
$$

which is always true.

It follows from equation (30) that our proposed stratified mixed RR model is superior to the one earlier considered by Kim and Warde (2005).

To have tangible idea about the performance of the proposed stratified estimator $\hat{\pi}_{mS}$ over Kim and Warde (2005) stratified estimator $\hat{\pi}_{kw}$ in case of two strata (i.e., $r = 2$), we have computed PRE by using the formula:

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

$$= \frac{\left[\sum_{k=1}^{2} w_k \left\{\pi_{S_k}(1-\pi_{S_k}) + \left(\frac{(1-Q_k)\{\lambda_k Q_k(1-\pi_{S_k})+(1-\lambda_k)\}}{Q_k^2}\right)\right\}^{1/2}\right]^2}{\left[\sum_{k=1}^{2} w_k \left\{\pi_{S_k}(1-\pi_{S_k}) + \left((1-\pi_{S_k})(1-Q_k)\left\{\frac{\lambda_k}{Q_k} + (1-\lambda_k)\right\}\right)\right\}^{1/2}\right]^2} \times 100, \tag{31}$$

where $V(\hat{\pi}_{kw})$ is given by Kim and Warde (2005, eq. (4.12):218), $\pi_S = w_1\pi_{S1} + w_2\pi_{S2}$.

We have obtained the values of the PREs for different cases of $\pi_S, \lambda,\ n,\ n_1,$ and P_1. Findings are shown in Table 2, and diagrammatic representation is given in Figure 2, respectively.

Table 2 exhibits that the PRE of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to the Kim and Wardes (2005) stratified estimator $\hat{\pi}_{kw}$ increases as sample size and value of P increase. Larger gain in efficiency is observed for small as well as moderately large sample sizes. However, the PRE is more than 100 for all parametric values considered here; therefore, the proposed estimator $\hat{\pi}_{mS}$ is better than the Kim and Warde's (2005) stratified estimator $\hat{\pi}_{kw}$. Also the values of the PRE $(\hat{\pi}_{mS}, \hat{\pi}_{kw})$ increase as the value of λ increase for fixed values of P_1. Figure 2 demonstrates that there is large gain in efficiency by using the suggested estimator $\hat{\pi}_{mS}$ based on the proposed stratified mixed RR model over the Kim and Warde's (2005) stratified estimator $\hat{\pi}_{kw}$ when the proportion of stigmatizing attribute is moderately large.

Discussion

In this article, we have envisaged a mixed RR model as well as its stratified version to estimate the proportion of qualitative sensitive character. It has been shown both theoretically and empirically that the proposed mixed RR model and its stratified version are always better than the Kim and Warde's (2005) mixed randomize response models with larger gain in efficiency. Thus, our recommendation is to prefer the proposed mixed RR model and its stratified version in practice.

Table 2. Percentage Relative Efficiency of the Proposed Stratified Estimator $\hat{\pi}_{mS}$ With Respect to Kim and Warde's (2005) Stratified Estimator $\hat{\pi}_{kw}$.

						$P_1 = Q_1 = Q_2$						
π_{S1}	π_{S2}	π_S	w_1	w_2	λ	0.9	0.91	0.92	0.93	0.94	0.95	0.96
0.88	0.92	0.896	0.6	0.4	0.1	349.31	339.00	328.76	318.58	308.44	298.33	288.23
			0.6	0.4	0.2	337.52	328.36	319.27	310.24	301.25	292.30	283.36
			0.6	0.4	0.3	325.55	317.57	309.66	301.80	293.98	286.21	278.45
			0.6	0.4	0.4	313.41	306.63	299.91	293.25	286.63	280.05	273.50
0.89	0.93	0.906	0.6	0.4	0.1	374.52	362.90	351.34	339.84	328.36	316.90	305.41
			0.6	0.4	0.2	361.31	350.97	340.69	330.46	320.27	310.09	299.91
			0.6	0.4	0.3	347.90	338.86	329.89	320.97	312.08	303.22	294.36
			0.6	0.4	0.4	334.27	326.57	318.93	311.34	303.79	296.26	288.76
0.90	0.94	0.916	0.6	0.4	0.1	405.96	392.69	379.47	366.29	353.12	339.94	326.71
			0.6	0.4	0.2	390.96	379.13	367.35	355.61	343.89	332.16	320.41
			0.6	0.4	0.3	375.72	365.36	355.05	344.78	334.53	324.30	314.05
			0.6	0.4	0.4	360.22	351.36	342.55	333.78	325.05	316.33	307.62
0.91	0.95	0.926	0.6	0.4	0.1	446.34	430.93	415.56	400.21	384.84	369.41	353.89
			0.6	0.4	0.2	429.03	415.26	401.54	387.83	374.11	360.36	346.55
			0.6	0.4	0.3	411.42	399.33	387.28	375.26	363.24	351.20	339.12
			0.6	0.4	0.4	393.48	383.11	372.78	362.48	352.19	341.91	331.61

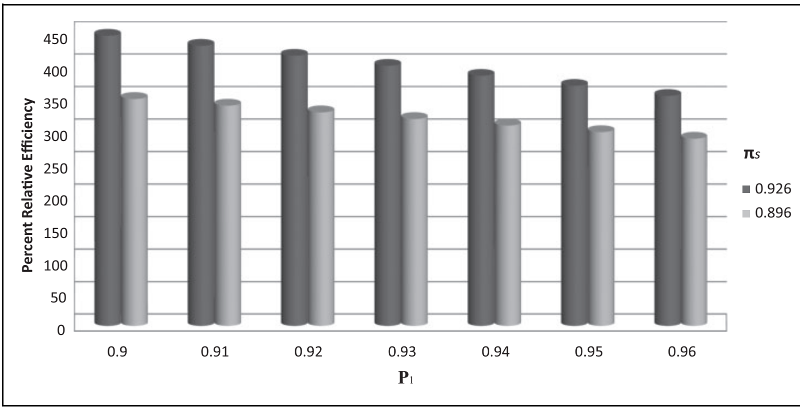

Figure 2. Percentage relative efficiency of the proposed stratified estimator $\hat{\pi}_{mS}$ with respect to Kim and Warde (2005) stratified estimator $\hat{\pi}_{kw}$.

Acknowledgements

The authors are thankful to the editor in chief—Professor Christopher Winship, and to the three anonymous learned referees for their valuable suggestions regarding improvement of the article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Chang, H. J. and K. C. Huang. 2001. "Estimation of Proportion and Sensitivity of a Qualitative Character." *Mertika* 53:269-80.

Chang, H. J., C. L. Wang, and K. C. Huang. 2004. "On Estimating the Proportion of a Qualitative Sensitive Character Using Randomized Response Sampling." *Quality and Quantity* 38:675-80.

Chaudhuri, A. and R. Mukerjee. 1987. "Randomized Response Technique: A Review." *Statistica Neerlandica* 41:27-44.

Chaudhuri, A. and R. Mukerjee. 1988. *Randomized Response: Theory and Techniques*. New York: Marcel-Dekker.

Chua, T. C. and A. K. Tsui. 2000. "Procuring Honest Responses Indirectly." *Journal of Statististical Planning Inference* 90:107-16.

Fox, J. A. and P. E. Tracy. 1986. *Randomized Response: A Method of Sensitive Surveys*. Newbury Park, CA: Sage.

Grewal, I. S., M. L. Bansal, and S. S. Sidhu. 2005-2006. "Population Mean Corresponding to Horvitz-Thompson's Estimator for Multi-characteristics Using Randomized Response Technique." *Model Assisted Statistics and Applications* 1:215-20.

Hedayat, A. S. and B. K. Sinha. 1991. *Design and Inference in Finite Population Sampling*. New York: Wiley.

Hong, K., J. Yum, and H. Lee. 1994. "A Stratified Randomized Response Technique." *Korean Journal of Applied Statistics* 7:141-47.

Hong, Z. 2005-2006. "Estimation of Mean in Randomized Response Surveys When Answers Are Incompletely Truthful." *Model Assisted Statistics and Applications* 1:221-30.

Horvitz, D. G., B. V. Shah, and W. R. Simmons. 1967. "The Unrelated Question Randomized Response Model." *Proceedings of the Social Statistics Section American Statistical Association* 65-72.

Huang, K. C. 2004. "A Survey Technique for Estimating the Proportion and Sensitivity in a Dichotomous Finite Population." *Statistica Neerlandica* 58:75-82.

Kim, J. and W. D. Warde. 2004. "A Stratified Warner Randomized Response Model." *Journal of Statistical Planning and Inference* 120:155-65.

Kim, J. M. and M. E. Elam. 2005. "A Two-stage Stratified Warner's Randomized Response Model Using Optimal Allocation." *Metrika* 61:1-7.

Kim, J. M. and W. D. Warde. 2005. "A Mixed Randomized Response Model." *Journal of Statistical Planning and Inference* 133:211-21.

Land, M., S. Singh, and S. A. Sedory. 2011. "Estimation of a Rare Attribute Using Poisson Distribution." *Statistics* 46:351-60. doi:10.1080/02331888.2010.524300.

Lanke, J. 1976. "On the Degree of Protection in Randomized Interview Internet." *Statistical Review* 44:80-83.

Mahajan, P. K., P. Sharma, and R. K. Gupta. 2007. "Optimum Stratification for Allocation Proportional to Strata Totals for Scrambled Response." *Model Assisted Statistics and Applications* 2:81-88.

Mangat, N. S. 1994. "An Improved Randomized Response Strategy." *Journal of Royal Statistical Society*, Series B, 56, 93-95.

Mangat, N. S., R. Singh, and S. Singh. 1997. "Violation of Respondent's Privacy in Moors Model—Its Rectification through a Random Group Strategy Response Model." *Communications in Statistics Theory and Methods* 3:243-55.

Mahmood, M., S. Singh, and S. Horn. 1998. "On the Confidentiality Guaranteed Under Randomized Response Sampling: A Comparison with Several New Techniques." *Biometrical Journal* 40:2, 237-42.

Moors, J. A. 1971. "Optimization of the Unrelated Question Randomized Response Model." *Journal of the American Statistical Association* 66:627-29.

Perri, P. F. 2008. "Modified Randomized Devices for Simmons' Model." *Model Assisted Statistics and Applications* 3:233-39.

Ryu, J. B., K. H. Hong, and G. S. Lee. 1993. *Randomized Response Model*. Seoul, Korea: Freedom Academy.

Ryu, J. B., J. M. Kim, T. Y. Heo, and C. G. Park. 2005-2006. "On Stratified Randomized Response Sampling." *Model Assisted Statistics and Applications* 1:31-36.

Singh, S. 2003. *Advanced Sampling Theory with Applications*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Singh, S., R. Singh, and N. S. Mangat. 2000. "Some Alternative Strategies to Moor's Model in Randomized Response Model." *Journal of Statistical Planning and Inference* 83:243-55.

Tracy, D. S. and N. S. Mangat. 1996. "Some Developments in Randomized Response Sampling During the Last Decade—A Follow Up of Review by Chaudhuri and Mukherjee." *Journal of Applied Statistical Science* 4:147-58.

Warner, S. L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63-69.

Author Biographies

Housila P. Singh is Professor and Dean in faculty of Sciences at School of Studies in Statistics, Vikram University, Ujjain, India.

Tanveer A. Tarray PhD, is a Research Scholar at School of Studies in Statistics, Vikram University, Ujjain, India.