International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 21(2), 2021, pp 281-296

Improved Randomized Response Model for Estimating the Population Proportion

Sunil Kumar[1*], Sanam Preet Kour[1], Monica Choudhary[1] and Rahul Sharma[2]

[1]Department of Statistics, University of Jammu, J&K, India
[2]Statistical Assistant, Govt. of Jammu and Kashmir, J&K, India

[*]Correspondence should be addressed to Sunil Kumar
(Email: sunilbhougal06@gmail.com)

[Received June 19, 2021; Accepted November 15, 2021]

Abstract

In this paper we have suggested a modified randomized response model and studied its properties. The conditions have been obtained under which the proposed randomized response model is always more efficient than the Warner (1965), Mangat and Singh (1990), Mangat (1994) and Gjestvang and Singh (2006) randomized response model. Numerical illustrations and graphical representations are also given in support of the present study.

Keywords: Randomized response model, Dichotomous population, innocuous variable, sensitive variable, privacy of respondent.

AMS Subject Classification: 62D05.

1. Introduction

In survey of human populations, questions requiring personal or controversial assertion often run into trouble in terms of resistance. It is often difficult to collect reliable data from interviewees and hard to raise the quality of response when the survey topic is sensitive viz necessary information regarding employee integrity, drug and alcohol use, sexual harassment, compliance with legal guidelines, adherence to company policies, and diversity in the workplace, e.g. Chaudhuri & Mukherjee (1988), Fox and Tracy (1986) and Grewal et al. (2003).

The seminal work providing a method for obtaining sensitive information with no risk to the respondent was developed by Warner (1965). There have been many variations of the randomized response technique since its introduction by Warner (1965). Of these, the "unrelated question" model is perhaps the most popular or frequently advocated. With this model, respondents are faced with paired questions or statements; one seeks to elicit a response about a sensitive issue or behaviour, whereas the other inquires about an unrelated and innocuous issue or behaviour.

Some other developments on randomized response sampling in recent years include Mangat and Singh (1990), Mangat (1994), Mohmood et al. (1998), Singh et al. (2000), Singh (2003), Chang et al. (2004), Huang (2004), Kim and Warde (2004), Kim and Elam (2005), Gjestvang and Singh (2006), Ryu et al. (2005-2006), Grewal et al. (2005-2006), Perri (2008), Land et al. (2011), Arnab and Thuto (2015), Arnab et al. (2012), Lee et al. (2014), Chaudhuri (2015), Batool and Shabbir (2016), Fox (2016), etc.

In this paper we briefly review some recent literature and develop a new randomized response model. The properties of the proposed model have been studied. It has been shown that the suggested model performs better than Warner, Mangat and Gjestvang & Singh's models. Numerical illustration is given in support of the proposed study.

2. Some previous work: An Overview

To measure reliable response when from respondents, an effective random device is needed so as to induce each respondent to give truthful answers to sensitive questions without exposing his/her identity to the interviewer. Warner (1965) pioneered randomized response procedure to produce trustworthy data for estimating the proportion π of the population belonging to a sensitive group. The case where the respondents in a population can be divided into two mutually exclusive groups: one group with stigmatizing or sensitive characteristic A and the other group without it. He made use of a randomization device, by using a deck of cards with each card having one of the following two statements:

(i) I belong to group A;

(ii) I do not belong to group A.

Each respondent in the sample is asked to select a card at random from well shuffled deck. Without showing the card to the interviewer, the interviewee answers the question 'Is the statement true for you?' For estimating π, the proportion of respondents in the population belonging to the sensitive group A, a simple random sample of n-respondents is selected with replacement from the population. Out of n-respondents, the number of respondent 'n_1' who answer "Yes" is binomially distributed with parameters $p_0\pi + (1-p_0)(1-\pi)$ and n, where p_0 and $(1-p_0)$ are the relative frequencies in the deck of cards. Thus the maximum likelihood unbiased estimator of π exists for $p_0 \neq 0.5$ and is given by

$$t_W = \frac{(n_1/n)-(1-p_0)}{2p_0-1}, \tag{1}$$

with variance

$$Var(t_W) = \frac{\pi(1-\pi)}{n} + \frac{p_0(1-p_0)}{n(2p_0-1)^2}. \tag{2}$$

Mangat and Singh (1990) suggested a two stage randomized response model. In the first stage each respondent is requested to use a randomization device R_1^* such as a deck of cards with each card having written one of the following statements:

(i) I belong to the sensitive group; and

(ii) Go to random device R_2^*.

The statements occurs with relative frequencies T_0 and $(1-T_0)$, respectively in the first device R_1^*. In the second stage, if directed by the outcome of R_1^*, the respondent is requested to use the randomization device R_2^* which is same as Warner (1965) device. Under the two-stage randomized response model, an unbiased estimator of the population proportion π is given by

$$t_{MS} = \frac{(n_1/n)-(1-T_0)(1-p_0)}{2p_0-1+2T_0(1-p_0)}, \tag{3}$$

with variance

$$Var(t_{MS}) = \frac{\pi(1-\pi)}{n} + \frac{(1-T_0)(1-p_0)\{1-(1-p_0)(1-T_0)\}}{n\{2p_0-1+2T_0(1-p_0)\}^2}. \tag{4}$$

Mangat (1994) suggested another randomized response model where each respondent is instructed to report "Yes" if he/she belongs to the sensitive group A; otherwise the respondent is instructed to use the Warner (1965) device. For this model, an unbiased estimator of the population proportion is given by

$$t_M = \frac{(n_1/n)-(1-p_0)}{p_0}, \tag{5}$$

with variance

$$Var(t_M) = \frac{\lambda_M(1-\lambda_M)}{np_0^2}. \tag{6}$$

where $\lambda_M = \pi + (1-\pi)(1-p_0)$.

Further, Gjestvang and Singh (2006) suggested a new procedure for randomized response technique, discuss as:

If the person belongs to the sensitive group A, then he/she is requested to use the randomization device R_1. Let α_1 and β_1 be any two positive real numbers chosen, such that $p = \frac{\alpha_1}{\alpha_1+\beta_1}$ and $1-p = \frac{\beta_1}{\alpha_1+\beta_1}$ are the probabilities in the randomization device R directing the selected respondent to report a scrambled response (or indirect response) as $1+\beta_1 S_1$ and $1+\alpha_1 S_1$, respectively, where S_1 is any non-directional scrambling variable; i.e. S_1 can take positive, zero and negative values. If the person does not belong to the sensitive group A, then he/she is requested to use the randomization device R_2. Let α_1 and β_2 be any two positive real numbers chosen, such that $T = \frac{\alpha_2}{\alpha_2+\beta_2}$ and $1-T = \frac{\beta_2}{\alpha_2+\beta_2}$ are the probabilities in the randomization device R_2 directing the selected respondent to report a scrambled response as $\beta_2 S_2$ and $-\alpha_2 S_2$, respectively, where S_2 is any non-directional scrambling variable. The distribution of the scrambling variables S_1 and S_2 may or may not be known. Thus, the unbiased estimator of π is given by

$$t_{GS} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{7}$$

where y_i be the random number (positive, zero or negative) that is reported by the i^{th} respondent through the device proposed; with variance

$$Var(t_{GS}) = \frac{\pi(1-\pi)}{n} + \frac{\alpha_2\beta_2}{n}, \tag{8}$$

under the assumptions that $Var(S_1) = \gamma_1^2$ and $Var(S_2) = \gamma_2^2$ are known; $\frac{\alpha_2\beta_2}{\alpha_1\beta_1} = \frac{\gamma_1^2+\theta_1^2}{\gamma_2^2+\theta_2^2}$ and $\gamma_2^2 + \theta_2^2 = 1$.

3. Proposed Randomized Response Device

In the proposed randomized response procedure, each respondent is asked to give answer to the direct question, “I am member of innocuous trait group”. If the respondent reply “Yes” to direct question, then he or she is instructed to go randomization device R_1^*, which is comprising of four statements (i) “I am a member of sensitive trait group”, (ii) “I am not a member of sensitive trait group”, (iii) “I am a member of innocuous trait group”, (iv) “I am not a member of innocuous trait group”. Let α_{1S}, β_{1S}, α_{1Inn} and β_{1Inn} be any positive real numbers chosen, such that $p_S = \frac{\alpha_{1S}}{\alpha_{1S}+\beta_{1S}}$; $(1-p_S) = \frac{\beta_{1S}}{\alpha_{1S}+\beta_{1S}}$, $p_{Inn} = \frac{\alpha_{1Inn}}{\alpha_{1Inn}+\beta_{1Inn}}$; $(1-p_{Inn}) = \frac{\beta_{1Inn}}{\alpha_{1Inn}+\beta_{1Inn}}$ are the probabilities in the randomization device R_1^* directing the selected respondent to report a scrambled response on sensitive and innocuous trait group as $1+\beta_{1S}S_{1S}$, $1-\alpha_{1S}S_{1S}$, $1+\beta_{1Inn}S_{1Inn}$ and $1-\alpha_{1Inn}S_{1Inn}$, respectively, where S_{1S} and S_{1Inn} are any non-directional scrambling variable which can take positive, zero and negative values.

If the person who is selected in the sample does not belong to innocuous trait group then he or she is instructed to use the randomization device R_2^*. Let $\alpha_{2S}, \beta_{2S}, \alpha_{2Inn}$ and β_{2Inn} be any positive real numbers chosen such that $T_S = \frac{\alpha_{2S}}{\alpha_{2S}+\beta_{2S}}$; $(1-T_S) = \frac{\beta_{2S}}{\alpha_{2S}+\beta_{2S}}$, $T_{Inn} = \frac{\alpha_{2Inn}}{\alpha_{2Inn}+\beta_{2Inn}}$; $(1-T_{Inn}) = \frac{\beta_{2Inn}}{\alpha_{2Inn}+\beta_{2Inn}}$ are the probabilities in the randomization device R_2^* directing the selected respondent to report on sensitive and innocuous trait group as $\beta_{2S}S_{2S}$, $-\alpha_{2S}S_{2S}$, $\beta_{2Inn}S_{2Inn}$ and $-\alpha_{2Inn}S_{2Inn}$, respectively, where S_{2S} and S_{2Inn} are any non-directional scrambling variable which can take positive, zero and negative values. Here, in this randomization procedure the distribution of S_{1S}, S_{1Inn}, S_{2S} and S_{2Inn} may or may not be known. As per Gjestvang and Singh (2006), the negative responses will not disclose the privacy of any respondent belonging to non-sensitive or sensitive groups because they come from both groups, but respondents reporting 'no' answers by using the Mangat (1994) model surely belong to a non-sensitive group and disclose their privacy. In proposed model, the negative responses will not disclose the privacy of any respondent belonging to sensitive or innocuous groups.

Based on above, let t_i be the random number (positive, zero or positive) that is reported by the i^{th} respondent, then the unbiased estimator of the population proportion t_S is given by

$$\hat{t}_S = \frac{1}{n}\sum_{i=1}^{n} t_i, \tag{9}$$

where

$$t_i = \begin{cases} 1+\beta_{1S}S_{1S} \text{ with probability } p_S t_S, \\ 1+\beta_{1Inn}S_{1Inn} \text{ with probability } p_{Inn} t_{Inn}, \\ 1-\alpha_{1S}S_{1S} \text{ with probability } (1-p_S)t_S, \\ 1-\alpha_{1Inn}S_{1Inn} \text{ with probability } (1-p_{Inn})t_{Inn}, \\ \beta_{2S}S_{2S} \text{ with probability } (1-t_S)T_S, \\ -\alpha_{2S}S_{2S} \text{ with probability } (1-t_S)(1-T_S), \\ \beta_{2Inn}S_{2Inn} \text{ with probability } (1-t_{Inn})T_{Inn}, \\ -\alpha_{2Inn}S_{2Inn} \text{ with probability } (1-t_{Inn})(1-T_{Inn}), \end{cases} \tag{10}$$

Let E_1 and E_2 denote the expected values over all possible samples and over the randomization device, thus one can get

$$E(\hat{t}_S) = E_1E_2\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) = \frac{1}{n}E_1\sum_{i=1}^{n} E_2(t_i). \tag{11}$$

For simplicity, let $E_1(S_{1S}) = \theta_{1S}$;$E_1(S_{2S}) = \theta_{2S}$;$E_1(S_{1Inn}) = \theta_{1Inn}$; $E_1(S_{2Inn}) = \theta_{2Inn}$; then

$$E_2(t_i) = t_S\{p_S(1+\beta_{1S}\theta_{1S}) + (1-p_S)(1-\alpha_{1S}\theta_{1S})\} + (1-t_S)\{T_S\beta_{2S}\theta_{2S} - (1-T_S)\alpha_{2S}\theta_{2S}\} + t_{Inn}\{p_{Inn}(1+\beta_{1Inn}\theta_{1Inn}) + (1-p_{Inn})(1-\alpha_{1Inn}\theta_{1Inn})\} + (1-t_{Inn})\{T_{Inn}\beta_{2Inn}\theta_{2Inn} - (1-T_{Inn})\alpha_{2Inn}\theta_{2Inn}\} = t_S + t_{Inn}. \tag{12}$$

Substitute the value of $E_2(t_i)$ from (12) in (11), one can obtain $\hat{t}_S$ as an unbiased estimator of the population proportion t_S.

Further, let us assume that $Var(S_{1S}) = \gamma_{1S}^2$; $Var(S_{2S}) = \gamma_{2S}^2$; $Var(S_{1Inn}) = \gamma_{1Inn}^2$ and $Var(S_{2Inn}) = \gamma_{2Inn}^2$ are known. Here, the responses are independent, the variance of $\hat{t}_S$ is given by

$$Var(\hat{t}_S) = Var\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(t_i). \tag{13}$$

Assume that V_1 and V_2 denote the variance over all possible samples and the variance over the randomization device respectively, one can obtain

$$\begin{aligned} V_2(t_i) = {} & E_2(t_i^2) - \{E_2(t_i)\}^2 \\ = {} & t_S\{p_S E_2(1+\beta_{1S}S_{1S})^2 + (1-p_S)E_2(1-\alpha_{1S}S_{1S})^2\} \\ & + (1-t_S)\{T_S E_2(\beta_{2S}S_{2S})^2 + (1-T_S)E_2(-\alpha_{2S}S_{2S})^2\} \\ & + t_{Inn}\{p_{Inn}E_2(1+\beta_{1Inn}S_{1Inn})^2 \\ & + (1-p_{Inn})E_2(1-\alpha_{1Inn}S_{1Inn})^2\} \\ & + (1-t_{Inn})\{T_{Inn}E_2(\beta_{2Inn}S_{2Inn})^2 \\ & + (1-T_{Inn})E_2(-\alpha_{2Inn}S_{2Inn})^2\} - (t_S + t_{Inn})^2 \\ = {} & t_S(1-t_S) + t_S(\gamma_{1S}^2 + \theta_{1S}^2)\{p_S\beta_{1S}^2 + (1-p_S)\alpha_{1S}^2\} \\ & + (1-t_S)(\gamma_{2S}^2 + \theta_{2S}^2)\{T_S\beta_{2S}^2 + (1-T_S)\alpha_{2S}^2\} + t_{Inn}(1-t_{Inn}) \\ & + t_{Inn}(\gamma_{1Inn}^2 + \theta_{1Inn}^2)\{p_{Inn}\beta_{1Inn}^2 + (1-p_{Inn})\alpha_{1Inn}^2\} \\ & + (1-t_{Inn})(\gamma_{2Inn}^2 + \theta_{2Inn}^2)\{T_{Inn}\beta_{2Inn}^2 + (1-T_{Inn})\alpha_{2Inn}^2\} \\ & - 2t_S t_{Inn} \end{aligned}$$

$$\begin{aligned} V_2(t_i) = {} & t_S(1-t_S) + t_{Inn}(1-t_{Inn}) - 2t_S t_{Inn} + t_S(\gamma_{1S}^2 + \theta_{1S}^2)\{p_S\beta_{1S}^2 + \\ & (1-p_S)\alpha_{1S}^2\} + t_{Inn}(\gamma_{1Inn}^2 + \theta_{1Inn}^2)\{p_{Inn}\beta_{1Inn}^2 + (1-p_{Inn})\alpha_{1Inn}^2\} + \\ & (1-t_S)(\gamma_{2S}^2 + \theta_{2S}^2)\{T_S\beta_{2S}^2 + (1-T_S)\alpha_{2S}^2\} + \\ & (1-t_{Inn})(\gamma_{2Inn}^2 + \theta_{2Inn}^2)\{T_{Inn}\beta_{2Inn}^2 + (1-T_{Inn})\alpha_{2Inn}^2\}, \end{aligned} \tag{14}$$

Let $\frac{\alpha_{2S}+\beta_{2S}}{\alpha_{1S}+\beta_{1S}} = \frac{\gamma_{1S}^2+\theta_{1S}^2}{\gamma_{2S}^2+\theta_{2S}^2}$; $\frac{\alpha_{2Inn}+\beta_{2Inn}}{\alpha_{1Inn}+\beta_{1Inn}} = \frac{\gamma_{1Inn}^2+\theta_{1Inn}^2}{\gamma_{2Inn}^2+\theta_{2Inn}^2}$; and $\gamma_{2S}^2 + \theta_{2S}^2 = 1 = \gamma_{2Inn}^2 + \theta_{2Inn}^2$. By using the assumptions and (14), one can obtain the variance of $\hat{t}_S$ as given by

$$Var(\hat{t}_S) = \frac{t_S(1-t_S)}{n} + \frac{t_{Inn}(1-t_{Inn})}{n} + \frac{\alpha_{2S}\beta_{2S}}{n} + \frac{\alpha_{2Inn}\beta_{2Inn}}{n} - 2t_S t_{Inn}. \tag{15}$$

The justification and the benefit of choosing γ_{ij}^2, θ_{ij}^2, α_{ij} and β_{ij} for $i = 1,2$; $j = S, Inn$; as mentioned earlier that the second term in the variance of the new estimator becomes free from the parameter of interest $t_j; j = S$ and Inn. Thus, the variance of $\hat{t}_S$ can always be made smaller than the variances of the Warner (1965), Mangat and Singh (1990), and Gjestvang and Singh (2009) models just by adjusting the values of α_{2j} and β_{2j}.

Specifically, we have the following results

a) If
$p_S(1+\beta_{1S}\theta_{1S}) + p_{Inn}(1+\beta_{1Inn}\theta_{1Inn}) + (1-p_S)(1-\alpha_{1S}\theta_{1S}) + (1-p_{Inn})(1-\alpha_{1Inn}\theta_{1Inn}) = p_0$, and
$T_S\beta_{2S}\theta_{2S} + T_{Inn}\beta_{2Inn}\theta_{2Inn} - (1-T_S)\alpha_{2S}\theta_{2S} - (1-T_{Inn})\alpha_{2Inn}\theta_{2Inn} = 1 - p_0$, then the proposed new model reduces to the Warner (1965) model.

b) If
$p_S(1+\beta_{1S}\theta_{1S}) + p_{Inn}(1+\beta_{1Inn}\theta_{1Inn}) + (1-p_S)(1-\alpha_{1S}\theta_{1S}) + (1-p_{Inn})(1-\alpha_{1Inn}\theta_{1Inn}) = (1-p_0)(1-T_0)$, and
$T_S\beta_{2S}\theta_{2S} + T_{Inn}\beta_{2Inn}\theta_{2Inn} - (1-T_S)\alpha_{2S}\theta_{2S} - (1-T_{Inn})\alpha_{2Inn}\theta_{2Inn} = 1-(1-p_0)(1-T_0)$, then the proposed new model reduces to Mangat and Singh (1990) model.

c) If
$p_S(1+\beta_{1S}\theta_{1S}) + p_{Inn}(1+\beta_{1Inn}\theta_{1Inn}) + (1-p_S)(1-\alpha_{1S}\theta_{1S}) + (1-p_{Inn})(1-\alpha_{1Inn}\theta_{1Inn}) = 1$, and
$T_S\beta_{2S}\theta_{2S} + T_{Inn}\beta_{2Inn}\theta_{2Inn} - (1-T_S)\alpha_{2S}\theta_{2S} - (1-T_{Inn})\alpha_{2Inn}\theta_{2Inn} = 1 - p_0$, then the proposed new model reduces to the Mangat (1994) model.

d) If
$p_S(1+\beta_{1S}\theta_{1S}) + p_{Inn}(1+\beta_{1Inn}\theta_{1Inn}) + (1-p_S)(1-\alpha_{1S}\theta_{1S}) + (1-p_{Inn})(1-\alpha_{1Inn}\theta_{1Inn}) = t(1+\beta_1\theta_1) + (1-t)(1-\alpha_1\theta_1)$, and
$T_S\beta_{2S}\theta_{2S} + T_{Inn}\beta_{2Inn}\theta_{2Inn} - (1-T_S)\alpha_{2S}\theta_{2S} - (1-T_{Inn})\alpha_{2Inn}\theta_{2Inn} = T\beta_2\theta_2 - (1-T)\alpha_2 S_2$, then the proposed model reduces to the Gjestvang and Singh (2006) model.

4. Relative Efficiency

It is noted that the values of α_{1S}, α_{2S}, β_{1S}, β_{2S}, α_{1Inn}, α_{2Inn}, β_{1Inn}, β_{2Inn}, γ_{1S}^2, γ_{2S}^2, γ_{1Inn}^2 and γ_{2Inn}^2 are predetermined before doing the survey and also assumed to be known. Note that θ_{1S}, θ_{2S}, θ_{1Inn} and θ_{2Inn} are non-directional. From (2), (4), (6), (8) and (15), one can get

(i) $Var(\hat{t}_S) \le Var(t_W)$, if

$$\{t_S(1-t_S) + t_{Inn}(1-t_{Inn}) + \alpha_{2S}\beta_{2S} + \alpha_{2Inn}\beta_{2Inn} - 2nt_S t_{Inn}\} \le (1-\pi) + \frac{p_0(1-p_0)}{(2p_0-1)^2} \tag{16}$$

ii) $Var(\hat{t}_S) \le Var(t_{MS})$, if

$$\{t_S(1-t_S) + t_{Inn}(1-t_{Inn}) + \alpha_{2S}\beta_{2S} + \alpha_{2Inn}\beta_{2Inn} - 2nt_S t_{Inn}\} \le \pi(1-\pi) + \frac{(1-T_0)(1-p_0)\{1-(1-p_0)(1-T_0)\}}{\{2p_0-1+2T_0(1-p_0)\}^2} \tag{17}$$

(iii) $Var(\hat{t}_S) \le Var(t_M)$, if

$$\{t_S(1-t_S) + t_{Inn}(1-t_{Inn}) + \alpha_{2S}\beta_{2S} + \alpha_{2Inn}\beta_{2Inn} - 2nt_S t_{Inn}\} \le \frac{\lambda_M(1-\lambda_M)}{p_0^2} \tag{18}$$

(iv) $Var(\hat{t}_S) \le Var(t_{GS})$, if

$$\{t_S(1-t_S) + t_{Inn}(1-t_{Inn}) + \alpha_{2S}\beta_{2S} + \alpha_{2Inn}\beta_{2Inn} - 2nt_S t_{Inn}\} \le \pi(1-\pi) + \alpha_2\beta_2 \tag{19}$$

The proposed model is more efficient than other considered models if the conditions (16)-(19) holds true. To see the magnitude of the gain in efficiency of the proposed randomized response model with respect to the existing model, we compute the relative efficiency of proposed model with respect to others, as

$$RE_1 = \frac{Var(t_W)}{Var(\hat{t}_S)} * 100\ ;\ RE_2 = \frac{Var(t_{MS})}{Var(\hat{t}_S)} * 100\ ;\ RE_3 = \frac{Var(t_M)}{Var(\hat{t}_S)} * 100\ ;\ RE_4 = \frac{Var(t_{GS})}{Var(\hat{t}_S)} * 100.$$

Results are shown in Table (1-4) and diagrammatic representations are also given in Figure (1-4).

Table 1: Relative efficiency of the proposed model with respect to Warner's model.

p_0			Relative efficiencies for the following values of π						
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	$t_S = 0.02;\ t_{Inn} = 0.02;\ \alpha_{2S} = 0.1;\ \beta_{2S} = 0.2;\ \alpha_{2Inn} = 0.01;\ \beta_{2Inn} = 0.01$								
0.7	2397.44	2517.09	2602.56	2653.846	2670.94	2653.85	2602.56	2517.09	2397.44
0.8	913.58	1033.24	1118.71	1169.99	1187.09	1169.99	1118.71	1033.24	913.58
0.9	394.23	513.89	599.36	650.64	667.74	650.64	599.36	513.89	394.23
	$t_S = t_{Inn} = \alpha_{2S} = \beta_{2S} = \alpha_{2Inn} = \beta_{2Inn} = 0.02$								
0.7	3577.81	3756.38	3883.93	3960.46	3985.97	3960.46	3883.93	3756.38	3577.81
0.8	1363.38	1541.95	1669.50	1746.03	1771.542	1746.03	1669.50	1541.95	1363.38
0.9	588.33	766.90	894.45	970.98	996.49	970.98	894.45	766.90	588.33
	$t_S = 0.6;\ t_{Inn} = 0.4;\ \alpha_{2S} = 0.2;\ \beta_{2S} = 0.1;\ \alpha_{2Inn} = 0.7;\ \beta_{2Inn} = 0.08$								
0.7	1845.36	1937.50	2003.29	2042.76	2055.92	2042.76	2003.29	1937.50	1845.39
0.8	703.22	795.32	861.11	900.58	913.74	900.58	861.11	795.32	703.22
0.9	303.45	395.56	461.35	500.82	513.98	500.82	461.35	395.56	303.45

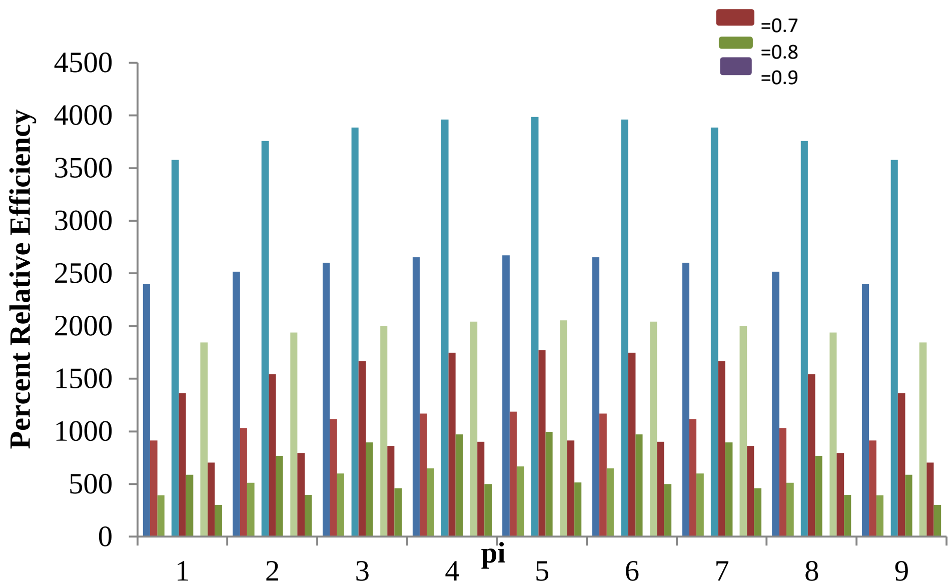

Figure 1: Relative efficiency of the proposed model with respect to Warner's model.

Table 2: Relative efficiency of the proposed model with respect to Mangat's model.

p_0	Relative efficiencies for the following values of π								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	$t_S = 0.02;\ t_{Inn} = 0.02;\ \alpha_{2S} = 0.1;\ \beta_{2S} = 0.2;\ \alpha_{2Inn} = 0.01;\ \beta_{2Inn} = 0.01$								
0.7	398.46	421.19	427.18	416.41	388.89	344.62	283.59	205.81	111.28
0.8	344.62	393.85	421.19	426.67	410.26	371.97	311.79	229.74	125.81
0.9	263.08	344.62	398.46	424.62	423.08	393.85	336.92	252.31	140.00
	$t_S = t_{Inn} = \alpha_{2S} = \beta_{2S} = \alpha_{2Inn} = \beta_{2Inn} = 0.02$								
0.7	594.64	628.57	637.50	621.43	580.36	514.29	423.21	307.14	166.07
0.8	514.29	587.76	628.57	636.73	612.24	555.10	465.31	342.86	187.76
0.9	392.60	514.29	594.64	633.67	631.38	587.76	502.81	376.53	208.93
	$t_S = 0.6;\ t_{Inn} = 0.4;\ \alpha_{2S} = 0.2;\ \beta_{2S} = 0.1;\ \alpha_{2Inn} = 0.7;\ \beta_{2Inn} = 0.08$								
0.7	306.71	324.21	328.82	320.53	299.34	265.26	218.29	158.42	85.66
0.8	265.26	303.16	324.21	328.42	315.79	286.32	240.00	176.84	96.84
0.9	202.50	265.26	306.71	326.84	325.66	303.16	259.34	194.21	107.76

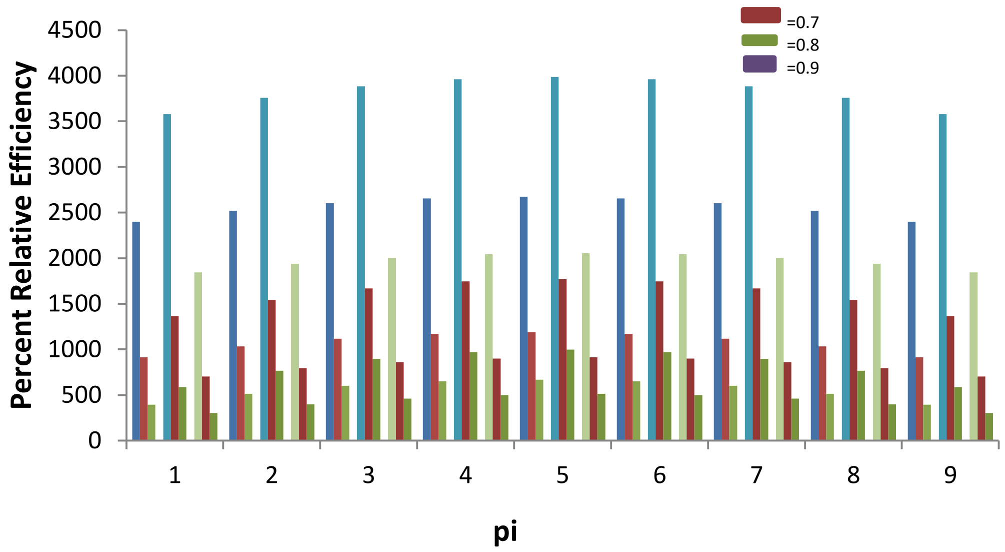

Figure 2: Relative efficiency of the proposed model with respect to Mangat's model.

Table 3: Relative efficiency of the proposed model with respect to Mangat and Singh's model.

p_0	Relative efficiencies for the following values of π								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	$t_S = 0.02;\ t_{Inn} = 0.02;\ \alpha_{2S} = 0.1;\ \beta_{2S} = 0.2;\ \alpha_{2Inn} = 0.01;\ \beta_{2Inn} = 0.01;\ T_0 = 0.2$								
0.7	2295.29	2414.95	2500.42	2551.71	2568.79	2551.71	2500.42	2414.95	2295.29
0.8	803.57	923.23	1008.70	1059.98	1077.08	1059.98	1008.70	923.23	803.57
0.9	350.92	470.58	556.05	607.33	624.42	607.33	556.05	470.58	350.92
	$t_S = t_{Inn} = \alpha_{2S} = \beta_{2S} = \alpha_{2Inn} = \beta_{2Inn} = T_0 = 0.02$								
0.7	3541.95	3720.52	3848.07	3924.60	3950.11	3924.60	3848.07	3720.52	3541.95
0.8	1346.46	1525.03	1652.58	1729.11	1754.62	1729.11	1652.58	1525.03	1346.46
0.9	581.94	760.52	888.07	964.59	990.11	964.59	888.07	760.52	581.94
	$t_S = 0.6;\ t_{Inn} = 0.4;\ \alpha_{2S} = 0.2;\ \beta_{2S} = 0.1;\ \alpha_{2Inn} = 0.7;\ \beta_{2Inn} = 0.08;\ T_0 = 0.5$								
0.7	2515.04	2607.14	2672.93	2712.41	2725.56	2712.41	2672.93	2607.14	2515.04
0.8	488.49	580.59	646.38	685.86	699.01	685.86	646.38	580.59	488.47
0.9	217.63	309.73	375.52	414.99	428.15	414.99	375.52	309.73	217.63

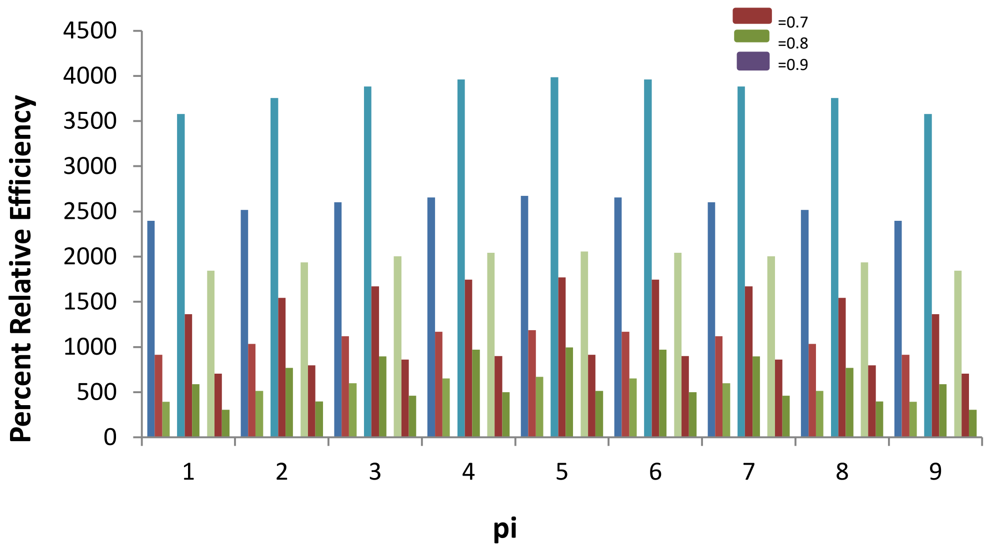

Figure 3: Relative efficiency of the proposed model with respect to Mangat and Singh's model.

Table 4: Relative efficiency of the proposed model with respect to Gjestvang and Singh's model.

$\alpha_{2S} = \alpha_2$	Relative efficiencies for the following values of π								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	$t_S = 0.02;\ t_{Inn} = 0.02;\ \beta_{2S} = 0.2;\ \alpha_{2Inn} = 0.01;\ \beta_{2Inn} = 0.01$								
0.1	188.03	307.69	393.16	444.44	461.54	444.44	393.16	307.69	188.03
0.3	152.28	223.35	274.11	304.57	314.72	304.57	274.11	223.35	152.28
0.5	137.18	187.73	223.83	245.49	252.71	245.49	223.83	187.73	137.18
0.7	128.85	168.07	196.08	212.89	218.49	212.89	196.08	168.07	128.85
0.9	123.57	155.61	178.49	192.22	196.79	192.22	178.49	155.61	123.57
$\beta_{2S} = \beta_2$	$t_S = 0.02;\ t_{Inn} = 0.02;\ \alpha_{2S} = 0.2;\ \alpha_{2Inn} = 0.01;\ \beta_{2Inn} = 0.01$								
0.1	206.19	350.52	453.61	515.46	536.08	515.46	453.61	350.52	206.19
0.3	175.18	277.37	350.37	394.16	408.76	394.16	350.37	277.37	175.18
0.5	158.19	237.29	293.79	327.68	338.98	327.68	293.79	237.29	158.19
0.7	147.47	211.98	258.06	285.71	294.93	285.71	258.06	211.98	147.47
0.9	140.08	194.55	233.46	256.81	264.59	256.81	233.46	194.55	140.08

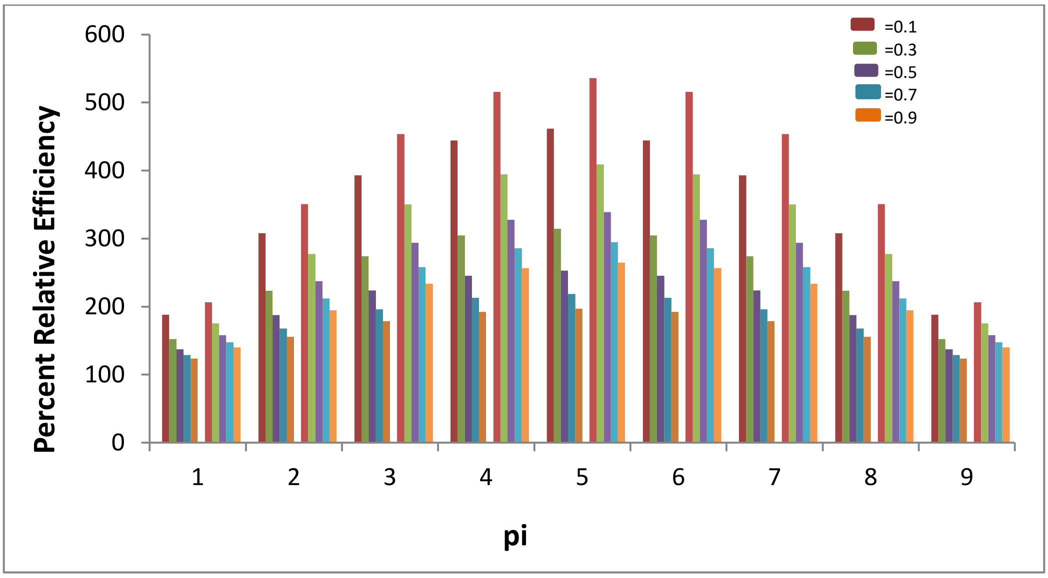

Figure 4: Relative efficiency of the proposed model with respect to Gjestvang and Singh's model.

Table (1 to 4) and figure (1 to 4) envisaged that the proposed estimator is always be more efficient than the estimators by Warner, Mangat & Singh, Mangat, and Gjestvang & Singh estimator's in different situations by considering different values of model's known parameters. In case when the value of π from table 2 is very high i.e. 0.8 and 0.9, the proposed estimator is less efficient than the Mangat's model. It is envisaged from Table 1, 3 and 4 that the proposed model's efficient increases with the increase in the value of π and reached its maximum then decrease in the similar manner with the increase in the value of π. Further, it is envisaged from Table 2 that the proposed model is most efficient as compared to Mangat's model for moderate values of π i.e. for $\pi = 0.3$ and 0.4.

5. Conclusion

In this paper, a new mixed randomized response model is proposed to estimate the proportion of qualitative sensitive character. It has been shown theoretically and empirically that the proposed mixed randomized response model is always better than Warner (1965), Mangat and Singh (1990), Mangat (1994), and Gjestvang and Singh (2006). Thus, our recommendation is to prefer the proposed mixed randomized response model in practice.

Reference

[1] Arnab, R. and Thuto, M. (2015). Randomized response techniques: A case study of the risky behaviours of students of a certain University. Model Assisted Statistics and Applications, 10(4), 421-430.

[2] Arnab, R., Singh, S. and North, D. (2012). Use of two decks of cards in randomized response techniques for complex survey designs. Communications in Statistics-Theory and Methods, 41, 3198-3210.

[3] Batool, F., and Shabbir, J. (2016). A two stage design for multivariate estimation of proportions. Communications in Statistics-Theory and Methods, Doi: 10.1080/03610926.2014.942435.

[4] Chang, H. J., Wang, C. L. and Huang, K. C. (2004). On estimating the proportion of a qualitative sensitive character using randomized response sampling. Quality and Quantity, 38, 675-680.

[5] Chaudhuri, A. and Mukherjee, R. (1988). Randomized Response: Theory and Techniques. Marcel- Dekker, New York, USA.

[6] Chaudhuri, A. (2015). Special issue: Warner's randomized response model. Model Assisted Statistics and Applications, 10(4), 277-457.

[7] Fox, J. A. and Tracy, P. E. (1986). Randomized Response: A Method for Sensitive Surveys, Newbury Park, CA: SAGE Publications.

[8] Fox, J. A. (2016). Randomized response and related methods: Surveying sensitive data. Second edition, SAGE.

[9] Grewal, I. S., Bansal, M. L. and Singh, S. (2003). Estimation of population mean of a stigmatized quantitative variable using double sampling. Statistica, 63, 79-88.

[10] Grewal I. S., Bansal M. L., and Sidhu, S. S. (2006). Population mean corresponding to Horvitz-Thompson's estimator for multi-characteristics using randomized response technique. Model Assisted Statistics and Applications, 1, 215–220.

[11] Huang, K. C. (2004). Survey technique for estimating the proportion and sensitivity in a dichotomous finite population. Statistica Neerlandica., 58, 75-82.

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

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

[14] Land, M., Singh, S. and Sedory, S. A. (2011). Estimation of a rare attribute using Poisson distribution. Statistics, Doi:10.1080/02331888.2010.524300.

[15] Lee, G. S., Hong, K. H., Kim, J. M., and Son, C. K. (2014). Estimation of the proportion of a sensitive attribute based on a two-stage randomized response model with stratified unequal probability sampling. Brazilian Journal of Probability and Statistics, 28(3), 381-408.

[16] Mangat, N. S. (1994). An improved randomized response strategy. J R. Statist. Soc. B, 56, 93-95.

[17] Mangat, S. and Singh, R. (1990). An alternative randomized response procedure. Biometrika, 77, 439-442.

[18] Mohammod M., Singh, S. and Horn, S. (1998). On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. Biometric Journal, 40(2), 237 – 242.

[19] Perri, P. F. (2008). Modified randomized devices for Simmons' model. Model Assisted Statistics Applications, 3(3), 233-239.

[20] Ryu, J. B., Kim, J. M., Heo, T. Y. and Park, C. G. (2005–2006). On stratified randomized response sampling. Model Assisted Statistics Applications, 1, 31–36.

[21] Singh, S., Singh, R. and Mangat, N. S. (2000). Some alternative strategies to Moor's model in randomized response model. Journal of Statistical Planning and Inference, 83, 243-255.

[22] Singh, S. (2003). Advanced Sampling Theory with Applications: How Michael" Selected" Amy. Dordrecht: Kluwer, Doi: 10.1007/978-94-007-0789-4.

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

[24] Gjestvang, C. R. and Singh, S. (2006). A new randomized response model. Journal of the Royal Statistical Society, Series B, 68(30), 523-530.