THREE-STAGE RANDOMIZED RESPONSE MODEL FOR ESTIMATING A RARE SENSITIVE ATTRIBUTE IN PROBABILITY PROPORTIONAL TO SIZE SAMPLING USING POISSON DISTRIBUTION

G. N. Singh, C. Singh[1], A. Kumar and S. Suman
Department of Mathematics and Computing
Indian Institute of Technology (Indian School of mines)
Dhanbad- 826004, India.

ABSTRACT
This paper has a great potential for estimating mean number of individuals in the population who possess a rare sensitive attributes using Poisson distribution for two situations: clustered population and stratified population with clusters are strata units. Properties of the proposed estimation procedures are deeply examined when the proportion of a rare unrelated non-sensitive attributes is assumed to be known as well as unknown. Empirical studies are accomplished to show the superiority of the resultant estimators over well-known contemporary estimators.

KEYWORDS: Randomized Response Model, Rare Sensitive Attribute, Rare Non-Sensitive Unrelated Attribute, Probability Proportional to Size (Pps) Sampling, Poisson Distribution.

MSC: 62D05

RESUMEN
Este artículo tiene un gran potencial para estimar el número medio de individuos en la población que posee atributos sensibles raros utilizando la distribución de Poisson para dos situaciones: población agrupada y la población estratificada con grupos son unidades de estratos. Las propiedades de los procedimientos de estimación propuestos se examinan en profundidad cuando se supone que la proporción de un atributo no relacionado, no sensible se supone que se conoce tanto como desconocido. Se realizan estudios empíricos para mostrar la superioridad de los estimadores resultantes sobre los estimadores contemporáneos conocidos.

PALABRAS CLAVE: Modelo de Respuestas Aleatorizadas, Atributos Raros No-Sensitivos No-relacionados, Muestreo de Probabilidades Proporcionales al Tamaño, Distribución de Poisson.

1. INTRODUCTION

In socio-economic surveys, when we deal with sensitive issues or confidential issues, such as questions related to regular payment of income tax, illegal use of drug, alcohol or smoking habit, suffering from any mental disorder etc., the actual answers of these questions are hidden or misguided by people participated in the surveys. Therefore, obtained such data are definitely open to error if surveys are conducted through classical methods which makes biased inference about study matter. In such practical cases to get rid of these kind of situations, Warner (1965) introduced an ingenious procedure to accumulate the information on stigmatized issues known as randomized response technique. This randomized response model requires the interviewee to give a “Yes” or “No” answers either to the sensitive question ‘A’ or to its compliment without revealing his actual status to the interviewer. To enhance the confidence and rectify the privacy of the respondents Greenberg et al. (1969) proposed unrelated question model or U model. The randomized response technique was further modified for different practical situations by Moors (1971), Cochran (1977), Fox and Tracy (1986), Chaudhuri and Mukherjee (1988), Mangat et al. (1992), Tracy and Osahan (1999), Singh et al. (2003) and Kim and Warde (2005) Kim and Elam (2007), Singh and Tarray (2012, 2015), Lee et al. (2016), Lee et al. (2018) and among others.
When study characteristics are sensitive in nature as well as rare in existence such as the number of persons suffering from Schizophrenia who are not taking medicine during treatment, the number of persons that attempted suicide twice, Land et al. (2012) provided the solution using Poisson distribution for estimating the mean number of persons following these type of attributes. Lee et al. (2014), Singh et al. (2019) used

[1] Corresponding author email: chandraketu.lko@gmail.com

probability proportional to size (pps) sampling scheme for the estimation of parameter of a rare sensitive attribute using Poisson distribution.
Singh et al. (2003) used the randomization device carrying three types of cards bearing statements: (i) "I belong to sensitive group A_1", (ii) "I belong to group A_2" and (iii) "Blank cards", with corresponding probabilities Q_1, Q_2 and Q_3 respectively, such that $\sum_{i=1}^{3} Q_i = 1$. In case the blank card is drawn by the respondent, he/she will report "no". The rest of the procedure remains as usual. The probability of "Yes" answer is given by $\theta_1 = Q_1\pi_1 + Q_2\pi_2$ where π_1 and π_2 are the true proportion of the rare sensitive attribute A_1 and the rare unrelated attribute A_2 in the population respectively.

From the above equation the estimator of π_1 is as $\hat{\pi}_1 = \frac{\hat{\theta}_1 - P_2\pi_2}{P_1}$ where $\hat{\theta}_1$ the observed proportion of "yes" answers in the sample. The variance of the estimator $\hat{\pi}_1$ is given as

$$V(\hat{\pi}_1) = \frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_1(1-P_1-2P_2\pi_2)}{nP_1} + \frac{P_2\pi_2[1-P_2\pi_2]}{nP_1^2}$$

Motivated with the above works, in this paper, we have made an attempt to extend Singh et al. (2003) unrelated randomized response model to three-stage unrelated randomized response procedure for estimating the mean number of individuals in the population who possess a rare sensitive attribute when the parameter of a rare non-sensitive unrelated attribute is known and unknown. The properties of the proposed estimators have been discussed when the parameter for the unrelated attribute is known and unknown. The present work is compared with the Lee et al. (2014) and Singh et al. (2019) works using numerical illustrations.

2. SAMPLING DESIGN AND NOTATIONS

Let $U = (U_1, U_2, ..., U_N)$ be the finite population of N clusters, which represent first-stage units, consists of $(M_1, M_2, ..., M_N)$ second-stage units. At the first-stage, we select a sample of n clusters using probability proportional to size sampling with replacement. At the second-stage, we select m_i $(i=1,2,...n)$, second-stage units from the i^{th} selected first-stage unit using simple random sampling with replacement. The following notations are used as

π_{i1} : The true proportion of the rare sensitive attribute A_1 in i^{th} cluster.

π_{i2} : The true proportion of the rare unrelated non-sensitive attribute A_2 in i^{th} cluster.

M_i : The size of the i^{th} cluster and $M_0 = \sum_{i=1}^{N} M_i$.

2.1 Estimation procedure when the rare non-sensitive unrelated attribute is known

Following afore mentioned two-stage sampling scheme and assuming the proportion of rare unrelated non-sensitive attribute is known, the responses from the elementary units in the second stage samples were collected using the extended three-stage Singh et al. (2003) unrelated randomized response procedure which consists the following statements for the i^{th} cluster:

First-stage randomization device R_1 consists of two statements

	Statements	Selection probability
Statement 1	Are you a member of a rare sensitive Group A_1 ?	T
Statement 2	Go to randomization device R_2	(1-T)

Second-stage randomization device R_2 consists of two statements

	Statements	Selection probability
Statement 1	Are you a member of a rare sensitive Group A_1 ?	P
Statement 2	Go to randomization device R_3	(1-P)

The randomization device R_3 used three statements which are same as Singh et al. (2003).
Following the above randomized response procedures, the probability θ_{i0} of "yes" answer in the i^{th} cluster is given by $\theta_{i0} = T_i\pi_{i1} + (1 - T_i)\left[P_i\pi_{i1} + (1 - P_i)\{Q_{1i}\pi_{i1} + Q_{2i}\pi_{i2}\}\right]$. Here π_{i2} is assumed to be known. Since A_1 and A_2 are rare attributes, hence for large m_i and small θ_{i0} $(\theta_{i0} \to 0)$, we have $m_i\theta_{i0} = \lambda_{i0} > 0$ is finite.
Let $y_{i1}, y_{i2}, ...,y_{im_i}$ be a random sample drawn from i^{th} cluster which follow Poisson distribution with mean λ_{i0}, so the estimator for the mean number of individuals with the rare sensitive characteristics, $\lambda_{i1}\left(\lambda_{i1} = m_i\pi_{i1}\right)$ is defined as

$$\hat{\lambda}_{i1} = \frac{1}{\left[T_i + (1 - T_i)P_i + (1 - T_i)(1 - P_i)Q_{1i}\right]}\left[\frac{1}{m_i}\sum_{i=1}^{m_i} y_{ij} - (1 - T_i)(1 - P_i)Q_{2i}\lambda_{i2}\right] \tag{1}$$

where $\lambda_{i2}\left(\lambda_{i2} = m_i\pi_{i2}\right)$ is the mean number of individuals who have the rare unrelated non-sensitive attribute in the i^{th} cluster. Hence in two-stage procedure, the estimator for the mean number of individuals with the rare sensitive attribute in the population is as follows:

$$\hat{\lambda}_{1ppzwr} = \frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1}}{p_i} = \frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i}{p_i}\left[\frac{1}{\left[T_i + (1 - T_i)P_i + (1 - T_i)(1 - P_i)Q_{1i}\right]}\left[\frac{1}{m_i}\sum_{i=1}^{m_i} y_{ij} - (1 - T_i)(1 - P_i)Q_{2i}\lambda_{i2}\right]\right] \tag{2}$$

Theorem 2.1.1 The estimator $\hat{\lambda}_{1ppzwr}$ is unbiased for the mean number of individuals (λ_1).

Proof: Since y_{ij} follows Poisson distribution with parameter $\lambda_{i0} = T_i\lambda_{i1} + (1 - T_i)\left[P_i\lambda_{i1} + (1 - P_i)\{Q_{1i}\lambda_{i1} + Q_{2i}\lambda_{i2}\}\right]$, therefore, $E\left(\hat{\lambda}_{1ppzwr}\right) = E_1E_2\left(\hat{\lambda}_{1ppzwr}\right)$, where E_1 and E_2 are the expectations over the first and second stage samples respectively.
Further, we have

$$E_1E_2\left(\hat{\lambda}_{1ppzwr}\right) = E_1E_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1}}{p_i}\right] = E_1\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i}{p_i}E_2\left(\hat{\lambda}_{i1}\right)\right]$$

Now,

$$E_2\left(\hat{\lambda}_{i1}\right) = \frac{1}{\left[T_i + (1 - T_i)P_i + (1 - T_i)(1 - P_i)Q_{1i}\right]}\left[\frac{1}{m_i}\sum_{i=1}^{m_i} E_2\left(y_{ij}\right) - (1 - T_i)(1 - P_i)Q_{2i}\lambda_{i2}\right]$$

Hence, finally we have

$$E_1E_2\left(\hat{\lambda}_{1ppzwr}\right) = E_1\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\lambda_{i1}}{p_i}\right] = \frac{1}{M_0}\sum_{i=1}^{N}p_i\frac{M_i\lambda_{i1}}{p_i} = \lambda_1$$

Hence, $\hat{\lambda}_{1ppzwr}$ is an unbiased estimator of λ_1.

Theorem 2.1.2 The variance of the unbiased estimator $\hat{\lambda}_{1ppzwr}$ is given as

$$V\left(\hat{\lambda}_{1ppzwr}\right) = \frac{1}{nM_0^2}\left[\sum_{i=1}^{N}p_i\left(\frac{M_i\lambda_{i1}}{p_i} - M_0\lambda_1\right)^2 + \sum_{i=1}^{N}\frac{M_i^{\,2}}{p_i}\frac{\varphi_i}{m_i}\right] \tag{3}$$

where $\varphi_i = \dfrac{\lambda_{i1}}{\left[T_i + (1 - T_i)P_i + (1 - T_i)(1 - P_i)Q_{1i}\right]} + \dfrac{(1 - T_i)(1 - P_i)Q_{2i}\lambda_{i2}}{n\left[T_i + (1 - T_i)P_i + (1 - T_i)(1 - P_i)Q_{1i}\right]^2}$

Proof: The variance of the estimator $\hat{\lambda}_{1ppzwr}$ is expressed as

$$V\left(\hat{\lambda}_{1ppzwr}\right) = V_1E_2\left(\hat{\lambda}_{1ppzwr}\right) + E_1V_2\left(\hat{\lambda}_{1ppzwr}\right) \tag{4}$$

where the variance of overall possible subsamples is denoted by V_2 and the variance of overall first-stage samples is denoted by V_1.
The first term in the expression of variance given in equation (4) is simplified as

$$V_1E_2\left(\hat{\lambda}_{1ppzwr}\right)=V_1E_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1}}{p_i}\right] = \frac{1}{nM_0^2}\sum_{i=1}^{N}p_i\left[\frac{M_i\lambda_{i1}}{p_i}-M_0\lambda_1\right]^2 \tag{5}$$

The second term is simplified as

$$E_1V_2\left(\hat{\lambda}_{1ppzwr}\right)=E_1V_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1}}{p_i}\right]=E_1\left[\frac{1}{\left(nM_0\right)^2}\sum_{i=1}^{n}\frac{M_i^{\ 2}}{p_i^2}\frac{1}{\left[T_i+\left(1-T_i\right)P_i+\left(1-T_i\right)\left(1-P_i\right)Q_{1i}\right]^2 m_i}\lambda_{i0}\right] = \frac{1}{nM_0^2}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\frac{\varphi_i}{m_i} \tag{6}$$

Thus, by the addition of the equation (5) and (6), we get the variance of the unbiased estimator $\hat{\lambda}_{1ppzwr}$ as given in equation (3).

Further an unbiased estimator of the variance of $\hat{\lambda}_{1ppzwr}$ is as follows

$$\hat{V}\left(\hat{\lambda}_{1ppzwr}\right)=\frac{1}{n\left(n-1\right)M_0^2}\sum_{i=1}^{n}\left(\frac{M_i\hat{\lambda}_{i1}}{p_i}-\hat{\lambda}_{1ppzwr}\right)^2 \tag{7}$$

Meanwhile, the size of the cluster is known and n clusters are chosen using probability proportional to size with replacement sampling scheme with probability $p_i=\frac{M_i}{M_0}$ for the ith cluster. Hence under the unequal probability sampling, the unbiased estimator of λ_1 given in equation (2) is written as:

$$\hat{\lambda}_{1ppswr}=\frac{1}{n}\sum_{i=1}^{n}\hat{\lambda}_{i1} \tag{8}$$

using $p_i=\frac{M_i}{M_0}$, from equation (3) and (7) the variance of the estimate $\hat{\lambda}_{1ppswr}$ and its estimate respectively are simplified as

$$V\left(\hat{\lambda}_{1ppswr}\right)=\frac{1}{nM_0}\left[\sum_{i=1}^{N}M_i\left(\lambda_{i1}-\lambda_1\right)^2+\sum_{i=1}^{N}\frac{M_i}{m_i}\varphi_i\right] \tag{9}$$

and

$$\hat{V}\left(\hat{\lambda}_{1ppswr}\right)=\frac{1}{n\left(n-1\right)}\sum_{i=1}^{n}\left[\hat{\lambda}_{i1}-\frac{\hat{\lambda}_{1ppswr}}{M_0}\right]^2$$

(10)

2.2 Estimation procedure when the proportion of rare non- sensitive unrelated attribute is unknown

In this section, we have estimated the mean number of persons in the population who are possessing a sensitive attribute when the proportion of rare non-sensitive unrelated attribute is unknown. The individuals selected in the sample are asked to answer "yes" or "no" using the following two randomization devices. The first randomization device is given as follows:

First-stage randomization device R_{11} consists of two statements

	Statements	Selection probability
Statement 1	Are you a member of a rare sensitive Group A_1 ?	T_{1i}
Statement 2	Go to randomization device R_{12}	$\left(1-T_{1i}\right)$

Second-stage randomization device R_{12} consists of two statements

	Statements	Selection Probability
Statement 1	Are you a member of a rare sensitive Group A_1 ?	P_{1i}
Statement 2	Go to randomization device R_{13}	$\left(1-P_{1i}\right)$

Third-stage randomization device R_{13} which uses three statements

Statements		Selection Probability
Statements 1	Are you a member of Group A_1 ?	Q_{1i}
Statements 2	Are you a member of a rare unrelated Group A_2 ?	Q_{2i}
Statements 3	Blank Cards	Q_{3i}

such that $\sum_{k=1}^{3} Q_{ki} = 1$. Next, the respondent is requested again to answer one of the same questions using second randomization device.

The second randomization device is given as follows:

First-stage randomization device R_{21} consists of two statements

Statements		Selection probability
Statements 1	Are you a member of a rare sensitive Group A_1 ?	T_{2i}
Statements 2	Go to randomization device R_{22}	$(1\text{-}T_{2i})$

Second-stage randomization device R_{22} consists of two statements

Statements		Selection probability
Statements 1	Are you a member of a rare sensitive Group A_1 ?	P_{2i}
Statements 2	Go to randomization device R_{23}	$(1\text{-}P_{2i})$

Third-stage randomization device R_{23} consists three statements

Statements		Selection probability
Statement 1	Are you a member of Group A_1 ?	Q_{4i}
Statement 2	Are you a member of a rare unrelated Group A_2 ?	Q_{5i}
Statement 3	Blank Cards	Q_{6i}

such that $\sum_{k=4}^{6} Q_{ki} = 1$. Based on the responses obtained through using two randomization devices, the probabilities of "yes" answer in the i[th] cluster are obtained as

$$\theta_{i1} = T_{1i}\pi_{i1} + (1 - T_{1i})\left[P_{1i}\pi_{i1} + (1 - P_{1i})\{Q_{1i}\pi_{i1} + Q_{2i}\pi_{i2}\}\right]$$

and

$$\theta_{i2} = T_{2i}\,\pi_{i1} + (1 - T_{2i})\left[P_{2i}\pi_{i1} + (1 - P_{2i})\{Q_{4i}\pi_{i1} + Q_{5i}\pi_{i2}\}\right]$$

Since A_1 and A_2 are the rare attributes in the population for the i[th] cluster, we have $m_i\theta_{i1} = \lambda_{i1}^*$ and $m_i\theta_{i2} = \lambda_{i2}^*$ are finite for large m_i and small θ_{i1} and θ_{i2}.

After computational procedures the following two equations are obtained as

$$\left[T_{1i} + (1 - T_{1i})P_{1i} + (1 - T_{1i})(1 - P_{1i})Q_{1i}\right]\hat{\lambda}_{i1u} + \left[(1 - T_{1i})(1 - P_{1i})Q_{2i}\right]\hat{\lambda}_{i2u} = \frac{1}{m_i}\sum_{i=1}^{m_i} y_{i1j} \tag{11}$$

and

$$\left[T_{2i} + (1 - T_{2i})P_{2i} + (1 - T_{2i})(1 - P_{2i})Q_{4i}\right]\hat{\lambda}_{i1u} + \left[(1 - T_{2i})(1 - P_{2i})Q_{5i}\right]\hat{\lambda}_{i2u} = \frac{1}{m_i}\sum_{i=1}^{m_i} y_{i2j} \tag{12}$$

where $\hat{\lambda}_{i1u}$ and $\hat{\lambda}_{i2u}$ are the estimates of the parameters λ_1 and λ_2. From the equations (11) and (12) the expressions of $\hat{\lambda}_{i1u}$ and $\hat{\lambda}_{i2u}$ are simplified as

$$\hat{\lambda}_{i1u} = \frac{1}{m_i A_i}\left[(1 - T_{2i})(1 - P_{2i})Q_{5i}\sum_{i=1}^{m_i} y_{i1j} - (1 - T_{1i})(1 - P_{1i})Q_{2i}\sum_{i=1}^{m_i} y_{i2j}\right] \tag{13}$$

and

$$\hat{\lambda}_{i2u}=\frac{1}{m_iB_i}\left[\left[T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{4i}\right]\sum_{i=1}^{m_i}y_{i1j}-\left[T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{1i}\right]\sum_{i=1}^{m_i}y_{i2j}\right]$$
(14)

where

$$A=\left[\left[\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{5i}\right]\left[T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{1i}\right]-\left[\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{2i}\right]\left[T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{4i}\right]\right]$$

$$\left[\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{5i}\right]\left[T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{1i}\right]\neq\left[\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{2i}\right]\left[T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{4i}\right]$$

$$B=\left[\left[T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{4i}\right]\left[\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{2i}\right]-\left[T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{1i}\right]\left[\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{5i}\right]\right]$$

$$\left[T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{4i}\right]\left[\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{2i}\right]\neq\left[T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{1i}\right]\left[\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{5i}\right]$$

Hence, finally the estimator for the parameter of the rare sensitive attribute λ_1 is proposed as

$$\hat{\lambda}_{1ppzwru}=\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1u}}{p_i}$$
(15)

Theorem 2.2.1 The suggested estimator $\hat{\lambda}_{1ppzwru}$ for the mean number of persons who have the rare sensitive attribute is unbiased.

Proof: Since y_{i1j} and y_{i2j} are iid Poisson variates with parameters λ^*_{i1} and λ^*_{i2} respectively, therefore, we have

$E\left(\hat{\lambda}_{1ppzwru}\right)=E_1E_2\left(\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1u}}{p_i}\right)$ and we have

$$E_2\left(\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1u}}{p_i}\right)=\frac{1}{m_iA_i}\left[\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{5i}\sum_{i=1}^{m_i}E_2\left(y_{i1j}\right)-\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{2i}\sum_{i=1}^{m_i}E_2\left(y_{i2j}\right)\right]=\frac{1}{m_iA_i}\left[(1-T_{2i})(1-P_{2i})Q_{5i}\sum_{j=1}^{m_i}\lambda*_{i1}-(1-T_{1i})(1-P_{1i})Q_{2i}\sum_{j=1}^{m_i}\lambda*_{i2}\right]=\lambda_{i1}$$

Finally, we get

Escribe aquí la ecuación.

$$E_1E_2\left(\hat{\lambda}_{1ppzwru}\right)=E_1\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\lambda_{i1}}{P_i}\right]=\frac{1}{M_0}\sum_{i=1}^{N}p_i\frac{M_i\lambda_{i1}}{p_i}=\lambda_1$$

Theorem 2.2.2 The variance of the estimator $\hat{\lambda}_{1ppzwru}$ is given by

$$V\left(\hat{\lambda}_{1ppzwru}\right)=\frac{1}{nM_0^2}\left[\sum_{i=1}^{N}p_i\left(\frac{M_i\lambda_{i1}}{p_i}-M_0\lambda_1\right)^2+\sum_{i=1}^{N}\frac{M_i^{\,2}}{p_i}\frac{\psi_i^{(12)}}{A_i^2m_i}\right]$$
(16)

where

$$\begin{aligned}\psi_i^{(12)}=\Big[\Big[&\left(1-T_{2i}\right)^2\left(1-P_{2i}\right)^2Q_{5i}^2\left\{T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{1i}\right\}+\left(1-T_{1i}\right)^2\left(1-P_{1i}\right)^2Q_{2i}^2\\&\left\{T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{4i}\right\}-2\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{5i}\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{2i}\\&\left\{T_{1i}+\left(1-T_{1i}\right)P_{1i}+\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{1i}\right\}\left\{T_{2i}+\left(1-T_{2i}\right)P_{2i}+\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{4i}\right\}\Big]\lambda_{i1}\\&+\Big[\left(1-T_{2i}\right)^2\left(1-P_{2i}\right)^2Q_{5i}^2\left\{\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{2i}\right\}+\left(1-T_{1i}\right)^2\left(1-P_{1i}\right)^2Q_{2i}^2\left\{\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{5i}\right\}\\&-2\left(1-T_{2i}\right)\left(1-P_{2i}\right)Q_{5i}\left(1-T_{1i}\right)\left(1-P_{1i}\right)Q_{2i}\left\{\left(1-P_{1i}\right)\left(1-T_{1i}\right)Q_{2i}\right\}\left\{\left(1-P_{2i}\right)\left(1-T_{2i}\right)Q_{5i}\right\}\Big]\lambda_{i2}\Big]\end{aligned}$$

Proof: The variance of the estimator $\hat{\lambda}_{1ppzwru}$ is written as

$$V\left(\hat{\lambda}_{1ppzwru}\right)=V_1E_2\left(\hat{\lambda}_{1ppzwru}\right)+E_1V_2\left(\hat{\lambda}_{1ppzwru}\right)$$
(17)

we get

$$V_1E_2\left(\hat{\lambda}_{1ppzwru}\right)=\frac{1}{nM_0^2}\sum_{i=1}^{N}P_i\left[\frac{M_i\lambda_{i1}}{p_i}-M_0\lambda_1\right]^2$$

(18)

and

$$E_1V_2\left(\hat{\lambda}_{1ppzwru}\right)=E_1V_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\lambda}_{i1u}}{p_i}\right]$$

$$=E_1V_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i}{p_i}\left\{\frac{1}{m_iA_i}\left((1-T_{2i})(1-P_{2i})Q_{5i}\sum_{i=1}^{m_i}y_{i1j}-(1-T_{1i})(1-P_{1i})Q_{2i}\sum_{i=1}^{m_i}y_{i2j}\right)\right\}\right]$$

$$=\frac{1}{nM_0^2}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\left[\frac{1}{A_i^2m_i}\left\{\frac{1}{m_iA_i}\begin{pmatrix}(1-T_{2i})^2(1-P_{2i})^2Q_{5i}^2\lambda_{i1}^*-(1-T_{1i})^2(1-P_{1i})^2Q_{2i}^2\lambda_{i2}^*\\ -2(1-T_{2i})(1-P_{2i})Q_{5i}(1-T_{1i})(1-P_{1i})Q_{2i}\lambda_{i12}^*\end{pmatrix}\right\}\right]$$

(19)

where

$$\lambda_{i1}^*=\left[T_{1i}+(1-T_{1i})P_{1i}+(1-P_{1i})(1-T_{1i})Q_{1i}\right]\lambda_{i1}+\left[(1-P_{1i})(1-T_{1i})Q_{2i}\right]\lambda_{i2}$$

$$\lambda_{i2}^*=\left[T_{2i}+(1-T_{2i})P_{2i}+(1-P_{2i})(1-T_{2i})Q_{4i}\right]\lambda_{i1}+\left[(1-P_{2i})(1-T_{2i})Q_{5i}\right]\lambda_{i2}$$

$$\lambda_{i12}^*=\begin{bmatrix}\left\{T_{1i}+(1-T_{1i})P_{1i}+(1-P_{1i})(1-T_{1i})Q_{1i}\right\}\left\{T_{2i}+(1-T_{2i})P_{2i}+(1-P_{2i})(1-T_{2i})Q_{4i}\right\}\lambda_{i1}\\ +\left\{(1-P_{1i})(1-T_{1i})Q_{2i}\right\}\left\{(1-P_{2i})(1-T_{2i})Q_{5i}\right\}\lambda_{i2}\end{bmatrix}$$

Thus, adding the equations (18) and (19), we get the variance of the unbiased estimator $\hat{\lambda}_{1ppzwru}$ as given in equation (16).
The estimator of the variance of $\hat{\lambda}_{1ppzwru}$ is obtained as

$$\hat{V}\left(\hat{\lambda}_{1ppzwru}\right)=\frac{1}{n(n-1)M_0^2}\sum_{i=1}^{n}\left(\frac{M_i\hat{\lambda}_{i1u}}{p_i}-\hat{\lambda}_{1ppzwru}\right)^2$$

(20)

The first stage sample is selected using PPSWR $\left(p_i=\frac{M_i}{M_0}\right)$. Hence the unbiased estimator for the mean number of persons who possess the rare sensitive attribute is as

$$\hat{\lambda}_{1ppswru}=\frac{1}{n}\sum_{i=1}^{n}\hat{\lambda}_{i1u}$$

(21)

and subsequently its variance is obtained as

$$V\left(\hat{\lambda}_{1ppswru}\right)=\frac{1}{nM_0}\left[\sum_{i=1}^{N}M_i\left(\lambda_{i1}-\lambda_1\right)^2+\sum_{i=1}^{N}M_i\frac{\psi_i^{(12)}}{m_i\gamma_i^2}\right]$$

(22)

and the estimate of the var $\left(\hat{\lambda}_{1ppswru}\right)$ is simplified as

$$\hat{V}\left(\hat{\lambda}_{1ppswru}\right)=\frac{1}{n(n-1)}\sum_{i=1}^{n}\left[\hat{\lambda}_{i1u}-\frac{\hat{\lambda}_{1ppswru}}{M_0}\right]^2$$

(23)

3. ESTIMATION PROCEDURE UNDER TWO-STAGE SAMPLING DESIGN WITH STRATIFICATION USING RANDOMIZED RESPONSE MODEL (RRM)

In this section, it is considered that population is stratified into L strata with N_h clusters in h^{th} (h=1,2,…,L) stratum. It is assumed that the size of the i^{th} cluster in h^{th} stratum is M_{hi} ($i=1,2,\ldots,N_h$) and in first stage a sample of n_h clusters are drawn using PPSWR with selection probability p_{hi} from h^{th} stratum. In second stage simple random samples using with replacement of sizes m_{hi} are selected from the i^{th} $(i=1,2,.....,n_h)$ cluster drawn from the h^{th} stratum.

3.1 Estimation procedure when the proportion of a rare non-sensitive unrelated attribute is known for a stratified population

In this section we extend the procedure as discussed in section 2.1 for stratified population, the probability that respondents answer "yes" in i^{th} sampled cluster of h^{th} stratum is defined as

$$\theta_{hic} = T_{hi}\pi_{hi1} + (1 - T_{hi})\left[P_{hi}\pi_{hi1} + (1 - P_{hi})\{Q_{1hi}\pi_{hi1} + Q_{2hi}\pi_{hi2}\}\right]$$

where T_{hi}, P_{hi}, Q_{1hi} and Q_{2hi} are the similar probabilities of being asked the questions as described in section 2.1. π_{hi1} and π_{hi2} are the population proportions of the rare sensitive and rare unrelated non-sensitive attribute, A_1 and A_2 respectively, hence for large m_{hi} and small θ_{hi0}, we have $m_{hi}\theta_{hi0} = \lambda_{hi0}$ is finite.

Let $y_{hi1}, y_{hi2}, ..., y_{him_{hi}}$ be a random sample of size m_{hi} from the Poisson distribution with mean λ_{hi0} from i^{th} cluster of h^{th} stratum, hence, an estimator for the mean number of individuals who possess a rare sensitive attribute λ_{hi1} in i^{th} cluster of h^{th} stratum is defined as follows

$$\hat{\lambda}_{hi1} = \frac{1}{\left[T_{hi} + (1 - T_{hi})P_{hi} + (1 - T_{hi})(1 - P_{hi})Q_{1hi}\right]}\left[\frac{1}{m_{hi}}\sum_{i=1}^{m_{hi}} y_{hij} - (1 - T_{hi})(1 - P_{hi})Q_{2hi}\lambda_{hi2}\right] \tag{24}$$

It may be seen that the estimator $\hat{\lambda}_{hi1}$ is unbiased for λ_{hi1}.

An estimator for the mean number of individuals who posses rare sensitive attribute λ_{h1} in stratum h is as follows

$$\hat{\lambda}_{h1} = \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1}}{p_{hi}} \tag{25}$$

where $M_{h0} = \sum_{i=1}^{N_h} M_{hi}$ .Finally, an estimator for the mean number of individuals who posses rare sensitive attribute λ_1 is defined as

$$\hat{\lambda}_{h1} = \sum_{h=1}^{L} W_h \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1}}{p_{hi}} \tag{26}$$

where $W_h = \frac{N_h}{N}$ and $N = \sum_{h=1}^{L} N_h$.

Theorem 3.1.1 The estimator $\hat{\lambda}_{1sppzwr}$ for the mean number of persons who have the rare sensitive attribute is unbiased.

Proof: Since $y_{hij}\,(j=1,2,...,m_{hi})$ are iid Poisson variates with parameter λ_{hi0}, hence $E(y_{hij}) = \lambda_{hi0}$ where

$$\lambda_{hi0} = T_{hi}\lambda_{hi1} + (1 - T_{hi})\left[P_{hi}\lambda_{hi1} + (1 - P_{hi})\{Q_{1hi}\lambda_{hi1} + Q_{2hi}\lambda_{hi2}\}\right]$$

$$E_1E_2(\hat{\lambda}_{1sppzwr}) = E_1E_2\left[\sum_{h=1}^{L} W_h \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1}}{p_{hi}}\right] = E_1\left[\sum_{h=1}^{L} W_h \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\lambda_{hi1}}{p_{hi}}\right] = \sum_{h=1}^{L} W_h \frac{1}{M_{h0}}\sum_{i=1}^{N_h} M_{hi}\lambda_{hi1} = \sum_{h=1}^{L} W_h\lambda_{h1} = \lambda_1$$

Theorem3.1.2 The variance of the unbiased estimator $\hat{\lambda}_{1sppzwr}$ is as follows

$$V\left(\hat{\lambda}_{1sppzwr}\right)=\sum_{h=1}^{L} W_h^2 \frac{1}{n_h M_{h0}^2}\left[\sum_{i=1}^{N_h} p_{hi}\left(\frac{M_{hi}\lambda_{hi1}}{p_{hi}}-M_{h0}\lambda_{h1}\right)^2+\sum_{i=1}^{N_h}\frac{M_{hi}{}^2}{p_{hi}}\frac{\varphi_{hi}}{m_{hi}}\right] \quad (27)$$

where $\varphi_i=\frac{\lambda_{hi1}}{\left[T_{hi}+\left(1-T_{hi}\right)P_{hi}+\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{1hi}\right]}+\frac{\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{2hi}\lambda_{hi2}}{n\left[T_{hi}+\left(1-T_{hi}\right)P_{hi}+\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{1hi}\right]^2}$

Proof

The variance is expressed as

$$V\left(\hat{\lambda}_{1sppzwr}\right)=V_1E_2\left(\hat{\lambda}_{1sppzwr}\right)+E_1V_2\left(\hat{\lambda}_{1sppzwr}\right) \quad (28)$$

where the variance of overall possible subsamples selections for a given set units is denoted by V_2 and the variance of overall first-stage samples is denoted by V_1

$$V_1E_2\left(\hat{\lambda}_{1sppzwr}\right)=V_1E_2\left[\sum_{h=1}^{L} W_h \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1}}{p_{hi}}\right]=\sum_{h=1}^{L} W_h^2 \frac{1}{n_h M_{h0}^2}\sum_{i=1}^{N_h} p_{hi}\left[\frac{M_{hi}\lambda_{hi1}}{p_{hi}}-M_{h0}\lambda_{h1}\right]^2 \quad (29)$$

Since $V\left(y_{hij}\right)=\lambda_{hi0}$, therefore, the second term is simplified as

$$E_1V_2\left(\hat{\lambda}_{1sppzwr}\right)=E_1V_2\left[\sum_{h=1}^{L} W_h \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1}}{p_{hi}}\right]=E_1\left[\sum_{h=1}^{L} W_h^2 \frac{1}{\left(n_h M_{h0}\right)^2}\sum_{i=1}^{n_h}\frac{M_{hi}{}^2 V_2\left(\hat{\lambda}_{hi1}\right)}{p_{hi}{}^2}\right]$$

$$=E_1\left[\sum_{h=1}^{L} W_h^2 \frac{1}{\left(n_h M_{h0}\right)^2}\sum_{i=1}^{n_h}\frac{M_{hi}{}^2}{p_{hi}^2}\frac{1}{\left[T_{hi}+\left(1-T_{hi}\right)P_{hi}+\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{1hi}\right]^2 m_{hi}}\lambda_{hi0}\right]$$

$$=E_1\left[\sum_{h=1}^{L} W_h^2 \frac{1}{\left(n_h M_{h0}\right)^2}\sum_{i=1}^{n_h}\frac{M_{hi}{}^2}{p_{hi}^2 m_{hi}}\left\{\frac{\lambda_{hi1}}{\left[T_{hi}+\left(1-T_{hi}\right)P_{hi}+\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{1hi}\right]}+\frac{\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{2hi}\lambda_{hi2}}{\left[T_{hi}+\left(1-T_{hi}\right)P_{hi}+\left(1-T_{hi}\right)\left(1-P_{hi}\right)Q_{1hi}\right]^2}\right\}\right]=\sum_{h=1}^{L} W_h^2 \frac{1}{n_h M_{h0}^2}\sum_{i=1}^{N_h}\frac{M_{hi}^2}{p_{hi}}\frac{\varphi_{hi}}{m_{hi}} \quad (30)$$

By adding the equations (29) and (30) we get the variance of $\hat{\lambda}_{1sppzwr}$ as given in equation (27).

An unbiased estimator of the variance of $\hat{\lambda}_{1sppzwr}$ is also given as

$$\hat{V}\left(\hat{\lambda}_{1sppzwr}\right)=\sum_{h=1}^{L} W_h^2 \frac{1}{n_h\left(n_h-1\right)M_{h0}^2}\sum_{i=1}^{n_h}\left(\frac{M_{hi}\hat{\lambda}_{hi1}}{p_{hi}}-\hat{\lambda}_{h1}\right)^2 \quad (31)$$

From the N_h clusters, n_h clusters are chosen depending on their sizes in h^{th} stratum using PPSWR, then the selection probability p_{hi} is defined as $\frac{M_{hi}}{M_{h0}}$. The unbiased estimator of λ_1 is given by

$$\hat{\lambda}_{1sppswr}=\sum_{h=1}^{L} W_h \frac{1}{n_h}\sum_{i=1}^{n_h}\hat{\lambda}_{hi1} \quad (32)$$

and its variance is simplified as

$$V\left(\hat{\lambda}_{1sppswr}\right)=\sum_{h=1}^{L} W_h^2 \frac{1}{n_h M_{h0}}\left[\sum_{i=1}^{N_h} M_{hi}\left(\lambda_{hi1}-\lambda_{h1}\right)^2+\sum_{i=1}^{N_h}\frac{M_{hi}}{m_{hi}}\varphi_{hi}\right] \quad (33)$$

and subsequently, the estimator for the variance of the estimator $\hat{\lambda}_{1sppswr}$ is simplified as

$$\hat{V}\left(\hat{\lambda}_{1sppswr}\right)=\sum_{h=1}^{L} W_h^2 \frac{1}{n_h\left(n_h-1\right)}\sum_{i=1}^{n_h}\left[\hat{\lambda}_{hi1}-\frac{\hat{\lambda}_{h1}}{M_0}\right]^2$$
(34)

3.2 Estimation procedure when the rare non-sensitive unrelated attribute is unknown under two-stage sampling design for stratified population

In this section π_{hi2} is assumed to be unknown. Respondents are asked to answer "yes" or "no" according to extended three-stage Singh et al. (2003) randomized response device. The probabilities of "yes" answer in i^{th} cluster of h^{th} stratum as follows

$$\theta_{hi1}=T_{1hi}\pi_{hi1}+\left(1-T_{1hi}\right)\left[P_{1hi}\pi_{hi1}+\left(1-P_{1hi}\right)\left\{Q_{1hi}\pi_{hi1}+Q_{2hi}\pi_{hi2}\right\}\right]$$

and

$$\theta_{hi2}=T_{2hi}\pi_{hi1}+\left(1-T_{2hi}\right)\left[P_{2hi}\pi_{hi1}+\left(1-P_{2hi}\right)\left\{Q_{4hi}\pi_{hi1}+Q_{5hi}\pi_{hi2}\right\}\right]$$

where T_{1hi}, T_{2hi}, P_{1hi}, P_{2hi} and Q_{1hi}, Q_{2hi}, Q_{4hi}, Q_{5hi} are the probabilities of being asked the questions same as described in section 2.2. Since the two attributes are rare in the population, therefore, $m_{hi}\theta_{hi1}=\lambda^*_{hi1}$ and $m_{hi}\theta_{hi2}=\lambda^*_{hi2}$ are finite for large m_{hi} and small θ_{hi1} and θ_{hi2}.

After derivational procedures, the following two equations are obtained as

$$\left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{1hi}\right]\hat{\lambda}_{hi1u}+\left[\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{2hi}\right]\hat{\lambda}_{hi2u}=\frac{1}{m_{hi}}\sum_{i=1}^{m_{hi}}y_{hi1j}$$
(35)

$$\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{4hi}\right]\hat{\lambda}_{hi1u}+\left[\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{5hi}\right]\hat{\lambda}_{hi2u}=\frac{1}{m_{hi}}\sum_{i=1}^{m_{hi}}y_{hi2j}$$
(36)

From the above equations, the estimators for λ_{hi1u} and λ_{hi2u} are obtained as

$$\hat{\lambda}_{hi1u}=\frac{1}{m_{hi}A_{hi}}\left[\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{5hi}\sum_{i=1}^{m_{hi}}y_{hi1j}-\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{2hi}\sum_{i=1}^{m_{hi}}y_{hi2j}\right]$$
(37)

and

$$\hat{\lambda}_{hi2u}=\frac{1}{m_{hi}B_{hi}}\left[\begin{array}{l}\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{4hi}\right]\sum_{i=1}^{m_{hi}}y_{hi1j}- \\ \left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{1hi}\right]\sum_{i=1}^{m_{hi}}y_{hi2j}\end{array}\right]$$
(38)

where

$$A_{hi}=\left[\left[\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{5hi}\left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{1hi}\right]-\left[\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{2hi}\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{4hi}\right]\right]\right]\right]$$

and

$$\left[\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{5hi}\left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{1hi}\right]\neq\left[\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{2hi}\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{4hi}\right]\right]\right]$$

$$B_{hi}=\left[\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{4hi}\right]\left[\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{2hi}\right]-\left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{1hi}\right]\left[\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{5hi}\right]\right]$$

$$\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{4hi}\right]\left[\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{2hi}\right]\neq\left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{1hi}\right]\left[\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{5hi}\right]$$

In h^{th} stratum, the estimator for the mean number of individuals who possess the rare sensitive attribute is as follows

$$\hat{\lambda}_{h1sppzwru}=\frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1u}}{p_{hi}} \qquad (39)$$

where p_{hi} is the selection probability in the i^{th} cluster under PPSWR scheme, and $M_{h0}=\sum_{i=1}^{N_h}M_{hi}$ therefore, the estimator for the mean number of persons who possess the rare sensitive attribute is given by

$$\hat{\lambda}_{1sppzwru}=\sum_{h=1}^{L}W_h\frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1u}}{p_{hi}} \qquad (40)$$

Theorem 3.2.1 The estimator for the mean number of individuals who possess the rare sensitive attribute, $\hat{\lambda}_{1sppzwru}$ is unbiased.

Proof: Since y_{hi1j} and y_{hi2j} are iid Poisson variates with parameters λ^*_{hi1} and λ^*_{hi2} respectively, therefore, we have $E\left(\hat{\lambda}_{1sppzwru}\right)=E_1E_2\left(\hat{\lambda}_{1sppzwru}\right)$

$$E_2\left(\hat{\lambda}_{hi1u}\right)=\frac{1}{m_{hi}A_{hi}}\left[(1-T_{2hi})(1-P_{2hi})Q_{5hi}\sum_{i=1}^{m_{hi}}E_2\left(y_{hi1j}\right)-(1-T_{1hi})(1-P_{1hi})Q_{2hi}\sum_{i=1}^{m_{hi}}E_2\left(y_{hi2j}\right)\right]=\frac{1}{m_{hi}A_{hi}}\left[(1-T_{2hi})(1-P_{2hi})Q_{5hi}\sum_{j=1}^{m_{hi}}\lambda^*_{hi1}-(1-T_{1hi})(1-P_{1hi})Q_{2hi}\sum_{j=1}^{m_{hi}}\lambda^*_{hi2}\right]=\lambda_{hi1}$$

Using the unbiased estimator $\hat{\lambda}_{hi1}$

$$E_1E_2\left(\hat{\lambda}_{1sppzwru}\right)=E_1E_2\left[\sum_{h=1}^{L}W_h\frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1u}}{P_{hi}}\right]=E_1\left[\sum_{h=1}^{L}W_h\frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\lambda_{hi1}}{P_{hi}}\right]=\lambda_1$$

Theorem 3.2.2 The variance of the unbiased estimator $\hat{\lambda}_{1sppzwru}$ is

$$V\left(\hat{\lambda}_{1sppzwru}\right)=\sum_{h=1}^{L}W_h^2\frac{1}{n_h M_{h0}^2}\left[\sum_{i=1}^{N_h}p_{hi}\left(\frac{M_{hi}\lambda_{hi1}}{p_{hi}}-M_{h0}\lambda_{h1}\right)^2+\sum_{i=1}^{N_h}\frac{{M_{hi}}^2}{p_{hi}}\frac{\psi_{hi}^{(12)}}{A_{hi}^2 m_{hi}}\right] \qquad (41)$$

where

$$\begin{aligned}\psi_{hi}^{(12)}=\Big[\Big[&(1-T_{2hi})^2(1-P_{2hi})^2Q_{5hi}^2\left\{T_{1hi}+(1-T_{1hi})P_{1hi}+(1-P_{1hi})(1-T_{1hi})Q_{1hi}\right\}+(1-T_{1hi})^2(1-P_{1hi})^2Q_{2hi}^2\\&\left\{T_{2hi}+(1-T_{2hi})P_{2hi}+(1-P_{2hi})(1-T_{2hi})Q_{4hi}\right\}-2(1-T_{2hi})(1-P_{2hi})Q_{5hi}(1-T_{1hi})(1-P_{1hi})Q_{2hi}\\&\left\{T_{1hi}+(1-T_{1hi})P_{1hi}+(1-P_{1hi})(1-T_{1hi})Q_{1hi}\right\}\left\{T_{2hi}+(1-T_{2hi})P_{2hi}+(1-P_{2hi})(1-T_{2hi})Q_{4hi}\right\}\Big]\lambda_{hi1}\\&+\Big[(1-T_{2hi})^2(1-P_{2hi})^2Q_{5hi}^2\left\{(1-P_{1hi})(1-T_{1hi})Q_{2hi}\right\}+(1-T_{1hi})^2(1-P_{1hi})^2Q_{2hi}^2\left\{(1-P_{2hi})(1-T_{2hi})Q_{5hi}\right\}\\&-2(1-T_{2hi})(1-P_{2hi})Q_{5hi}(1-T_{1hi})(1-P_{1hi})Q_{2hi}\left\{(1-P_{1hi})(1-T_{1hi})Q_{2hi}\right\}\left\{(1-P_{2hi})(1-T_{2hi})Q_{5hi}\right\}\Big]\lambda_{hi2}\Big]\end{aligned}$$

Proof: The $V\left(\hat{\lambda}_{1sppzwru}\right)$ is decomposed as

$$V\left(\hat{\lambda}_{1sppzwru}\right)=V_1E_2\left(\hat{\lambda}_{1sppzwru}\right)+E_1V_2\left(\hat{\lambda}_{1sppzwru}\right) \qquad (42)$$

Since, $\hat{\lambda}_{hi1u}$ is an unbiased estimator, therefore,

$$V_1E_2\left(\hat{\lambda}_{1sppzwru}\right)=\sum_{h=1}^{L}W_h^2\frac{1}{n_h M_{h0}^2}\sum_{i=1}^{N_h}P_{hi}\left[\frac{M_{hi}\lambda_{hi1}}{p_{hi}}-M_{h0}\lambda_{h1}\right]^2 \qquad (43)$$

The Second term is obtained as

$$E_1V_2\left(\hat{\lambda}_{1sppzwru}\right)=E_1V_2\left[\sum_{h=1}^{L}W_h\frac{1}{n_hM_{h0}}\sum_{i=1}^{n_h}\frac{M_{hi}\hat{\lambda}_{hi1u}}{p_{hi}}\right]$$

$$=E_1\left[\sum_{h=1}^{L}W_h^2\frac{1}{\left(n_hM_{h0}\right)^2}\sum_{i=1}^{n_h}\frac{M_{hi}^2}{p_{hi}^2}\left\{\frac{1}{A_{hi}^2m_{hi}^2}\left(\begin{array}{l}\left(1-T_{2hi}\right)^2\left(1-P_{2hi}\right)^2Q_{5hi}^2\sum_{j=1}^{m_{hi}}\lambda_{hi1}^*-\left(1-T_{1hi}\right)^2\left(1-P_{1hi}\right)^2Q_{2hi}^2\sum_{j=1}^{m_{hi}}\lambda_{hi2}^*\\-2\left(1-T_{2hi}\right)\left(1-P_{2hi}\right)Q_{5hi}\left(1-T_{1hi}\right)\left(1-P_{1hi}\right)Q_{2hi}\sum_{j=1}^{m_{hi}}\lambda_{hi12}^*\end{array}\right)\right\}\right]$$

$$=\sum_{h=1}^{L}W_h^2\frac{1}{n_hM_{h0}^2}\sum_{i=1}^{N_h}\frac{M_{hi}^2}{p_{hi}^2}\left(\frac{\psi_{hi}^{(12)}}{A_{hi}^2m_{hi}}\right)$$

(44)

where

$$\lambda_{hi1}^*=\left[T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{1hi}\right]\lambda_{hi1}+\left[\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{2hi}\right]\lambda_{hi2}$$

$$\lambda_{hi2}^*=\left[T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{4hi}\right]\lambda_{hi1}+\left[\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{5hi}\right]\lambda_{hi2}$$

$$\lambda_{hi12}^*=\left[\begin{array}{l}\left\{T_{1hi}+\left(1-T_{1hi}\right)P_{1hi}+\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{1hi}\right\}\left\{T_{2hi}+\left(1-T_{2hi}\right)P_{2hi}+\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{4hi}\right\}\lambda_{hi1}\\+\left\{\left(1-P_{1hi}\right)\left(1-T_{1hi}\right)Q_{2hi}\right\}\left\{\left(1-P_{2hi}\right)\left(1-T_{2hi}\right)Q_{5hi}\right\}\lambda_{hi2}\end{array}\right]$$

Thus, by the addition of the equations (43) and (44), we get the variance of the unbiased estimator, $\hat{\lambda}_{1sppzwru}$ as given in equation (41).

An unbiased estimator of the variance of $\hat{\lambda}_{1sppzwru}$ is as follows

$$\hat{V}\left(\hat{\lambda}_{1sppzwru}\right)=\sum_{h=1}^{L}W_h^2\frac{1}{n_h\left(n_h-1\right)M_{h0}^2}\sum_{i=1}^{n_h}\left(\frac{M_{hi}\hat{\lambda}_{hi1u}}{p_{hi}}-\hat{\lambda}_{h1u}\right)^2$$

(45)

Under PPSWR scheme the unbiased estimator for the mean number of persons who possess the rare sensitive attribute is

$$\hat{\lambda}_{1sppswru}=\sum_{h=1}^{L}W_h\frac{1}{n_h}\sum_{i=1}^{n_h}\hat{\lambda}_{hi1u}$$

(46)

and its variance is simplified as

$$V\left(\hat{\lambda}_{1sppswru}\right)=\sum_{h=1}^{L}W_h^2\frac{1}{n_hM_{h0}}\left[\sum_{i=1}^{N_h}M_{hi}\left(\lambda_{hi1}-\lambda_{h1}\right)^2+\sum_{i=1}^{N_h}M_{hi}\frac{\psi_{hi}^{(12)}}{m_{hi}\gamma_{hi}^2}\right]$$

(47)

Subsequently, the estimator for the variance of $\hat{\lambda}_{1sppswru}$ is given as

$$\hat{V}\left(\hat{\lambda}_{1sppswru}\right)=\sum_{h=1}^{L}W_h^2\frac{1}{n_h\left(n_h-1\right)}\sum_{i=1}^{n_h}\left[\hat{\lambda}_{hi1u}-\frac{\hat{\lambda}_{hiu}}{M_{h0}}\right]^2$$

(48)

4. EFFICIENCY COMPARISON

To show the dominance of the proposed estimators the empirical comparison are made over Lee et al. (2014) and Singh et al. (2019) estimators. The percent relative efficiencies of the proposed estimators are calculated with respect to Lee et al. (2014) and Singh et al. (2019) estimators for the corresponding situations using the formula:

$$PRE_1=\frac{V\left(\text{Lee et al. estimator}\right)}{V\left(\text{Prooposed estimator}\right)}x100\,. \qquad PRE_2=\frac{V\left(\text{Singh et al. estimator}\right)}{V\left(\text{Prooposed estimator}\right)}x100$$

To calculate the percent relative efficiencies, of the proposed estimators over Lee et al. (2014) and Singh et al. (2019) estimators for corresponding situations a finite population is assumed to have five clusters $(N=5)$ with sizes $(1000, 2000, 2000, 3000, 4000)$ for M_i $(i=1,2,....,5)$ respectively. Two clusters $(n=2)$ are selected PPS with replacement. In PPSWR, the probabilities with which the clusters are selected are calculated using the cluster sizes as $p_i = \frac{M_i}{M_0}$ where $M_0 = \sum_{i=1}^{5} M_i = 12000$. For cases II and IV, let us assume that a population is stratified into two strata $(L=2)$ and there are two clusters $(N_1=2)$ with sizes $M_{11}=1000$, $M_{12}=2000$ in the first stratum whereas three clusters $(N_2=3)$ with sizes $M_{2i} = (2000, 3000, 4000)$ for $i=1,2,3$ in the second stratum. A cluster is selected from each stratum $(n_1 = n_2 = 1)$ .Population is divided into two strata with stratum weights $W_1=0.4$ and $W_2=0.6$. In both procedures, we assume that 10% units are selected in the samples from each cluster and the parameters for the rare unrelated attribute which was assumed to be known are taken as 1.

Table: 1 Percent relative efficiency with respect to Lee et al. (2014), when rare non-sensitive unrelated attribute is known.

	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	T = 0.40 P = 0.80 Q_1 = 0.60 Q_2 =0.20	0.50 0.70 0.50 0.25	0.60 0.60 0.40 0.30	0.70 0.50 0.30 0.35	0.80 0.40 0.20 0.40
Case 1	1	1	1	1	1	286.60	395.09	596.13	1043.70	2384.50
Case 2	1	1	1	1	2	394.37	399.61	409.99	433.40	500.97
Case 3	1	1	1	2	1	297.98	303.36	313.83	337.17	404.33
Case 4	1	1	1	2	2	347.45	350.29	355.85	368.23	403.57
Case 5	1	1	2	1	1	203.55	209.15	219.80	243.22	310.30
Case 6	1	1	2	1	2	299.01	301.91	307.51	319.92	355.24
Case 7	1	1	2	2	1	249.93	252.87	258.49	270.88	306.09
Case 8	1	1	2	2	2	298.85	300.88	304.78	313.37	337.65
Case 9	1	2	1	1	1	203.55	209.15	219.80	243.22	310.30
Case 10	1	2	1	1	2	299.01	301.91	307.51	319.92	355.24
Case 11	1	2	1	2	1	249.93	252.87	258.49	270.88	306.09
Case 12	1	2	1	2	2	298.85	300.88	304.78	313.37	337.65
Case 13	1	2	2	1	1	201.61	204.61	210.29	222.70	257.89
Case 14	1	2	2	1	2	266.33	268.39	272.31	280.92	305.19
Case 15	1	2	2	2	1	233.41	235.48	239.42	248.01	272.23
Case 16	1	2	2	2	2	274.29	275.90	278.96	285.63	304.32
Case 17	2	1	1	1	1	109.12	114.94	125.77	149.28	216.27
Case 18	2	1	1	1	2	250.57	253.53	259.18	271.61	306.91
Case 19	2	1	1	2	1	201.61	204.61	210.29	222.70	257.89
Case 20	2	1	1	2	2	266.33	268.39	272.31	280.92	305.19
Case 21	2	1	2	1	1	153.29	156.35	162.08	174.52	209.68
Case 22	2	1	2	1	2	233.81	235.89	239.85	248.46	272.72
Case 23	2	1	2	2	1	200.94	203.05	207.01	215.62	239.82
Case 24	2	1	2	2	2	249.82	251.44	254.51	261.19	279.87

Case 25	2	2	1	1	1	153.29	156.35	162.08	174.52	209.68
Case 26	2	2	1	1	2	233.81	235.89	239.85	248.46	272.72
Case 27	2	2	1	2	1	200.94	203.05	207.01	215.62	239.82
Case 28	2	2	1	2	2	249.82	251.44	254.51	261.19	279.87
Case 29	2	2	2	1	1	168.47	170.61	174.60	183.22	207.41
Case30	2	2	2	1	2	225.34	226.98	230.07	236.76	255.43
Case 31	2	2	2	2	1	200.60	202.25	205.35	212.03	230.68
Case 32	2	2	2	2	2	239.85	241.21	243.77	249.28	264.59

Table: 2 Percent relative efficiency with respect to Lee et al. (2014), when rare non-sensitive unrelated attribute is known under stratified population.

	λ_{111}	λ_{121}	λ_{211}	λ_{221}	λ_{231}	T = 0.40 P = 0.80 Q_1 = 0.60 Q_2 =0.20	0.50 0.70 0.50 0.25	0.60 0.60 0.40 0.30	0.70 0.50 0.30 0.35	0.80 0.40 0.20 0.40
Case 1	1	1	1	1	1	257.94	355.58	536.51	939.34	2146.10
Case 2	1	1	1	1	2	102.52	104.22	107.41	114.34	134.08
Case 3	1	1	1	2	1	103.34	105.60	109.82	119.00	145.16
Case 4	1	1	1	2	2	101.54	102.56	104.46	108.56	120.15
Case 5	1	1	2	1	1	104.95	108.30	114.54	128.14	166.90
Case 6	1	1	2	1	2	101.79	102.98	105.19	109.97	123.46
Case 7	1	1	2	2	1	102.14	103.57	106.21	111.93	128.08
Case 8	1	1	2	2	2	101.26	102.09	103.63	106.91	116.13
Case 9	1	2	1	1	1	103.82	106.38	111.16	121.54	151.05
Case 10	1	2	1	1	2	101.64	102.73	104.74	109.09	121.33
Case 11	1	2	1	2	1	101.93	103.20	105.57	110.67	125.03
Case 12	1	2	1	2	2	101.20	101.98	103.43	106.51	115.15
Case 13	1	2	2	1	1	102.33	103.88	106.74	112.91	130.29
Case 14	1	2	2	1	2	101.33	102.21	103.82	107.26	116.88
Case 15	1	2	2	2	1	101.51	102.49	104.31	108.19	119.05
Case 16	1	2	2	2	2	101.05	101.72	102.96	105.59	112.90
Case 17	2	1	1	1	1	107.43	112.41	121.69	141.84	199.19
Case 18	2	1	1	1	2	102.04	103.40	105.91	111.32	126.57
Case 19	2	1	1	2	1	102.51	104.17	107.25	113.88	132.58
Case 20	2	1	1	2	2	101.39	102.30	103.97	107.54	117.55
Case 21	2	1	2	1	1	103.24	105.39	109.36	117.93	142.09
Case 22	2	1	2	1	2	101.57	102.60	104.50	108.56	119.91
Case 23	2	1	2	2	1	101.82	103.01	105.20	109.89	123.00
Case 24	2	1	2	2	2	101.18	101.94	103.34	106.30	114.54
Case 25	2	2	1	1	1	102.76	104.58	107.94	115.18	135.51

Case 26	2	2	1	1	2	101.47	102.43	104.19	107.94	118.42
Case 27	2	2	1	2	1	101.68	102.77	104.77	109.06	121.01
Case 28	2	2	1	2	2	101.13	101.85	103.18	105.99	113.78
Case 29	2	2	2	1	1	101.95	103.22	105.56	110.54	124.46
Case30	2	2	2	1	2	101.24	102.03	103.49	106.58	115.15
Case 31	2	2	2	2	1	101.37	102.26	103.87	107.30	116.81
Case 32	2	2	2	2	2	101.00	101.64	102.80	105.24	111.97

Table: 3 Percent relative efficiency with respect to Lee et al. (2014), when rare non-sensitive unrelated attribute is unknown.

	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	(T,P) = (0.80,0.60)	(0.70,0.50)	(0.60,0.40)	(0.50,0.30)
						Q_1 = 0.15	0.20	0.25	0.30
						Q_2 =0.35	0.40	0.45	0.50
						Q_4 = 0.10	0.15	0.20	0.25
						Q_5 =0.70	0.65	0.60	0.55
Case 1	1	1	1	1	1	8749.70	5261.50	1734.00	193.29
Case 2	1	1	1	1	2	248.03	269.09	260.83	145.31
Case 3	1	1	1	2	1	306.58	334.40	316.94	152.50
Case 4	1	1	1	2	2	199.17	213.84	210.59	135.44
Case 5	1	1	2	1	1	380.52	415.92	383.70	160.10
Case 6	1	1	2	1	2	206.49	222.17	218.37	137.39
Case 7	1	1	2	2	1	233.83	253.04	246.29	142.21
Case 8	1	1	2	2	2	182.19	194.51	192.53	131.26
Case 9	1	2	1	1	1	375.44	410.41	379.47	159.92
Case 10	1	2	1	1	2	204.69	220.14	216.53	137.11
Case 11	1	2	1	2	1	231.70	250.65	244.18	141.95
Case 12	1	2	1	2	2	180.99	193.14	191.25	131.02
Case 13	1	2	2	1	1	250.68	272.06	263.41	145.59
Case 14	1	2	2	1	2	184.21	196.84	194.80	132.13
Case 15	1	2	2	2	1	200.71	215.58	212.18	135.69
Case 16	1	2	2	2	2	170.36	181.01	179.75	128.05
Case 17	2	1	1	1	1	623.81	675.11	570.34	172.55
Case 18	2	1	1	1	2	223.15	241.04	235.64	140.85
Case 19	2	1	1	2	1	261.27	283.87	273.61	146.64
Case 20	2	1	1	2	2	189.63	202.99	200.48	133.16
Case 21	2	1	2	1	1	296.16	322.89	307.44	151.79
Case 22	2	1	2	1	2	194.55	208.60	205.79	134.64
Case 23	2	1	2	2	1	215.44	232.27	227.49	138.74
Case 24	2	1	2	2	2	176.23	187.71	186.10	129.65
Case 25	2	2	1	1	1	292.67	319.04	304.23	151.54
Case 26	2	2	1	1	2	193.00	206.85	204.18	134.37

Case 27	2	2	1	2	1	213.65	230.25	225.68	138.48
Case 28	2	2	1	2	2	175.15	186.48	184.95	129.42
Case 29	2	2	2	1	1	225.29	243.45	237.79	141.13
Case30	2	2	2	1	2	177.37	189.03	187.41	130.27
Case 31	2	2	2	2	1	190.98	204.52	201.90	133.41
Case 32	2	2	2	2	2	166.37	176.45	175.39	126.86

Table: 4 Percent relative efficiency with respect to Lee et al. (2014), when rare non-sensitive unrelated attribute is unknown under stratified population.

	λ_{111}	λ_{121}	λ_{211}	λ_{221}	λ_{231}	(T,P) = (0.80,0.60) Q_1 = 0.15 Q_2 =0.35 Q_4 = 0.10 Q_5 =0.70	(0.70,0.50) 0.20 0.40 0.15 0.65	(0.60,0.40) 0.25 0.45 0.20 0.60	(0.50,0.30) 0.30 0.50 0.25 0.55
Case 1	1	1	1	1	1	8749.70	5261.50	1734.00	193.29
Case 2	1	1	1	1	2	257.00	279.17	269.74	146.72
Case 3	1	1	1	2	1	318.36	347.45	327.87	153.82
Case 4	1	1	1	2	2	204.36	219.73	216.04	136.61
Case 5	1	1	2	1	1	397.89	434.86	398.58	161.42
Case 6	1	1	2	1	2	212.53	229.02	224.67	138.68
Case 7	1	1	2	2	1	241.03	261.15	253.55	143.47
Case 8	1	1	2	2	2	186.24	199.13	196.87	132.31
Case 9	1	2	1	1	1	325.46	355.44	334.87	155.34
Case 10	1	2	1	1	2	201.42	216.43	213.10	136.38
Case 11	1	2	1	2	1	224.71	242.77	237.08	140.71
Case 12	1	2	1	2	2	180.36	192.43	190.59	130.88
Case 13	1	2	2	1	1	239.52	259.50	252.25	143.76
Case 14	1	2	2	1	2	183.27	195.76	193.78	131.89
Case 15	1	2	2	2	1	198.22	212.75	209.57	135.16
Case 16	1	2	2	2	2	170.45	181.11	179.85	128.10
Case 17	2	1	1	1	1	412.77	461.24	406.72	136.60
Case 18	2	1	1	1	2	193.48	210.92	206.88	123.45
Case 19	2	1	1	2	1	222.02	243.87	236.26	127.38
Case 20	2	1	1	2	2	171.93	185.12	183.13	120.93
Case 21	2	1	2	1	1	241.82	267.36	256.55	128.54
Case 22	2	1	2	1	2	174.21	188.05	185.87	120.91
Case 23	2	1	2	2	1	190.86	207.35	203.63	123.87
Case 24	2	1	2	2	2	162.37	173.76	172.46	119.36
Case 25	2	2	1	1	1	218.66	240.50	233.28	126.25
Case 26	2	2	1	1	2	168.10	180.87	179.17	119.82

Case 27	2	2	1	2	1	182.33	197.40	194.51	122.55
Case 28	2	2	1	2	2	158.63	169.37	168.31	118.58
Case 29	2	2	2	1	1	186.85	203.00	199.66	122.74
Case30	2	2	2	1	2	158.97	169.96	168.88	118.34
Case 31	2	2	2	2	1	169.08	181.69	179.92	120.55
Case 32	2	2	2	2	2	153.05	162.72	161.96	117.55

Table: 5 Percent relative efficiency with respect to Singh et al. (2019), when rare non-sensitive unrelated attribute is known.

	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	T = 0.40 P = 0.80 Q_1 = 0.60 Q_2 =0.20	0.50 0.70 0.50 0.25	0.60 0.60 0.40 0.30	0.70 0.50 0.30 0.35	0.80 0.40 0.20 0.40
Case 1	1	1	1	1	1	112.6362	117.2294	127.4816	148.4051	191.8888
Case 2	1	1	1	1	2	385.8219	385.5119	385.6491	386.6943	389.4146
Case 3	1	1	1	2	1	289.4765	289.3363	289.6129	290.7153	293.3587
Case 4	1	1	1	2	2	342.8091	342.6913	342.8271	343.4687	345.046
Case 5	1	1	2	1	1	195.0446	195.1239	195.5823	196.7658	199.3281
Case 6	1	1	2	1	2	294.3646	294.3082	294.4949	295.1593	296.7138
Case 7	1	1	2	2	1	245.2959	245.2836	245.5065	246.1854	247.7194
Case 8	1	1	2	2	2	295.5453	295.5061	295.6518	296.1593	297.3395
Case 9	1	2	1	1	1	195.0446	195.1239	195.5823	196.7658	199.3281
Case 10	1	2	1	1	2	294.3646	294.3082	294.4949	295.1593	296.7138
Case 11	1	2	1	2	1	245.2959	245.2836	245.5065	246.1854	247.7194
Case 12	1	2	1	2	2	295.5453	295.5061	295.6518	296.1593	297.3395
Case 13	1	2	2	1	1	196.9744	197.0267	197.3031	198.006	199.5161
Case 14	1	2	2	1	2	263.022	263.0137	263.1851	263.704	264.8728
Case 15	1	2	2	2	1	230.1057	230.1172	230.3048	230.8302	231.9894
Case 16	1	2	2	2	2	271.6634	271.6525	271.7938	272.2293	273.2128
Case 17	2	1	1	1	1	100.6128	100.9116	101.5517	102.8163	105.2976
Case 18	2	1	1	1	2	245.9201	245.9251	246.1628	246.8498	248.3817
Case 19	2	1	1	2	1	196.9744	197.0267	197.3031	198.006	199.5161
Case 20	2	1	1	2	2	263.022	263.0137	263.1851	263.704	264.8728
Case 21	2	1	2	1	1	148.6528	148.7698	149.0998	149.8266	151.3128
Case 22	2	1	2	1	2	230.4987	230.5214	230.7183	231.2488	232.406
Case 23	2	1	2	2	1	197.6379	197.6818	197.8962	198.4337	199.5809
Case 24	2	1	2	2	2	247.185	247.1934	247.3507	247.7933	248.7697
Case 25	2	2	1	1	1	148.6528	148.7698	149.0998	149.8266	151.3128
Case 26	2	2	1	1	2	230.4987	230.5214	230.7183	231.2488	232.406
Case 27	2	2	1	2	1	197.6379	197.6818	197.8962	198.4337	199.5809

Case 28	2	2	1	2	2	247.185	247.1934	247.3507	247.7933	248.7697
Case 29	2	2	2	1	1	165.17	165.2465	165.4877	166.0371	167.1724
Case30	2	2	2	1	2	222.7065	222.7343	222.9076	223.3574	224.3267
Case 31	2	2	2	2	1	197.9735	198.0134	198.1966	198.6503	199.6137
Case 32	2	2	2	2	2	237.629	237.6451	237.7919	238.1875	239.0487

Table: 6 Percent relative efficiency with respect to Singh et al. (2019), when rare non-sensitive unrelated attribute is known under stratified population.

	λ_{111}	λ_{121}	λ_{211}	λ_{221}	λ_{231}	T = 0.40 P = 0.80 Q_1 = 0.60 Q_2 =0.20	0.50 0.70 0.50 0.25	0.60 0.60 0.40 0.30	0.70 0.50 0.30 0.35	0.80 0.40 0.20 0.40
Case 1	1	1	1	1	1	105.3221	113.8116	130.7842	165.0799	241.2719
Case 2	1	1	1	1	2	100.0887	100.2395	100.5498	101.1729	102.4877
Case 3	1	1	1	2	1	100.1176	100.3174	100.7285	101.5539	103.2962
Case 4	1	1	1	2	2	100.0562	100.1512	100.3458	100.7341	101.5462
Case 5	1	1	2	1	1	100.1744	100.4704	101.0793	102.3017	104.8831
Case 6	1	1	2	1	2	100.0654	100.176	100.4026	100.8547	101.8001
Case 7	1	1	2	2	1	100.0783	100.2106	100.4818	101.0226	102.1539
Case 8	1	1	2	2	2	100.0478	100.1283	100.2924	100.6182	101.2942
Case 9	1	2	1	1	1	100.1362	100.3671	100.8414	101.7914	103.7908
Case 10	1	2	1	1	2	100.0607	100.1633	100.373	100.7907	101.6619
Case 11	1	2	1	2	1	100.0713	100.1916	100.4378	100.928	101.9505
Case 12	1	2	1	2	2	100.0457	100.1227	100.2795	100.5901	101.2333
Case 13	1	2	2	1	1	100.0863	100.2319	100.5298	101.1229	102.3604
Case 14	1	2	2	1	2	100.051	100.1367	100.3114	100.6574	101.3739
Case 15	1	2	2	2	1	100.0575	100.1543	100.3515	100.742	101.5508
Case 16	1	2	2	2	2	100.041	100.1099	100.2496	100.5251	101.0921
Case 17	2	1	1	1	1	100.2652	100.7139	101.6347	103.479	107.3651
Case 18	2	1	1	1	2	100.0757	100.2034	100.4647	100.985	102.0703
Case 19	2	1	1	2	1	100.0928	100.2494	100.5697	101.2075	102.5382
Case 20	2	1	1	2	2	100.053	100.1421	100.3237	100.6834	101.4282
Case 21	2	1	2	1	1	100.1199	100.3223	100.7361	101.5599	103.2792
Case 22	2	1	2	1	2	100.0601	100.1612	100.3672	100.7753	101.6203
Case 23	2	1	2	2	1	100.0695	100.1863	100.4243	100.8957	101.872
Case 24	2	1	2	2	2	100.0463	100.1239	100.2815	100.5921	101.2315
Case 25	2	2	1	1	1	100.1033	100.2773	100.6327	101.3389	102.809
Case 26	2	2	1	1	2	100.0566	100.1518	100.3455	100.7285	101.5199
Case 27	2	2	1	2	1	100.0646	100.1732	100.3941	100.831	101.7337
Case 28	2	2	1	2	2	100.0446	100.1193	100.2708	100.5691	101.1817
Case 29	2	2	2	1	1	100.0752	100.2016	100.4586	100.967	102.0177

Case30	2	2	2	1	2	100.049	100.1311	100.2975	100.6253	101.2985
Case 31	2	2	2	2	1	100.0544	100.1454	100.3302	100.6939	101.4409
Case 32	2	2	2	2	2	100.0405	100.1083	100.2453	100.514	101.0629

Table: 7 Percent relative efficiency with respect to Singh et al. (2019), when rare non-sensitive unrelated attribute is unknown.

	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	(T,P) = (0.80, 0.60) Q_1 = 0.15 Q_2 =0.35 Q_4 = 0.10 Q_5 =0.70	(0.70, 0.50) 0.20 0.40 0.15 0.65	(0.60, 0.40) 0.25 0.45 0.20 0.60	(0.50, 0.30) 0.30 0.50 0.25 0.55
Case 1	1	1	1	1	1	317.1773	347.9824	296.6734	165.6544
Case 2	1	1	1	1	2	103.8131	108.2456	119.4773	132.0345
Case 3	1	1	1	2	1	105.378	111.501	126.3394	137.1911
Case 4	1	1	1	2	2	102.6284	105.6448	113.4868	125.1911
Case 5	1	1	2	1	1	107.1698	115.3355	134.2924	142.4279
Case 6	1	1	2	1	2	102.7813	106.0059	114.3836	126.5075
Case 7	1	1	2	2	1	103.5274	107.564	117.8158	129.9732
Case 8	1	1	2	2	2	102.2003	104.7144	111.3121	122.2677
Case 9	1	2	1	1	1	107.0106	115.0316	133.7469	142.2701
Case 10	1	2	1	1	2	102.7255	105.8949	114.1486	126.2912
Case 11	1	2	1	2	1	103.4611	107.4329	117.5458	129.7698
Case 12	1	2	1	2	2	102.1629	104.6396	111.1497	122.0844
Case 13	1	2	2	1	1	103.8957	108.4083	119.8071	132.2525
Case 14	1	2	2	1	2	102.2196	104.7863	111.5446	122.8166
Case 15	1	2	2	2	1	102.6762	105.7401	113.6897	125.3877
Case 16	1	2	2	2	2	101.8962	104.057	109.7659	120.0075
Case 17	2	1	1	1	1	113.2746	127.7798	151.736	156.1407
Case 18	2	1	1	1	2	103.1951	106.9067	116.4555	128.9195
Case 19	2	1	1	2	1	104.2253	109.056	121.1117	133.082
Case 20	2	1	1	2	2	102.3878	105.1224	112.2699	123.5974
Case 21	2	1	2	1	1	105.0528	110.8689	125.121	136.6174
Case 22	2	1	2	1	2	102.4847	105.3583	112.8746	124.5845
Case 23	2	1	2	2	1	103.0596	106.5587	115.548	127.5391
Case 24	2	1	2	2	2	102.05	104.3871	110.5386	121.1414
Case 25	2	2	1	1	1	104.9441	110.6572	124.71	136.4171
Case 26	2	2	1	1	2	102.4367	105.2626	112.6695	124.3765
Case 27	2	2	1	2	1	103.004	106.4484	115.3166	127.3392
Case 28	2	2	1	2	2	102.0163	104.3195	110.3911	120.9661
Case 29	2	2	2	1	1	103.2617	107.0385	116.7296	129.1377
Case30	2	2	2	1	2	102.0505	104.4146	110.6603	121.5213

Case 31	2	2	2	2	1	102.4298	105.2062	112.4501	123.7865
Case 32	2	2	2	2	2	101.7961	103.8378	109.2416	119.1766

Table: 8 Percent relative efficiency with respect to Singh et al. (2019), when rare non-sensitive unrelated attribute is unknown under stratified population.

	λ_{111}	λ_{121}	λ_{211}	λ_{221}	λ_{231}	(T,P) = (0.80,0.60) Q_1 = 0.15 Q_2 =0.35 Q_4 = 0.10 Q_5 =0.70	(0.70,0.50) 0.20 0.40 0.15 0.65	(0.60,0.40) 0.25 0.45 0.20 0.60	(0.50,0.30) 0.30 0.50 0.25 0.55
Case 1	1	1	1	1	1	575.7526	611.7634	509.2232	280.0789
Case 2	1	1	1	1	2	107.3981	115.3679	136.7298	175.8264
Case 3	1	1	1	2	1	109.5572	119.8155	146.0035	180.1744
Case 4	1	1	1	2	2	103.9625	108.4086	120.5136	145.9725
Case 5	1	1	2	1	1	114.7629	130.1074	167.7549	205.5738
Case 6	1	1	2	1	2	104.8076	110.1025	124.6044	155.9994
Case 7	1	1	2	2	1	105.6105	111.8138	128.3499	157.7915
Case 8	1	1	2	2	2	102.9886	106.4036	115.7325	136.1905
Case 9	1	2	1	1	1	111.1737	122.9665	153.2979	195.1239
Case 10	1	2	1	1	2	104.3331	109.1169	122.3217	152.6744
Case 11	1	2	1	2	1	104.9615	110.4662	125.3095	154.1131
Case 12	1	2	1	2	2	102.7848	105.9709	114.712	134.5784
Case 13	1	2	2	1	1	106.2571	113.0652	131.4484	167.0636
Case 14	1	2	2	1	2	103.236	106.8708	116.9415	141.2045
Case 15	1	2	2	2	1	103.5617	107.5917	118.5592	141.7063
Case 16	1	2	2	2	2	102.2328	104.8318	111.9429	128.0244
Case 17	2	1	1	1	1	122.012	144.1819	194.4524	223.1247
Case 18	2	1	1	1	2	105.4576	111.4286	127.7044	161.56
Case 19	2	1	1	2	1	106.528	113.6859	132.5902	163.9701
Case 20	2	1	1	2	2	103.2617	106.9666	117.0834	139.0769
Case 21	2	1	2	1	1	108.7947	118.2228	142.8988	181.1968
Case 22	2	1	2	1	2	103.859	108.1619	120.0103	147.016
Case 23	2	1	2	2	1	104.3485	109.2265	122.3739	147.9671
Case 24	2	1	2	2	2	102.5474	105.4927	113.5333	131.2656
Case 25	2	2	1	1	1	107.3582	115.2984	136.5222	174.6651
Case 26	2	2	1	1	2	103.5412	107.4965	118.4494	144.5634
Case 27	2	2	1	2	1	103.9403	108.3708	120.4053	145.3081
Case 28	2	2	1	2	2	102.3947	105.166	112.7581	130.0038
Case 29	2	2	2	1	1	104.7814	110.0578	124.4744	155.1951
Case30	2	2	2	1	2	102.747	105.8668	114.507	135.522

Case 31	2	2	2	2	1	102.9715	106.3746	115.6506	135.6712
Case 32	2	2	2	2	2	101.9587	104.2653	110.5559	124.5546

5. INTERPRETATIONS OF RESULTS

The following interpretations may be read out from Tables1-8:

1. From the Table 1- 8, it may be noticed that the obtained percent relative efficiencies of the proposed estimators are always found greater than 100 for all considered cases which shows that the proposed estimators are more efficient than Lee at al. (2014) and Singh et al. (2019) estimators.
2. From Table 1, 2, 6, the values of percent relative efficiencies are increasing for increasing values of T, Q_2 and decreasing values of P, Q_1.
3. From Table 3,4, the values of percent relative efficiencies do not follow any specific pattern for increasing values of T, Q_1, Q_2, Q_4 and decreasing values of P, Q_5.
4. From Table 5, the percent relative efficiencies do not follow any specific pattern for increasing values of T, Q_2 and decreasing values of P, Q_1.
5. From Table 7, 8, it may be seen that for increasing values of T, Q_1, Q_2, Q_4 and decreasing values of P, Q_5, the values of percent relative efficiencies are increasing except case 1.

6. CONCLUSIONS AND RECOMMENDATIONS

Following these results, it may be concluded that the proposed estimators based on randomized response model when characteristics under the study concerns to the stigmatized issues, are rewarding in the terms of percent relative efficiencies and dominate over Lee et al. (2014) and Singh et al. (2019) estimators. Thus, the suggested estimators in this work may be utilize effectively to handle the problems of untruthful response or non-response arises due to sensitive nature of characteristics and may be recommended to the survey practitioners for their practical uses.

Acknowledgements: Authors are thankful to Indian Institute of Technology (Indian School of Mines), Dhanbad for providing the necessary support to carry out the present research work, authors are also grateful to the honorable Reviewers, Editor for their valuable suggestions which helped us in improving the quality of the paper.

REFERENCES

[1] BOUZA-HERRERA, C. N. (2016): Behavior of Some Scrambled Randomized Response Models under Simple Random Sampling, Ranked Set Sampling and Rao–Hartley–Cochran Designs. In **Handbook of Statistics**, 209-220. Elsevier, Amsterdam.

[2] BOUZA-HERRERA, C. N. (2018): Data gathering, analysis and protection of privacy through randomized response techniques: qualitative and quantitative human traits. **Investigación Operacional,** *39*, 151-152.

[3] BOUZA, C. (2010): Behavior of a randomized response procedure under unequal selection model of Chaudhuri-Stenger for insensitive variables under ranked set sampling. **Advances and Applications in Statistical Sciences,** *4*, 1365-144.

[4] CHAUDHURi A. and MUKERJEE R., (1988): **Randomized Response: Theory and Techniques**, Marcel Dekker, New York.

[5] COCHRAN W G. (1977): **Sampling technique**. 3rd ed. John Wiley and Sons, New York.

[6] FOX J.A. and TRACY P.E., (1986): **Randomized response: A method for sensitive surveys**. Sage publications, Beverly Hills.

[7] GREENBERG, B. G., ABUL-ELA, A. L. A., SIMMONS, W. R., and HORVITZ, D. G., (1969): The unrelated question randomized response model: Theoretical framework. **Journal of the American Statistical Association,** 64, 520-539.

[8] KIM, J. M., and WARDE, W. D., (2005): Some new results on the multinomial randomized response model. **Communications in Statistics - Theory and Methods**, 34, 847-856.

[9] KIM, J. M., and ELAM, M. E., (2007): A stratified unrelated question randomized response model. **Statistical Papers**, 48, 215-233.

[10] LAND, M., SINGH, S., and SEDORY, S. A., (2012): Estimation of a rare sensitive attribute using Poisson distribution. **Statistics**, 46, 351-360.

[11] LEE, G. S., UHM, D., and KIM, J. M., (2014): Estimation of a rare sensitive attribute in probability proportional to size measures using Poisson distribution. **Statistics,** 48, 685-709.

[12] LEE, G. S., HONG, K. H., and SON, C. K., (2016): A stratified two-stage unrelated randomized response model for estimating a rare sensitive attribute based on the Poisson distribution. **Journal of Statistical Theory and Practice**, 10, 239-262.

[13] LEE, G. S., HONG, K. H., and SON, C. K., (2018): A new stratified three-stage unrelated randomized response model for estimating a rare sensitive attribute based on the Poisson distribution. **Communications in Statistics - Theory and Methods**, 1-26.

[14] MANGAT, N.S., SINGH, R. and SINGH, S., (1992): An improved unrelated question randomized response strategy. **Calcutta Statistical Association Bulletin,** 42, 277-281.

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

[16] SINGH, G. N., SINGH, C., and SUMAN, S. (2019): Estimation of a rare sensitive attribute for two-stage randomized response model in probability proportional to size sampling using Poisson probability distribution. **Statistics**, 53, 387-394.

[17] SINGH, H. P., and TARRAY, T. A., (2015): A revisit to the Singh, Horn, Singh and Mangat's randomization device for estimating a rare sensitive attribute using Poisson distribution. **Model Assisted Statistics and Applications**, 10, 129-138.

[18] SINGH, H. P., and T. A. Tarray. (2012): A Stratified unknown repeated trials in randomized response sampling. **Communications of the Korean Statistical Society** 19, 751–9.

[19] SINGH, S., HORN, S., SINGH, R. and MANGAT, N.S., (2003): On the use of modified randomization device for estimating the prevalence of a sensitive attribute. **Statistics in Transition**, 6, 515-522.

[20] TRACY, D. S., and OSAHAN, S. S. (1999). An improved randomized response technique. **Pakistan Journal of Statistics-All Series-**, 15, 1-6.

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