Journal of Sciences, Islamic Republic of Iran 23(4): 347-356 (2012) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

New Randomized Response Procedures

Z. Hussain[1,2,*] and J. Shabbir[2]

[1]*Department of Statistics, Faculty of Sciences, King Abdulaziz University 80203, Jeddah 21589, Saudi Arabia*
[2]*Department of Statistics, Faculty of Sciences, Quaid-i-Azam University 45320, Islamabad 44000, Pakistan*

Received: 16 January 2012 / Revised: 3 December 2012 / Accepted: 7 January 2013

Abstract

This article focuses on the estimation of population proportion when the study variable is sensitive in nature. Two implicit randomized response techniques are proposed where the unrelated trait can be chosen subjectively. In addition to unbiased estimation of population proportion and variance, an empirical study is conducted to inspect the relative efficiency facet of the proposed techniques. The cases of positive binomial and negative binomial sampling are also studied. The proposed techniques are exposed to be better at the job than the accustomed randomized response dealings in binomial sampling. Further, it is established that negative binomial sampling may result in more precise estimation of population proportion using the proposed techniques.

Keywords: Randomized response; Estimation of proportion; Sensitive attribute; Dichotomous population and Relative efficiency

Introduction

Survey techniques are applied more or less in every field of scientific and social studies ranging from physical sciences to economics, from business studies to bioinformatics, from educational behaviors to reliability engineering etc. Collection of unswerving information has come out as an exigent concern in socio-economic and behavioral studies reason being making dependable and compelling inferences primarily depends upon the dependability of the data. Warner [32], for the first time considered this concern and projected an inventive and original technique, called Randomized Response Technique (RRT), to elicit truthful data for estimating proportion of a sensitive trait. Warner's method consists of two randomized questions pertaining to the possession of a sensitive attribute A or its non-sensitive complement $\bar{A}$. The idea of randomizing the response was further improved by Greenberg et al. [10] to the use of unrelated trait, say Y, where a selected respondent is asked about the possession of A or Y. For a review of a rich amount of available literature on RRT one can refer to Fox and Tracy [8], Chaudhuri and Mukerjee [6] and Tracy and Mangat [31]. Variants of Warner's RRT have been suggested by a number of researchers. Greenberg et al. [10], Kuk [17], Mangat and Singh [21,22], Mangat [20], Mangat et al. [23], Mangat et al. [24], Mahmood et al. [19], Bhargava and Singh [4], Singh et al. [28], Singh et al. [30], Chang and Huang [5], Gupta et al. [11], Christofides [7], Kim and Warde [18], Huang [14] and Hussain et al. [15] Hussain and Shabbir [16] are some of the many to be cited. It has been reported by Huang [14] that accustomed RRTs have some limitations. For example, some respondents may refuse to answer at all because the statements in a given randomized response techniques are essentially

* Corresponding author, Tel.: +92(51)90642185, E-mail: zhlangah@yahoo.com

direct questions about the possession of a sensitive trait. In the forced response model some respondents may feel embarrassed to report simply a *yes* response. In the unrelated question randomized response models there is a requirement of two independent sub-samples and optimal allocation of sample size into two sub-samples depends upon the unknown value of π. Further, the relative efficiency is related to the values of π_Y as it requires to have π_Y on the same side of 0.5 as is the π. The maximum efficiency is achieved when $|\pi_Y - 0.5|$ is maximum. In practice, it is difficult to have such an unrelated attribute Y for which $|\pi_Y - 0.5|$ is a maximum. Also, because π is unknown, the selection of unrelated attribute becomes more difficult.

In many practical situations, generally, multiple sensitive items are studied. It may happen that some of the items are very rare (abundant) with very small (large) population proportions. For such items the probability of a *yes* response through a given RRT turns out to be very small (large). Obviously, in these situations we may have a small (large) number of *yes* responses which is not desirable from privacy point of view. For example, in a psychological/medical study the number of patients who evade tax may be very small. In such cases the probability of having an estimate outside the [0,1] interval is increased. To avoid such cases, Mangatand Singh [21] suggested applying negative binomial sampling with Warner [32] RRT. Singh and Mathur [29] extended the study by Mangat and Singh [21] and suggested several upper bounds on the variance of the estimator. The application of negative binomial sampling to unrelated trait RRTs cannot be found in literature. Moreover, through many studies, it has been established that unrelated trait RRTs perform relatively better (see Greenberg et al. [10], Mahmood et al. [19], and Huang [14], etc.). Therefore, the problem of studying the unrelated trait RRTs becomes more apparent and demanding. In this paper, we study the unrelated trait RRTs and improve them further by using negative binomial sampling.

The contribution of this paper is twofold in the sense that we suggest two new unrelated trait RRTs and study their performance under binomial and negative binomial sampling methods. There are two adavantages of the proposed RRTs: we do not need the two subsamples, and the unrelated trait may be chosen arbitrarily. Also, the proposed randomized response procedures circumvent the difficulties pointed out by Huang [14]. Iin addition, proposed estimators yeild a moderate number of *yes* responses to maintain privacy and consequently obtaining an estimate in the interval [0,1]. Further, three upper bounds of the variances of the proposed estimators are also given and compared with each other along with the exact variance. The proposed techniques are studied using binomial sampling and compared with that of Mangat et al. [21], Mahmood et al. [19] and Bhargava and Singh [4] randomized response techniques. The proposed techniques are also studied using negative binomial sampling and then comparison of positive and negative binomial sampling methods is made for some values of the design parameters.

The organization of the paper is as follows. In sections 2 and 3, we present the proposed techniques assuming binomial and negative binomial sampling designs. Comparisons are made in section 4 followed by conclusion in section 5.

The Proposed Techniques Assuming Binomial Sampling

This section presents two new techniques for estimating the population proportion of a sensitive attribute.

Technique I

Consider a dichotomous population $U = \{u_1, u_2, ..., u_N\}$ in which every u_i can be classified either to a sensitive group A or to its compliment, $\bar{A}$. The focus of the study lies in the estimation of the population proportion of the u_i's which are actually classified in the sensitive group A. Let Y be an unrelated trait. In the proposed technique a simple random sample of size n is drawn with replacement from the population. The proposed procedure consists of two types of statements. With probability p_1 the respondent is asked to answer to the statement (a) "I possess both the attributes A and Y" and with probability $(1-p_1)$ to answer (b) "I posses the attribute A and do not possess the attribute Y". The statement randomly selected by the interviewee is unseen to the surveyor. Let $P(.)$ be the probability of a particular eventthen $P(A) = \pi$. Through the suggested technique, the probability of obtaining a *yes* answer is given by

$$\beta_1 = p_1\{P(A \cap Y)\} + (1-p_1)\{P(A \cap Y^c)\}$$

$$\beta_1 = (1-p_1)P(A) + (2p_1 - 1)\{P(A \cap Y)\}. \qquad (1)$$

Now form (1) it is obvious that to have probability of *yes* (β_1) unconnected to the trait Y the coefficient of

$P\left(A \cap Y\right)$ must be zero, which will be the case when $p_1 = 0.5$. Consequently, we have

$$\beta_1 = \frac{1}{2}\pi . \quad (2)$$

Using (2) and method of moments, an unbiased estimator of π is proposed as

$$\hat{\pi}_1 = 2\hat{\beta}_1 , \quad (3)$$

Where $\hat{\beta}_1 = \frac{n_1}{n}$ and n_1 is the number of *yes* responses in the sample.

The variance of $\hat{\pi}_1$ is given by

$$Var\left(\hat{\pi}_1\right) = \frac{4\beta_1\left(1-\beta_1\right)}{n} . \quad (4)$$

Substituting the value of $\left(\beta_1\right)$ from (2) in (4) we get

$$Var\left(\hat{\pi}_1\right) = \frac{4\left(\frac{\pi}{2}\right)\left(1-\frac{\pi}{2}\right)}{n} = \frac{\pi\left(2-\pi\right)}{n} = \frac{\pi\left(1-\pi\right)}{n} + \frac{\pi}{n} , \quad (5)$$

which is unbiasedly estimated by

$$\hat{V}ar\left(\hat{\pi}_1\right) = \frac{4\hat{\beta}_1\left(1-\hat{\beta}_1\right)}{n-1} . \quad (6)$$

Technique II

The second technique works in a similar manner as the first one except a minor difference in the statements. The statements, in the Technique II are: (c) "I Possess the attribute Y and do not possess the attribute A" and (d) " I do not possess both the attributes A and Y". The rest of the things in Techniques I and II are identical. Now, the probability of a *yes* answer is given by

$$\beta_2 = p_2\left\{P\left(A^c \cap Y\right)\right\} + \left(1-p_2\right)\left\{P\left(A^c \cap Y^c\right)\right\}$$

$$\beta_2 = \left(1-p_2\right)\left\{1-P\left(A\right)\right\} + \left(2p_2-1\right)\left\{P\left(A^c \cap Y\right)\right\} . \quad (7)$$

As in the Technique I, to have β_2 unconnected to the attribute Y, the coefficient of $P\left(A^c \cap Y\right)$ must be zero, which is the case when $p_2 = 0.5$. As a consequence β_2 is given by

$$\beta_2 = \left(1-p_2\right)\left\{1-P\left(A\right)\right\} = \left(0.5\right)\left\{1-\pi\right\} . \quad (8)$$

From (8), we have

$$\pi = \frac{0.5-\beta_2}{0.5} . \quad (9)$$

Thus using (9) and moment method of estimation, we have an unbiased estimator of the population proportion given by

$$\hat{\pi}_2 = \frac{0.5-\hat{\beta}_2}{0.5} , \quad (10)$$

with variance, given by

$$Var\left(\hat{\pi}_2\right) = \frac{4\beta_2\left(1-\beta_2\right)}{n} = \frac{\pi\left(1-\pi\right)}{n} + \frac{\left(1-\pi\right)}{n} . \quad (11)$$

An unbiased estimator of $Var\left(\hat{\pi}_2\right)$ is given by

$$\hat{V}ar\left(\hat{\pi}_2\right) = \frac{4\hat{\beta}_2\left(1-\hat{\beta}_2\right)}{n-1} . \quad (12)$$

Proposed Techniques Assuming Negative Binomial Sampling

From (2) it is obvious that when the population proportion π is very small (which may be the case in most of practical situations) the value of β_1 will be small and β_2 will be large. For such cases, the number of *yes* responses in the sample will be small for not so large n. However, having a small number of *yes* responses may not be desirable from practical point of view. In order to avoid this we may use negative binomial sampling where sampling is continued until a fixed number m of *yes* responses are obtained. Here the sample size n is not fixed in advance. By considering the two techniques discussed above, using negative binomial sampling, unbiased estimators of π are defined as given in (2) and (10) but here $\hat{\beta}_j = \frac{m-1}{n-1}, \ j = 1, 2.$ To derive the variance of these two estimators under negative binomial sampling we use the following lemma as given in the Best [2].

Lemma 3.1: If $\hat{\beta} \sim Negative\ Binomial\left(m, \beta\right)$ then

$$E\left(\hat{\beta}\right)^2 = \left(1-\beta\right)\left(m-1\right)$$
$$\left[\left\{\sum_{t=2}^{m-1}\frac{\left(-1\right)^t}{\left(m-t\right)}\left(\frac{\beta}{1-\beta}\right)^t\right\} - \left(-1\right)^m\left(\frac{\beta}{1-\beta}\right)^m \log_e \beta\right]. \quad (13)$$

Thus, the variances of the estimators now becomes

$$Var\left(\hat{\pi}_j\right) = 4\left[\left(1-\beta_j\right)\left(m-1\right)\left[\left\{\sum_{t=2}^{m-1}\frac{(-1)^t}{(m-t)}\left(\frac{\beta_j}{1-\beta_j}\right)^t\right\} - (-1)^m\left(\frac{\beta_j}{1-\beta_j}\right)^m \log_e \beta_j\right] - \beta_j^2\right]. \quad (14)$$

It is to be noted that for the existence of variance expression in (14) we must fix $m > 2$. It is obvious from (14) that as m increases it becomes tedious to have a numerical value of the variances through (14). However, following Sathe [27], Pathak and Sathe [25] and Sahai [26] different upper bound of the variances in (14) can be found.

Sathe [27] reported following upper bound for the variance of negative binomial estimator

$$UBV_1\left(\hat{\beta}\right) = \frac{2\beta^2\left(1-\beta\right)}{m-2\left(1-\beta\right)+\sqrt{\left(m-2\left(1-\beta\right)\right)^2+4\beta\left(1-\beta\right)}}. \quad (15)$$

Sahai [26] derived the upper bound for variance of the negative binomial estimator as given by

$$UBV_2\left(\hat{\beta}\right) = \frac{\beta}{6m}\left[\sqrt{A^2-12m\beta B}-A\right], \quad (16)$$

where

$$A = \left[\left\{m^2+\left(3\beta-1\right)m-3\beta\left(1-\beta\right)\right\}-\frac{6\left(1-\beta\right)^2}{\left(m+1\right)}\right]$$

and

$$B = \left\{\frac{(m-1)}{(m+1)}\left(1-\beta\right)-\left(m+2\right)\right\}\left(1-\beta\right).$$

The upper bound for the variance due to Pathak and Sathe [25] is given by

$$UBV_3\left(\hat{\beta}\right) = \frac{\beta^2\left(1-\beta\right)}{m}\left[1+\frac{2\left(1-\beta\right)}{\left(m-2\right)} - \frac{12\beta(1-\beta)}{(m-2)\left[(m+3\beta-2)+\left\{(m+5\beta-4)^2-16\beta(1-\beta)\right\}^{0.5}\right]}\right]. \quad (17)$$

Thus using (15), (16) and (17) in (14) the different upper bounds of the variance of unbiased estimators of π obtained through Techniques 1 and 2 are now given by

$$UBV_1\left(\hat{\pi}_j\right) = \frac{8\beta_j^2\left(1-\beta_j\right)}{m-2\left(1-\beta_j\right)+\sqrt{\left(m-2\left(1-\beta_j\right)\right)^2+4\beta_j\left(1-\beta_j\right)}}, \quad (18)$$

$$UBV_2\left(\hat{\pi}_j\right) = \frac{4\beta_j}{6m}\left[\sqrt{A_j^2-12m\beta_j B_j}-A_j\right], \quad (19)$$

where A_j and B_j, $(j = 1, 2)$ are defined as earlier, and

$$UBV_3\left(\hat{\pi}_j\right) = \frac{4\beta_j^2\left(1-\beta_j\right)}{m}\left[1+\frac{2\left(1-\beta_j\right)}{\left(m-2\right)} - \frac{12\beta_j\left(1-\beta_j\right)}{(m-2)\left[\left(m+3\beta_j-2\right)+\left\{\left(m+5\beta_j-4\right)^2-16\beta_j\left(1-\beta_j\right)\right\}^{0.5}\right]}\right]. \quad (20)$$

It is to be mentioned that the values of the exact variance of the estimator $\hat{\pi}_2$ and its different upper bounds are calculated numerically for different values of π and are given in the Table 1 (similarly, the upper bounds on the variance of $\hat{\pi}_1$ can be calculated). From Table 1, it is observed that for $\pi = 0.01$ and 0.05, all the three upperbounds $UBV_1\left(\hat{\pi}_2\right)$, $UBV_2\left(\hat{\pi}_2\right)$ and $UBV_3\left(\hat{\pi}_2\right)$ are exactly equal to the actual variance, $Var\left(\hat{\pi}_2\right)$, over a wide range of m. For $\pi = 0.1$, and $m \le 12$, $UBV_1\left(\hat{\pi}_2\right)$ and $UBV_2\left(\hat{\pi}_2\right)$ are equal to the actual variance followed by $UBV_1\left(\hat{\pi}_2\right)$. For $\pi = 0.1$ and $12 < m \le 25$ all the three upperbounds are equal to actual variances. When $\pi = 0.15$ and $m = 25$ all the upperbounds are equal to true variance. When $5 \le m < 25$, $UBV_2\left(\hat{\pi}_2\right)$ is closer to the true variance followed by $UBV_1\left(\hat{\pi}_2\right)$ and $UBV_3\left(\hat{\pi}_2\right)$. Similarly when $\pi = 0.2, 0.25$, $UBV_2\left(\hat{\pi}_2\right)$ is closer to the true variance followed by $UBV_1\left(\hat{\pi}_2\right)$ and $UBV_3\left(\hat{\pi}_2\right)$. From the above observation a general conclusion can be made that for $0.01 \le \pi \le 0.25$ and $5 \le m \le 25$, $UBV_2\left(\hat{\pi}_2\right)$ is the best approximation of the $Var\left(\hat{\pi}_2\right)$ compared to $UBV_1\left(\hat{\pi}_2\right)$ and $UBV_3\left(\hat{\pi}_2\right)$. Therefore, to calculate the true variance and relative efficiency of $\hat{\pi}_2$, $UBV_2\left(\hat{\pi}_2\right)$ may be used in practice.

Table 1. Values of exact variance of $\hat{\pi}_2$ under negative binomial sampling and its upper bounds for different values of π and m

	π = 0.01				**π = 0.05**			
m	$Var(\hat{\pi})$	$UBV_1(\hat{\pi})$	$UBV_2(\hat{\pi})$	$UBV_3(\hat{\pi})$	$Var(\hat{\pi})$	$UBV_1(\hat{\pi})$	$UBV_2(\hat{\pi})$	$UBV_3(\hat{\pi})$
5	0.000033	0.000033	0.000033	0.000033	0.000793	0.000797	0.000795	0.000788
6	0.000025	0.000025	0.000025	0.000025	0.000600	0.000601	0.000600	0.000601
7	0.000020	0.000020	0.000020	0.000020	0.000482	0.000482	0.000482	0.000483
8	0.000017	0.000017	0.000017	0.000017	0.000402	0.000403	0.000402	0.000403
9	0.000014	0.000014	0.000014	0.000014	0.000345	0.000346	0.000345	0.000346
10	0.000012	0.000012	0.000012	0.000012	0.000303	0.000303	0.000303	0.000303
11	0.000011	0.000011	0.000011	0.000011	0.000269	0.000269	0.000269	0.000269
12	0.000010	0.000010	0.000010	0.000010	0.000242	0.000242	0.000242	0.000243
13	0.000009	0.000009	0.000009	0.000009	0.000220	0.000221	0.000220	0.000221
14	0.000008	0.000008	0.000008	0.000008	0.000202	0.000202	0.000202	0.000202
15	0.000008	0.000008	0.000008	0.000008	0.000187	0.000187	0.000187	0.000187
16	0.000007	0.000007	0.000007	0.000007	0.000173	0.000173	0.000173	0.000174
17	0.000007	0.000007	0.000007	0.000007	0.000162	0.000162	0.000162	0.000162
18	0.000006	0.000006	0.000006	0.000006	0.000152	0.000152	0.000152	0.000152
19	0.000006	0.000006	0.000006	0.000006	0.000143	0.000143	0.000143	0.000143
20	0.000006	0.000006	0.000006	0.000006	0.000135	0.000135	0.000135	0.000135
21	0.000005	0.000005	0.000005	0.000005	0.000128	0.000128	0.000128	0.000128
22	0.000005	0.000005	0.000005	0.000005	0.000122	0.000122	0.000122	0.000122
23	0.000005	0.000005	0.000005	0.000005	0.000116	0.000116	0.000116	0.000116
24	0.000005	0.000005	0.000005	0.000005	0.000111	0.000111	0.000111	0.000111
25	0.000004	0.000004	0.000004	0.000004	0.000106	0.000106	0.000106	0.000106
	π = 0.1				**π = 0.15**			
5	0.003025	0.003050	0.003034	0.002975	0.006493	0.006562	0.006517	0.006324
6	0.002299	0.002311	0.002304	0.002309	0.004940	0.004995	0.004974	0.004993
7	0.001855	0.001859	0.001856	0.001863	0.004017	0.004031	0.004020	0.004044
8	0.001553	0.001555	0.001553	0.001559	0.003370	0.003378	0.003371	0.003390
9	0.001335	0.001337	0.001335	0.001340	0.002902	0.002907	0.002903	0.002916
10	0.001171	0.001172	0.001171	0.001174	0.002548	0.002551	0.002548	0.002558
11	0.001043	0.001043	0.001043	0.001045	0.002270	0.002273	0.002271	0.002278
12	0.000940	0.000940	0.000940	0.000942	0.002047	0.002049	0.002048	0.002054
13	0.000855	0.000856	0.000855	0.000857	0.001864	0.001866	0.001864	0.001869
14	0.000785	0.000785	0.000785	0.000786	0.001711	0.001712	0.001711	0.001715
15	0.000725	0.000725	0.000725	0.000726	0.001581	0.001582	0.001581	0.001584
16	0.000673	0.000674	0.000673	0.000674	0.001470	0.001470	0.001470	0.001472
17	0.000629	0.000629	0.000629	0.000630	0.001373	0.001373	0.001373	0.001375
18	0.000590	0.000590	0.000590	0.000590	0.001288	0.001288	0.001288	0.001290
19	0.000555	0.000555	0.000555	0.000556	0.001213	0.001213	0.001213	0.001214
20	0.000525	0.000525	0.000525	0.000525	0.001146	0.001146	0.001146	0.001147
21	0.000497	0.000497	0.000497	0.000498	0.001086	0.001087	0.001086	0.001087
22	0.000473	0.000473	0.000473	0.000473	0.001033	0.001033	0.001033	0.001033
23	0.000450	0.000450	0.000450	0.000450	0.000984	0.000984	0.000984	0.000985
24	0.000430	0.000430	0.000430	0.000430	0.000939	0.000939	0.000939	0.000940
25	0.000411	0.000411	0.000411	0.000411	0.000899	0.000899	0.000899	0.000899
	π = 0.2				**π = 0.25**			
5	0.011016	0.011153	0.011058	0.010647	0.016429	0.016656	0.016493	0.015815
6	0.008314	0.008528	0.008484	0.008528	0.012073	0.012791	0.012714	0.012801
7	0.006871	0.006900	0.006876	0.006931	0.010324	0.010376	0.010334	0.010434
8	0.005773	0.005793	0.005779	0.005820	0.008678	0.008726	0.008700	0.008777
9	0.004981	0.004991	0.004982	0.005013	0.007509	0.007527	0.007511	0.007567
10	0.004377	0.004384	0.004378	0.004401	0.006605	0.006618	0.006607	0.006649
11	0.003904	0.003909	0.003904	0.003922	0.005896	0.005905	0.005897	0.005928
12	0.003523	0.003526	0.003523	0.003536	0.005323	0.005330	0.005324	0.005348
13	0.003209	0.003212	0.003210	0.003220	0.004852	0.004857	0.004852	0.004871
14	0.002947	0.002949	0.002947	0.002955	0.004457	0.004461	0.004458	0.004473
15	0.002724	0.002726	0.002724	0.002731	0.004122	0.004125	0.004122	0.004134
16	0.002533	0.002534	0.002533	0.002538	0.003833	0.003836	0.003833	0.003843
17	0.002366	0.002367	0.002367	0.002371	0.003583	0.003584	0.003583	0.003591
18	0.002221	0.002221	0.002221	0.002224	0.003362	0.003364	0.003363	0.003369
19	0.002092	0.002092	0.002092	0.002095	0.003168	0.003169	0.003168	0.003174
20	0.001977	0.001977	0.001977	0.001980	0.002995	0.002996	0.002995	0.003000
21	0.001874	0.001875	0.001874	0.001876	0.002839	0.002840	0.002839	0.002843
22	0.001781	0.001782	0.001781	0.001783	0.002699	0.002700	0.002699	0.002703
23	0.001697	0.001698	0.001697	0.001699	0.002572	0.002573	0.002572	0.002576
24	0.001621	0.001621	0.001621	0.001623	0.002457	0.002457	0.002457	0.002460
25	0.001551	0.001551	0.001551	0.001553	0.002351	0.002352	0.002351	0.002354

Now we study the relative efficiency features of the proposed techniques relative to some usual randomized response techniques.

Efficiency Comparisons and Discussion

We, now, compare the proposed RRTs under the two cases of sampling, namely, binomial and negative binomial sampling.

(a) Case of Binomial Sampling

(i) $\hat{\pi}_1$ versus $\hat{\pi}_2$

The proposed estimator $\hat{\pi}_1$ will be relatively more efficient than the second proposed estimator $\hat{\pi}_2$ if

$$Var(\hat{\pi}_1) < Var(\hat{\pi}_2). \quad (21)$$

Using (5) and (11) in (21) we see that the inequality (21) holds when $\pi < 0.5$, which implies that the $\hat{\pi}_1$ will be more precise as compared to $\hat{\pi}_2$ for $\pi < 0.5$. On the other hand, $\hat{\pi}_2$ will be more efficient than $\hat{\pi}_1$ when $\pi > 0.5$. It is quite clear that the two estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ will be equally good at $\pi = 0.5$.

(ii) Proposed estimators ($\hat{\pi}_1$ and $\hat{\pi}_2$) versus Mahmood et al. estimator

We compare our proposed estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ with the Mahmood et al. [19] estimator depending upon the value of the population proportion π. Mahmood et al. [19] actually presented three estimators and indicated one as the best of them. We take this best one for our comparison purposes. The minimum variance expression of the Mahmood et al. [19] estimator , say $\hat{\pi}_3$, is given by

$$Var(\hat{\pi}_3)_{\min} = \frac{\left[\sqrt{\beta_3(1-\beta_3)} + |p_2 - p_3|\sqrt{\pi_Y(1-\pi_Y)}\right]^2}{np_1^2}, \quad (22)$$

where $\beta_3 = p_1\pi + p_2(1-\pi_Y) + p_3\pi_Y$, and p_1, p_2, p_3 are pre-assigned probabilities of randomly selecting the statements concerning the possession of A, Y^c, and Y, respectively. An empirical study is undertaken to see the variation in extent of relative efficiency by fixing the practicable values of the parameters. The Relative Efficiency (*RE*) of the proposed estimators with respect to Mahmood et al. [19] procedure is defined as

$$RE_1 = \begin{cases} \dfrac{Var(\hat{\pi}_3)_{\min}}{Var(\hat{\pi}_2)}, & \text{if } \pi \geq 0.5 \\[2ex] \dfrac{Var(\hat{\pi}_3)_{\min}}{Var(\hat{\pi}_1)}, & \text{if } \pi < 0.5. \end{cases}$$

We have chosen $p_1 = 0.5$ and values of p_3 are taken as $i \times (1-p_1)/9,\ i = 1,2,3,4$ against the whole range of π. It is observed that for $p_3 = i \times (1-p_1)/9, i = 5,6,7,8$ the values of RE_1 are the mirror image of the values when $i = 1,2,3,4$. Therefore, for the sake of brevity, we have not provided the values of RE_1 for $i = 1,2,3,4$. The values of RE_1 are presented in Table 2 which clearly shows the better performance of the proposed estimators as compared to the Mahmood et al. [19] procedure. It is observed that for a fixed value of π_Y the RE_1 decreases when $|\pi - 0.5|$ increases. Also, RE_1 increases, for fixed values of π_Y and π, if $|p_3 - 0.5|$ increases. The magnitude of RE_1 ranges from 1.42 to 9.25.

(iii) Proposed estimators versus Mangat et al. and Bhargava and Singh estimators

The variance expression of Mangat et al. [23] estimator, say $\hat{\pi}_4$, is given by

$$Var(\hat{\pi}_4) = \frac{\pi(1-\pi)}{n} + \frac{\pi p_3}{n(p_1 - p_2)} + \frac{p_2(1-p_2)}{n(p_1-p_2)^2}, \quad (23)$$

where p_1 and p_2 are the pre-assigned probabilities of choosing a question concerning the membership in A, A^c and p_3 proportion of the sampled respondents are requested to say just *no*. The variance of Bhargava and Singh [4] estimator, say $\hat{\pi}_5$, is given by

$$Var(\hat{\pi}_5) = \frac{\pi(1-\pi)}{n} + \frac{\pi p_3}{n(p_1 - p_2)} + \frac{p_1(1-p_1)}{n(p_1-p_2)^2}, \quad (24)$$

where p_1 and p_2 are same as that of Mangat et al. [23] procedure, and p_3 is the probability of reporting just a *yes* answer.

It has been reported by Bhargava and Singh [4] that their estimator is better than Mangat et al. [23] estimator if $\pi > 0.5$. So we have defined the *RE* of our proposed estimators depending upon the values of the π. When $\pi \geq 0.5$, we compare our second estimator $\hat{\pi}_2$ with Bhargava and Singh [4] estimator, $\hat{\pi}_5$. Otherwise, we compare $\hat{\pi}_1$ with $\hat{\pi}_4$. That is, the *RE* of the proposed

Table 2. The values of RE_1 of the estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ with respect to $\hat{\pi}_3$

π_Y	π								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
					p_3 = 1/8				
0.1	7.95	4.22	2.95	2.29	1.86	2.04	2.34	2.91	4.61
0.3	9.25	5.03	3.60	2.86	2.40	2.73	3.26	4.31	7.42
0.5	8.96	5.00	3.67	2.99	2.57	2.99	3.67	5.00	8.96
0.7	7.42	4.31	3.26	2.73	2.40	2.86	3.60	5.03	9.25
0.9	4.61	2.91	2.34	2.04	1.86	2.29	2.95	4.22	7.95
					p_3 = 2/8				
0.1	6.96	3.76	2.66	2.09	1.73	1.93	2.25	2.88	4.74
0.3	7.73	4.24	3.06	2.45	2.07	2.37	2.84	3.77	6.53
0.5	7.50	4.21	3.10	2.53	2.17	2.53	3.10	4.21	7.50
0.7	6.53	3.77	2.84	2.37	2.07	2.45	3.06	4.24	7.73
0.9	4.74	2.88	2.25	1.93	1.73	2.09	2.66	3.76	6.96
					p_3 = 3/8				
0.1	5.95	3.27	2.36	1.88	1.58	1.79	2.13	2.78	4.70
0.3	6.31	3.51	2.56	2.07	1.76	2.02	2.44	3.25	5.66
0.5	6.17	3.48	2.57	2.10	1.81	2.10	2.57	3.48	6.17
0.7	5.66	3.25	2.44	2.02	1.76	2.07	2.56	3.51	6.31
0.9	4.70	2.78	2.13	1.79	1.58	1.88	2.36	3.27	5.95
					p_3 = 4/8				
0.1	4.93	2.77	2.04	1.66	1.42	1.63	1.97	2.62	4.54
0.3	5.02	2.84	2.10	1.71	1.47	1.70	2.06	2.76	4.82
0.5	4.97	2.83	2.10	1.72	1.48	1.72	2.10	2.83	4.97
0.7	4.82	2.76	2.06	1.70	1.47	1.71	2.10	2.84	5.02
0.9	4.54	2.62	1.97	1.63	1.42	1.66	2.04	2.77	4.93

Table 3. The values of RE_2 of the estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ with respect to $\hat{\pi}_4$ and $\hat{\pi}_5$

p_3	π								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
1/18	81.80	84.33	88.91	96.23	107.66	96.23	88.91	84.33	81.80
2/18	20.44	21.05	22.15	23.91	26.66	23.91	22.15	21.05	20.44
3/18	9.08	9.33	9.79	10.52	11.66	10.52	9.79	9.33	9.08
4/18	5.10	5.23	5.46	5.83	6.41	5.83	5.46	5.23	5.10
5/18	3.26	3.33	3.46	3.66	3.98	3.66	3.46	3.33	3.26
6/18	2.26	2.30	2.37	2.48	2.66	2.48	2.37	2.30	2.26
7/18	1.66	1.68	1.71	1.77	1.87	1.77	1.71	1.68	1.66
9/18	1.26	1.27	1.29	1.31	1.35	1.31	1.29	1.27	1.26

Table 4. The values of *RE* of proposed estimator $\hat{\pi}_2$ under negative binomial sampling relative to binomial sampling for different values of n, π and m

	n = 25				n = 35			
m	π = 0.01	π = 0.05	π = 0.1	π = 0.15	π = 0.01	π = 0.05	π = 0.1	π = 0.15
5	1.065	1.065	2.512	1.709	17.228	3.511	1.794	1.221
6	1.450	1.450	3.306	2.247	22.933	4.647	2.362	1.605
7	1.695	1.695	4.098	2.763	28.643	5.785	2.927	1.974
8	2.017	2.017	4.894	3.294	34.354	6.925	3.496	2.353
9	2.331	2.331	5.692	3.825	40.067	8.066	4.066	2.732
10	2.649	2.649	6.490	4.357	45.780	9.208	4.636	3.112
11	2.968	2.968	7.289	4.889	51.493	10.350	5.207	3.492
12	3.287	3.287	8.088	5.421	57.206	11.492	5.777	3.872
13	3.607	3.607	8.887	5.954	62.920	12.634	6.348	4.253
14	3.926	3.926	9.687	6.487	68.634	13.776	6.919	4.633
15	4.246	4.246	10.486	7.019	74.348	14.919	7.490	5.014
16	4.565	4.565	11.286	7.552	80.062	16.061	8.061	5.394
17	4.885	4.885	12.085	8.085	85.775	17.204	8.632	5.775
18	5.205	5.205	12.885	8.618	91.489	18.347	9.204	6.156
19	5.524	5.524	13.685	9.151	97.204	19.489	9.775	6.537
20	5.844	5.844	14.484	9.684	102.918	20.632	10.346	6.917
21	6.164	6.164	15.284	10.217	108.632	21.775	10.917	7.298
22	6.484	6.484	16.084	10.750	114.346	22.917	11.489	7.679
23	6.803	6.803	16.884	11.284	120.060	24.060	12.060	8.060
24	7.123	7.123	17.684	11.817	125.774	25.203	12.631	8.441
25	7.443	7.443	18.483	12.350	131.488	26.345	13.202	8.821
	n = 50				n = 100			
5	12.059	2.458	1.256	0.855	6.030	1.229	0.628	0.427
6	16.053	3.253	1.653	1.124	8.027	1.626	0.827	0.562
7	20.050	4.049	2.049	1.382	10.025	2.025	1.024	0.691
8	24.048	4.848	2.447	1.647	12.024	2.424	1.224	0.823
9	28.047	5.646	2.846	1.912	14.023	2.823	1.423	0.956
10	32.046	6.445	3.245	2.178	16.023	3.223	1.623	1.089
11	36.045	7.245	3.645	2.444	18.022	3.622	1.822	1.222
12	40.044	8.044	4.044	2.711	20.022	4.022	2.022	1.355
13	44.044	8.844	4.444	2.977	22.022	4.422	2.222	1.488
14	48.044	9.644	4.843	3.243	24.022	4.822	2.422	1.622
15	52.043	10.443	5.243	3.510	26.022	5.222	2.622	1.755
16	56.043	11.243	5.643	3.776	28.022	5.621	2.821	1.888
17	60.043	12.043	6.043	4.043	30.021	6.021	3.021	2.021
18	64.043	12.843	6.442	4.309	32.021	6.421	3.221	2.155
19	68.042	13.642	6.842	4.576	34.021	6.821	3.421	2.288
20	72.042	14.442	7.242	4.842	36.021	7.221	3.621	2.421
21	76.042	15.242	7.642	5.109	38.021	7.621	3.821	2.554
22	80.042	16.042	8.042	5.375	40.021	8.021	4.021	2.688
23	84.042	16.842	8.442	5.642	42.021	8.421	4.221	2.821
24	88.042	17.642	8.842	5.908	44.021	8.821	4.421	2.954
25	92.042	18.442	9.242	6.175	46.021	9.221	4.621	3.087

estimators relative to Mangat et al. [23] and Bhargava and Singh [4] estimators is defined as

$$RE_2 = \begin{cases} \dfrac{Var(\hat{\pi}_5)}{Var(\hat{\pi}_2)}, & \text{if } \pi \geq 0.5 \\ \dfrac{Var(\hat{\pi}_4)}{Var(\hat{\pi}_1)}, & \text{if } \pi < 0.5. \end{cases}$$

To find the numerical values of RE_2 of the proposed estimators, same values of the parameters are taken as that were fixed in calculating RE_1. The values of RE_2 are not affected by the different values of π_Y. The values of RE_2 obtained without using the parameter π_Y are presented in Table 3. From Table 3, it is observed that for a given p_3, RE_2 decreases when $|\pi - 0.5|$ increases and RE_2 is maximum when $\pi = 0.5$. It is also observed that, over the whole range of π, RE_2 decreases when p_3 increases. In general, proposed estimators are relatively more efficient than the Bhargava [4] and Mangat et al. [23] estimators for all the values of π and p_3 fixed in Table 3.

(b) Case of Negative Binomial Sampling

To see the effect of sampling design we, now, compare thenegative binomial sampling and binomial sampling designs using variances of theproposed estimator $\hat{\pi}_2$. To calculate the *RE* of the estimator $\hat{\pi}_2$ under negative binomial sampling relative to binomial sampling we use (11) and (14). The *RE* results are given in Table 4. Form Table 4, it is observed that under proposed Technique II, negative binomial sampling is more efficient than binomial sampling when the population proportion π is small. In particular, for a fixed n and π, *RE* increases when m increases. To achieve maximum efficiency a larger m should be fixed.

Results

Two new randomized response techniques are proposed where unrelated characteristic may be chosen arbitrarily . These techniques are seen to be more efficient than the techniques suggested by Mangat et al. [23], Mahmood et al. [19] and Bhargava and Singh [4] under binomial sampling (as can be seen from Tables 2 and 3). In addition to being more precise estimators, the proposed estimators do not have the weak points associated with the usual RR techniques. To avoid the possibility of having an estimate outside the interval [0,1] the use of negative binomial sampling is suggested and the Technique II is compared under the two types of sampling namely binomial and negative binomial. The negative binomial sampling is observed as the more efficient sampling. Similar results are observed for Technique I and therefore are not presented in this paper. Moreover, three different upper boundson the variance of negative binomial estimator, $\hat{\pi}_2$, have been studied and it is observed that these upper bounds are sufficiently accurate when m is larger and these can serve the purpose of calculating the variance. When m is moderate or small the upper bound proposed by Sahai [26] may be preferred. To sum up, we conclude that the newly suggested estimators are more practicable and efficient and can be easily applied in any sensitive survey.

Acknowledgements

The authors are highly grateful to the referees for guiding towards the improvement of earlier draft of this article.

References

1. Antonak, R. F. and Livneh, H. Randomized response technique: A review and proposed extension to disability attitude research. *Genetic, Social and General Psychology Monographs*, B 121(1): **97**-145 (1995).
2. Best, D. J. The variance of the inverse binomial estimator. *Biometrika*, **61**: 385-386 (1974).
3. Bouza, C. N., Herrera, C. and Pasha, G. M. A review of randomized response procedures: the qualitative variable case. *Rev InvOper*, **31**(3): 240-247 (2010).
4. Bhargava, M. and Singh, R. A modified randomization device for Warner's model. *Statistica*, **60**: 315-321 (2000).
5. Chang, H. J. and Huang, K.C. Estimation of proportion and sensitivity of a qualitative character. *Metrika*, **53**: 269-280 (2001).
6. Chaudhuri, A. and Mukerjee, R. *Randomized Response: Theory and Techniques*. Marcel Dekker, New York, (1988).
7. Christofides, T. C. A generalized randomized response technique. *Metrika* **57**: 195-200 (2003).
8. Fox, J. and Tracy, P. *Randomized Response; A Method for Sensitive Surveys*. Sage, CA, (1986).
9. Gjestvang, C. R. and Singh, S. A new randomized response Model. *J Roy. Stat Soc, B*, **68**(3): 523-530.
10. Greenberg, B. G., Abul-Ela Abdel-Latif, A., Simmons, W. R. and Horvitz, D. G. The unrelated question RR model: Theoretical framework. *J Am Stat Assoc*, **64**: 52-539 (1969).
11. Gupta, S., Gupta, B. and Singh, S. Estimation of sensitivity level of personal interview survey questions. *J Stat Plan Infer*, **100**: 239-247 (2002).
12. Guerrieroa, M. and Sandri, M. F. A note on the comparison of some randomized response procedures. *J*

Stat Plan Infer, **137**: 2184-2190 (2007).

13. Housila, P. S. and Mathur, N. On Inverse binomial randomized response technique. *J IndSocAgril Statist*, **59**(3): 192-198 (2005).
14. Huang, K. C. Estimation of sensitive data from a dichotomous population. *Stat Pap*, **47**: 149-156 (2005).
15. Hussain, Z., Shabbir, J. and Gupta S. An alternative to Ryu et al. randomized response model. *J StaManag Sys,***10**(4): 511-517 (2007).
16. Hussain, Z. and Shabbir, J. Three stage quantitative randomized response model. *J Prob and Stat Sci,* **8** (2): 223-235 (2010).
17. Kuk, A. Y. C. Asking sensitive questions indirectly. *Biometrika*, **77**: 436-438 (1990).
18. Kim, J. M., Warde, D. W. A stratified Warner's randomized response model. *J. Stat Plan Infer*, 1**20**(1-2): 155-165 (2004).
19. Mahmood, M. Singh, S. and Horn, S. On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. *Biometrical J*, **40**: 237-242 (1998).
20. Mangat, N. S. An Improved randomized response strategy. *J Roy stat Soc B*, **56**: 93-95 (1994).
21. Mangat, N. S., Singh, R. An alternative approach to randomized response survey. *Statistica, LI* (3), 327-332 (1991).
22. Mangat, N. S., Singh, R. An alternative randomized response procedure. *Biometrika*, **77**: 439-442 (1990).
23. Mangat, N. S., Singh, R. and Singh, S. Violation of respondent's privacy in Moor's model- its rectification through a random group strategy. *Commun Stat TheoMeth* **26**: 743-754 (1997).
24. Mangat, N. S., Singh, S. and Singh, R. On the use of a modified randomization device in Warner's model. *J In Soc Stat Oper Res*, **16**: 65-69 (1995).
25. Pathak, P. K. and Sathe, Y. S. A new variance formula for unbiased estimation in inverse binomial sampling. *Sankhya B*, **46**(3): 93-95 (1984).
26. Sahai, A. (1983). On a systematic sharpening of variance bounds of MVU estimator in inverse binomial sampling. *Statistica XLIII*, **4**: 621-624 (1983).
27. Sathe, Y. S. Sharper variance for unbiased estimation in inverse sampling. *Biometrika*, **64**: 425-426 (1977).
28. Singh, S., Singh, R. and Mangat, N. S. Some alternative strategies to Moor's model in randomized response sampling. *J Stat Plan Infer*, **83**: 243-255 (2000).
29. Singh, H. P., Mathur, N. On inverse binomial randomized response technique. *J In SocAgril Statist*. **59**(3), 192-198 (2005).
30. Singh, S., Mahmood, M. and Tracy, D. C. Estimation of mean and Variance of stigmatized quantitative variable using distinct units in randomized response sampling. *Stat Pap*, **42**: 403-411 (2001).
31. Tracy, D. and Mangat, N. Some development in randomized response sampling during the last decade-a follow up of review by Chaudhuri and Mukerjee. *J Appl Stat Sci*, **4**: 533-544 (1996).
32. Warner, S. L. Randomized response: a survey technique for eliminating evasive answer bias. *J Am Stat Assoc*, **60**: 63-69 (1965).