Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(2) 2016 97–110

Estimation of Population Mean under Scrambled Response in Sample Surveys

G.N. Singh, S. Suman and M. Khetan
Indian School of Mines, Dhanbad

Received 07 April 2016; Revised 20 April 2016; Accepted 21 April 2016

SUMMARY

The present manuscript is an attempt to analyze the consequences of non–response in estimation procedure of population mean. Using the sub-sampling of non-respondent technique, families of factor type estimators have been proposed to estimate the population mean when variable under study represents the sensitive characteristic. Properties of the proposed families of estimators are examined under different randomized response models. The measures of privacy protection and efficiencies of the suggested families of estimators have been calculated and their performances are examined. Results are interpreted and suitable recommendations have been made to the survey practitioners.

Keywords: Auxiliary information, Factor type estimator, Non-response, Scramble response, Bias, Mean square error.

1. INTRODUCTION

The problem of non-response is one of the most frequent and widely realized phenomenon in sample surveys. The usual approach to overcome the non-response problem is to contact the non-respondents and obtain the information as much as possible. Hansen and Hurwitz (1946) initiated a technique to handle the non-response problem in sample surveys by taking a subsample from the non-responding units of the sample and collecting information by personal interviews from the units in subsample. Finally all available information are calibrated for producing the reliable estimates of population parameters. There are several schemes for selecting subsamples from the non-responding units of the sample. Hansen and Hurwitz (1946) assumed that the sample units who refused to respond at first contact would give full response at the time of second contact. Generally in human surveys, especially when we deal with stigmatized character, it is hard to get direct response from the sampled units, no matter how many times one contact.

When survey is related to sensitive issues like drug addiction, alcoholism, gambling, abortion, sex abuse, history of income tax evasion etc. many respondents either refuse to participate or give false or evasive responses. Technique that protect confidentiality and privacy (anonymity) may be useful in reducing the non-response bias and getting reliable information from respondents, such technique is known as randomized response technique (RRT). Warner (1965) introduced this technique for estimating the proportion of person bearing a sensitive attribute in a population, based on a simple random sample of individuals drawn from the population under with replacement scheme. Subsequently, Fox and Tracy (1986), Mangat and Singh (1990), Kuk (1990), Shabbir and Gupta (2005), Christofides (2003), Diana and Perri (2009) among others have contributed to this area and modified Warner's (1965) technique in several directions for qualitative characteristics.

Corresponding author: G.N. Singh
E-mail address: gnsingh_ism@yahoo.com

Greenberg *et al.* (1969) and Eriksson (1973) discussed the situations when response to a sensitive question results in a quantitative variable. Following the idea of quantitative sensitive response, Pollock and Bek (1976), Eichhorn and Hayre (1983), Bar-Lev *et al.* (2004), Saha (2007), Diana and Perri (2011) and Diana *et al.* (2013 b) among others introduced the scrambled randomized response technique (SRRT).

In survey sampling, direct response techniques for collecting information about non-sensitive characters make massive use of information on auxiliary variables to improve sampling design and to achieve higher precision in estimates of population parameters. Various techniques have been developed to estimate the parameters related to non-sensitive characteristics in presence of auxiliary information, while very few procedures have been suggested to improve the randomized response techniques for sensitive characteristics under similar situations. Some notable works for sensitive characteristics which utilize supplementary (auxiliary) information for improving the randomized response techniques may be referred as Yan (2005), Dinna and Perri (2009) and Dinna *et al*. (2013 b) among others.

Motivated by the works on the scramble randomized response techniques and following Hansen and Hurwitz (1946), we have proposed families of factor type estimators to estimate the population mean of quantitative sensitive variable using information on non-sensitive auxiliary variable under the assumption that the person who refuses to respond on first contact gives scramble response on second contact. The objective of this work is also to provide a mechanism to the respondents which ensures them for their confidentiality protection and gives enthusiasm for their true responses. Numerical comparisons are carried out in terms of efficiency and privacy protection with some well-known scrambled response models. Results have been analyzed and followed by suitable recommendations.

2. MODIFIED HANSEN AND HURWITZ (1946) TECHNIQUE FOR NON-RESPONSE

Let $U = \{U_1, U_2, ---, U_N\}$ be a finite population of size N associated with study and auxiliary quantitative variables Y and X respectively and assume the values Y_i and X_i for the i^{th} (i = 1, 2, ... , N) unit of the population. To estimate the population mean of study variable Y, a simple random sample without replacement (SRSWOR) of size n is drawn from the population U and it is assumed that the sample suffers from the problems of non-response. Since non-response is observed, therefore, we assume that the whole population is divided into two classes, those who will respond on first attempt, known as response class and those who will not respond, known as non-response class. Let N_1 and N_2 be the number of units in the population that belongs to response class and non-response class respectively where $N_1 + N_2 = N$. Let n_1 and n_2 be the number of responding and non-responding units respectively in the sample such that $n_1 + n_2 = n$. Following the Hansen and Hurwitz (1946) technique, a subsample under SRSWOR of size $h = \frac{n_2}{f_h}, f_h > 1$ is drawn from the non-responding units of the sample for direct interview. Hansen and Hurwitz (1946) suggested a weighted unbiased estimator of population mean $\bar{Y}$ as

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_h \tag{2.1}$$

where,

$$w_1 = \frac{n_1}{n}, w_2 = \frac{n_2}{n}, \bar{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1} \text{ and } \bar{y}_h = \frac{\sum_{i=1}^{h} y_i}{h}.$$

When nature of study variable Y is related to sensitive issues, it is quite difficult to make sure that all h units of subsample respond through direct interviews at second call. And if they do, then their responses may be untruthful, exaggerated or misleading. To overcome such situations, Hansen and Hurwitz (1946) technique has been modified with assumption that one group of people give direct response at first attempt and non-respondents group give scrambled response at second attempt during sample survey. The scrambled responses are used at second attempt to evoke responses truthfully and secure the privacy protection of respondents.

To estimate population mean of sensitive characteristics, use of many randomized response devices have been discussed in survey literatures. Most of these use a coding mechanism of the true response on Y, i.e. the respondents are asked to algebraically perturb the true value of Y through one or more random numbers generated from known scramble distribution. For the second call in modified Hansen and Hurwitz (1946) technique, these aforesaid available randomized devices have been used for getting the response on Y. Z denotes the scrambled response of true value of sensitive characteristic Y. U_1 and U_2 be two random variables unrelated to Y but not necessarily they are independent, whose distributions are known as well as their means (μ_{u_1}, μ_{u_2}), variances $(\sigma^2_{u_1}, \sigma^2_{u_2})$ and covariance $\sigma_{u_1 u_2}$ are also known. Following, Diana *et al.* (2013 b), the randomized response linear model is given as

$$Z = U_1 Y + U_6 \tag{2.2}$$

We can have $E_R(Z) = \mu_{u_1} Y + \mu_{u_2}$

and $V_R(Z) = \sigma^2_{u_1} Y + \sigma^2_{u_2} + 2\sigma_{u_1 u_2} Y,$

where (E_R, V_R) denote expectation and variance under randomized mechanism.

Let $\hat{y}_i$ be a suitable transformation on second call of scramble response z_i for the i^{th} unit whose expectation coincides with the true response y_i under the randomized mechanism, we get

$$\hat{y}_i = \frac{z_i - \mu_{u_2}}{\mu_{u_1}} \tag{2.3}$$

with variance

$$V_R(\hat{y}_i) = \frac{\sigma^2_{u_1} y_i^2 + \sigma^2_{u_2} + 2\sigma_{u_1 u_2} y_i}{\mu^2_{u_1}} = \theta_i \text{ (say)} \tag{2.4}$$

Hence, the modified version of Hansen and Hurwitz (1946) estimator is given as

$$\hat{\bar{y}}^* = w_1 \bar{y}_1 + w_2 \hat{\bar{y}}_h$$

where,

$$\hat{\bar{y}}_h = \frac{\sum_{i=1}^{h} \hat{y}_i}{h} \tag{2.5}$$

Remark 2.1: The scramble response model is used under the assumptions (i) The interviewer is totally unaware of the random numbers U_1 and U_2 generated by the respondent, which is used for scrambling the true response Y and (ii) The interviewer has complete knowledge of the scrambling distributions of U_1 and U_2. These assumptions provide greater confidence among the respondents about their privacy protection.

2.1 Properties of Modified Hansen and Hurwitz (1946) Estimator

1. The modified Hansen and Hurwitz (1946) estimator $\hat{\bar{y}}^*$ defined in equation (2.5) is an unbiased estimator of $\bar{Y}$.
2. The variance of the estimator $\hat{\bar{y}}^*$ is given as;

$$Var\left(\hat{\bar{y}}^*\right) = f_1 S_Y^2 + \psi S^2_{Y_{(2)}} + \frac{f_h N_2}{nN}\left(\frac{\sigma^2_{u_1}\mu_{2,y} + \sigma^2_{u_2} + 2\sigma_{u_1u_2}\bar{Y}_2}{\mu^2_{u_1}}\right) \tag{2.6}$$

where

$$S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2,\; S^2_{Y(2)} = \frac{1}{N_2-1}\sum_{i=1}^{N_2}\left(Y_i - \bar{Y}_2\right)^2,$$

$$f_h = \frac{n_2}{h},\; f_1 = \left(\frac{1}{\text{n}} - \frac{1}{\text{N}}\right), \psi = \frac{f_h - 1}{n}W_2,\; W_2 = \frac{N_2}{N}$$

and $\mu_{2,y} = \text{S}^2_{Y(2)} + \bar{Y}_2^2$.

The randomized response of study variable possessing the sensitive characteristics may be considered under different randomized response model such as additive, multiplicative and mixed etc. Four known models; additive, multiplicative and mixed types have been discussed in this work.

To show these models, let us denote V_1 and V_2 as two mutually independent scrambled variables and also independent of Y. Table 1 summaries four examples of linear randomized response models considering different situations of the randomized response model shown in equation (2.2).

Table 1. The particular cases of scramble response model shown in equation (2.2)

Model	Authors	U_1	U_2	Scramble Response Z
M_1: Additive Model	Pollock and Bek (1976)	1	V_2	$Z = Y + V_1$
M_2: Multiplicative model	Eichorn and Hayre (1983)	V_1	0	$Z = YV_2$
M_3: Mixed model 1	Saha (2007)	V_1	V_2	$Z = V_1(Y + V_2)$
M_4: Mixed model 2	Dinna *et al.* (2010)	$(1-\alpha)V_1$	$\alpha V_1 V_2$	$z = V_1((1-\alpha)Y + \alpha Y_2)$ where $\alpha \in (0, 1)$

The variance of modified Hansen and Hurwitz (1946) estimator $\hat{\bar{y}}^*$ under different scramble models shown in Table 1 have been derived and presented in Table 2.

Table 2. The variances of $\hat{\bar{y}}^*$ under different models

Model	$Var\left(\hat{\bar{y}}^*_{M_j}\right)\ j = 1, 2, 3, 4.$
M_1: Additive Model	$f_1\ S_Y^2 + \ \psi S^2_{Y_{(2)}} + \frac{f_h N_2}{nN}\sigma^2_{v2}$
M_2: Multi-plicative model	$f_1\ S_Y^2 + \ \psi S^2_{Y_{(2)}} + \frac{f_h N_2}{nN}\left(\frac{\sigma^2_{v_1}\mu_{2,y}}{\mu^2_{v_1}}\right)$
M_3: Mixed model 1	$f_1\ S_Y^2 + \ \psi S^2_{Y_{(2)}} + \frac{f_h N_2}{nN}\left(\frac{\sigma^2_{v_1}\mu_{2,y} + \mu^2_{v_1}\sigma^2_{v_2} + 2\sigma^2_{v_1}\mu_{v_2}\bar{Y}_2 + \sigma^2_{v_1}(\mu^2_{v_2} + \sigma^2_{v_2})}{\mu^2_{v_1}}\right)$
M_4: Mixed model 2	$+\frac{f_h N_2}{nN}\left(\frac{\alpha^2\left\{\sigma^2_{v_1}\left(\mu^2_{v_2} + \sigma^2_{v_2}\right) + \mu^2_{v_1}\sigma^2_{v_2}\right\} + (1-\alpha)^2\sigma^2_{v_1}\mu_{2,y} + 2\alpha(1-\alpha)\bar{Y}_2\sigma^2_{v_1}\mu_{v_2}}{(1-\alpha)^2\mu^2_{v_1}}\right)$

3. PROPOSED FAMILY OF ESTIMATORS OF POPULATION MEAN USING NON-SENSITIVE AUXILIARY VARIABLE

It is cumbersome to estimate the population parameter using information on auxiliary variable for the situations when both study and auxiliary variables are representing sensitive characters and suffered from non-response. To estimate the population mean $\bar{Y}$, families of estimators $T_1(d)$ and $T_2(d)$ have been proposed in presence of non-response for two different cases when (i) non-response occurs only in study variable Y which represents a sensitive characteristic and (ii) non-response occurs in study variable Y and auxiliary variable X which represents a non-sensitive characteristic. It is assumed that Y and X are highly positively correlated.

***Case 1*: When non-response occurs only in study variable**

For this case, the modified Hansen and Hurwitz (1946) technique has been implemented in study variable Y to reduce the effect of non-response in survey data. The proposed family of factor type estimators to estimate the population mean $\bar{Y}$ is given as

$$T_1(d) = \hat{\bar{y}}^* \left\{ \frac{(A+C)\bar{X} + f\ B\,\bar{x}}{(A + f\ B)\bar{X} + \ C\,\bar{x}} \right\} \tag{3.1}$$

***Case 2*: When non-response occurs in study and auxiliary variables both**

In this case, the modified Hansen and Hurwitz (1946) technique has been implemented for study variable whereas Hansen and Hurwitz (1946) technique is implemented for auxiliary variable to reduce the effect of non-response in survey data. The proposed family of factor type estimators to estimate the population mean $\bar{Y}$ is given as

$$T_2(d) = \hat{\bar{y}}^* \left\{ \frac{(A+C)\bar{X} + f\ B\,\bar{x}^*}{(A + f\ B)\bar{X} + \ C\,\bar{x}^*} \right\} \tag{3.2}$$

It is obvious that for different choices of d the families of estimators $T_1(d)$ and $T_2(d)$ converge to different well known estimators.

3.1 Properties of the Proposed Families of Estimators $T_1(d)$ and $T_2(d)$

To study the detailed behaviors of the proposed families of estimators $T_1(d)$ and $T_2(d)$, the expressions of bias and mean square errors of proposed families of estimators are derived up to the first order of approximations under large sample assumption and produced in following theorems:

Theorem 1: The bias of the families of estimators $T_1(d)$ and $T_2(d)$ up to the first order of approximations are obtained as

$$B\{T_1(d)\} = \bar{Y}\, f_1\, \phi(d)\left(\rho_{YX} C_X C_Y - \phi_2(d) C_X^2\right) \quad (3.3)$$

and $B\{T_2(d)\} = \bar{Y}\ \phi(d)$

$$\left\{ f_1\left(\rho_{YX} C_X C_Y - \phi_2(d) C_X^2\right) + \psi\left(\frac{S_{YX_{(2)}}}{\bar{Y}\,\bar{X}} - \phi_2(d)\frac{S^2_{X_{(2)}}}{\bar{X}^2}\right)\right\} \quad (3.4)$$

where

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i = \bar{X})^2,$$

$$C_Y = \frac{S_Y}{Y},\ C_X = \frac{S_X}{X}$$

$$S^2_{X(2)} = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2}(x_i - \bar{X})^2,$$

$$S^2_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}_2\right),$$

$$S^2_{XY(2)} = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}_2\right)$$

$$\phi_1(d) = \frac{fB}{(A + fB + C)},\ \phi_2(d) = \frac{C}{(A + fB + C)},$$

$$\phi(d) = \phi_1(d) - \phi_2(d)$$

and ρ_{YX} is the correlation coefficient between study variable Y and auxiliary variable X.

Theorem 2: The mean square errors of families of estimators $T_1(d)$ and $T_2(d)$ to the first order of approximations are obtained as

$$MSE\{T_1(d)\} = \bar{Y}^2 f_1\left(C_Y^2 + \phi^2(d) C_X^2 + 2\phi(d)\rho_{YX} C_X C_Y\right)$$

$$+ \psi\, S^2_{Y_{(2)}} + \frac{f_h N_2}{nN}\left(\frac{n^2_{u_1} n_{2,y} + n^2_{u2} + 2 n_{u_1 u_2}\bar{Y}_2}{n^2_{u_1}}\right) \quad (3.5)$$

and

$$MSE\{T_2(d)\}$$

$$= \bar{Y}^2\left[f_1\left(C_Y^2 + \phi^2(d) C_X^2 + 2\phi(d)\rho_{YX} C_X C_Y\right)\right.$$

$$\left. + \psi\left(\frac{S^2_{Y(2)}}{\bar{Y}^2} + \phi^2(d)\frac{S^2_{X(2)}}{\bar{X}^2} + 2\phi(d)\frac{S_{YX(2)}}{\bar{Y}\,\bar{X}}\right)\right]$$

$$+ \frac{f_h N_2}{nN}\left(\frac{\sigma^2_{u_1}\mu_{2,y} + \sigma^2_{u_2} + 2\sigma_{u_1 u_2}\bar{Y}_2}{\mu^2_{u_1}}\right) \quad (3.6)$$

3.2 Optimum Choices of *d*

The families of estimators $T_1(d)$ and $T_2(d)$ are the functions of unknown constant d, therefore, to get the optimum choices of d corresponding to the two families of estimators, the corresponding mean square errors $MSE\{T_1(d)\}$ and $MSE\{T_2(d)\}$ derived in equations (3.5) and (3.6) respectively are minimized with respect to d, the respective optimum values of d are given as:

$$\phi_{opt}(d) = -\frac{\rho_{YX} C_Y}{C_X} = C_1 \text{ (say)} \quad (3.7)$$

and

$$\phi^*_{opt}(d) = -\frac{\bar{X}}{\bar{Y}}\left(\frac{f_1 S_{YX} + \psi S_{YX(2)}}{f_1 S_X^2 + \psi S^2_{X(2)}}\right) = C_2 \text{ (say)} \quad (3.8)$$

It is visible in equations (3.7) and (3.8) that the optimum choices $\phi_{opt}(d)$ and $\phi^*_{opt}(d)$ are the functions of stigmatized variable Y and auxiliary variable X in terms of population parameters. For practical applications the corresponding parameters may be estimated by their respective sample estimates. Substituting the optimum values $\phi_{opt}(d)$ and $\phi^*_{opt}(d)$ in equations (3.5) and (3.6) respectively, the minimum mean square errors of families of estimators $T_1(d)$ and $T_2(d)$ are obtained as

$$MSE^*\{T_1(d)\} = \bar{Y}^2\ f_1(1 - \rho^2_{YX}) C_Y^2 + \Psi\ S^2_{Y_{(2)}}$$

$$+ \frac{f_h N_2}{nN}\left(\frac{f^2_{u_1}\Psi_{2,y} + \sigma^2_{u_2} + 2\sigma_{u_1 u_2}\bar{Y}_2}{\mu^2_{u_1}}\right) \quad (3.9)$$

and

$$MSE^{*}\ \{T_2(d)\}=\left[\left(f_1S_Y^2+\psi S_{Y(2)}^2\right)-\frac{\left(f_1S_{YX}+\psi S_{YX(2)}\right)^2}{f_1S_X^2+\psi S_{X(2)}^2}\right]+\frac{f_hN_2}{nN}\left(\frac{\sigma_{u_1}^2\mu_{2,y}+\sigma_{u_2}^2+2\sigma_{u_1u_2}\bar{Y}_2}{\mu_{u_1}^2}\right) \quad (3.10)$$

***Remark 3.1*:** Equations (3.7) and (3.8) are cubic equations in d, for any known values of C_1 and C_2 at most there will be three values of d^* for which $MSE\{T_1(d)\}$ and $MSE\{T_2(d)\}$ attain their minimum. Therefore, a criterion needs to be set up for an appropriate choice of d^* among three possible values. Hence the criterion is set out as, we will choose that value of d^* which makes the absolute bias of the estimator smallest.

3.3 Properties of Proposed Families of Estimators Under Different Randomized Response Models

In this section, the properties of proposed families of estimators $T_1(d)$ and $T_2(d)$ have been studied under different randomized response models M_1, M_2, M_3 and M_4 to get the ideas about their efficiencies and level of privacy protection. The mean square errors of the proposed families of estimators have been obtained under said models and shown in Tables 3-4. The optimum mean square errors of the proposed families of estimators have been also obtained under said models and shown in Tables 5-6.

Table 3. MSE of the family of estimators $T_1(d)$ under different models

Model	$MSE\ \{T_{1M_j}(d)\}\ j=1,2,3,4.$
M_1: Additive Model	$\bar{Y}^2 f_1(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_xC_Y)+\Psi\ S_{Y_{(2)}}^2+\frac{f_kN_2}{nN}\sigma_{v2}^2$
M_2: Multiplicative model	$\bar{Y}^2 f_1(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_xC_Y)+\Psi\ S_{Y_{(2)}}^2+\frac{f_hN_2}{nN}\left(\frac{\sigma_{v1}^2\mu_{2,Y}}{\mu_{v1}^2}\right)$
M_3: Mixed model 1	$\bar{Y}^2 f_1\left(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_XC_Y\right)+\psi S_{y_{(2)}}^2+\frac{f_hN_2}{nN}\left(\frac{\sigma_{v_1}^2\mu_{2,y}+\mu_{v_1}^2\sigma_{v_2}^2+2\sigma_{v_1}^2\mu_{v_2}\bar{Y}_2+\sigma_{v_1}^2(\mu_{v_2}^2+\sigma_{v_2}^2)}{\mu_{v_1}^2}\right)$
M_4: Mixed model 2	$\bar{Y}^2 f_1(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_xC_Y)+\Psi\ S_{Y_{(2)}}^2+\frac{f_hN_2}{nN}\left(\frac{\alpha^2\{\sigma_{v1}^2(\mu_{v2}^2+\sigma_{v2}^2)+\mu_{v1}^2\sigma_{v2}^2\}+(1-\alpha)^2\sigma_{v1}^2\mu_{2y}+2\alpha(1-\alpha)\bar{Y}_2\sigma_{v1}^2\mu_{v2}}{(1-\alpha)^2\mu_{v1}^2}\right)$

Table 4. MSE of the family of estimators $T_2(d)$ under different models

Model	$MSE\ \{T_{2M_j}(d)\}\ j=1,2,3,4.$
M_1: Additive Model	$\bar{Y}^2\left[f_1\left(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_XC_Y\right)+\psi\left(\frac{S_{Y(2)}^2}{\bar{Y}^2}+\phi^2(d)\frac{S_{X(2)}^2}{\bar{X}^2}+2\phi(d)\frac{S_{YX(2)}}{\bar{Y}\bar{X}}\right)\right]+\frac{f_hN_2}{nN}\sigma_{v2}^2$

M_2: Multiplicative model	$\bar{Y}^2\left[f_1\left(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_XC_Y\right)+\ \psi\ \left(\frac{S^2_{Y(2)}}{\bar{Y}^2}+\phi^2(d)\frac{S^2_{X(2)}}{\bar{X}^2}+2\phi(d)\frac{S_{YX(2)}}{\bar{Y}\ \bar{X}}\right)\right]+\frac{f_hN_2}{nN}\left(\frac{\sigma^2_{v_1}\mu_{2,y}}{\mu^2_{v_1}}\right)$
M_3: Mixed model 1	$\bar{Y}^2\left[f_1\left(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_XC_Y\right)+\ \psi\ \left(\frac{S^2_{Y(2)}}{\bar{Y}^2}+\phi^2(d)\frac{S^2_{X(2)}}{\bar{X}^2}+2\phi(d)\frac{S_{YX(2)}}{\bar{Y}\ \bar{X}}\right)\right]$ $+\frac{f_hN_2}{nN}\left(\frac{\sigma^2_{v_1}\mu_{2,y}+\mu^2_{v_1}\sigma^2_{v_2}+2\sigma^2_{v_1}\mu_{v_2}\bar{Y}_2+\sigma^2_{v_1}(\mu^2_{v_2}+\sigma^2_{v_2})}{\mu^2_{v_1}}\right)$
M_4: Mixed model 2	$\bar{Y}^2\left[f_1\left(C_Y^2+\phi^2(d)C_X^2+2\phi(d)\rho_{YX}C_XC_Y\right)+\ \psi\ \left(\frac{S^2_{Y(2)}}{\bar{Y}^2}+\phi^2(d)\frac{S^2_{X(2)}}{\bar{X}^2}+2\phi(d)\frac{S_{YX(2)}}{\bar{Y}\ \bar{X}}\right)\right]$ $+\frac{f_hN_2}{nN}\left(\frac{\alpha^2\left\{\sigma^2_{v_1}\left(\mu^2_{v_2}+\sigma^2_{v_2}\right)+\mu^2_{v_1}\sigma^2_{v_2}\right\}+(1-\alpha)^2\sigma^2_{v_1}\mu_{2,y}+2\alpha(1-\alpha)\bar{Y}_2\sigma^2_{v_1}\mu_{v_2}}{(1-\alpha)^2\mu^2_{v_1}}\right)$

Table 5. Optimum MSE of the family of estimators $T_1(d)$ under different models

Model	$MSE^*\left\{T_{1M_j}(d)\right\}\ \ j=1,2,3,4.$
M_1: Additive Model	$\bar{Y}^2\ f_1(1-\rho^2_{YX})C_Y^2+\Psi S^2_{Y(2)}+\frac{f_hN_2}{nN}\sigma^2_{v2}$
M_2: Multiplicative model	$\bar{Y}^2 f_1\left(1\text{-}\rho^2_{YX}\right)C_Y^2+\Psi S^2_{Y(2)}+\frac{f_hN_2}{nN}\left(\frac{\sigma^2_{v_1}\ \mu_{2,y}}{\mu^2_{v_1}}\right)$
M_3: Mixed model 1	$\bar{Y}^2 f_1\left(1\text{-}\rho^2_{YX}\right)C_Y^2+\psi S^2_{Y(2)}+\frac{f_hN_2}{nN}\left(\frac{\sigma^2_{v_1}\mu_{2,y}+\mu^2_{v_1}\sigma^2_{v_2}+2\sigma^2_{v_1}\mu_{v_2}\bar{Y}_2+\sigma^2_{v_1}(\mu^2_{v_2}+\sigma^2_{v_2})}{\mu^2_{v_1}}\right)$
M_4: Mixed model 2	$\bar{Y}^2 f_1\left(1\text{-}\rho^2_{YX}\right)C_Y^2+\Psi S^2_{Y(2)}$ $+\frac{f_hN_2}{nN}\left(\frac{\alpha^2\{\sigma^2_{v1}(\mu^2_{v2}+\sigma^2_{v2})+\mu^2_{v1}\sigma^2_{v2}\}+(1-\alpha)^2\sigma^2_{v1}\mu_{2y}+2\alpha(1-\alpha)\bar{Y}_2\sigma^2_{v1}\mu_{v2}}{(1-\alpha)^2\mu^2_{v1}}\right)$

Table 6. Optimum MSE of the family of estimators $T_2(d)$ under different models

Model	$MSE^*\left\{T_{2M_j}(d)\right\}\ \ j=1,2,3,4.$
M_1: Additive Model	$\left[\left(f_1S_Y^2+\psi S^2_{Y(2)}\right)\text{-}\frac{\left(f_1S_{YX}+\psi S_{YX(2)}\right)^2}{f_1S_X^2+\psi S^2_{X(2)}}\right]+\frac{f_hN_2}{nN}\sigma^2_{v2}$

M_2: Multiplicative model	$\left[\left(f_1S_Y^2+\psi S_{Y(2)}^2\right)-\dfrac{\left(f_1S_{YX}+\psi S_{YX(2)}\right)^2}{f_1S_X^2+\psi S_{X(2)}^2}\right]+\dfrac{f_hN_2}{nN}\left(\dfrac{\sigma_{v_1}^2\mu_{2,y}}{\mu_{v_1}^2}\right)$
M_3: Mixed model 1	$\left[\left(f_1S_Y^2+\psi S_{Y(2)}^2\right)-\dfrac{\left(f_1S_{YX}+\psi S_{YX(2)}\right)^2}{f_1S_X^2+\psi S_{X(2)}^2}\right]$ $+\dfrac{f_hN_2}{nN}\left(\dfrac{\sigma_{v_1}^2\mu_{2,y}+\mu_{v_1}^2\sigma_{v_2}^2+2\sigma_{v_1}^2\mu_{v_2}\overline{Y}_2+\sigma_{v_1}^2\left(\mu_{v_2}^2+\sigma_{v_2}^2\right)}{\mu_{v_1}^2}\right)$
M_4: Mixed model 2	$\left[\left(f_1S_Y^2+\psi S_{Y(2)}^2\right)-\dfrac{\left(f_1S_{YX}+\psi S_{YX(2)}\right)^2}{f_1S_X^2+\psi S_{X(2)}^2}\right]$ $+\dfrac{f_hN_2}{nN}\left(\dfrac{\alpha^2\left\{\sigma_{v_1}^2\left(\mu_{v_2}^2+\sigma_{v_2}^2\right)+\mu_{v_1}^2\sigma_{v_2}^2\right\}+(1-\alpha)^2\sigma_{v_1}^2\mu_{2,y}+2\alpha(1-\alpha)\overline{Y}_2\sigma_{v_1}^2\mu_{v_2}}{(1-\alpha)^2\mu_{v_1}^2}\right)$

4. EFFICIENCY COMPARISON AND PRIVACY PROTECTION MEASURE

The performance of the estimators are assessed based on their efficiencies with respect to other estimators when characteristic under study is non-sensitive in nature, but efficiencies of the estimators are not enough to determine the performances of the estimators when characteristic under the study is sensitive (stigmatize) in nature. The confidentiality protection of the respondent is another important issue in this case, because the number of respondents in the survey will increase when the level of protection of confidentiality of respondent increases.

In practice the respondents are more concerned with high confidentiality about their true response and the researchers generally are more interested in producing more precise estimates of population parameters. In randomized response mechanism, confidentiality of respondent's response is described in terms of measure of privacy protection. An index which is used to measure privacy should indicate how closely the original values of the perturbed sensitive variable Y can be estimated. The closer these values, the higher the privacy disclosure. The efficiencies of the estimators are described in terms of their variances or mean square errors. Scramble randomized devices which ensure high performance in terms of efficiency are usually less protective of confidentiality. The privacy and efficiency are inversely proportional to each other. Therefore, a valid comparison of estimator with other estimators under a randomized response mechanism or the comparison of the estimator under different randomized response models should necessarily take into account both efficiency and privacy protection into considerations.

In survey literatures, different methods are deliberated to measure the confidentiality. Some of them model-based, depending on the full distribution of the sensitive variable and others are design-based independent of the distributions. In a framework of design-based approach, Dinna and Perri (2011) and Zhimin *et al.* (2010) considered the square of the correlation coefficient ρ_{yz}^2 as a measure to compare different randomized response models in terms of privacy protection. Diana and Perri (2010) and Diana *et al.* (2013 a) suggested the multiple correlation coefficient as a normalized privacy protection measure for the case of simple random sampling. For the proposed families of factor type estimators $T_1(d)$ and $T_2(d)$, we have considered the normalized privacy protection measure as discussed by above authors and shown as

$$\tau = 1-\rho_{y.zx}^2 = 1-\frac{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{zx}}{1-\rho_{zx}^2}$$

It is observed that as the values of τ becomes higher, more the privacy is protected and greater cooperation from respondents is expected while for lesser values of τ less cooperation is expected from the respondents.

5. NUMERICAL ILLUSTRATION

Empirical studies have been performed to get the ideas about possible trade-off between efficiency and privacy protection using Monte Carlo simulation method. For empirical studies, we have considered a population of size N = 100,000 units with 40% weight of missing values. Simple random sampling without replacement (SRSWOR) scheme is considered to draw a sample of size n = 500 units. Auxiliary variable X is generated using gamma probability model with parameters a = 2.2 and b = 3.5. We considered the study variable Y is positively correlated to the auxiliary variable X and has relation $y_i = Rx_i + \varepsilon x_i^g$; where $\varepsilon \sim N(0,1)$, R = 2.0 and g = 1.5. Two scrambled random variables U_1 and U_2 are generated independently from uniform probability model $U(0,1)$.

The behaviors of the families of estimators $\bar{y}_F(d)$, $\bar{y}_F^*(d)$, $\bar{y}_F^{**}(d)$, $T_1(d)$ and $T_2(d)$ are explored for different values of d and f_h in terms of their mean square errors under different scramble response models M_1, M_2, M_3 and M_4. Here $\bar{y}_F(d)$ is the Singh and Shukla (1987) family of factor type estimators under complete response situation. The estimator $\bar{y}_F^*(d)$ is the family of factor type estimators when non-response occurs only in study variable Y. The estimator $\bar{y}_F^{**}(d)$ be the family of factor type estimators when non-response occurs in study variable Y as well as in auxiliary variable X. The mean square errors of the families of estimators $\bar{y}_F^*(d)$ and $\bar{y}_F^{**}(d)$ up to the first order of approximations are given as

$$MSE\left(\bar{y}_F^*(d)\right) = \bar{Y}^2 f_1\left(C_X^2+\phi^2(d)C_Y^2 + 2\phi(d)\rho_{YX}C_XC_Y\right) + \psi S_{Y(2)}^2 \quad (5.1)$$

and

$$MSE\left(\bar{y}_F^{**}(d)\right) = \bar{Y}^2\left[f_1\left(C_X^2+\phi^2(d)C_Y^2+2\phi(d)\rho_{YX}C_XC_Y\right) + \psi\left(\frac{S_{Y(2)}^2}{\bar{Y}^2}+\frac{S_{X(2)}^2}{\bar{X}^2}+2\phi(d)\frac{S_{YX(2)}}{\bar{Y}\,\bar{X}}\right)\right] \quad (5.2)$$

The measures of privacy protection τ have been calculated for above foresaid randomized response models and shown in Table-14.

For different choices of weights of missing values such as 10% 20%, 30%, 40% etc., it could be seen that the behaviors of the proposed families of estimators remain same. Hence, for convenience the results are displayed only for 40% weights of missing values. The bold values in Tables 7–11 represent the mean square errors of the proposed families of estimators under non-response with randomized response mechanism. The MSE's of the set of estimators ($\bar{y}_F(d)$, $\bar{y}_F^*(d)$, $T_1(d)$, $\bar{y}_F^{**}(d)$, $T_2(d)$) are calculated and shown respectively in the Table 7 when scramble response for sensitive characteristic is observed and linear randomized response model is considered as it is shown in equation (2.2).

Table 7. MSE's × 1000 of the estimators under model (2.2)

d	f_h	2	3	4	5
1	$MSE\{\bar{y}_F(1)\}$	1.3667	1.3667	1.3667	1.3667
	$MSE\{\bar{y}_F^*(1)\}$	2.4925	3.6182	4.7439	5.8696
	$MSE\{T_1(1)\}$	**10.3422**	**15.3927**	**20.4433**	**25.4938**
	$MSE\{\bar{y}_F^{**}(1)\}$	1.9227	2.4786	3.0345	3.5905
	$MSE\{T_2(1)\}$	**9.7724**	**14.2531**	**18.7339**	**23.2147**

2	$MSE\{\bar{y}_F(2)\}$	7.0640	7.0640	7.0640	7.0640
	$MSE\{\bar{y}_F^*(2)\}$	8.1897	9.3154	10.4412	11.5669
	$MSE\{T_1(2)\}$	**16.0394**	**21.0899**	**26.1405**	**31.1911**
	$MSE\{\bar{y}_F^{**}(2)\}$	9.9033	12.7425	15.5818	18.4211
	$MSE\{T_2(2)\}$	**17.7529**	**24.5171**	**31.2812**	**38.0453**
3	$MSE\{\bar{y}_F(3)\}$	2.7730	2.7730	2.7730	2.7730
	$MSE\{\bar{y}_F^*(3)\}$	3.8988	5.0245	6.1502	7.2760
	$MSE\{T_1(3)\}$	**11.7485**	**16.799**	**21.8496**	**26.9001**
	$MSE\{\bar{y}_F^{**}(3)\}$	3.8931	5.0131	6.1331	7.2531
	$MSE\{T_2(3)\}$	**11.7427**	**16.7876**	**21.8324**	**26.8773**
4	$MSE\{\bar{y}_F(4)\}$	2.7873	2.7873	2.7873	2.7873
	$MSE\{\bar{y}_F^*(4)\}$	3.9131	5.0388	6.1645	7.2902
	$MSE\{T_1(4)\}$	**11.7627**	**16.8133**	**21.8639**	**26.9144**
	$MSE\{\bar{y}_F^{**}(4)\}$	3.9131	5.0388	6.1645	7.2902
	$MSE\{T_2(4)\}$	**11.7627**	**16.8133**	**21.8639**	**26.9144**

5.1 Numerical Illustrations Under Different Randomized Response Models

In this section, the performances of proposed families of estimators are assessed for different choices of d and f_h under scramble response models M_1, M_2, M_3, and M_4. The mean square errors of set of families of estimators

$$(\bar{y}_F(d),\ \bar{y}_F^*(d),\ T_1(d),\ \bar{y}_F^{**}(d), T_2(d))$$

are calculated when scramble response is observed for sensitive characteristic under

1. additive randomized response model M_1 and shown in Table 8,
2. multiplicative randomized response model M_2 and shown in Table 9,
3. mixed randomized response model M_3 and shown in Table 10, and
4. mixed randomized response model M_4 and shown in Table 11.

The mixed randomized response model M_4 and its algebraic calculations involve the constant, $0 \le \alpha \le 1$, therefore, the mean square errors of set of families of estimators

$$(\bar{y}_F(d),\ \bar{y}_F^*(d),\ T_1(d),\ \bar{y}_F^{**}(d),\ T_2(d))$$

are calculated for different choices of α, d and f_h under this model. For α = 1, the mean square error does not exist under the model M_4, therefore, the privacy protection measure for this model has not been considered at the set choice of α. The behaviors of the families of estimators remain same for different values of f_h, hence, for convenience the results are displayed only for f_h = 2.

The optimum mean square errors of the proposed families of estimators $T_1(d)$ and $T_2(d)$ are also calculated under the aforesaid models and shown in Tables 12-13.

Table 8. MSE's × 1000 of the estimators under model M_1

d	f_h	**2**	**3**	**4**	**5**
1	$MSE\{\bar{y}_F(1)\}$	1.3943	1.3943	1.3943	1.3943
	$MSE\{\bar{y}_F^*(1)\}$	2.5385	3.6828	4.827	5.9712
	$MSE\{T_1(1)\}$	**2.6718**	**3.8827**	**5.0936**	**6.3045**
	$MSE\{\bar{y}_F^{**}(1)\}$	1.9642	2.5342	3.1041	3.6741
	$MSE\{T_2(1)\}$	**2.0975**	**2.7341**	**3.3707**	**4.0073**
2	$MSE\{\bar{y}_F(2)\}$	7.1552	7.1552	7.1552	7.1552
	$MSE\{\bar{y}_F^*(2)\}$	8.2995	9.4437	10.588	11.7322
	$MSE\{T_1(2)\}$	**8.4328**	**9.6436**	**10.8545**	**12.0654**
	$MSE\{\bar{y}_F^{**}(2)\}$	10.0294	12.9035	15.7776	18.6517
	$MSE\{T_2(2)\}$	**10.1627**	**13.1034**	**16.0442**	**18.985**

3	$MSE\{\bar{y}_F(3)\}$	2.8220	2.8220	2.8220	2.8220
	$MSE\{\bar{y}_F^*(3)\}$	3.9662	5.1104	6.2547	7.3989
	$MSE\{T_1(3)\}$	**4.0995**	**5.3104**	**6.5213**	**7.7321**
	$MSE\{\bar{y}_F^{**}(3)\}$	3.9604	5.0989	6.2374	7.3758
	$MSE\{T_2(3)\}$	**4.0937**	**5.2988**	**6.5039**	**7.7090**
4	$MSE\{\bar{y}_F(4)\}$	2.8364	2.8364	2.8364	2.8364
	$MSE\{\bar{y}_F^*(4)\}$	3.9806	5.1249	6.2691	7.4134
	$MSE\{T_1(4)\}$	**4.1139**	**5.3248**	**6.5357**	**7.7466**
	$MSE\{\bar{y}_F^{**}(4)\}$	3.9806	5.1249	6.2691	7.4134
	$MSE\{T_2(4)\}$	**4.1139**	**5.3248**	**6.5357**	**7.7466**

Table 9. MSE's $\times$ 1000 of the estimators under model M_2

d	f_h	2	3	4	5
1	$MSE\{\bar{y}_F(1)\}$	1.3567	1.3567	1.3567	1.3567
	$MSE\{\bar{y}_F^*(1)\}$	2.4508	3.5449	4.6391	5.7332
	$MSE\{T_1(1)\}$	**4.0272**	**5.9096**	**7.7919**	**9.6742**
	$MSE\{\bar{y}_F^{**}(1)\}$	1.8744	2.3922	2.9100	3.4277
	$MSE\{T_2(1)\}$	**3.4509**	**4.7568**	**6.0628**	**7.3688**
2	$MSE\{\bar{y}_F(2)\}$	7.0609	7.0609	7.0609	7.0609
	$MSE\{\bar{y}_F^*(2)\}$	8.1551	9.2492	10.3433	11.4375
	$MSE\{T_1(2)\}$	**9.7315**	**11.6138**	**13.4962**	**15.3785**
	$MSE\{\bar{y}_F^{**}(2)\}$	9.854	12.6471	15.4402	18.2332
	$MSE\{T_2(2)\}$	**11.4304**	**15.0117**	**18.5930**	**22.1743**

3	$MSE\{\bar{y}_F(3)\}$	2.7735	2.7735	2.7735	2.7735
	$MSE\{\bar{y}_F^*(3)\}$	3.8676	4.9618	6.0559	7.1500
	$MSE\{T_1(3)\}$	**5.4441**	**7.3264**	**9.2087**	**11.0911**
	$MSE\{\bar{y}_F^{**}(3)\}$	3.8619	4.9504	6.0388	7.1272
	$MSE\{T_2(3)\}$	**5.4383**	**7.315**	**9.1916**	**11.0683**
4	$MSE\{\bar{y}_F(4)\}$	2.7878	2.7878	2.7878	2.7878
	$MSE\{\bar{y}_F^*(4)\}$	3.8819	4.9761	6.0702	7.1643
	$MSE\{T_1(4)\}$	**5.4583**	**7.3407**	**9.223**	**11.1054**
	$MSE\{\bar{y}_F^{**}(4)\}$	3.8819	4.9761	6.0702	7.1643
	$MSE\{T_2(4)\}$	**5.4583**	**7.3407**	**9.223**	**11.1054**

Table 10. MSE's × 1000 of the estimators under model M_3

d	***fh***	**2**	**3**	**4**	**5**
1	$MSE\{\bar{y}_F(1)\}$	1.3994	1.3994	1.3994	1.3994
	$MSE\{\bar{y}_F^*(1)\}$	2.5308	3.6622	4.7936	5.925
	$MSE\{T_1(1)\}$	**5.1059**	**7.5248**	**9.9438**	**12.3627**
	$MSE\{\bar{y}_F^{**}(1)\}$	1.9632	2.5271	3.091	3.6548
	$MSE\{T_2(1)\}$	**4.5383**	**6.3897**	**8.2412**	**10.0926**
2	$MSE\{\bar{y}_F(2)\}$	7.1662	7.1662	7.1662	7.1662
	$MSE\{\bar{y}_F^*(2)\}$	8.2976	9.429	10.5604	11.6919
	$MSE\{T_1(2)\}$	**10.8727**	**13.2917**	**15.7106**	**18.1296**
	$MSE\{\bar{y}_F^{**}(2)\}$	10.0215	12.8767	15.732	18.5872
	$MSE\{T_2(2)\}$	**12.5966**	**16.7394**	**20.8822**	**25.025**
3	$MSE\{\bar{y}_F(3)\}$	2.8192	2.8192	2.8192	2.8192
	$MSE\{\bar{y}_F^*(3)\}$	3.9506	5.082	6.2135	7.3449
	$MSE\{T_1(3)\}$	**6.5257**	**8.9447**	**11.3637**	**13.7826**
	$MSE\{\bar{y}_F^{**}(3)\}$	3.9449	5.0706	6.1962	7.3219
	$MSE\{T_2(3)\}$	**6.5200**	**8.9332**	**11.3464**	**13.7596**

4	$MSE\{\bar{y}_F(4)\}$	2.8337	2.8337	2.8337	2.8337
	$MSE\{\bar{y}_F^*(4)\}$	3.9651	5.0965	6.2279	7.3593
	$MSE\{T_1(4)\}$	**6.5402**	**8.9591**	**11.3781**	**13.7971**
	$MSE\{\bar{y}_F^{**}(4)\}$	3.9651	5.0965	6.2279	7.3593
	$MSE\{T_2(4)\}$	**6.5402**	**8.9591**	**11.3781**	**13.7971**

Table 11. MSE's × 1000 of the estimators under model M_4 for f_h =2

d	α	0	0.2	0.4	0.6	0.8
1	$MSE\{\bar{y}_F(1)\}$	1.3693	1.3693	1.3693	1.3693	1.3693
	$MSE\{\bar{y}_F^*(1)\}$	2.459	2.4836	2.4670	2.4936	2.4527
	$MSE\{T_1(1)\}$	**2.5904**	**2.8044**	**3.1865**	**4.3397**	**10.2462**
	$MSE\{\bar{y}_F^{**}(1)\}$	1.9246	1.9348	1.919	1.9337	1.913
	$MSE\{T_2(1)\}$	**2.0559**	**2.2556**	**2.6384**	**3.7798**	**9.7065**
2	$MSE\{\bar{y}_F(2)\}$	6.9137	6.9137	6.9137	6.9137	6.9137
	$MSE\{\bar{y}_F^*(2)\}$	8.0035	8.028	8.0114	8.0381	7.9971
	$MSE\{T_1(2)\}$	**8.1348**	**8.3488**	**8.7309**	**9.8842**	**15.7906**
	$MSE\{\bar{y}_F^{**}(2)\}$	9.6662	9.7142	9.6790	9.7281	9.6646
	$MSE\{T_2(2)\}$	**9.7975**	**10.035**	**10.3985**	**11.5742**	**17.458**
3	$MSE\{\bar{y}_F(3)\}$	2.7227	2.7227	2.7227	2.7227	2.7227
	$MSE\{\bar{y}_F^*(3)\}$	3.8125	3.837	3.8205	3.8471	3.8062
	$MSE\{T_1(3)\}$	**3.9438**	**4.1578**	**4.5400**	**5.6932**	**11.5997**
	$MSE\{\bar{y}_F^{**}(3)\}$	3.8070	3.8314	3.8149	3.8415	3.8006
	$MSE\{T_2(3)\}$	**3.9383**	**4.1522**	**4.5344**	**5.6876**	**11.5941**
4	$MSE\{\bar{y}_F(4)\}$	2.7366	2.7366	2.7366	2.7366	2.7366
	$MSE\{\bar{y}_F^*(4)\}$	3.8263	3.8509	3.8343	3.8610	3.8201
	$MSE\{T_1(4)\}$	**3.9577**	**4.1717**	**4.553853**	**5.7071**	**11.6135**
	$MSE\{\bar{y}_F^{**}(4)\}$	3.8264	3.8509	3.8344	3.861	3.8201
	$MSE\{T_2(4)\}$	**3.9577**	**4.1717**	**4.5539**	**5.7071**	**11.6135**

Table 12. Optimum MSE's × 1000 of the proposed families of estimators T_1(d) and T_2(d) under different models

Model	f_h	**2**	**3**	**4**	**5**
Model (2.2)	$MSE^*\{T_1(d)\}$	10.5716	15.7347	20.8977	26.0608
	$MSE^*\{T_2(d)\}$	9.9791	14.5496	19.1200	23.6905
M_1	$MSE^*\{T_1(d)\}$	2.6552	3.8589	5.0625	6.2662
	$MSE^*\{T_2(d)\}$	2.0833	2.7150	3.3467	3.9783
M_2	$MSE^*\{T_1(d)\}$	4.1595	6.1405	8.1215	10.1026
	$MSE^*\{T_2(d)\}$	3.5431	4.9078	6.2724	7.6370
M_3	$MSE^*\{T_1(d)\}$	5.0536	7.4736	9.8937	12.3138
	$MSE^*\{T_2(d)\}$	4.4786	6.3236	8.1687	10.0137

Table 13. Optimum MSE's × 1000 of the proposed family of estimators $T_1(d)$ and $T_2(d)$ Under Model M_4

α	f_h	2	3	4	5
0.0	$MSE^*\{T_1(d)\}$	2.6605	3.8730	5.0856	6.2982
	$MSE^*\{T_2(d)\}$	2.0776	2.7072	3.3369	3.9666
0.2	$MSE^*\{T_1(d)\}$	2.8305	4.1205	5.4106	6.7006
	$MSE^*\{T_2(d)\}$	2.2440	2.9476	3.6512	4.3547
0.4	$MSE^*\{T_1(d)\}$	3.2565	4.7731	6.2897	7.8063
	$MSE^*\{T_2(d)\}$	2.6606	3.5813	4.5020	5.4227
0.6	$MSE^*\{T_1(d)\}$	4.3572	6.4180	8.4789	10.5398
	$MSE^*\{T_2(d)\}$	3.7717	5.2471	6.7225	8.1979
0.8	$MSE^*\{T_1(d)\}$	10.3022	15.3404	20.3787	25.4169
	$MSE^*\{T_2(d)\}$	9.6861	14.1082	18.5302	22.9522

Table 14. τ 's for different models

τ	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\tau(M_1)$	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
$\tau(M_2)$	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286	0.286
$\tau(M_3)$	**0.349**	0.349	0.349	0.349	0.349	0.349	0.349	0.349	0.349	0.349
$\tau(M_4)$	0.289	0.295	0.302	0.312	0.325	0.343	0.367	0.399	0.437	**0.473**

6. INTERPRETATIONS OF RESULTS

From Tables 7 – 10, it is observed that, (i) The family of estimators $\bar{y}_F(d)$ in absence of non-response has the lowest mean square error for all values of d and f_h. Under non-response without randomized mechanism, the families of estimators $\bar{y}_F^*(d)$ and $\bar{y}_F^{**}(d)$ have greater mean square errors than $\bar{y}_F(d)$ but lesser than the families of estimators $T_1(d)$ and $T_2(d)$ proposed under randomized response mechanism. These results indicate that the proposed families of estimators $T_1(d)$ and $T_2(d)$ have the high values of mean square errors under non-response with randomized response mechanism.

(ii) The mean square errors of the proposed families of estimators $T_1(d)$ and $T_2(d)$ increase for fixed values of d with the increase in the values of f_h.

(iii) Since the study variable is positively correlated with auxiliary variable, therefore, for the fixed values of f_h, the minimum values of mean square errors of the proposed families of estimators $T_1(d)$ and $T_2(d)$ have been observed when one chooses $d = 1$ which is the case of ratio estimator.

From Tables 8–10, it has been also observed that the mean square errors for proposed families of estimators in presence of non-response with randomized response mechanism are in increasing order for the fixed values of f_h and d with respect to randomized response models M_1, M_2 and M_3 respectively. In other words, the proposed families of estimators have maximum mean square errors under model M_3 whereas they have minimum mean square errors under model M_1. Further, it may be noticed from Table 14 that $\tau(M_1)$ is close to zero which indicates that there is almost no privacy protection under model M_1. Among others $\tau(M_3)$ is higher, which indicates that privacy is well protected under model M_3. Hence, it may be concluded that in privacy protection perspective, the model M_3 is preferable over the models M_1 and M_2.

From Table 11, a similar interpretations may be given as discussed for Tables 7-10 for the variations in the values of α.

From Table 12, the minimum values of optimum mean square errors of the proposed families of estimators $T_1(d)$ and $T_2(d)$ have been observed under randomized response model M_1 while their maximum values are observed under model M_3 for fixed values of f_h. It is noticed that optimum mean squares of the proposed families of estimators are increasing with the increase in the values of f_h.

For Table 13, a similar interpretations may be given as discussed for Table 12, for the variations in the values of α.

The values of measures of privacy protection $\tau(M_1), \tau(M_2), \tau(M_3)$ and $\tau(M_4)$ have been calculated for the different scramble randomized response models M_1, M_2, M_3, and M_4 and presented in Table 14. It is observed that $\tau(M_1)$ has minimum value (0.052) which indicates the lesser privacy protection if one uses the model M_1 and $\tau(M_3)$ has the highest value (0.349) among models M_1, M_2 and M_3 which indicates that privacy is well protected if one is considering the model M_3. Hence, it may be concluded that the model M_3 is preferable over the models M_1 and M_2 in terms of privacy protection.

It is also observed that the values of $\tau(M_4)$ and the mean square errors of the proposed family of estimators $T_1(d)$ and $T_2(d)$ are increasing with the increase in the values of α. Therefore, the proposed families of estimators $T_1(d)$ and $T_2(d)$ are preferable for higher values of α when scramble response is considered under model M_4 in terms of privacy protection. The values of measure of privacy protection τ is observed maximum in the range $0.6 \le \alpha < 1$ under the model M_4, therefore, the proposed families of estimators $T_1(d)$ and $T_2(d)$ are preferable with model M_4 over other models if the values of α exceeds 0.6, otherwise, the proposed families of estimators are preferable with model M_3 over other models.

7. CONCLUSION AND RECOMMENDATIONS

In studies involving sensitive characteristics of a population, the privacy protection of respondents is an important issue for getting reliable responses in survey data. From this perspective, the proposed families of estimators $T_1(d)$ and $T_2(d)$ are giving encouraging results. Therefore, the proposed families of estimators may be recommended to the survey practitioners if they are planning to gather information on parameters related to stigmatized characters using survey sampling.

ACKNOWLEDGEMENTS

Authors are thankful to the Indian School of Mines, Dhanbad for providing financial and necessary infrastructural support to carry out the present research work.

REFERENCES

Bar-Lev, S.K., Bobovitch, E. and Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika*, **60(3)**, 255-260.

Christofides, T.C. (2003). A generalized randomized response technique. *Metrika*, 5**7(2)**, 195-200.

Diana, G. and Perri, P.F. (2009). Estimating a sensitive proportion through randomized response procedures based on auxiliary information. *Statistical Papers*, **50(3)**, 661-672.

Diana, G. and Perri, P.F. (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *J. Appl. Statist.*, **37(11)**, 1875-1890.

Diana, G. and Perri, P.F. (2011). A class of estimators for quantitative sensitive data. *Statistical Papers*, **52(3)**, 633-650.

Diana, G., Giordan, M. and Perri, P.F. (2013 a). Randomized response surveys: A note on some privacy protection measures. *Model Assist. Statist. Appl.*, **8(1)**, 19-28.

Diana, G., Riaz, S. and Shabbir, J. (2013 b). Hansen and Hurwitz estimator with scrambled response on the second call. *J. Appl. Statist.*, **41(3)**, 596-611.

Eichhorn, B.H. and Hayre, L.S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *J. Statist. Plann. Inf.*, **7(4)**, 307-316.

Eriksson, S.A. (1973). A new model for randomized response. *Inter. Statist.Rev./Revue Internationale de Statistique*, 101-113.

Fox, J.A., and Tracy, P.E. (1986). *Randomized Response: A Method for Sensitive Surveys.* SAGE Publications Inc.

Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R. and Horvitz, D.G. (1969). The unrelated question randomized response model: Theoretical framework. *J. Amer. Statist. Assoc.*, **64(326)**, 520-539.

Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.*, **41(236)**, 517-529.

Kuk, A.Y. (1990). Asking sensitive questions indirectly. *Biometrika*, **77(2)**, 436-438.

Mangat, N.S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, **77(2)**, 439-442.

Pollock, K.H. and Bek, Y. (1976). A comparison of three randomized response models for quantitative data. *J. Amer. Statist. Assoc.*, **71(356)**, 884-886.

Saha, A. (2007). A simple randomized response technique in complex surveys. *Metron*, **65**, 59-66.

Shabbir, J. and Gupta, S. (2005). On modified randomized device of Warner's model. *Pak. J. Statist.-All Series*, **21(1)**, 123-129.

Singh, V.K. and Shukla, D. (1987). One parameter family of factor-type ratio estimators. *Metron*, **45(1-2),** 273-283.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.*, **60(309)**, 63-69.

Yan, Z. (2005). Ratio method of estimation of population proportion using randomized response technique. *Model Assist. Statist. Appl.*, **1(2)**, 125-130.

Zhimin, H., Zaizai, Y. and Lidong, W. (2010). Combination of the additive and multiplicative models at the estimation stage. In : *2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*.