Estimation of a rare sensitive attribute by using Poisson probability distribution and consideration of respondent privacy protection in randomized response model

By Amod Kumar

Estimation of a rare sensitive attribute by using Poisson probability distribution and consideration of respondent privacy protection in randomized response model

G. N. Singh, Amod Kumar and Gajendra K. Vishwakarma

Department of Applied Mathematics,

Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India

Email: gnsingh_ism@yahoo.com, amod.ism01@gmail.com, vishwagk@rediffmail.com

Abstract

A new two-stage unrelated randomized response model is proposed for the estimation of mean number of individuals who possess a rare sensitive attribute in a given population by utilizing Poisson probability distribution, when the proportion of rare non-sensitive unrelated attribute is known and unknown. The properties of the proposed model are examined. The two-stage unrelated randomized response model provides more efficient estimator of the mean number of individuals in the population with sensitive attribute than the contemporary models. The procedure also introduces the measure of privacy protection of respondents and compares randomized response models in term of efficiency and privacy protection. Numerical illustrations are presented to support the theoretical results and suitable recommendations are put forward to the survey statisticians/practitioners.

Keywords: Poisson probability distribution, unrelated randomized response model, sensitive attribute, estimation of proportion, privacy protection.

Mathematics Subject Classification: 62D05

1. Introduction

Often social surveys include sensitive issues for enquiry which involves stigmatized subject such as use of illegal drugs, homosexuality, tax evasion, abortion et cetera which people do not like to disclose to others. Direct questions about sensitive issues often yield untruthful responses or non-response, which are major problems in sample surveys. Non-response/untruthful response introduces bias in survey estimates. To cope with such issues, Warner (1965) initiated a randomized response technique for estimating the proportion π_s of the population possessing a sensitive attribute. Warner (1965) provided a Bernoulli random device with outcomes possess a sensitive attribute or not to the respondents selected using simple random sampling with replacement for reporting whether is unobservable to the interviewer, favours their group or not. Greenberg et al. (1969) have modified the Warner (1965) model which addresses the two questions: one being the sensitive question and the other being unrelated non-sensitive and suggested an unrelated question randomized response model. Moors (1971), Cochran (1977), Fox and Tracy (1986), Chaudhuri and Mukherjee (1988), Hedayat and Sinha (1991), Ryu et al. (1993), Singh and Mangat (1996), Tracy and Mangat (1996), Tracy and Osahan (1999), Singh (2003), Singh et al. (2003), and Kim and Warde (2005) improved Warner (1965) procedure to obtain more efficient estimator of sensitive proportion. Mangat and Singh (1990), Kim et al. (1992), and Mangat (1992) have developed two-stage related and unrelated randomized response models to improve the efficiency of resultant estimates.

Land et al. (2012) suggested an estimator for estimating the mean number of persons who possess a rare sensitive attribute by utilizing the Poisson probability distribution. In their model, the randomized device consist two questions: (a) do you belongs to rare sensitive attribute A? and (b) do you belong to rare non-sensitive unrelated attribute Y? Lee et al. (2013), Lee et al. (2016), and Singh and Tarray (2014, 2015) modified Land et al. (2012)

model to estimate the mean number of individuals by utilizing Poisson probability distribution with respect to simple and stratified random sampling.

We consider the problem where the number of individuals who possess a rare sensitive attribute is very small and hence a large sample is required to estimate the mean number of such individuals. We propose a new two-stage unrelated randomized response model to obtain the more efficient estimator of the mean numbers of persons who possess a rare sensitive attribute by utilizing Poisson probability distribution. We discuss two situations, when the population proportion of the rare non-sensitive unrelated attribute is known and unknown. We shall examine the performance of the estimators in term of efficiency with respect to Land et al. (2012) and Singh and Tarray (2015) estimators. Numerical illustrations have been carried out to support the theoretical results along with recommendations to the survey statisticians/practitioners.

2. The Proposed Model

Singh et al. (2003) proposed an unrelated question model for estimating the population proportion π_s of sensitive attribute which is alternative to Greenberg et al. (1969) model in the sense that the randomized device used in Singh et al. (2003) model has three outcomes (i) "i belong to sensitive group A" with probability P_1, (ii) "i belong to non-sensitive group Y" with probability P_2, and (iii) "blank cards" with probability P_3 such that $\sum_{i=1}^{3} P_i = 1$. If blank card is drawn by the respondent, he/she will report no. The rest of the procedure remains as usual. The probability θ_s of a yes answer is:

$$\theta_s = P_1\pi_s + P_2\pi_y . \quad (1)$$

We extend Singh et al. (2003) model in two-stage unrelated randomized response model. A sample of size n is selected from a finite population of size N by using simple random sampling with replacement scheme. Each respondent selected in the sample is instructed to use the first stage randomized device R_1 which consist two statements (i) "i belong to sensitive group A" with probability T_1, and (ii) "Go to the randomized device R_2" with probability $(1-T_1)$. The second stage randomized device R_2 has three statements which is similar as in Singh et al. (2003) model. The probability θ_A of yes answers from the respondents using randomized devices R_1 and R_2 is:

$$\theta_A = T_1\pi_s + (1-T_1)\{P_1\pi_s + P_2\pi_y\}, \tag{2}$$

where π_s denote the true proportion of yes answer from the rare sensitive group and π_y denote the true proportion of yes answer from the rare non-sensitive unrelated group in the population.

An estimator of the population proportion π_s is:

$$\hat{\pi}_A = \frac{\hat{\theta}_A - (1-T_1)P_2\pi_y}{[T_1 + (1-T_1)P_1]}, \tag{3}$$

where $\hat{\theta}_A$ is the sample proportion of yes response. The expected value of $\hat{\pi}_A$ is:

$$E(\hat{\pi}_A) = \frac{E(\hat{\theta}_A) - (1-T_1)P_2\pi_y}{[T_1 + (1-T_1)P_1]} = \pi_s. \tag{4}$$

with variance

$$V(\hat{\pi}_A) = \frac{\pi_s(1-\pi_s)}{n} + \frac{\pi_s[1-\{T_1+(1-T_1)P_1\} - 2P_2(1-T_1)\pi_y]}{n\{T_1+(1-T_1)P_1\}} + \frac{P_2(1-T_1)\pi_y[1-(1-T_1)P_2\pi_y]}{n\{T_1+(1-T_1)P_1\}^2} \tag{5}$$

Substituting $T_1 = 0$ in Eq. (2), the proposed model reduces to Singh et al. (2003) unrelated randomized response model and when $T_1 = 0$ and $P_2 = (1 - P_1)$, the proposed model reduces to Greenberg et al. (1969) unrelated randomized response model.

Using Eq. (2) we estimate mean number of individuals who possess a rare sensitive attribute by utilizing Poisson probability distribution, when the proportion of the rare non-sensitive unrelated attribute is known and unknown.

2.1. Estimation of rare sensitive attribute when the proportion of a rare non-sensitive unrelated attribute is known:

Let π_s be the true proportion of the rare sensitive attribute A in a finite population of size N. For examples, the proportion of AIDS patients who continue having affairs with strangers, the proportion of persons who have witnessed a murder and the proportion of persons who are told by their doctors that they will not survive long due to a ghastly disease et cetera. Since the attribute under consideration is rare in nature, therefore a large sample of size n (say $n \to \infty$) is drawn from the population using simple random sample with replacement scheme such that $n\pi_s = \lambda_s > 0$ as $\pi_s \to 0$. π_y is the true proportion of the population having the rare non-sensitive unrelated attribute Y such that for $n \to \infty$ and $\pi_y \to 0$, we have $n\pi_y = \lambda_y$ $(\lambda_y > 0)$ which is known. The probability θ_A of obtaining yes answer in the proposed procedure is:

$$\theta_A = T_1\pi_s + (1 - T_1)\{P_1\pi_s + P_2\pi_y\}. \quad (6)$$

As both attributes A and Y are very rare in the population, hence for $n \to \infty$ and $\theta_A \to 0$, we have $n\theta_A = \lambda_A$ (finite), where

$$\lambda_A = T_1\lambda_s + (1 - T_1)\{P_1\lambda_s + P_2\lambda_y\}. \quad (7)$$

Let $y_1, y_2, ..., y_n$ be a random sample of n observations from the Poisson probability distribution with parameter λ_A. The likelihood function of the random sample of n observations is given as:

$$L(y_1, y_2, ..., y_n / \lambda_A) = \prod_{i=1}^{n} \frac{exp^{-\lambda_A} \lambda_A^{y_i}}{y_i!}. \quad (8)$$

Taking natural logarithm on both sides of Eq. (8) and setting $\partial logL / \partial \lambda_s$ to zero, the maximum likelihood estimator of λ_s is:

$$\hat{\lambda}_s = \frac{1}{\left[P_1 + T_1(1 - P_1)\right]}\left[\frac{1}{n}\sum_{i=1}^{n} y_i - P_2(1 - T_1)\lambda_y\right]. \quad (9)$$

We have the following theorems.

Theorem 1. The estimator $\hat{\lambda}_s$ is an unbiased estimator of the parameter λ_s.

$$E(\hat{\lambda}_s) = \lambda_s . \quad (10)$$

Proof. Since $y_1, y_2, ..., y_n$ are independent and identically distributed Poisson varieties with parameter λ_A, we have

$$E(\hat{\lambda}_s) = E\left[\frac{1}{\left[P_1 + T_1(1 - P_1)\right]}\left\{\frac{1}{n}\sum_{i=1}^{n} y_i - P_2(1 - T_1)\lambda_y\right\}\right]$$

$$= \frac{1}{\left[P_1 + T_1(1 - P_1)\right]}\left\{\frac{1}{n}\sum_{i=1}^{n} E(y_i) - P_2(1 - T_1)\lambda_y\right\}$$

$$= \frac{1}{\left[P_1 + T_1(1 - P_1)\right]}\left\{\frac{1}{n}\sum_{i=1}^{n} \lambda_A - P_2(1 - T_1)\lambda_y\right\}$$

$$= \lambda_s,$$

which proves the theorem.

Theorem 2. The variance of the unbiased estimator $\hat{\lambda}_s$ is:

$$V\left(\hat{\lambda}_s\right)=\frac{\lambda_s}{n\left[P_1+T_1\left(1-P_1\right)\right]}+\frac{P_2\left(1-T_1\right)\lambda_y}{n\left[P_1+T_1\left(1-P_1\right)\right]^2}. \qquad (11)$$

Proof. Since $y_1, y_2,..., y_n$ are independent and identically distributed Poisson varieties with parameter λ_A and samples are drawn independently, we have

$$V\left(\hat{\lambda}_s\right)=V\left[\frac{1}{\left[P_1+T_1\left(1-P_1\right)\right]}\left\{\frac{1}{n}\sum_{i=1}^{n}y_i-P_2\left(1-T_1\right)\lambda_y\right\}\right]$$

$$=\frac{1}{n^2\left[P_1+T_1\left(1-P_1\right)\right]^2}\sum_{i=1}^{n}V\left(y_i\right)$$

$$=\frac{1}{n^2\left[P_1+T_1\left(1-P_1\right)\right]^2}\sum_{i=1}^{n}\lambda_A$$

$$=\frac{\lambda_s}{n\left[P_1+T_1\left(1-P_1\right)\right]}+\frac{P_2\left(1-T_1\right)\lambda_y}{n\left[P_1+T_1\left(1-P_1\right)\right]^2},$$

hence, it is proved.

Theorem 3. An unbiased estimator of the variance of the estimator $\hat{\lambda}_s$ is:

$$\hat{V}\left(\hat{\lambda}_s\right)=\frac{1}{n^2\left[P_1+T_1\left(1-P_1\right)\right]^2}\left(\sum_{i=1}^{n}y_i\right). \qquad (12)$$

Proof. Taking the expectation on both sides of Eq. (12), we have

$$E\left[\hat{V}\left(\hat{\lambda}_s\right)\right]=\frac{1}{n^2\left[P_1+T_1\left(1-P_1\right)\right]^2}E\left(\sum_{i=1}^{n}y_i\right)$$

$$=\frac{1}{n^2\left[P_1+T_1\left(1-P_1\right)\right]^2}\left(\sum_{i=1}^{n}\lambda_A\right)$$

$$=\frac{\lambda_s}{n\left[P_1+T_1\left(1-P_1\right)\right]}+\frac{P_2\left(1-T_1\right)\lambda_y}{n\left[P_1+T_1\left(1-P_1\right)\right]^2},$$

which completes the proof.

Substituting $T_1 = 0$ and $P_2 = (1 - P_1)$ in Eq. (6), the proposed model reduces to Land et al. (2012) unrelated randomized response model and for $T_1 = 0$, the proposed model reduces to Singh and Tarray (2015) model.

When the proportion of a rare non-sensitive unrelated attribute was known, Land et al. (2012) and Singh and Tarray (2015) showed that the variances of unbiased estimators $\hat{\lambda}_L$ and $\hat{\lambda}_{ST}$ by using Poisson probability distribution are:

$$V\left(\hat{\lambda}_L\right) = \frac{\lambda_s}{nP_1} + \frac{(1 - P_1)\lambda_y}{nP_1^2}. \tag{13}$$

and

$$V\left(\hat{\lambda}_{ST}\right) = \frac{\lambda_s}{nP_1} + \frac{P_2\lambda_y}{nP_1^2}. \tag{14}$$

2.2. Comparison of Efficiency

The estimator $\hat{\lambda}_s$ is always more efficient than that of Land et al. (2012) estimator $\hat{\lambda}_L$ if

$$V\left(\hat{\lambda}_L\right) > V\left(\hat{\lambda}_s\right),$$

Which gives the condition, when

$$(1 - P_1)\left[\lambda_y\left\{1 + \frac{T_1(1 - P_1)}{P_1}\right\} + T_1\lambda_s\right] - \left[\frac{P_1P_2(1 - T_1)\lambda_y}{P_1 + T_1(1 - P_1)}\right] > 0,$$

The estimator $\hat{\lambda}_s$ is always more efficient than that of Singh and Tarray (2015) estimator $\hat{\lambda}_{ST}$ if

$$V\left(\hat{\lambda}_{ST}\right) > V\left(\hat{\lambda}_s\right),$$

Which is true, if

$$(1 - P_1)T_1\lambda_s + P_2\lambda_y\left[1 + \frac{T_1(1 - P_1)}{P_1} - \frac{P_1(1 - T_1)}{P_1 + T_1(1 - P_1)}\right] > 0,$$

To have a tangible idea about the efficacious performance of the estimator $\hat{\lambda}_s$ over Land et al. (2012) and Singh and Tarray (2015) estimators, we compute the percent relative efficiencies (PREs) PRE $(\hat{\lambda}_s, \hat{\lambda}_L)$ and PRE $(\hat{\lambda}_s, \hat{\lambda}_{ST})$ for different choices of parametric combinations. We vary λ_s and λ_y from 0.5 to 1.5 by step of 0.5; T_1 from 0.9 to 0.1 by decrement 0.2. The percent relative efficiencies presented in Tables 1 and 2 are calculated for different values of probabilities P_1, P_2 and P_3. Tables 1 and 2 present the cases where the estimator $\hat{\lambda}_s$ is more efficient than the usual estimators. The percent relative efficiencies calculate by utilizing the following formulas.

$$PRE\left(\hat{\lambda}_s, \hat{\lambda}_L\right) = \frac{V\left(\hat{\lambda}_L\right)}{V\left(\hat{\lambda}_s\right)} \times 100. \tag{15}$$

and

$$PRE\left(\hat{\lambda}_s, \hat{\lambda}_{ST}\right) = \frac{V\left(\hat{\lambda}_{ST}\right)}{V\left(\hat{\lambda}_s\right)} \times 100. \tag{16}$$

The results, which are worth discussing, are presented in the following points:

(i). Tables 1 and 2 showed that for all the parametric combinations the values of percent relative efficiencies are substantially exceeding 100, which indicate that the estimator $\hat{\lambda}_s$ is uniformly better than Land et al. (2012) and Singh and Tarray (2015) estimators.

(ii). However, the values of percent relative efficiencies decreasing with the decrease in the values of T_1 while the values of P_1, P_2, P_3, λ_s and λ_y are fixed.

(iii). Tables 1 and 2 visible that the values of percent relative efficiencies are showing increasing trend with the increasing values of λ_y.

(iv). It may also be seen that with the increase in the values of λ_s there is a decreasing pattern in the values of percent relative efficiencies, which is an obvious phenomena.

TABLE 1: Percent relative efficiency of the proposed estimator $\hat{\lambda}_s$ with respect to Land et al. (2012) estimator $\hat{\lambda}_L$, when the proportion of the rare non-sensitive unrelated attribute is known

P_1	P_2	P_3	λ_s	λ_y	T_1				
					0.9	0.7	0.5	0.3	0.1
0.60	0.20	0.20	0.50	0.50	261.22	228.84	197.53	167.44	138.75
				1.00	358.40	301.15	248.88	201.60	159.28
				1.50	451.76	365.28	290.90	227.36	173.55
			1.00	0.50	211.13	189.10	167.32	145.82	124.68
				1.00	261.22	228.84	197.53	167.44	138.75
				1.50	310.30	266.11	224.56	185.80	150.03
			1.50	0.50	194.20	175.27	156.44	137.73	119.19
				1.00	227.94	202.64	177.77	153.44	129.74
				1.50	261.22	228.84	197.53	167.44	138.75
0.70	0.15	0.15	0.50	0.50	194.94	176.96	159.40	142.30	125.72
				1.00	249.62	219.70	191.68	165.57	141.38
				1.50	302.69	258.75	219.45	184.42	153.31
			1.00	0.50	166.97	154.04	141.21	128.50	115.91
				1.00	194.94	176.96	159.40	142.30	125.72
				1.50	222.49	198.82	176.17	154.58	134.12
			1.50	0.50	157.55	146.16	134.81	123.50	112.26
				1.00	176.34	161.80	147.44	133.29	119.36
				1.50	194.94	176.96	159.40	142.30	125.72

continue

0.80	0.10	0.10	0.50	0.50	151.57	142.33	133.22	124.26	115.45
				1.00	180.07	165.67	151.87	138.67	126.07
				1.50	208.00	187.65	168.75	151.20	134.94
			1.00	0.50	137.11	130.11	123.14	116.20	109.31
				1.00	151.57	142.33	133.22	124.26	115.45
				1.50	165.89	154.18	142.78	131.72	121.01
			1.50	0.50	132.25	125.95	119.65	113.38	107.12
				1.00	141.95	134.22	126.56	118.96	111.42
				1.50	151.57	142.33	133.22	124.26	115.45

FIGURE 1. Pictorial representation of percent relative efficiency of the proposed estimator $\hat{\lambda}_s$ with respect to Land et al. (2012) estimator $\hat{\lambda}_L$, when $P_1 = 0.60, 0.70, 0.80$ and $\lambda_y = 0.50$.

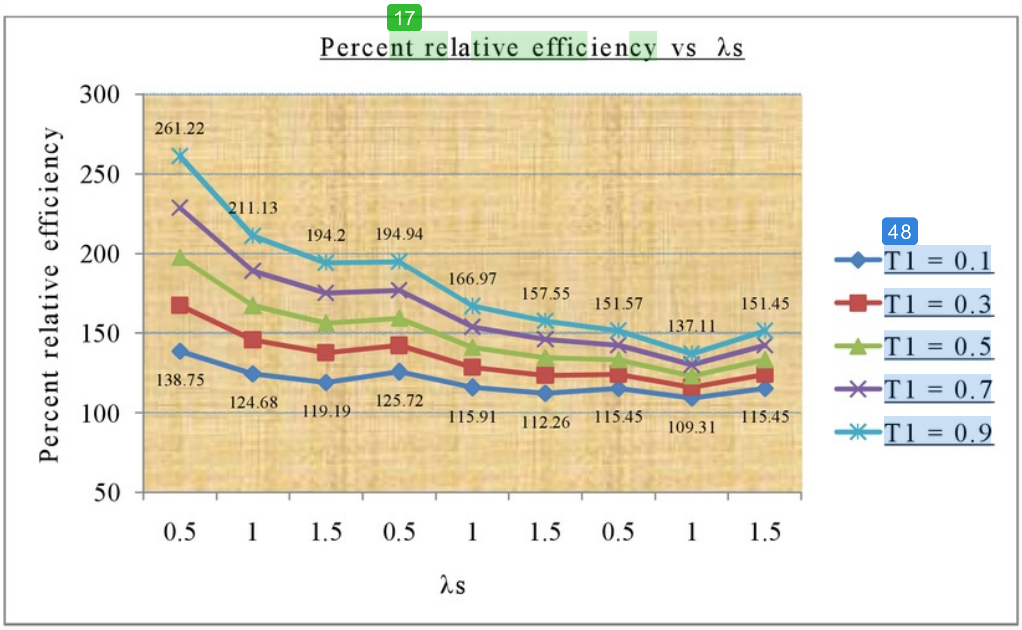


TABLE 2: Percent relative efficiency of the proposed estimator $\hat{\lambda}_s$ with respect to Singh and Tarray (2015) estimator $\hat{\lambda}_{ST}$, when the proportion of the rare non-sensitive unrelated attribute is known

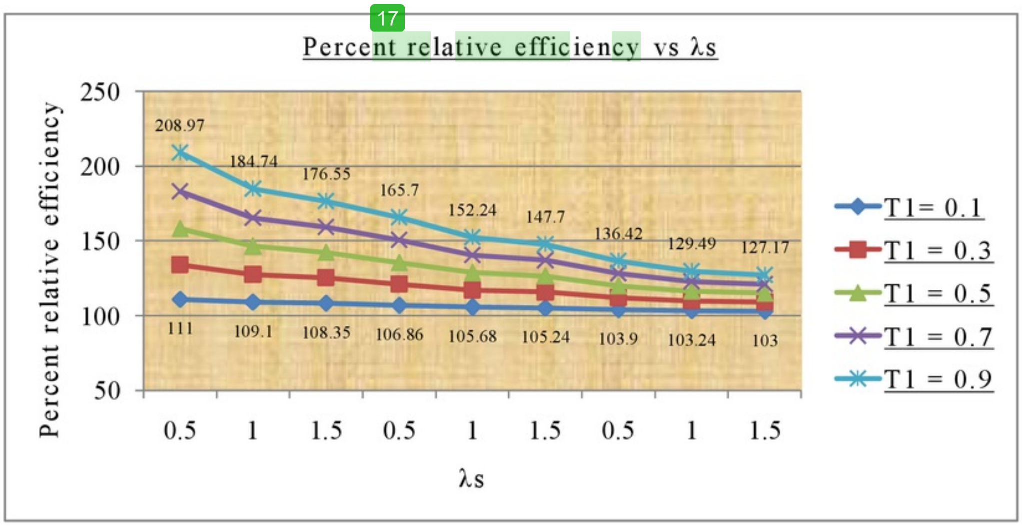
P_1	P_2	P_3	λ_s	λ_y	T_1				
					0.9	0.7	0.5	0.3	0.1
0.60	0.20	0.20	0.50	0.50	208.97	183.07	158.02	133.95	111.00
				1.00	256.00	215.11	177.77	144.00	113.77
				1.50	301.17	243.52	193.93	151.57	115.70
			1.00	0.50	184.74	165.47	146.40	127.59	109.10
				1.00	208.97	183.07	158.02	133.95	111.00
				1.50	232.72	199.58	168.42	139.35	112.52
			1.50	0.50	176.55	159.34	142.22	125.21	108.35
				1.00	192.87	171.46	150.42	129.83	109.78
				1.50	208.97	183.07	158.02	133.95	111.00
0.70	0.15	0.15	0.50	0.50	165.70	150.41	135.49	120.96	106.86
				1.00	192.02	169.00	147.44	127.36	108.75
				1.50	217.56	185.98	157.73	132.55	110.19
			1.00	0.50	152.24	140.45	128.75	117.16	105.68
				1.00	165.70	150.41	135.49	120.96	106.86
				1.50	178.96	159.92	141.70	124.34	107.88
			1.50	0.50	147.70	137.02	126.38	115.78	105.24
				1.00	156.75	143.82	131.06	118.48	106.10
				1.50	165.70	150.41	135.49	120.96	106.86

continue

0.80	0.10	0.10	0.50	0.50	136.42	128.09	119.90	111.83	103.90
				1.00	150.06	138.06	126.56	115.56	105.06
				1.50	163.43	147.44	132.58	118.80	106.02
			1.00	0.50	129.49	122.88	116.30	109.75	103.24
				1.00	136.42	128.09	119.90	111.83	103.90
				1.50	143.27	133.15	123.31	113.76	104.51
			1.50	0.50	127.17	121.10	115.05	109.02	103.00
				1.00	131.81	124.63	117.52	110.46	103.47
				1.50	136.42	128.09	119.90	111.83	103.90

From Table 1, if λ_s is equal to 0.50; then there are 45 choices of the parameters $P_i\,(i = 1, 2, 3)$, T and λ_y such that the estimator $\hat{\lambda}_s$ is more efficient. These choices result in a percent relative efficiencies level from minimum 115.45 to a maximum 451.76. At the same time for $\lambda_s = 0.50$, Table 2 shows that the percent relative efficiencies value of the retain results range between 103.90 and 301.17. Figures 1 and 2 present the values of percent relative efficiencies with corresponding $P_i\,(i = 1, 2, 3)$ and T for the fixed value of $\lambda_y = 0.50$, and for all the values of $0.50 \le \lambda_s \le 1.50$ with a step 0.50. A close look on Figures 1 and 2 indicate that the values in Figure 1 remain greater than Figure 2 for all situations. We conclude that the estimator $\hat{\lambda}_s$ with respect to Singh and Tarray (2015) estimator are outperformed by the estimator $\hat{\lambda}_s$ with respect to Land et al. (2012) estimator. Similar results observe for other practicable parameters. Further, based on Figures 1 and 2, an investigator could make a choice of parameters such that the more efficiency expected from the proposed model than from Land et al. (2012) and Singh and Tarray (2015) models. The rest of the results can be read out from the Tables 1 and 2.

FIGURE 2. Pictorial representation of percent relative efficiency of the proposed estimator $\hat{\lambda}_s$ with respect to Singh and Tarray (2015) estimator $\hat{\lambda}_{ST}$, when $P_1 = 0.60, 0.70, 0.80$ and $\lambda_y = 0.50$.



3. Estimation of rare sensitive attribute when the proportion of a rare non-sensitive unrelated attribute is unknown:

A large sample of size n respondents is drawn from the population by using simple random sampling with replacement scheme and two random devices $D_i\left(R_{i1}, R_{i2}\right)$, $i = 1, 2$ are provided to the respondents. Each respondent selected in the sample is asked to reply yes or no answer by using the following two-stage randomized device. The randomized device R_{11} of D_1 used in the first-stage presents two statements (i) "I belong to sensitive group A" with probability T_1 and (ii) "Go to the randomized device R_{12}" with probability $(1 - T_1)$. In second-stage the randomized device R_{21} of D_1 presents three statements (i) "I belong to sensitive group A" with probability P_1, (ii) "I belong to rare unrelated group Y" with probability P_2, and (iii) "Blank cards" with probability P_3 such that $\sum_{i=1}^{3} P_i = 1$. Again, the second random device D_2 follows the same procedures as described for the random device D_1 with alternative probabilities T_2 and $(1-T_2)$ in first-stage and P_4, P_5 and P_6 in second-stage such that $\sum_{i=4}^{6} P_i = 1$. Following, these two random devices $D_i\,(i=1,2)$, the probabilities of yes answers are:

$$\theta_1 = T_1\pi_s + (1-T_1)\{P_1\pi_s + P_2\pi_y\}, \tag{17}$$

and

$$\theta_2 = T_2\pi_s + (1-T_2)\{P_4\pi_s + P_5\pi_y\}, \tag{18}$$

where π_s is the true population proportion of the rare sensitive attribute A and π_y is the true population proportion of the rare non-sensitive unrelated attribute Y.

Since A and Y are very rare attributes, we set $n\theta_1 = \lambda_1^*$ and $n\theta_2 = \lambda_2^*$ are finite, assuming that, as $n \to \infty$, $\theta_1 \to 0$ and $\theta_2 \to 0$. By following Section 2, we obtain

$$\left[P_1 + T_1(1-P_1)\right]\hat{\lambda}_s + P_2(1-T_1)\hat{\lambda}_y = \frac{1}{n}\sum_{i=1}^{n} y_{1i}, \tag{19}$$

$$\left[P_4 + T_2(1-P_4)\right]\hat{\lambda}_s + P_5(1-T_2)\hat{\lambda}_y = \frac{1}{n}\sum_{i=1}^{n} y_{2i}, \tag{20}$$

where y_{1i} denote the first response and y_{2i} denote the second response from the i^{th} respondent.

Solving Eq. (19) and (20), we get the unbiased estimators of the λ_s and λ_y as:

$$\hat{\lambda}_s^* = \frac{1}{nA}\left[P_5(1-T_2)\sum_{i=1}^{n} y_{1i} - P_2(1-T_1)\sum_{i=1}^{n} y_{2i}\right], \tag{21}$$

and

$$\hat{\lambda}_y^* = \frac{1}{nB}\left[\{P_4 + T_2(1-P_4)\}\sum_{i=1}^{n} y_{1i} - \{P_1 + T_1(1-P_1)\}\sum_{i=1}^{n} y_{2i}\right], \tag{22}$$

where

$$A = \left(P_1P_5 - P_1P_5T_1 + P_5T_1\right)\left(1 - T_2\right) - \left(P_2P_4 - P_2P_4T_2 + P_2T_2\right)\left(1 - T_1\right),$$

$$\left(P_1P_5 - P_1P_5T_1 + P_5T_1\right)\left(1 - T_2\right) \neq \left(P_2P_4 - P_2P_4T_2 + P_2T_2\right)\left(1 - T_1\right).$$

and

$$B = \left(P_2P_4 - P_2P_4T_2 + P_2T_2\right)\left(1 - T_1\right) - \left(P_1P_5 - P_1P_5T_1 + P_5T_1\right)\left(1 - T_2\right),$$

$$\left(P_2P_4 - P_2P_4T_2 + P_2T_2\right)\left(1 - T_1\right) \neq \left(P_1P_5 - P_1P_5T_1 + P_5T_1\right)\left(1 - T_2\right).$$

Thus, we establish the following theorems.

Theorem 4. The estimator $\hat{\lambda}_s^*$ is an unbiased estimator of the parameter λ_s; the estimator $\hat{\lambda}_y^*$ is an unbiased estimator of the parameter λ_y that is

$$E\left(\hat{\lambda}_s^*\right) = \lambda_s. \tag{23}$$

and

$$E\left(\hat{\lambda}_y^*\right) = \lambda_y. \tag{24}$$

Proof. Since $y_{11}, y_{12}, \dots, y_{1n}$ is independent and identically distributed Poisson varieties with the parameter λ_1^*; $y_{21}, y_{22}, \dots, y_{2n}$ is independent and identically distributed Poisson varieties with the parameter λ_2^*, we have

$$E\left(\hat{\lambda}_s^*\right) = \frac{1}{nA} E\left[P_5\left(1 - T_2\right)\sum_{i=1}^{n} y_{1i} - P_2\left(1 - T_1\right)\sum_{i=1}^{n} y_{2i}\right]$$

$$= \frac{1}{nA}\left[P_5\left(1 - T_2\right)\sum_{i=1}^{n} E\left(y_{1i}\right) - P_2\left(1 - T_1\right)\sum_{i=1}^{n} E\left(y_{2i}\right)\right]$$

$$= \frac{1}{nA}\left[P_5\left(1 - T_2\right)\sum_{i=1}^{n} \lambda_1^* - P_2\left(1 - T_1\right)\sum_{i=1}^{n} \lambda_2^*\right]$$

$$= \frac{nA\lambda_s}{nA} = \lambda_s.$$

Similarly for the rare non-sensitive unrelated attribute, we obtain

$$E\left(\hat{\lambda}_y^*\right) = \frac{nB\lambda_y}{nB} = \lambda_y .$$

Thus the theorem is proved.

Theorem 5. The variances of the estimators $\hat{\lambda}_s^*$ and $\hat{\lambda}_y^*$ for the rare sensitive and rare non-sensitive unrelated attributes A and Y, are

$$\begin{aligned} V\left(\hat{\lambda}_s^*\right) = & \frac{\left[\left\{P_1+T_1\left(1-P_1\right)\right\}\left(1-T_2\right)^2 P_5^2+\left\{P_4+T_2\left(1-P_4\right)\right\}\left(1-T_1\right)^2 P_2^2\right]\lambda_s}{nA^2} \\ & -2\frac{P_2P_5\left(1-T_1\right)\left(1-T_2\right)\left\{P_1+T_1\left(1-P_1\right)\right\}\left\{P_4+T_2\left(1-P_4\right)\right\}\lambda_s}{nA^2} \\ & +\frac{P_2P_5\left(1-T_1\right)\left(1-T_2\right)\left\{P_2\left(1-T_1\right)+P_5\left(1-T_2\right)\right\}\lambda_y}{nA^2}-2\frac{P_2^2P_5^2\left(1-T_1\right)^2\left(1-T_2\right)^2\lambda_y}{nA^2} \end{aligned} \quad . \tag{25}$$

and

$$\begin{aligned} V\left(\hat{\lambda}_y^*\right) = & \frac{\left\{P_1+T_1\left(1-P_1\right)\right\}\left\{P_4+T_2\left(1-P_4\right)\right\}\left[\left\{P_1+T_1\left(1-P_1\right)\right\}+\left\{P_4+T_2\left(1-P_4\right)\right\}\right]\lambda_s}{nB^2} \\ & -2\frac{\left\{P_1+T_1\left(1-P_1\right)\right\}^2\left\{P_4+T_2\left(1-P_4\right)\right\}^2\lambda_s}{nB^2} \\ & +\frac{\left[\left\{P_1+T_1\left(1-P_1\right)\right\}^2 P_5\left(1-T_2\right)+\left\{P_4+T_2\left(1-P_4\right)\right\}^2 P_2\left(1-T_1\right)\right]\lambda_y}{nB^2} \\ & -2\frac{P_2P_5\left(1-T_1\right)\left(1-T_2\right)\left\{P_1+T_1\left(1-P_1\right)\right\}\left\{P_4+T_2\left(1-P_4\right)\right\}\lambda_y}{nB^2} \end{aligned} \quad . \tag{26}$$

Proof. Since $y_{11}, y_{12},..., y_{1n}$ is independent and identically distributed Poisson varieties with the parameter λ_1^*; $y_{21}, y_{22},..., y_{2n}$ is independent and identically distributed Poisson varieties with the parameter λ_2^* and both responses are not independent, we have

$$\begin{aligned} V\left(\hat{\lambda}_s^*\right) &= V\left[\frac{1}{nA}\left\{P_5\left(1-T_2\right)\sum_{i=1}^{n} y_{1i} - P_2\left(1-T_1\right)\sum_{i=1}^{n} y_{2i}\right\}\right] \\ &= \frac{1}{n^2A^2}\left\{P_5^2\left(1-T_2\right)^2\sum_{i=1}^{n} V\left(y_{1i}\right)+P_2^2\left(1-T_1\right)^2\sum_{i=1}^{n} V\left(y_{2i}\right)\right. \\ & \qquad \left. -2P_2P_5\left(1-T_1\right)\left(1-T_2\right)\sum_{i=1}^{n} Cov\left(y_{1i}, y_{2i}\right)\right\} \end{aligned}$$

$$= \frac{1}{n^2 A^2} \left\{ P_5^2 (1 - T_2)^2 \sum_{i=1}^{n} \lambda_1^* + P_2^2 (1 - T_1)^2 \sum_{i=1}^{n} \lambda_2^* - 2 P_2 P_5 (1 - T_1)(1 - T_2) \sum_{i=1}^{n} \lambda_{12}^* \right\}, \tag{27}$$

where

$$\lambda_1^* = V(y_{1i}) = E(y_{1i}) = \left[P_1 + T_1 (1 - P_1) \right] \lambda_s + P_2 (1 - T_1) \lambda_y, \tag{28}$$

$$\lambda_2^* = V(y_{2i}) = E(y_{2i}) = \left[P_4 + T_2 (1 - P_4) \right] \lambda_s + P_5 (1 - T_2) \lambda_y \tag{29}$$

and

$$\lambda_{12}^* = Cov(y_{1i}, y_{2i}) = E(y_{1i} y_{2i}) - E(y_{1i}) E(y_{2i})$$

$$= \{P_1 + T_1(1 - P_1)\}\{P_4 + T_2(1 - P_4)\}(\lambda_s^2 + \lambda_s) + P_2(1 - T_1) P_5 (1 - T_2)(\lambda_y^2 + \lambda_y)$$
$$+ \{P_1 + T_1(1 - P_1)\} P_5 (1 - T_2) \lambda_s \lambda_y + \{P_4 + T_2(1 - P_4)\} P_2 (1 - T_1) \lambda_s \lambda_y$$
$$- \left[\{P_1 + T_1(1 - P_1)\} \lambda_s + P_2(1 - T_1) \lambda_y \right] \left[\{P_4 + T_2(1 - P_4)\} \lambda_s + P_5 (1 - T_2) \lambda_y \right]$$

$$= \{P_1 + T_1(1 - P_1)\}\{P_4 + T_2(1 - P_4)\} \lambda_s + P_2 P_5 (1 - T_1)(1 - T_2) \lambda_y. \tag{30}$$

Substituting the Eq. (28), (29) and (30) in Eq. (27), we get the variance of the estimator $\hat{\lambda}_s^*$ as given in Eq. (25). Following the same procedure as in theorem 5, we can get the variance of the estimator $\hat{\lambda}_y^*$ as given in Eq. (26).

The unbiased estimators of the variances of the estimators $\hat{\lambda}_s^*$ and $\hat{\lambda}_y^*$ are

$$\hat{V}(\hat{\lambda}_s^*) = \frac{\left[\{P_1 + T_1(1 - P_1)\}(1 - T_2)^2 P_5^2 + \{P_4 + T_2(1 - P_4)\}(1 - T_1)^2 P_2^2 \right] \hat{\lambda}_s}{nA^2}$$
$$- 2 \frac{P_2 P_5 (1 - T_1)(1 - T_2)\{P_1 + T_1(1 - P_1)\}\{P_4 + T_2(1 - P_4)\} \hat{\lambda}_s}{nA^2}$$
$$+ \frac{P_2 P_5 (1 - T_1)(1 - T_2)\{P_2(1 - T_1) + P_5(1 - T_2)\} \hat{\lambda}_y}{nA^2} - 2 \frac{P_2^2 P_5^2 (1 - T_1)^2 (1 - T_2)^2 \hat{\lambda}_y}{nA^2}.$$

and

$$\hat{V}\left(\hat{\lambda}_y^*\right)=\frac{\{P_1+T_1(1-P_1)\}\{P_4+T_2(1-P_4)\}\left[\{P_1+T_1(1-P_1)\}+\{P_4+T_2(1-P_4)\}\right]\hat{\lambda}_s}{nB^2}$$
$$-2\frac{\{P_1+T_1(1-P_1)\}^2\{P_4+T_2(1-P_4)\}^2\hat{\lambda}_s}{nB^2}$$
$$+\frac{\left[\{P_1+T_1(1-P_1)\}^2P_5(1-T_2)+\{P_4+T_2(1-P_4)\}^2P_2(1-T_1)\right]\hat{\lambda}_y}{nB^2}$$
$$-2\frac{P_2P_5(1-T_1)(1-T_2)\{P_1+T_1(1-P_1)\}\{P_4+T_2(1-P_4)\}\hat{\lambda}_y}{nB^2}$$

When the proportion of a rare non-sensitive unrelated attributes was unknown, Land et al. (2012) and Singh and Tarray (2015) derived the variances of the unbiased estimators $\hat{\lambda}_{IL}$ and $\hat{\lambda}_{IST}$ by using Poisson probability distribution as:

$$V\left(\hat{\lambda}_{IL}\right)=\frac{1}{n(P_1-P_4)^2}\left[\left\{P_1(1-P_4)^2+P_4(1-P_1)^2-2P_1P_4(1-P_1)(1-P_4)\right\}\lambda_s+\left\{(1-P_1)(1-P_4)(2-P_1-P_4)-2(1-P_1)^2(1-P_4)^2\right\}\lambda_y\right]. \quad (31)$$

and

$$V\left(\hat{\lambda}_{IST}\right)=\frac{1}{n(P_1P_5-P_2P_4)^2}\left[\left\{P_1P_5^2+P_4P_2^2-2P_1P_2P_4P_5\right\}\lambda_s+\left\{P_2P_5^2+P_5P_2^2-2P_2^2P_5^2\right\}\lambda_y\right]. \quad (32)$$

4. Comparison of Efficiency

Eq. (25), (31), and (32) yield no clear analytical comparison between $\hat{\lambda}_s^*$ and $\hat{\lambda}_{IL}$ and between $\hat{\lambda}_s^*$ and $\hat{\lambda}_{IST}$. We compared the estimator $\hat{\lambda}_s^*$ with respect to Land et al. (2012) estimator $\hat{\lambda}_{IL}$ and Singh and Tarray (2015) estimator $\hat{\lambda}_{IST}$ in term of percent relative efficiencies (PREs). The formulas of percent relative efficiencies are:

$$PRE\left(\hat{\lambda}_s^*,\hat{\lambda}_{IL}\right)=\frac{V\left(\hat{\lambda}_{IL}\right)}{V\left(\hat{\lambda}_s^*\right)}\times 100. \quad (33)$$

and

$$PRE\left(\hat{\lambda}_s^*, \hat{\lambda}_{IST}\right) = \frac{V\left(\hat{\lambda}_{IST}\right)}{V\left(\hat{\lambda}_s^*\right)} \times 100 . \quad (34)$$

In this case we consider that the parametric values in first random device are same as given in subsection 2.2 but we change the parametric values of T_1 and the second random device as presented in Tables 3 and 4.

The following interpretation may be read-out from the present points.

(i). From Tables 3 and 4 clear that for all parametric combination the values of percent relative efficiency are substantially exceeding 100, which indicate that the estimator $\hat{\lambda}_s^*$ is uniformly dominating over Land et al. (2012) and Singh and Tarray (2015) estimators.

(ii). Further when the values of T_1 of presenting the question related to rare sensitive attribute in first-stage random device is increasing from 0.5-0.9, we observe the increase in the values of percent relative efficiency while we observe the zig-zag trend for the values of T_2.

One observation that can be made from Figures 3-4 are that as the values of λ_s change from 0.50 to1.50 with corresponding $\lambda_y = 0.50$ and $P_1 = 0.60$, the percent relative efficiency notably differ very much from each other, but when P_1 is closer to one, then the values of percent relative efficiency does not differ much. We report results only for the choice of $\lambda_y = 0.50$. Other results can be easily obtained by changing the values of parameters. However, it follows that the estimator $\hat{\lambda}_s^*$ with respect to Singh and Tarray (2015) estimator $\hat{\lambda}_{IST}$ is outperformed by the estimator $\hat{\lambda}_s^*$ with respect to Land et al. (2012) estimator $\hat{\lambda}_{IL}$ when the proportion of the rare non-sensitive unrelated attribute is unknown.

TABLE 3. Percent relative efficiency of the proposed estimator $\hat{\lambda}_s^*$ with respect to Land et al. (2012) estimator $\hat{\lambda}_{IL}$, when the proportion of the rare non-sensitive unrelated attribute is unknown

P_1	P_2	P_4	P_5	T_2	λ_s	λ_y	T_1				
							0.5	0.6	0.7	0.8	0.9
0.60	0.20	0.30	0.35	0.5	0.50	0.50	144.26	221.37	295.31	357.12	403.65
				0.5		1.00	160.25	260.15	366.88	465.31	545.36
				0.5		1.50	170.96	288.72	425.06	561.35	680.23
				0.4	1.00	0.50	174.51	224.14	267.78	303.53	331.29
				0.4		1.00	193.81	256.79	315.24	365.28	405.43
				0.4		1.50	209.49	284.88	358.22	423.63	477.81
				0.3	1.50	0.50	191.64	227.58	259.00	285.39	306.89
				0.3		1.00	210.45	254.65	294.48	328.74	357.15
				0.3		1.50	227.01	279.33	327.85	370.58	406.63
0.70	0.15	0.40	0.30	0.5	0.50	0.50	148.70	197.50	239.96	273.66	298.58
				0.5		1.00	167.58	233.82	296.92	350.95	393.26
				0.5		1.50	181.26	262.50	345.66	421.69	484.53
				0.4	1.00	0.50	160.55	190.25	215.17	235.11	250.49
				0.4		1.00	178.38	216.58	250.00	277.62	299.42
				0.4		1.50	193.61	240.11	282.35	318.37	347.50
				0.3	1.50	0.50	167.49	188.90	207.11	222.16	234.33
				0.3		1.00	183.03	209.37	232.28	251.56	267.36
				0.3		1.50	197.15	228.48	256.34	280.20	300.00

continue

0.80	0.10	0.50	0.25	0.5	0.50	0.50	139.44	161.40	179.26	193.15	203.55
				0.5		1.00	157.37	188.42	215.24	237.09	254.05
				0.5		1.50	171.85	211.64	247.82	278.62	303.36
				0.4	1.00	0.50	138.43	151.67	162.53	171.23	178.06
				0.4		1.00	151.46	168.49	182.82	194.53	203.85
				0.4		1.50	163.27	184.18	202.20	217.21	229.34
				0.3	1.50	0.50	139.08	148.85	157.09	163.93	169.53
				0.3		1.00	149.26	161.14	171.30	179.82	186.86
				0.3		1.50	158.84	172.92	185.11	195.45	204.06

FIGURE 3. Pictorial representation of percent relative efficiency of the proposed estimator $\hat{\lambda}_s^*$ with respect to Land et al. (2012) estimator $\hat{\lambda}_{IL}$, when $P_I = 0.60, 0.70, 0.80$ and $\lambda_y = 0.50$.

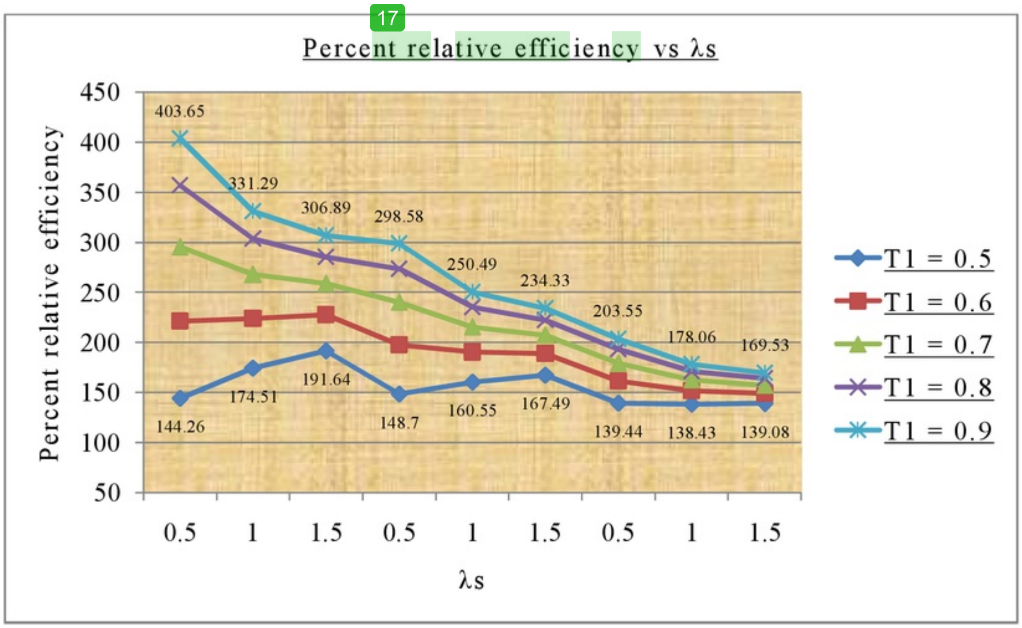


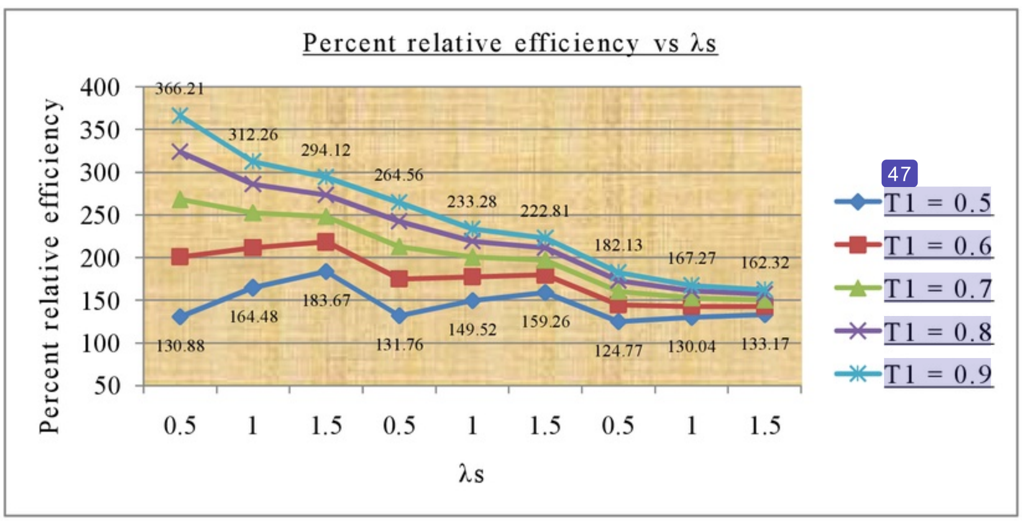
TABLE 4. Percent relative efficiency of the proposed estimator $\hat{\lambda}_s^*$ with respect to Singh and Tarray (2015) estimator $\hat{\lambda}_{IST}$, when the proportion of the rare non-sensitive unrelated attribute is unknown

P_1	P_2	P_4	P_5	T_2	λ_s	λ_y	T_1				
							0.5	0.6	0.7	0.8	0.9
0.60	0.20	0.30	0.35	0.5	0.50	0.50	130.88	200.83	267.92	324.00	366.21
				0.5		1.00	138.79	225.31	317.75	402.99	472.32
				0.5		1.50	144.09	243.34	358.26	473.12	573.32
				0.4	1.00	0.50	164.48	211.26	252.39	286.09	312.26
				0.4		1.00	175.83	232.97	286.00	331.40	367.82
				0.4		1.50	185.05	251.64	316.43	374.21	422.07
				0.3	1.50	0.50	183.67	218.11	248.22	273.51	294.12
				0.3		1.00	195.52	236.58	273.58	305.42	331.81
				0.3		1.50	205.95	253.42	297.44	336.20	368.91
0.70	0.15	0.40	0.30	0.5	0.50	0.50	131.76	175.00	212.63	242.48	264.56
				0.5		1.00	139.13	194.11	246.50	291.36	326.48
				0.5		1.50	144.46	209.21	275.48	336.09	386.16
				0.4	1.00	0.50	149.52	177.18	200.39	218.96	233.28
				0.4		1.00	158.06	191.91	221.51	245.99	265.31
				0.4		1.50	165.35	205.07	241.14	271.91	296.78
				0.3	1.50	0.50	159.26	179.61	196.92	211.24	222.81
				0.3		1.00	167.34	191.42	212.37	230.00	244.44
				0.3		1.50	174.69	202.45	227.13	248.28	265.82

continue

0.80	0.10	0.50	0.25	0.5	0.50	0.50	124.77	144.41	160.39	172.82	182.13
				0.5		1.00	131.14	157.02	179.36	197.58	211.71
				0.5		1.50	136.29	167.85	196.54	220.98	240.60
				0.4	1.00	0.50	130.04	142.47	152.68	160.85	167.27
				0.4		1.00	135.52	150.76	163.58	174.05	182.39
				0.4		1.50	140.49	158.48	173.99	186.90	197.34
				0.3	1.50	0.50	133.17	142.52	150.41	156.95	162.32
				0.3		1.00	137.78	148.75	158.12	165.99	172.48
				0.3		1.50	142.12	154.71	165.62	174.87	182.58

FIGURE 4. Pictorial representation of percent relative efficiency of the proposed estimator $\hat{\lambda}_s^*$ with respect to Singh and Tarray (2015) estimator $\hat{\lambda}_{IST}$, when $P_1 = 0.60, 0.70, 0.80$ and $\lambda_y = 0.50$.



When conducting personal interview on sensitive issues or highly personal questions, a major concern in randomized response technique is how to procure solid estimate of the population proportion of individuals who possess to stigmatizing character while at the same time the privacy of respondents is protected. So the degree of privacy is an essential part of the randomized response procedure and practice. Leysieffer and Warner (1976), Lanke (1976), Anderson (1977), Flinger et al. (1977), Nayak (1994), Yan and Nie (2004), Bhargava and Singh (2002), Guerriero and Sandri (2007), Giordano and Perri (2012), and among others have discussed both the efficiency and privacy by comparing the variance of different estimators under equal levels of confidentiality. Zhimin and Zaizai (2012) introduced a new intuitive alternative jeopardy measure of privacy protection and compared Warner (1965), Greenberg et al. (1969), Mangat and Singh (1990), Kuk (1990), and Mangat (1994) models in terms of efficiency and privacy protection. They have also mentioned that the efficiency and respondents privacy protection do not necessarily move in opposite directions. Bose (2016), Ryz and Grest (2016), and Bose and Dihidar (2018) have also discussed about respondent privacy measures for different situations. The brief review of some privacy measures including Zhimin and Zaizai (2012) are given below in the following section 5.

5. Review of Some Privacy Measures

5.1. Leysieffer and Warner (1976) Jeopardy Measure

Consider a dichotomous population is divided into two groups that is sensitive group A with unknown proportion π_A and its complement A^c with unknown proportion $(1-\pi_A)$. Let a dichotomous response model is with a typical response R is "yes" (say, y) or "no" (say, n). The conditional probabilities $P(R/A)$ and $P(R/A^c)$ that a response R comes from an individual of groups A and A^c. These probabilities are called the design probabilities of a randomized response survey. They are controlled by the researcher and are known both to the interviewee and the interviewer. Using these design probabilities Leysieffer and Warner (1976) introduced the measure of jeopardy carried by response R about A and A^c, that is

$$g(R/A) = \frac{P(R/A)}{P(R/A^c)} \; ; \quad g(R/A^c) = \frac{P(R/A^c)}{P(R/A)}, \tag{35}$$

and the values of $g(R/A)$ $\left[g(R/A^c)\right]$ greater than unity indicate that response $y(n)$ is jeopardizing with respect to $A(A^c)$, in the sense that with this response individuals of group $A(A^c)$ are more likely to belong to group $A(A^c)$ than to group $A^c(A)$. They have also mentioned that an unbiased estimator of π_A exist if and only if

$$P(y/A) - P(y/A^c) \neq 0 . \tag{36}$$

Suppose, without loss generality, that $P(y/A) > P(y/A^c)$, so that

$$g(y/A) > 1 \quad \text{and} \quad g(n/A^c) > 1 .$$

For the sake of efficiency, one needs as large magnitudes as possible for $g(y/A)$ and $g(n/A^c)$ and both above unity. Hence, from the practical point of view, regarding protection of privacy, one fix maximal allowable level of $g(y/A)$ is k_1 and $g(n/A^c)$ is k_2. The problem now becomes one of constrained, that is

$$1 < g(y/A) \leq k_1 \quad \text{and} \quad 1 < g(n/A^c) \leq k_2 .$$

Hence, jeopardy degree lower than k_1 and k_2 would ensure respondents a higher degree of protection, but this would necessarily increase the variance of the estimators.

5.2. Lanke (1976) Jeopardy Measure

When A^c is not stigmatizing, Lanke (1976) proposed another measure of privacy protection and made the argument that the respondents want to hide their membership in group A but not membership in group A^c. So, based on this focus on "maximum suspicion of belong to group A". Lanke (1976) introduced a measure of protection as

$$L = max\{P(A/y), P(A/n)\}. \quad (37)$$

The smaller value of $L = max\{P(A/y), P(A/n)\}$, the more the privacy is protected.

5.3. Zhimin and Zaizai (2012) Jeopardy Measure

Zhimin and Zaizai (2012) introduced an another measure of privacy protection based on the idea that the posterior probabilities $P(A/R)$ and $P(A^c/R)$ of respondents belong to groups A and A^c by giving the response R. These are the revealing probabilities. By Bayes's rule,

$$P(R/A) = \frac{P(R)P(A/R)}{P(A)}, \; P(R/A^c) = \frac{P(R)P(A^c/R)}{P(A^c)}, \text{ and}$$

$$\frac{P(R/A)}{P(R/A^c)} = \frac{P(A^c)}{P(A)} \frac{P(A/R)}{P(A^c/R)}.$$

Following Leysieffer and Warner (1976) and without loss of generality, the probability of a yes response by using random device is

$$\lambda = P(y/A)\pi_A + P(y/A^c)(1 - \pi_A)$$

$$= [P(y/A) - P(y/A^c)]\pi_A + P(y/A^c).$$

If a sample of n individuals is drawn from the population by using simple random sampling with replacement. The unbiased estimator of π_A is

$$\hat{\pi}_A = \frac{\hat{\lambda} - P(y/A^c)}{P(y/A) - P(y/A^c)}, \quad (38)$$

which is defined if and only if $P(y/A) - P(y/A^c) \neq 0$ and $\hat{\lambda} = n_1/n$ is sample proportion of yes answer in the sample.

The variance of unbiased estimator $\hat{\pi}_A$ is obtained as (Zhimin and Zaizai (2012))

$$V(\hat{\pi}_A) = \frac{\pi_A^2 (1-\pi_A)^2}{n\left[\pi_A - P(A/y)\right]\left[P(A/n) - \pi_A\right]}. \quad (39)$$

It is clear from Eq. (39) that

$$\frac{\partial V(\hat{\pi}_A)}{\partial P(A/y)} > 0 \quad \text{and} \quad \frac{\partial V(\hat{\pi}_A)}{\partial P(A/n)} < 0.$$

For the sake of efficiency, one needs the higher level for $P(A/y)$ and lower level for $P(A/n)$.

Zhimin and Zaizai (2012) have also mentioned that the response R (say "yes" or "no") is non-jeopardizing if and only if revealing probabilities of a respondent being perceived as belonging to group A based on his response R are equal to π_A. To be useful, the design cannot be totally non-jeopardizing.

The revealing probabilities $P(A/y)$ and $P(A/n)$ which departures from $P(A)$ could be treated as a measure of revelation of secrecy. By Bayes' rule

$$P(A/y) = \frac{\pi_A}{\pi_A + (1-\pi_A)\left(P(y/\bar{A})/P(y/A)\right)},$$

and

$$P(A/n) = \frac{\pi_A}{\pi_A + (1-\pi_A)\left(P(n/\bar{A})/P(n/A)\right)}.$$

Letting $\tau(y) = P(y/A)/P(y/\bar{A})$ and $\tau(n) = P(n/A)/P(n/\bar{A})$, we have

$$P(A/R) = \frac{\pi_A}{\pi_A + (1-\pi_A)(1/\tau(R))}. \quad (40)$$

The closer the value of $P(A/R)$ to π_A, this is equivalent to the closer $\tau(R)$ is to unity; the higher is the protection to a respondent's privacy in the proposed procedure. The $\tau(y)$

and $\tau(n)$ are quantities at the investigator's disposal and a function only of the design probabilities not of π_A. For a dichotomous response model, these quantities are directly related to each other if one increases the other must be decreases. Zhimin and Zaizai (2012) introduced a new measure of privacy protection of respondents as

$$M(R) = \left| 1 - \frac{1}{2}\left[\tau(y) + \tau(n)\right] \right|. \tag{41}$$

The closer the value of $M(R)$ to zero, the more the privacy is protected and the term $M(R)$ contains the two alternative forms of R.

To shown the performance of the proposed two-stage unrelated randomized response model over Greenberg et al. (1969) and Singh et al. (2003) unrelated randomized response models in term of efficiency and privacy protection of respondents. We consider the privacy protection measure proposed by Zhimin and Zaizai (2012).

6. Efficiency vs Privacy Protection of the Unrelated Randomized Response Models

Now, we study the measures of privacy protection and the efficiency of unrelated randomized response models.

6.1. Greenberg et al. (1969) Model

Greenberg et al. (1969) proposed an unrelated randomized response model in which each of the n respondents, selected by using simple random sampling with replacement scheme, required the response to answer the questions without revealing to interviewer "I belong to sensitive group A" with probability P_1 and "I belong to non-sensitive group Y" with probability $(1 - P_1)$, where both the questions are unrelated and the second one is a completely harmless question. The design probabilities are

$$P(y/A) = P_1 + (1 - P_1)\pi_Y \text{ and } P(y/A^c) = (1 - P_1)\pi_Y,$$

where π_Y is the true proportion of the rare non-sensitive unrelated group in the population.

The values of $P(A/y)$, $P(A/n)$, and the measure of privacy protection $M_G(R)$ obtain by following the Eq. (40) and (41).

$$P(A/y) = \frac{\pi_A}{\pi_A + (1 - \pi_A)\left[(1 - P_1)\pi_Y / \{P_1 + (1 - P_1)\pi_Y\}\right]}. \tag{42}$$

$$P(A/n) = \frac{\pi_A}{\pi_A + (1 - \pi_A)\left[\{1 - (1 - P_1)\pi_Y\} / \{1 - P_1 - (1 - P_1)\pi_Y\}\right]}. \tag{43}$$

and

$$M_G(R) = \left| \frac{P_1\left[1 - 2\pi_Y(1 - P_1)\right]}{2\pi_Y(1 - P_1)\left[1 - \pi_Y(1 - P_1)\right]} \right|. \tag{44}$$

The efficiency of Greenberg et al. (1969) estimator $\hat{\pi}_G = \left[\hat{\theta}_G - (1 - P_1)\pi_Y\right]/P_1$ is

$$V(\hat{\pi}_G) = \frac{\pi_A(1 - \pi_A)}{n} + \frac{\pi_A(1 - P_1)(1 - 2\pi_Y)}{nP_1} + \frac{\pi_Y(1 - P_1)\left[1 - (1 - P_1)\pi_Y\right]}{nP_1^2}. \tag{45}$$

6.2. Singh et al. (2013) Model

Singh et al. (2003) proposed an unrelated RR model which is alternative to Greenberg et al. (1969) model in the sense that the randomized device used in Singh et al. (2003) model has three outcomes (i) "I belong to sensitive group A" with probability P_1, (ii) "I belong to non-sensitive group Y" with probability P_2, and (iii) "Blank cards" with probability P_3 such that $\sum_{i=1}^{3} P_i = 1$.

The design probabilities are

$$P(y/A) = P_1 + P_2\pi_Y \text{ and } P(y/A^c) = P_2\pi_Y,$$

where π_Y denotes the true proportion of the rare non-sensitive unrelated group in the population.

The values of $P(A/y)$, $P(A/n)$, and the measure of privacy protection $M_S(R)$ obtain by following the Eq. (40) and (41).

$$P(A/y)=\frac{\pi_A}{\pi_A+(1-\pi_A)\left[P_2\pi_Y/\{P_1+P_2\pi_Y\}\right]}. \tag{46}$$

$$P(A/n)=\frac{\pi_A}{\pi_A+(1-\pi_A)\left[\{1-P_2\pi_Y\}/\{1-P_1-P_2\pi_Y\}\right]}. \tag{47}$$

and

$$M_S(R)=\left|\frac{P_1[1-2P_2\pi_Y]}{2P_2\pi_Y[1-P_2\pi_Y]}\right|. \tag{48}$$

The efficiency of Singh et al. (2003) estimator $\hat{\pi}_S=\left[\hat{\theta}_S-P_2\pi_Y\right]/P_1$ is

$$V(\hat{\pi}_S)=\frac{\pi_A(1-\pi_A)}{n}+\frac{\pi_A(1-P_1-2P_2\pi_Y)}{nP_1}+\frac{P_2\pi_Y[1-P_2\pi_Y]}{nP_1^2}. \tag{49}$$

6.3. Proposed Model

By following the procedure as given in section 2, the design probabilities of the proposed model are

$$P(y/A)=T_1+(1-T_1)(P_1+P_2\pi_Y) \text{ and } P(y/A^c)=(1-T_1)P_2\pi_Y,$$

where π_Y is the true proportion of the rare non-sensitive unrelated group in the population.

The values of $P(A/y)$, $P(A/n)$, and the measure of privacy protection $M_A(R)$ obtain by following the Eq. (40) and (41).

$$P(A/y)=\frac{\pi_A}{\pi_A+(1-\pi_A)\left[(1-T_1)P_2\pi_Y/\{T_1+(1-T_1)(P_1+P_2\pi_Y)\}\right]}. \tag{50}$$

$$P(A/n)=\frac{\pi_A}{\pi_A+(1-\pi_A)\left[\{1-(1-T_1)P_2\pi_Y\}/\{1-T_1-(1-T_1)(P_1+P_2\pi_Y)\}\right]}. \tag{51}$$

and

$$M_A(R)=\left|\frac{T_1+(1-T_1)\left[P_1-2P_2\pi_Y\{P_1(1-T_1)+T_1\}\right]}{2P_2(1-T_1)\pi_Y\left[1-P_2(1-T_1)\pi_Y\right]}\right|. \tag{52}$$

The efficiency of the proposed estimator $\hat{\pi}_A=\left[\hat{\theta}_A-P_2(1-T_1)\pi_Y\right]/\left[T_1+P_1(1-T_1)\right]$ is

$$V(\hat{\pi}_A)=\frac{\pi_A(1-\pi_A)}{n}+\frac{\pi_A\left[1-\{T_1+P_1(1-T_1)\}-2P_2(1-T_1)\pi_Y\right]}{n\{T_1+P_1(1-T_1)\}}+\frac{P_2(1-T_1)\pi_Y\left[1-P_2(1-T_1)\pi_Y\right]}{n\{T_1+P_1(1-T_1)\}^2}. \tag{53}$$

7. Comparisons of Unrelated Randomized Response Models

We have made efficiency comparisons of the proposed model with Greenberg et al. (1969) and Singh et al. (2003) models in the following theorems:

Theorem 6. The Greenberg et al. (1969) model and Singh et al. (2003) model have the same privacy protection and efficiency for $P_2=(1-P_1)$.

Theorem 7. The Greenberg et al. (1969) model and proposed model have the same privacy protection and efficiency if $T_1=0$ and $P_2=(1-P_1)$.

Theorem 8. The Singh et al. (2003) model and proposed model have the same privacy protection and efficiency if $T_1=0$.

Comparing two designs, the more protective model has the value of $M(R)$ closer to zero. We shall now specify when one unrelated randomized response model is more protective than another by accomplish a numerical study.

8. Numerical Study

To have some idea about the efficiency and for the measure of privacy protection of respondents in randomized response models, we choose the designs parameters in this manner that the value of $P(A/R)$ close to π_A and $M(R)$ close to zero. Tables 5, 6, and 7 present the values of $P(A/R)$, $M(R)$, and variance $V(.)$ for the previously discussed models along with the different design parametric values.

For the selected values of the design parameters, the results appear very interesting and may be guide for researchers in finding a suitable randomized response model. However, as mention in section 5.3, the measure of privacy protection $M(R)$ is a function only of the design probabilities of the randomized response model and this measure itself may be used to compare the various randomized response models between them with respect to the privacy protection. For instance, from Tables 5, 6, and 7 we observe that the proposed model with $P_1 = 0.05$, $P_2 = 0.90$ and $T_1 = 0.4$ indicates more respondents' privacy protection and better performance in terms of efficiency than Singh et al. (2003) and Greenberg et al. (1969) models when $P_1 = 0.20$, $P_2 = 0.60$ and $P_1 = 0.40$. It is observed that the efficiency and respondents' privacy protection is not always in conflict. Further, we see that Singh et al. (2003) and the proposed models provide a higher degree of privacy protection with $\pi_Y = 0.7$ and $T_1 = 0.3$ then Greenberg et al. (1969) model but less efficient. The proposed model is also more protective when $T_1 = 0.1$ and $\pi_Y = 0.3$ than existing models but not more efficient. However, sometimes, the value of $M(R)$ closer to zero, this is not equivalent the value of $P(A/R)$ closer to π_A. The rest of the results can be read out from the summarized presented in Tables 5, 6, and 7.

TABLE 5. Greenberg et al. (1969) model values of $P(A/y)$, $P(A/n)$, $nV(\hat{\pi}_G)$, and $M(R)$ for different values of π_A, π_Y and P_I

P_I	π_Y		π_A				$M(R)$
			0.1	0.3	0.5	0.7	
		P(A/y)	0.111	0.325	0.529	0.723	
0.1	0.9	P(A/n)	0.050	0.168	0.321	0.525	0.201
		$nV(\hat{\pi}_G)$	14.760	13.440	12.040	10.560	
		P(A/y)	0.132	0.370	0.578	0.761	
0.25	0.9	P(A/n)	0.025	0.090	0.187	0.350	0.199
		$nV(\hat{\pi}_G)$	3.360	3.000	2.560	2.040	
		P(A/y)	0.162	0.427	0.635	0.802	
0.40	0.9	P(A/n)	0.014	0.052	0.115	0.233	0.064
		$nV(\hat{\pi}_G)$	1.522	1.402	1.202	0.922	
		P(A/y)	0.207	0.502	0.702	0.846	
0.55	0.9	P(A/n)	0.008	0.031	0.070	0.150	0.216
		$nV(\hat{\pi}_G)$	0.821	0.810	0.719	0.548	
		P(A/y)	0.114	0.331	0.536	0.730	
0.1	0.7	P(A/n)	0.075	0.238	0.421	0.630	0.055
		$nV(\hat{\pi}_G)$	23.04	22.440	21.760	21.000	

continue

0.25	0.7	$P(A/y)$	0.140	0.387	0.596	0.775	0.025
		$P(A/n)$	0.050	0.168	0.321	0.525	
		$nV(\hat{\pi}_G)$	3.960	3.840	3.640	3.360	
0.40	0.7	$P(A/y)$	0.178	0.455	0.661	0.820	0.131
		$P(A/n)$	0.033	0.117	0.236	0.420	
		$nV(\hat{\pi}_G)$	1.553	1.552	1.472	1.312	
0.55	0.7	$P(A/y)$	0.233	0.540	0.733	0.865	0.471
		$P(A/n)$	0.021	0.077	0.164	0.315	
		$nV(\hat{\pi}_G)$	0.770	0.825	0.799	0.694	
0.1	0.5	$P(A/y)$	0.119	0.343	0.550	0.740	0.020
		$P(A/n)$	0.083	0.259	0.450	0.656	
		$nV(\hat{\pi}_G)$	24.840	24.960	25.000	24.960	
0.25	0.5	$P(A/y)$	0.156	0.416	0.625	0.795	0.133
		$P(A/n)$	0.062	0.204	0.375	0.583	
		$nV(\hat{\pi}_G)$	3.840	3.960	4.000	3.960	
0.40	0.5	$P(A/y)$	0.205	0.500	0.700	0.844	0.381
		$P(A/n)$	0.045	0.155	0.300	0.500	
		$nV(\hat{\pi}_G)$	1.402	1.522	1.562	1.522	

continue

		P(A/y)	0.276	0.596	0.775	0.889	
0.55	0.5	P(A/n)	0.031	0.110	0.225	0.403	0.867
		$nV(\hat{\pi}_G)$	0.666	0.786	0.826	0.786	
		P(A/y)	0.132	0.370	0.578	0.761	
0.1	0.3	P(A/n)	0.087	0.270	0.463	0.668	0.116
		$nV(\hat{\pi}_G)$	20.160	21.000	21.760	22.440	
		P(A/y)	0.190	0.475	0.678	0.831	
0.25	0.3	P(A/n)	0.070	0.225	0.403	0.612	0.394
		$nV(\hat{\pi}_G)$	3.000	3.360	3.640	3.840	
		P(A/y)	0.263	0.580	0.763	0.882	
0.40	0.3	P(A/n)	0.053	0.180	0.338	0.544	0.867
		$nV(\hat{\pi}_G)$	1.072	1.312	1.472	0.867	
		P(A/y)	0.360	0.685	0.835	0.922	
0.55	0.3	P(A/n)	0.038	0.135	0.266	0.459	1.719
		$nV(\hat{\pi}_G)$	0.508	0.694	0.799	0.825	

Figures 5, 6, and 7 show the behaviour of the values of variance and privacy protection as the value of π_y change regardless of changes in the values of the other parameters $P_i\,(i = 1, 2)$, T_1 and π_A. Figures 5, 6, and 7 shows that as the value of P_1 is close to 0.1, the privacy protection of respondents of the retained cases remaining close to zero, but the variances takes on very large value except only the proposed model. However, as the value of

P_1 is close to one, the privacy of respondents is jeopardy only on Greenberg et al. (1969) model. With P_1 close to one, it could be difficult to protect the respondent privacy in Greenberg et al. (1969) technique. It is interesting to note from Figures 6 and 7 that the value of P_1 remains either close to zero or one, there seems to be no restriction on the choice of the parameters. From Figure 7, it seems that when P_1 is close to zero and P_2 is close to one, and then there is a combination of T_1, π_A and π_y which could lead to more protection and smaller variance of the proposed model. Following the discussion of the results we conclude that the proposed model is more efficient and more protective for any given level of the proportion between 0.3-0.9 of a non-sensitive characteristic π_y in a population.

FIGURE 5. Relationship between variance of the Greenberg et al. (1969) model and $M(R)$ for different values of π_A, π_Y and P_1.

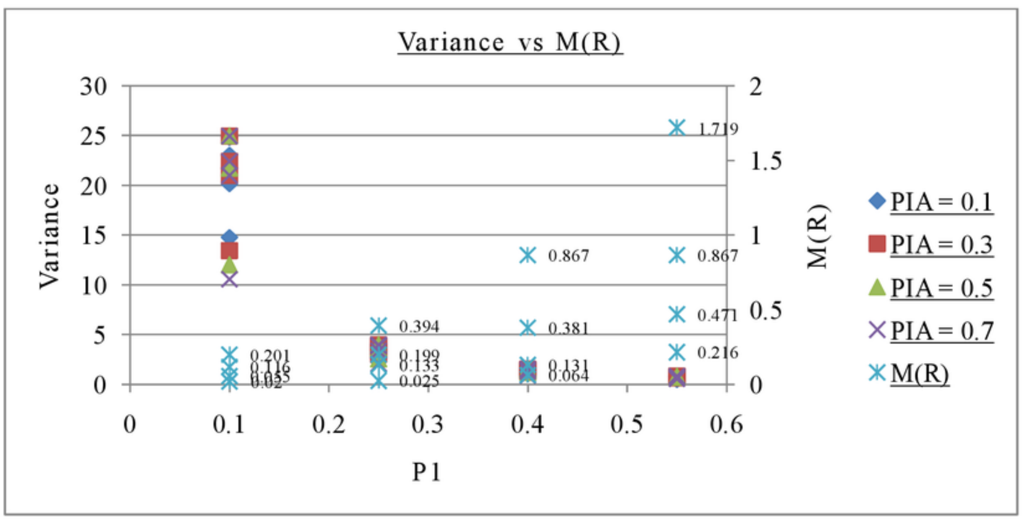


TABLE 6. Singh et al. (2003) model values of $P(A/y)$, $P(A/n)$, $nV(\hat{\pi}_S)$, and $M(R)$ for different values of π_A, π_Y and P_i $(i = 1, 2, 3)$

P_1	P_2	π_Y		π_A 0.1	0.3	0.5	0.7	$M(R)$
			$P(A/y)$	0.132	0.370	0.578	0.761	
0.20	0.60	0.9	$P(A/n)$	0.059	0.195	0.361	0.568	0.032
			$nV(\hat{\pi}_S)$	6.160	6.000	5.760	5.440	
			$P(A/y)$	0.120	0.346	0.553	0.742	
0.15	0.70	0.9	$P(A/n)$	0.062	0.203	0.372	0.581	0.083
			$nV(\hat{\pi}_S)$	10.176	9.750	9.243	8.656	
			$P(A/y)$	0.112	0.328	0.532	0.726	
0.10	0.80	0.9	$P(A/n)$	0.066	0.216	0.391	0.600	0.109
			$nV(\hat{\pi}_S)$	19.710	18.750	17.710	16.590	
			$P(A/y)$	0.105	0.312	0.515	0.712	
0.05	0.90	0.9	$P(A/n)$	0.075	0.240	0.424	0.632	0.100
			$nV(\hat{\pi}_S)$	60.310	57.750	55.110	52.390	
			$P(A/y)$	0.140	0.387	0.596	0.775	
0.20	0.60	0.7	$P(A/n)$	0.067	0.219	0.395	0.604	0.065
			$nV(\hat{\pi}_S)$	6.160	6.240	6.240	6.160	

continue

			$P(A/y)$	0.126	0.358	0.566	0.752	
0.15	0.70	0.7	$P(A/n)$	0.072	0.232	0.413	0.622	0.006
			$nV(\hat{\pi}_s)$	11.110	11.056	10.923	10.710	
			$P(A/y)$	0.115	0.335	0.541	0.733	
0.10	0.80	0.7	$P(A/n)$	0.079	0.248	0.435	0.643	0.024
			$nV(\hat{\pi}_s)$	24.510	24.190	23.790	23.310	
			$P(A/y)$	0.107	0.316	0.519	0.715	
0.05	0.90	0.7	$P(A/n)$	0.087	0.270	0.463	0.668	0.027
			$nV(\hat{\pi}_s)$	92.710	91.590	90.390	89.110	
			$P(A/y)$	0.156	0.416	0.625	0.795	
0.20	0.60	0.5	$P(A/n)$	0.073	0.234	0.416	0.625	0.190
			$nV(\hat{\pi}_s)$	5.440	5.760	6.000	6.160	
			$P(A/y)$	0.137	0.379	0.588	0.769	
0.15	0.70	0.5	$P(A/n)$	0.078	0.247	0.434	0.642	0.098
			$nV(\hat{\pi}_s)$	10.301	10.621	10.861	11.021	
			$P(A/y)$	0.122	0.348	0.555	0.744	
0.10	0.80	0.5	$P(A/n)$	0.084	0.263	0.454	0.660	0.041
			$nV(\hat{\pi}_s)$	24.190	24.510	24.750	24.910	

continue

0.05	0.90	0.5	$P(A/y)$	0.109	0.322	0.526	0.721	0.010
			$P(A/n)$	0.091	0.280	0.476	0.679	
			$nV(\hat{\pi}_s)$	99.190	99.510	99.750	99.910	
0.20	0.60	0.3	$P(A/y)$	0.190	0.475	0.678	0.831	0.433
			$P(A/n)$	0.077	0.244	0.430	0.638	
			$nV(\hat{\pi}_s)$	4.000	4.560	5.040	5.440	
0.15	0.70	0.3	$P(A/y)$	0.160	0.423	0.631	0.800	0.262
			$P(A/n)$	0.082	0.257	0.447	0.654	
			$nV(\hat{\pi}_s)$	7.750	8.443	9.056	9.590	
0.10	0.80	0.3	$P(A/y)$	0.136	0.377	0.586	0.767	0.142
			$P(A/n)$	0.088	0.271	0.464	0.669	
			$nV(\hat{\pi}_s)$	18.750	19.710	20.590	21.390	
0.05	0.90	0.3	$P(A/y)$	0.116	0.336	0.542	0.734	0.058
			$P(A/n)$	0.093	0.285	0.482	0.684	
			$nV(\hat{\pi}_s)$	79.750	81.510	83.190	84.790	

TABLE 7. The proposed model values of $P(A/y)$, $P(A/n)$, $nV(\hat{\pi}_S)$, and $M(R)$ for different values of π_A, π_Y, $P_i\,(i=1,2,3)$ and T_1

P_1	P_2	T_1	π_Y		π_A				$M(R)$
					0.1	0.3	0.5	0.7	
				$P(A/y)$	0.224	0.527	0.722	0.858	
0.20	0.60	0.4	0.9	$P(A/n)$	0.025	0.090	0.187	0.350	0.417
				$nV(\hat{\pi}_S)$	0.867	0.923	0.898	0.793	
				$P(A/y)$	0.203	0.496	0.696	0.842	
0.15	0.70	0.4	0.9	$P(A/n)$	0.023	0.083	0.175	0.331	0.254
				$nV(\hat{\pi}_S)$	1.019	1.038	0.978	0.837	
				$P(A/y)$	0.186	0.469	0.673	0.828	
0.10	0.80	0.4	0.9	$P(A/n)$	0.020	0.075	0.159	0.307	0.127
				$nV(\hat{\pi}_S)$	1.179	1.158	1.057	0.876	
				$P(A/y)$	0.173	0.446	0.653	0.814	
0.05	0.90	0.4	0.9	$P(A/n)$	0.017	0.065	0.140	0.276	0.024
				$nV(\hat{\pi}_S)$	1.347	1.280	1.133	0.906	
				$P(A/y)$	0.217	0.516	0.714	0.853	
0.20	0.60	0.3	0.7	$P(A/n)$	0.040	0.139	0.273	0.467	0.436
				$nV(\hat{\pi}_S)$	1.155	1.263	1.290	1.237	

continue

0.15	0.70	0.3	0.7	P(A/y)	0.195	0.483	0.685	0.835	0.282
				P(A/n)	0.040	0.141	0.277	0.472	
				$nV(\hat{\pi}_s)$	1.441	1.516	1.511	1.426	
0.10	0.80	0.3	0.7	P(A/y)	0.177	0.454	0.660	0.819	0.167
				P(A/n)	0.041	0.143	0.281	0.477	
				$nV(\hat{\pi}_s)$	1.789	1.826	1.782	1.659	
0.05	0.90	0.3	0.7	P(A/y)	0.163	0.429	0.637	0.804	0.080
				P(A/n)	0.042	0.146	0.286	0.483	
				$nV(\hat{\pi}_s)$	2.221	2.212	2.122	1.950	
0.20	0.60	0.2	0.5	P(A/y)	0.217	0.517	0.714	0.853	0.513
				P(A/n)	0.055	0.184	0.344	0.551	
				$nV(\hat{\pi}_s)$	1.541	1.750	1.879	1.928	
0.15	0.70	0.2	0.5	P(A/y)	0.192	0.478	0.681	0.833	0.349
				P(A/n)	0.058	0.192	0.357	0.564	
				$nV(\hat{\pi}_s)$	2.096	2.291	2.406	2.441	
0.10	0.80	0.2	0.5	P(A/y)	0.172	0.445	0.652	0.814	0.231
				P(A/n)	0.061	0.201	0.370	0.578	
				$nV(\hat{\pi}_s)$	2.894	3.071	3.168	3.185	

continue

0.05	0.90	0.2	0.5	$P(A/y)$	0.156	0.416	0.625	0.795	0.145
				$P(A/n)$	0.064	0.211	0.384	0.593	
				$nV(\hat{\pi}_S)$	4.106	4.260	4.333	4.326	
0.20	0.60	0.1	0.3	$P(A/y)$	0.232	0.539	0.731	0.864	0.697
				$P(A/n)$	0.068	0.222	0.399	0.608	
				$nV(\hat{\pi}_S)$	1.963	2.365	2.688	2.931	
0.15	0.70	0.1	0.3	$P(A/y)$	0.199	0.490	0.691	0.839	0.476
				$P(A/n)$	0.073	0.233	0.415	0.623	
				$nV(\hat{\pi}_S)$	3.030	3.479	3.848	4.138	
0.10	0.80	0.1	0.3	$P(A/y)$	0.172	0.446	0.652	0.814	0.318
				$P(A/n)$	0.077	0.245	0.431	0.638	
				$nV(\hat{\pi}_S)$	4.979	5.497	5.935	6.293	
0.05	0.90	0.1	0.3	$P(A/y)$	0.150	0.406	0.614	0.788	0.202
				$P(A/n)$	0.082	0.257	0.447	0.653	
				$nV(\hat{\pi}_S)$	9.093	9.722	10.271	10.740	

9. Conclusion

We extend Singh et al. (2003) model into two-stage unrelated randomized response model for estimation the mean number of individuals in a given population who possess to sensitive attribute by utilizing Poisson probability distribution. The model θ_A includes most of the randomized response models currently present in the literature based on unrelated-question methods. Apart from the analytical comparisons which provide the conditions under which the estimator $\hat{\lambda}_s$ is more efficient than Land et al. (2012), and Singh and Tarray (2015) estimators. The estimator $\hat{\lambda}_s^*$ of the mean number of individuals is also more efficient than Land et al. (2012), and Singh and Tarray (2015) estimators when the proportion of the rare non-sensitive unrelated attribute is unknown. The model θ_A with modifications performs better in terms of efficiency and privacy protection of respondents than Singh et al. (2003) and Greenberg et al. (1969) models. We recommend utilizing two-stage unrelated randomized response model in sampling surveys practice when a researcher deal with a rare sensitive characteristic. It incurs no additional sampling cost, it is more protective of privacy protection, and it yields are more efficient estimators of the mean number of individuals λ_s and the sensitive proportion π_s.

FIGURE 6. Relationship between variance of the Singh et al. (2003) model and $M(R)$ for different values of π_A, π_Y and $P_i\,(i = 1, 2, 3)$.

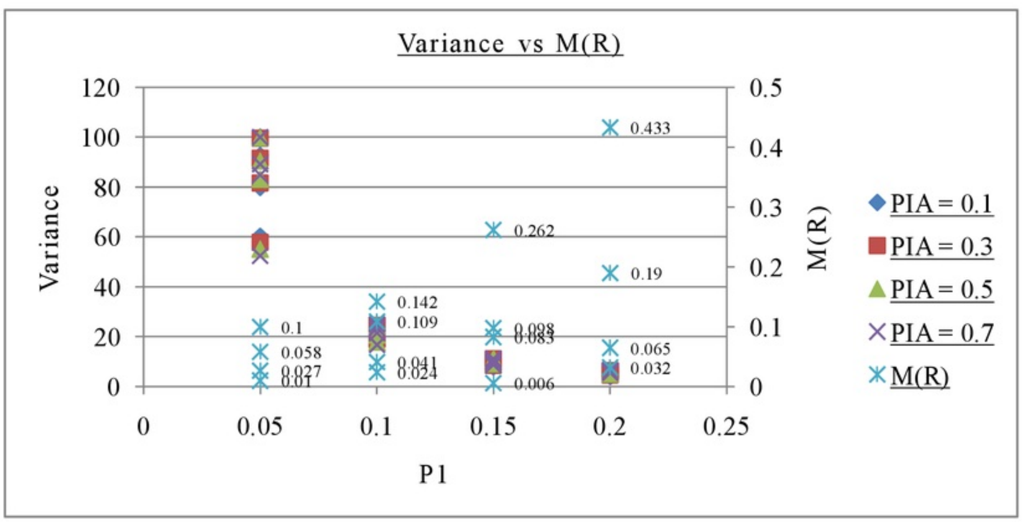
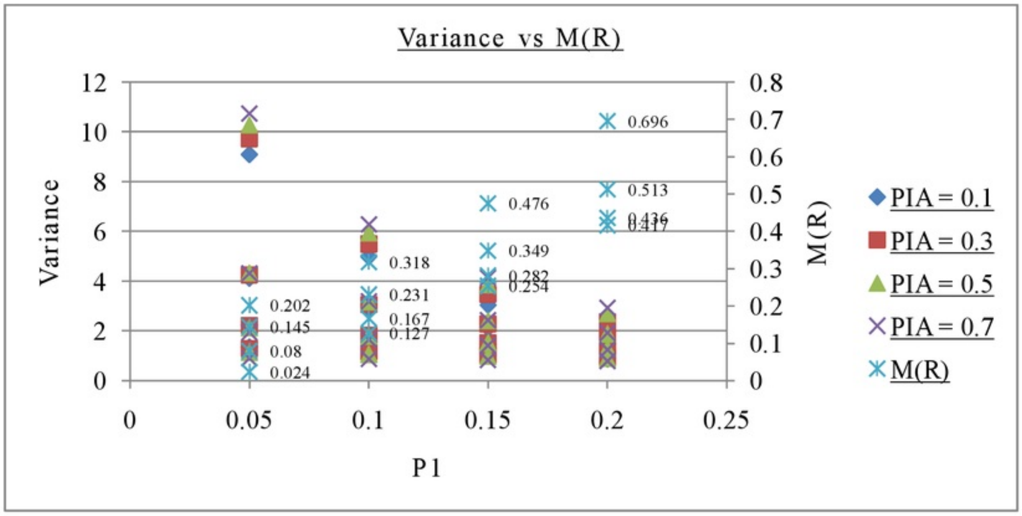


FIGURE 7. Relationship between variance of the proposed model and $M(R)$ for different values of π_A, π_Y and $P_i\ (i = 1, 2, 3)$.



References

Anderson, H. (1977). Estimation of a proportion through randomized response. *International Statistical Review*, *44*(2): 213-217.

Bhargava, M., and Singh, R. (2002). On the efficiency comparison of certain randomized response strategies. *Metrika*, *55*(3): 191-197.

Bose, M. (2016). Measures of respondent privacy in randomized response surveys. *Handbook of Statistics*, *34*: 341-351.

Bose, M., and Dihidar, K. (2018). Privacy protection measures for randomized response surveys on stigmatizing continuous variables. *Journal of Applied Statistics*, DOI: 1080/02664763.2018.1440540.

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

Cochran, W. G. (1977). *Sampling Technique* (3rd Edition). New York: John Wiley and Sons, USA.

Fligner, M. A., Policello, G. E. II., and Singh, J. (1977). A comparison of two randomized response survey method with consideration for the level of respondent protection. *Communication in Statistics – Theory and Methods*, *6*(15): 1511-1524.

Fox, J. A., and Tracy, P. E. (1986). *Randomized Response: A method of Sensitive Surveys*. Newbury Park, CA: SEGE Publications.

Giordano, S., and Perri, P. F. (2012). Efficiency comparison of unrelated question models based on same privacy protection degree. *Statistical Papers*, *53*(4): 987-999.

Greenberg, B., Abul-ela, A.-L. A., Simmons, W. R., and Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of American Statistical Association*, *64*(326): 520-539.

Guerriero, M., and Sandri, M. F. (2007). A note on the comparison of some randomized response procedures. *Journal of Statistical Planning and Inference*, *137*(7): 2184-2190.

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

Kim, J. H., Ryu, J. B., and Lee, G. S. (1992). A new two-stage randomized response model. *Korean Journal of Applied Statistics*, *5*(2): 157-167.

Kim, J. M., and Warde, W. D. (2005). A mixed randomized response model. *Journal of Statistical Planning and Inference*, *133*(1): 211-221.

Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika*, 77(2): 436-443.

Land, M., Singh, S., and Sedory, S. A. (2012). Estimation of a rare sensitive attribute using Poisson distribution. *Statistics*, *46*(3): 351-360.

Lanke, J. (1976). On the degree of protection in randomized interviews. *International Statistical Review*, *44*(2): 197-203

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*(2): 239-262.

Lee, G. S., Uhm, D., and Kim, J. M. (2013). Estimation of a rare sensitive attribute in a stratified sample using Poisson distribution. *Statistics: A Journal of Theoretical and Applied* Statistics, *47*(3): 575-589.

Leysieffer, F. W., and Warner, S. L. (1976). Respondent jeopardy and optimal designs in randomized response models. *Journal of American Statistical Association*, *71*(355): 649-656.

Mangat, N. S. (1992). Two-stage randomized response sampling procedure using unrelated question. *Journal of Indian Society Agricultural Statistics*, *44*(1): 82-87.

Mangat, N. S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society B*, *56*(1): 93-95.

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

Moors, J. J. A. (1971). Optimization of the unrelated question randomized response model. *Journal of the American Statistical Association*, *66*(335): 627-629.

Nayak, T. K. (1994). On randomized response surveys for estimating a proportion. *Communication in Statistics-Theory and Methods*, *23*(11): 3303-3321.

Ryu, J. B., Hong, K. H., and Lee, G. S. (1993). Randomized response model. *Freedom Academy, Seoul, Korea.*

Ryz, L., and Grest, L. (2016). A new era in data protection. *Computer Fraud & Security*, *2016*(3): 18-20.

Singh, H. P., and Tarray, T. A. (2014). A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution. *Statistics and Probability Letters*, *90*: 42-45.

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*(2): 129-138.

Singh, R., and Mangat, N. S. (1996). *Elements of survey sampling*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

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

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*(4): 515-522.

Tracy, D. S., and Mangat, N. S. (1996). Some developments in randomized response sampling during the last decade – A follow up of review by Chaudhuri and Mukherjee. *Journal of Applied Statistical Science*, *4*(2/3): 147-158.

Tracy, D. S., and Osahan, S. S. (1999). An improved randomized response technique. *Pakistan Journal of statistics*, *15*(1): 1-6.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*(309): 63-69.

Yan, Z. Z., and Nie, Z. K. (2004). A fair comparison of the randomized response strategies. *Acta Mathematica Scientia*, *24A*(3): 362-368.

Zhimin, H., and Zaizai, Y. (2012). Measure of privacy in randomized response model. *Quality & Quantity*, *46*(4): 1167-1180.

Estimation of a rare sensitive attribute by using Poisson probability distribution and consideration of respondent privacy protection in randomized response model

ORIGINALITY REPORT

31%

SIMILARITY INDEX

PRIMARY SOURCES

1	Hong Zhimin. "Measure of privacy in randomized response model", Quality & Quantity, 03/23/2012 Crossref	738 words — 7%
2	G. N. Singh, Amod Kumar, Gajendra K. Vishwakarma. "Estimation of a rare sensitive attribute in a stratified two-stage unrelated randomized response model by using Poisson probability distribution", Statistics, 2018 Crossref	349 words — 3%
3	Sabrina Giordano. "Efficiency comparison of unrelated question models based on same privacy protection degree", Statistical Papers, 10/01/2011 Crossref	161 words — 1%
4	Housila P. Singh, Tanveer A. Tarray. "A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution", Statistics & Probability Letters, 2014 Crossref	141 words — 1%
5	Su, Shu-Ching, Stephan A. Sedory, and Sarjinder Singh. "Adjusted KUK'S Model using Two Non-Sensitive Characteristics Unrelated to the Sensitive Characteristic", Communication in Statistics- Theory and Methods, 2016. Crossref	124 words — 1%
6	G. N. Singh, S. Suman. "A modified two-stage randomized response model for estimating the	111 words — 1%

proportion of stigmatized attribute", Journal of Applied Statistics, 2018
Crossref

7	G. N. Singh, Amod Kumar, Gajendra K. Vishwakarma. "Some alternative additive randomized response models for estimation of population mean of quantitative sensitive variable in the presence of scramble variable", Communications in Statistics - Simulation and Computation, 2018 Crossref	95 words — 1%
8	www.tandfonline.com Internet	83 words — 1%
9	G. N. Singh, C. Singh, S. Suman. "Estimation of a rare sensitive attribute for two-stage randomized response model in probability proportional to size sampling using Poisson probability distribution", Statistics, 2019 Crossref	82 words — 1%
10	Margaret Land. "Estimation of a rare sensitive attribute using Poisson distribution", Statistics, 2011 Crossref	76 words — 1%
11	G. N. Singh, C. Singh, S. Suman. "Revisit of a randomized response model for estimating a rare sensitive attribute under probability proportional to size sampling using Poisson probability distribution", Communications in Statistics - Theory and Methods, 2018 Crossref	68 words — 1%
12	Singh, Housila P., and Tanveer A. Tarray. "A Sinuous Stratified Unrelated Question Randomized Response Model", Communication in Statistics- Theory and Methods, 2016. Crossref	57 words — 1%
13	M. Bose. "Measures of Respondent Privacy in Randomized Response Surveys", Elsevier BV, 2016 Crossref	52 words — < 1%
14	Singh, Housila P., and Tanveer A. Tarray. "A Two - Stage Land	

et al.'s randomized response model for estimating a rare sensitive attribute using Poisson distribution", Communication in Statistics- Theory and Methods, 2016.
Crossref
46 words — < 1%

15 Tanveer A. Tarray, Housial P. Singh. "A randomization device for estimating a rare sensitive attribute in stratified sampling using Poisson distribution", Afrika Matematika, 2018
Crossref
45 words — < 1%

16 "Combining Soft Computing and Statistical Methods in Data Analysis", Springer Nature, 2010
Crossref
42 words — < 1%

17 www.stacommunications.com
Internet
42 words — < 1%

18 Garib N. Singh, Surbhi Suman, Chandraketu Singh. "Estimation of a rare sensitive attribute in two-stage sampling using a randomized response model under Poisson distribution", Mathematical Population Studies, 2019
Crossref
42 words — < 1%

19 Hong Zhimin. "A Note of Proposed Privacy Measures in Randomized Response Models", Advances in Intelligent and Soft Computing, 2010
Crossref
41 words — < 1%

20 Amod Kumar, G. N. Singh, Gajendra K. Vishwakarma. "An Efficient Survey Technique for Estimating the Proportion and Sensitivity Attributes in a Dichotomous Finite Population", Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 2019
Crossref
39 words — < 1%

21 Tarray, Tanveer A., and Housila P. Singh. "An optional randomized response model for estimating a rare sensitive attribute using Poisson distribution", Communication in Statistics- Theory and Methods, 2016.
Crossref
35 words — < 1%

22	journals.sagepub.com Internet	33 words — < 1%
23	Kaizhi Yu, Hong Zou. " The Combined Poisson INMA( ) Models for Time Series of Counts ", Journal of Applied Mathematics, 2015 Crossref	31 words — < 1%
24	Surbhi Suman, G. N. Singh. "An ameliorated stratified two-stage randomized response model for estimating the rare sensitive parameter under Poisson distribution", Statistics, 2019 Crossref	30 words — < 1%
25	Housila P. Singh, Swarangi M. Gorey. "Use of weights in mixed randomized response model", Behaviormetrika, 2018 Crossref	30 words — < 1%
26	Xing Long Pan, Guo He, Chao Jie Zhang, Ting Feng Ming, Xiao Chuan Wang. "Research on Modeling and Simulating of Discrete Event System Based on Petri Net", Advanced Engineering Forum, 2012 Crossref	29 words — < 1%
27	Lee, Cheon-Sig, Shu-Ching Su, Katrina Mondragon, Veronica I. Salinas, Monique L. Zamora, Stephen Andrew Sedory, and Sarjinder Singh. "Comparison of Cramer-Rao lower bounds of variances for at least equal protection of respondents : Randomized response sampling", Statistica Neerlandica, 2015. Crossref	29 words — < 1%
28	index.stat.gov.pl Internet	28 words — < 1%
29	icstckerala.com Internet	26 words — < 1%
30	Giancarlo Diana. "A class of estimators for quantitative sensitive data", Statistical Papers, 08/20/2009	25 words — < 1%

Crossref

31 S.-C. Su, C.-S. Lee, S.A. Sedory, S. Singh. "Estimation of Means of Two Rare Sensitive Characteristics", Elsevier BV, 2016 — 24 words — < 1%
Crossref

32 siba-ese.unile.it — 23 words — < 1%
Internet

33 Arijit Chaudhuri, Tasos C. Christofides. "Indirect Questioning in Sample Surveys", Springer Nature, 2013 — 21 words — < 1%
Crossref

34 Kuo-Chung Huang. "Constructing optimal randomized response designs with consideration for the level of privacy protection", Statistica Neerlandica, 8/2007 — 20 words — < 1%
Crossref

35 www.stat.gov.pl — 18 words — < 1%
Internet

36 Reggie Davidrajuh. "Modeling Discrete-Event Systems with GPenSIM", Springer Nature, 2018 — 17 words — < 1%
Crossref

37 Geoffrey H. Moore. "Analysis", Challenge, 2015 — 17 words — < 1%
Crossref

38 Tanveer Ahmad Tarray, Housila Prasad Singh, Saadia Masood. "An Endowed Randomized Response Model for Estimating a Rare Sensitive Attribute Using Poisson Distribution", Trends in Applied Sciences Research, 2019 — 17 words — < 1%
Crossref

39 Manoj Bhargava. "On the efficiency comparison of certain randomized response strategies", Metrika, 06/01/2002 — 16 words — < 1%
Crossref

40 Nikolay I. Kolev. "Detonation waves caused by chemical reactions or by melt-coolant interactions", — 16 words — < 1%

Multiphase Flow Dynamics, 2007
Crossref

---

41 epdf.tips
Internet
16 words — < 1%

---

42 Gi-Sung Lee, Ki-Hak Hong, Chang-Kyoon Son. "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, 2015
Crossref
16 words — < 1%

---

43 Singh, Housila P., and Tanveer A. Tarray. "An Improvement Over Kim and Elam Stratified Unrelated Question Randomized Response Mode Using Neyman Allocation", Sankhya B, 2015.
Crossref
16 words — < 1%

---

44 Gi-Sung Lee, Ki-Hak Hong, Chang-Kyoon Son. "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, 2018
Crossref
16 words — < 1%

---

45 Housila P. Singh, Tanveer A. Tarray. "An Efficient Alternative Mixed Randomized Response Procedure", Sociological Methods & Research, 2014
Crossref
16 words — < 1%

---

46 Jong-Min Kim, Matthew E. Elam. "A stratified unrelated question randomized response model", Statistical Papers, 2007
Crossref
15 words — < 1%

---

47 docslide.us
Internet
15 words — < 1%

---

48 Han Qiu, YuFeng Li, Peng Yi, JiangXing Wu. "A New Look at Buffer Sizing for Core Routers", 2006 First International Conference on Communications and
15 words — < 1%

Networking in China, 2006
Crossref

49 rivista-statistica.unibo.it
Internet
14 words — < 1%

50 Hussain, Zawar, Javid Shabbir, Zahid Pervez, Said Farooq Shah, and Manzoor Khan. "Generalized Geometric distribution of order k: A flexible choice to randomize the response", Communications in Statistics - Simulation and Computation, 2016.
Crossref
14 words — < 1%

51 Singh, Housila P., and Tanveer A. Tarray. "An Adroit Stratified Unrelated Question Randomized Response Model using Neyman Allocation", Sri Lankan Journal of Applied Statistics, 2014.
Crossref
14 words — < 1%

52 G. N. Singh, C. Singh, S. Suman, A. Kumar. "A two-stage unrelated randomized response model for estimating a rare sensitive attribute in probability proportional to size sampling using Poisson distribution", Communications in Statistics - Theory and Methods, 2017
Crossref
14 words — < 1%

53 graemeblair.com
Internet
13 words — < 1%

54 Leysieffer, Frederick W., and Stanley L. Warner. "Respondent Jeopardy and Optimal Designs in Randomized Response Models", Journal of the American Statistical Association, 1976.
Crossref
13 words — < 1%

55 ww2.amstat.org
Internet
13 words — < 1%

56 Emanuela Petracci. "Universal solutions for the classical dynamical Yang–Baxter equation and the Maurer–Cartan equations", Journal of Physics A: Mathematical and General, 2004
Crossref
12 words — < 1%

57 file.scirp.org
Internet
12 words — < 1%

58 Chua, T.C.. "Procuring honest responses indirectly", Journal of Statistical Planning and Inference, 20000901
Crossref
12 words — < 1%

59 Tapan K. Nayak. "On randomized response surveys for estimating a proportion", Communications in Statistics - Theory and Methods, 1994
Crossref
12 words — < 1%

60 Huihui Li. "Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations", Theoretical and Applied Genetics, 01/2008
Crossref
12 words — < 1%

61 H. Dominguez, B. Hribar Lee, V. Vlachy, O. Pizio. "Adsorption of electrolyte in a templated hard-sphere matrix. Predictions of the continuum replica Ornstein–Zernike approach", Physica A: Statistical Mechanics and its Applications, 2003
Crossref
12 words — < 1%

62 Lee, Gi-Sung, Ki-Hak Hong, and Chang-Kyoon Son. "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, 2015.
Crossref
12 words — < 1%

63 edoc.pub
Internet
11 words — < 1%

64 Housila P. Singh. "Unknown Repeated Trials in the Unrelated Question Randomized Response Model", Biometrical Journal, 07/2004
Crossref
10 words — < 1%

Giancarlo Diana, Pier Francesco Perri. "Estimating a sensitive

65	proportion through randomized response procedures based on auxiliary information", Statistical Papers, 2007 Crossref	10 words — < 1%
66	Son, Chang-Kyoon, and Jong-Min Kim. "Calibration Estimator of a Rare Sensitive Attribute with Poisson Distribution", Communication in Statistics- Theory and Methods, 2014. Crossref	10 words — < 1%
67	www.math.zju.edu.cn Internet	10 words — < 1%
68	download.atlantis-press.com Internet	10 words — < 1%
69	Wu, C.-H.. "Randomized response model in a matched pair study", Journal of Statistical Planning and Inference, 20080701 Crossref	10 words — < 1%
70	www.tamuk.edu Internet	10 words — < 1%
71	Giancarlo Diana. "Estimating a sensitive proportion through randomized response procedures based on auxiliary information", Statistical Papers, 06/2009 Crossref	9 words — < 1%
72	Singh, Housila P., and Tanveer A. Tarray. "A Stratified Tracy and Osahan's Two - Stage Randomized Response Model", Communication in Statistics- Theory and Methods, 2015. Crossref	9 words — < 1%
73	Lee, Gi-Sung, Daiho Uhm, and Jong-Min Kim. "Estimation of a rare sensitive attribute in probability proportional to size measures using Poisson distribution", Statistics, 2014. Crossref	9 words — < 1%

74 Internet — 9 words — < 1%

75 www.msandri.it
Internet — 9 words — < 1%

76 cran.r-project.org
Internet — 8 words — < 1%

77 link.springer.com
Internet — 8 words — < 1%

78 shareok.org
Internet — 8 words — < 1%

79 Arijit Chaudhuri. "Protection of privacy in efficient application of randomized response techniques", Statistical Methods and Applications, 04/26/2008
Crossref — 8 words — < 1%

80 G. N. Singh, Amod Kumar, Gajendra K. Vishwakarma. "DEVELOPMENT OF CHAIN-TYPE EXPONENTIAL ESTIMATORS FOR POPULATION VARIANCE IN TWOPHASE SAMPLING DESIGN IN PRESENCE OF RANDOM NON-RESPONS", Statistics in Transition New Series, 2019
Crossref — 8 words — < 1%

81 www.cvmt.auc.dk
Internet — 8 words — < 1%

82 ddd.uab.cat
Internet — 8 words — < 1%

83 www.ijbssnet.com
Internet — 8 words — < 1%

84 search.crossref.org
Internet — 8 words — < 1%

85 Chaudhuri, . "Miscellaneous Techniques, Applications, and Conclusions", Statistics A Series of Textbooks and Monographs, 2010. — 7 words — < 1%

Crossref

86	Massimo Guerriero, Marco F. Sandri. "A note on the comparison of some randomized response procedures", Journal of Statistical Planning and Inference, 2007 Crossref	7 words — < 1%
87	Guerriero, M.. "A note on the comparison of some randomized response procedures", Journal of Statistical Planning and Inference, 20070701 Crossref	7 words — < 1%
88	Oluseun Odumade. "Improved Bar-Lev, Bobovitch, and Boukai Randomized Response Models", Communications in Statistics - Simulation and Computation, 03/2009 Crossref	7 words — < 1%
89	Abdelfatah, Sally, and Reda Mazloum. "Efficient Estimation in a Two-Stage Randomized Response Model", Mathematical Population Studies, 2015. Crossref	7 words — < 1%
90	Sat Gupta, Samridhi Mehta, Javid Shabbir, Sadia Khalil. "A unified measure of respondent privacy and model efficiency in quantitative RRT models", Journal of Statistical Theory and Practice, 2018 Crossref	6 words — < 1%
91	Sarjinder Singh. "Advanced Sampling Theory with Applications", Springer Nature, 2003 Crossref	6 words — < 1%
92	Kluwer Texts in the Mathematical Sciences, 1996. Crossref	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		