Metrika (2001) 53: 269–280

Metrika

Estimation of proportion and sensitivity of a qualitative character

Horng-Jinh Chang, Kuo-Chung Huang

Department of Management Sciences, Tamkang University, Tamsui, Taipei, Taiwan 25137, R.O.C. (E-mail: chj@mail.tku.edu.tw)

Received: October 2000

Abstract. In this paper, a simple and obvious procedure is presented that allows to estimate π, the population proportion of a sensitive group, in addition to T, the probability that the respondents belonging to the sensitive group tell the truth whenever questioning directly. Properties of the estimators of π and T as well as sample size allocation are studied. And, efficiency comparisons are carried out to investigate the performance of the proposed method. It is found that the proposed strategy is more efficient than Warner's (1965) strategy, and has an additional advantage of deciding the optimal survey technique for practical situations.

Key words: Direct response; randomized response; relative efficiency

1. Introduction

Social surveys sometimes include sensitive or threatening issues of enquiry that people like to hide from others. Direct questioning of respondents about them is likely to result either in non-response or in deliberate incorrect answer. Social stigma and fear of reprisals often lead respondents to give biased, misleading or even erroneous responses when approached with a direct response (DR) survey method. Even for the reason of merely unwillingness to reveal secrets to strangers, many individuals attempt to avoid certain questions put to them by interviewers. Consider a dichotomous population in which every person belongs either to a sensitive group A or to the non-sensitive complement $\bar{\mathrm{A}}$. The problem of interest is to estimate the population proportion π of individuals who are members of A. Let T be the probability that the respondents belonging to A report the truth. The respondents belonging to the non-sensitive group $\bar{\mathrm{A}}$ have no reason to tell a lie. For a DR survey of size n, the interviewee is asked if he/she is a member of A. Then, we have a direct estimator

$$\hat{\pi}_D = \frac{\sum_{i=1}^{n} X_i}{n}, \tag{1.1}$$

with mean square error given by

$$\mathrm{MSE}(\hat{\pi}_D) = \frac{\theta_D(1-\theta_D)}{n} + \pi^2(1-T)^2, \tag{1.2}$$

where $X_i = 1(0)$ if the ith interviewee responds yes (no), and $\theta_D = \pi T$.

In order to improve respondent cooperation and to encourage honest responses, Warner (1965) suggested the following procedure called randomized response (RR) technique. Instead of a DR method, a randomization device used to collect sample information consists of two statements: (a) 'I am a member of group A' and (b) 'I am not a member of group A', with probabilities P_W and $(1 - P_W)$ respectively. Following this device, the respondent chooses a statement unobserved by the interviewer, and then simply gives a 'yes' or 'no' answer depending on his/her actual status. Let $\overline{Y}$ be the proportion of 'yes' answers in a random sample of n respondents. By the method of moments, Warner obtained

$$\hat{\pi}_W = \frac{\overline{Y} - (1-P_W)}{(2P_W - 1)}, \quad P_W \neq \frac{1}{2}, \tag{1.3}$$

as an estimator of π, where $Y_i = 1(0)$ according as the ith respondent answers yes (no). The estimator $\hat{\pi}_W$ is unbiased and its variance is given by

$$\mathrm{Var}(\hat{\pi}_W) = \frac{\theta_W(1-\theta_W)}{(2P_W - 1)^2 n}, \tag{1.4}$$

where $\theta_W = \pi P_W + (1-\pi)(1-P_W)$. It is demonstrated useful for reducing response error when potentially sensitive topics, such as the prevalence of illegal activities, incidence of domestic violence, sexual practice or other issues that are unethical or socially unacceptable, are included in surveys of human populations. Following the lead of Warner (1965), a number of improvements and enhancements have been suggested in the literatures. See Fox and Tracy (1986), Chaudhuri and Mukherjee (1987, 1988) and Hedayat and Sinha (1991) for the reviews, and Kuk (1990), Mangat and Singh (1990), Mangat (1994), Mangat, Singh and Singh (1997), Mahmood, Singh and Horn (1998) and Singh, Singh and Mangat (2000) for some recent developments.

In general, the probability T is a measurement instrument of the sensitivity. And, it has a primary use in appraising the efficiency of different survey plans. One may use a simple formula for ascertaining whether a RR technique is advantageous in efficiency relative to a DR scheme. However, the probability T is unknown in actual practice, and there is no feasible sample analogue for this parameter so far. Therefore, in an attempt to overcome this difficulty we suggest an alternative survey strategy which makes it possible to estimate the unknown parameters π and T simultaneously. The resulting estimators and properties of biases as well as mean square errors are presented in Section 2. Moreover, sample size allocations in the presence of some practical objectives are studied in Section 3. In Section 4, with respect to the criterion of mean square error, numerical studies are worked out to demonstrate the superiority of the proposed method. The concluding section of the article consists of a concise summary.

2. Proposed method

In the proposed method, two independent sub-samples of size n_j, $j = 1, 2$, are drawn from the population using simple random sampling with replacement such that $n_1 + n_2 = n$, the total sample size is required. The person in the jth sub-sample is instructed to reply a direct question concerning whether he/she bears A or not. If answer 'no', the respondent is required to use a randomization device R_j consisting of two statements (a) and (b), with probabilities P_j and $(1 - P_j)$ respectively. The whole procedure is completed by the respondent without revealing the question selected. Notice that the respondents in the non-sensitive group have no reason to tell a lie, and thus are completely truthful in their answers whether DR or Warner's procedure is provided. It is supposed that the respondents belonging to A state honest responses in full under Warner's model but with the probability T following the usual DR procedure.

On the basis of the proposed procedure, the proportion for the respondents who respond 'yes' for either one instruction in the jth sub-sample is given by

$$\theta_j = \pi T + \pi(1 - T)P_j + (1 - \pi)(1 - P_j), \quad j = 1, 2. \tag{2.1}$$

By means of the method of moments, an estimator of population proportion π can be obtained as

$$\hat{\pi} = \frac{(1 - P_2)\bar{Z}_1 - (1 - P_1)\bar{Z}_2}{P_1 - P_2}, \quad P_1 \neq P_2, \tag{2.2}$$

and an estimator of T is given by

$$\hat{T} = \frac{(1 - 2P_2)\bar{Z}_1 - (1 - 2P_1)\bar{Z}_2 - (P_1 - P_2)}{(1 - P_2)\bar{Z}_1 - (1 - P_1)\bar{Z}_2}, \tag{2.3}$$

where $\bar{Z}_j = n_j^{-1} \sum_{i=1}^{n_j} Z_{ij}$, the estimator of θ_j, is the observed proportion of 'yes' answers reported by the respondents in the jth sub-sample, and Z_{ij} is the Bernoulli variable with parameter θ_j, $i = 1, 2, \ldots, n_j$, $j = 1, 2$, respectively.

We now derive the properties of the estimator $\hat{\pi}$. For this, we have the following theorems.

Theorem 1: *$\hat{\pi}$ is an unbiased estimator of π.*

Proof: In view of the obvious results that $E(\bar{Z}_1) = \theta_1$ and $E(\bar{Z}_2) = \theta_2$, we have

$$E(\hat{\pi}) = E\left[\frac{(1-P_2)\bar{Z}_1 - (1-P_1)\bar{Z}_2}{P_1 - P_2}\right] = \frac{(1-P_2)\theta_1 - (1-P_1)\theta_2}{P_1 - P_2} = \pi.$$

Hence the theorem.

Theorem 2: *The variance of the estimator $\hat{\pi}$ is given by*

$$\mathrm{Var}(\hat{\pi}) = \frac{1}{(P_1 - P_2)^2}\left[\frac{(1-P_2)^2\theta_1(1-\theta_1)}{n_1} + \frac{(1-P_1)^2\theta_2(1-\theta_2)}{n_2}\right]. \quad (2.4)$$

Proof: Since the sub-samples are drawn independently and $\bar{Z}_j$ follows the binomial distribution $B(n_j, \theta_j), j = 1, 2$, respectively,

$$\begin{aligned}\mathrm{Var}(\hat{\pi}) &= \frac{1}{(P_1 - P_2)^2}[(1-P_2)^2 \mathrm{Var}(\bar{Z}_1) + (1-P_1)^2 \mathrm{Var}(\bar{Z}_2)] \\ &= \frac{1}{(P_1 - P_2)^2}\left[\frac{(1-P_2)^2\theta_1(1-\theta_1)}{n_1} + \frac{(1-P_1)^2\theta_2(1-\theta_2)}{n_2}\right].\end{aligned}$$

This completes the proof.

Furthermore, let us define

$$d_1 = (1-2P_2)\bar{Z}_1 - (1-2P_1)\bar{Z}_2 - (P_1 - P_2) \quad \text{and}$$

$$d_2 = (1-P_2)\bar{Z}_1 - (1-P_1)\bar{Z}_2,$$

then the estimator $\hat{T}$ can be written as $\hat{T} = d_1/d_2$. Clearly, it can be verified that

$$E(d_1) = (P_1 - P_2)\pi T \quad \text{and} \quad E(d_2) = (P_1 - P_2)\pi,$$

and thus we have $T = E(d_1)/E(d_2)$. Moreover, we define the following quantities:

$$e_1 = \frac{d_1 - E(d_1)}{E(d_1)} \quad \text{and} \quad e_2 = \frac{d_2 - E(d_2)}{E(d_2)},$$

assuming that $|e_2| < 1$ so that the function $(1 + e_2)^{-1}$ can be validly expanded as power series. Then the estimator $\hat{T}$ can be expressed in terms of e_1 and e_2 as

$$\hat{T} = T[1 + e_1 - e_2 - e_1e_2 + e_2^2 + O_p(n^{-3/2})], \quad (2.5)$$

and the estimation error is, therefore, given by

$$\hat{T} - T = T[e_1 - e_2 - e_1e_2 + e_2^2 + O_p(n^{-3/2})]. \quad (2.6)$$

Before obtaining the bias and mean square error of the estimator $\hat{T}$, we need the following expectations:

$$E(e_1) = E(e_2) = 0$$

$$E(e_1 e_2) = \frac{1}{(P_1 - P_2)^2 \pi^2 T}\left[\frac{(1-2P_2)(1-P_2)\theta_1(1-\theta_1)}{n_1} + \frac{(1-2P_1)(1-P_1)\theta_2(1-\theta_2)}{n_2}\right]$$

$$E(e_1^2) = \frac{1}{(P_1 - P_2)^2 \pi^2 T^2}\left[\frac{(1-2P_2)^2\theta_1(1-\theta_1)}{n_1} + \frac{(1-2P_1)^2\theta_2(1-\theta_2)}{n_2}\right]$$

$$E(e_2^2) = \frac{1}{(P_1 - P_2)^2 \pi^2}\left[\frac{(1-P_2)^2\theta_1(1-\theta_1)}{n_1} + \frac{(1-P_1)^2\theta_2(1-\theta_2)}{n_2}\right]. \quad (2.7)$$

The principal properties of the estimator $\hat{T}$ are outlined in the following theorems.

Theorem 3: *Up to terms of order $O(n^{-1})$, the estimator $\hat{T}$ is biased with magnitude of bias given by*

$$\mathrm{Bias}(\hat{T}) = \frac{1}{(P_1 - P_2)^2 \pi^2}\left\{\frac{(1-P_2)[T(1-P_2) + (2P_2 - 1)]\theta_1(1-\theta_1)}{n_1} + \frac{(1-P_1)[T(1-P_1) + (2P_1 - 1)]\theta_2(1-\theta_2)}{n_2}\right\}. \quad (2.8)$$

Proof: From (2.6), ignoring terms with power in e_i's higher than the second, taking expectation and then substituting the corresponding expected values in (2.7), it follows that

$$\begin{aligned}
\mathrm{Bias}(\hat{T}) &= TE(e_1 - e_2 - e_1 e_2 + e_2^2) \\
&= T\left\{\frac{-1}{(P_1 - P_2)^2 \pi^2 T}\left[\frac{(1-2P_2)(1-P_2)\theta_1(1-\theta_1)}{n_1} + \frac{(1-2P_1)(1-P_1)\theta_2(1-\theta_2)}{n_2}\right]\right. \\
&\quad \left. + \frac{1}{(P_1 - P_2)^2 \pi^2}\left[\frac{(1-P_2)^2\theta_1(1-\theta_1)}{n_1} + \frac{(1-P_1)^2\theta_2(1-\theta_2)}{n_2}\right]\right\} \\
&= \frac{1}{(P_1 - P_2)^2 \pi^2}\left\{\frac{(1-P_2)[T(1-P_2) + (2P_2 - 1)]\theta_1(1-\theta_1)}{n_1} + \frac{(1-P_1)[T(1-P_1) + (2P_1 - 1)]\theta_2(1-\theta_2)}{n_2}\right\},
\end{aligned}$$

which proves the theorem.

Theorem 4: *Up to terms of order $O(n^{-1})$, the mean square error of the estimator $\hat{T}$ is given by*

$$\mathrm{MSE}(\hat{T}) = \frac{1}{(P_1 - P_2)^2\pi^2}\left\{\frac{[T(1-P_2)+(2P_2-1)]^2\theta_1(1-\theta_1)}{n_1} + \frac{[T(1-P_1)+(2P_1-1)]^2\theta_2(1-\theta_2)}{n_2}\right\}. \tag{2.9}$$

Proof: Squaring the formula (2.6), omitting terms with power in e_i's higher than the second and then taking expectation, we have

$$\mathrm{MSE}(\hat{T}) = T^2 E(e_1^2 - 2e_1e_2 + e_2^2). \tag{2.10}$$

On substituting the corresponding expected values in (2.7) into (2.10), and then after some algebraic simplifications yields the theorem.

3. Sample size allocations

The appropriate sample size allocations for various objectives in actual survey sampling are presented in this section.

Case I. In practical situations, the total sample size n is fixed from a consideration of available resources. One may then concern with the selection of n_1 and n_2 so that $\mathrm{Var}(\hat{\pi})$, as in (2.4), is minimized subject to $n_1 + n_2 = n$. Application of the Cauchy-Schwarz inequality to this constrained optimization problem gives the optimum choices of n_1 and n_2 as

$$n_1 = \frac{(1-P_2)\sqrt{\theta_1(1-\theta_1)}}{(1-P_2)\sqrt{\theta_1(1-\theta_1)} + (1-P_1)\sqrt{\theta_2(1-\theta_2)}} \cdot n \tag{3.1}$$

and

$$n_2 = \frac{(1-P_1)\sqrt{\theta_2(1-\theta_2)}}{(1-P_2)\sqrt{\theta_1(1-\theta_1)} + (1-P_1)\sqrt{\theta_2(1-\theta_2)}} \cdot n, \tag{3.2}$$

for which the minimum variance is given by

$$\mathrm{Var}(\hat{\pi}) = \frac{[(1-P_2)\sqrt{\theta_1(1-\theta_1)} + (1-P_1)\sqrt{\theta_2(1-\theta_2)}]^2}{(P_1-P_2)^2 n}. \tag{3.3}$$

Case II. If one is interested in the minimum mean square error of the estimator $\hat{T}$ with the constraint $n_1 + n_2 = n$, then by means of the Cauchy-Schwarz inequality, the sample size allocation for which $\mathrm{MSE}(\hat{T})$ attains its minimum is given by

$$\frac{n_1}{n_2} = \frac{|T(1-P_2)+(2P_2-1)|\sqrt{\theta_1(1-\theta_1)}}{|T(1-P_1)+(2P_1-1)|\sqrt{\theta_2(1-\theta_2)}}, \tag{3.4}$$

and the minimum mean square error is equal to

$$\mathrm{MSE}(\hat{T})$$
$$= \frac{[|T(1-P_2)+(2P_2-1)|\sqrt{\theta_1(1-\theta_1)}+|T(1-P_1)+(2P_1-1)|\sqrt{\theta_2(1-\theta_2)}]^2}{(P_1-P_2)^2\pi^2 n}. \tag{3.5}$$

Case III. Here we consider the situation that the investigator seeks to estimate π and T simultaneously precisely and then is attempted to minimize VM (say), the product of $\mathrm{Var}(\hat{\pi})$ and $\mathrm{MSE}(\hat{T})$. From (2.4) and (2.9), the expression for VM is given by

$$VM = \frac{1}{(P_1-P_2)^4\pi^2}\left[\frac{(1-P_2)^2\theta_1(1-\theta_1)}{n_1}+\frac{(1-P_1)^2\theta_2(1-\theta_2)}{n_2}\right]$$
$$\times\left\{\frac{[T(1-P_2)+(2P_2-1)]^2\theta_1(1-\theta_1)}{n_1}\right.$$
$$\left.+\frac{[T(1-P_1)+(2P_1-1)]^2\theta_2(1-\theta_2)}{n_2}\right\}. \tag{3.6}$$

Through a simple application of the Cauchy-Schwarz inequality, we have

$$VM \geq \frac{1}{(P_1-P_2)^4\pi^2}\left[\frac{(1-P_2)|T(1-P_2)+(2P_2-1)|\theta_1(1-\theta_1)}{n_1}\right.$$
$$\left.+\frac{(1-P_1)|T(1-P_1)+(2P_1-1)|\theta_2(1-\theta_2)}{n_2}\right]^2, \tag{3.7}$$

with equality if and only if

$$\frac{1-P_2}{|T(1-P_2)+(2P_2-1)|}=\frac{1-P_1}{|T(1-P_1)+(2P_1-1)|}, \tag{3.8}$$

The expression for the minimum value of VM is, therefore, given by

$$VM = \frac{1}{(P_1-P_2)^4\pi^2 n^2}[\sqrt{(1-P_2)|T(1-P_2)+(2P_2-1)|\theta_1(1-\theta_1)}$$
$$+\sqrt{(1-P_1)|T(1-P_1)+(2P_1-1)|\theta_2(1-\theta_2)}]^4, \tag{3.9}$$

when n_1 and n_2 are chosen to satisfy the relation

$$\frac{n_1}{n_2}=\frac{(1-P_2)\sqrt{\theta_1(1-\theta_1)}}{(1-P_1)\sqrt{\theta_2(1-\theta_2)}}. \tag{3.10}$$

4. Efficiency comparisons

With respect to the mean square error criterion, the efficiency aspect of the proposed strategy is investigated in relation to Warner's (1965) technique as well as direct response method. It seems difficult to achieve the elegant conditions, we, therefore, resort to the empirical studies for some practical choices of parameters. Without loss of generality, it is assumed that $P_1 > P_2$.

4.1 Comparison of the proposed method with Warner's model

The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to Warner's estimator $\hat{\pi}_W$ is defined as

$$RE_1 = \frac{\theta_W(1-\theta_W)(P_1-P_2)^2}{(2P_W-1)^2[(1-P_2)\sqrt{\theta_1(1-\theta_1)}+(1-P_1)\sqrt{\theta_2(1-\theta_2)}]^2}. \quad (4.1)$$

It is remarkable that the value of P_W should be chosen as close to unity as practicable, and one should take P_1 as large and P_2 as small as possible. Hence, we simply set $P_W = P_1$ and $P_2 = 1 - P_1$ to compare them at the same level of protection of the respondents. Moreover, the values of P_1 used are 0.7, 0.8 and 0.9, and π as well as T are taken as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. These magnitudes help the investigator to choose the better strategy in different situations. The relative efficiency figures thus computed are shown in Table 1.

From Table 1, it is interesting to find that $\hat{\pi}$ is invariably more efficient than $\hat{\pi}_W$ implying that it is always beneficial to use the proposed technique. In other words, the proposed survey method is substantiated superior to Warner's (1965) model.

4.2 Comparison of the proposed method with direct response method

The expression for the relative efficiency of the proposed estimator $\hat{\pi}$ in relation to direct estimator $\hat{\pi}_D$ can be written as

$$RE_2 = \frac{[\theta_D(1-\theta_D)+n\pi^2(1-T)^2](P_1-P_2)^2}{[(1-P_2)\sqrt{\theta_1(1-\theta_1)}+(1-P_1)\sqrt{\theta_2(1-\theta_2)}]^2}. \quad (4.2)$$

Herein, we consider the same choices of parameters as employed in Table 1. Notice that RE_2 also depends upon total sample size, and thereupon the values of n are taken to be 1000 and 2000. Tables 2 and 3 are respectively appropriate for the cases where total sample size n is 1000 and 2000.

From these tables, the efficiency of the proposed strategy is considerably higher than that of DR method for most of the practical circumstances. For example, it is observed that for $P_1 = 0.9$, the value of relative efficiency ranges from 0.797 to 5732, and the value of RE_2 is less than unity only in 0.62% of the cases. For $P_1 = 0.7$ and $P_1 = 0.8$, which are most useful and practical values of P_1, the relative amounts of $RE_2 < 1$ are 4.93% and 2.47% respectively. Moreover, it is seen that the use of the proposed estimator is limited for

Table 1. The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to Warner's estimator $\hat{\pi}_W$.

P_1	T	π								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.7	0.1	1.000	1.000	1.001	1.001	1.002	1.003	1.004	1.006	1.008
	0.2	1.000	1.002	1.003	1.006	1.008	1.012	1.017	1.024	1.033
	0.3	1.001	1.003	1.007	1.013	1.019	1.028	1.040	1.055	1.077
	0.4	1.002	1.006	1.013	1.023	1.035	1.051	1.073	1.102	1.144
	0.5	1.003	1.010	1.021	1.036	1.056	1.083	1.118	1.168	1.241
	0.6	1.004	1.014	1.030	1.053	1.083	1.124	1.179	1.259	1.383
	0.7	1.005	1.020	1.042	1.074	1.118	1.177	1.262	1.388	1.598
	0.8	1.007	1.026	1.056	1.100	1.161	1.247	1.374	1.575	1.946
	0.9	1.009	1.033	1.073	1.132	1.216	1.339	1.532	1.867	2.591
0.8	0.1	1.000	1.000	1.001	1.001	1.002	1.002	1.003	1.005	1.008
	0.2	1.000	1.001	1.003	1.004	1.006	1.009	1.014	1.021	1.032
	0.3	1.001	1.003	1.006	1.010	1.015	1.021	1.031	1.047	1.073
	0.4	1.002	1.006	1.011	1.018	1.027	1.039	1.057	1.084	1.133
	0.5	1.003	1.009	1.018	1.028	1.043	1.062	1.091	1.136	1.217
	0.6	1.004	1.013	1.026	1.042	1.064	1.093	1.136	1.206	1.336
	0.7	1.006	1.018	1.036	1.059	1.090	1.133	1.196	1.300	1.506
	0.8	1.007	1.025	1.049	1.081	1.124	1.184	1.275	1.432	1.764
	0.9	1.010	1.032	1.064	1.107	1.167	1.252	1.386	1.627	2.201
0.9	0.1	1.000	1.000	1.000	1.001	1.001	1.001	1.002	1.004	1.007
	0.2	1.000	1.001	1.002	1.002	1.004	1.005	1.008	1.014	1.027
	0.3	1.001	1.002	1.004	1.006	1.008	1.012	1.018	1.030	1.059
	0.4	1.002	1.004	1.007	1.010	1.015	1.022	1.033	1.053	1.102
	0.5	1.003	1.007	1.011	1.017	1.024	1.035	1.052	1.084	1.159
	0.6	1.004	1.010	1.017	1.025	1.036	1.051	1.076	1.122	1.234
	0.7	1.006	1.015	1.024	1.036	1.051	1.073	1.107	1.172	1.333
	0.8	1.008	1.020	1.033	1.050	1.070	1.100	1.148	1.238	1.468
	0.9	1.010	1.026	1.045	1.067	1.096	1.138	1.203	1.329	1.667

some specific parameters. As far as the efficiency of $\hat{\pi}$ is concerned, we observe that it decreases with P_1, π or n whichever decreases or with increasing T if other parameters remain unchanged. Thus, in using $\hat{\pi}$ in place of $\hat{\pi}_D$, more gain in efficiency is expected either for smaller value of T or for any one of P_1, π or n is larger.

5. Concluding remarks

Randomized response techniques are attractive mechanisms for counteracting fears in response and providing with valid statistical inferences concerning a population. However, there was no way to guess about the magnitude of T so as to judge the extent of bias involved and the effect on the accuracy in estimation. If, instead, the proposed survey procedure is employed, one has a way of estimating π unbiasedly and getting an admissible estimator for T. Thus unlike other RR strategies where estimations can only supply with the population proportion, the proposed technique results in more efficient estimator of π as compared to Warner's (1965) model, and in a great advantage of esti-

Table 2. The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to direct estimator $\hat{\pi}_D$ for $n = 1000$.

P_1	T	π								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.7	0.1	5.783	22.03	47.94	83.62	129.9	188.4	261.9	354.2	471.8
	0.2	4.579	17.44	37.99	66.37	103.3	150.3	209.6	285.0	382.1
	0.3	3.518	13.40	29.23	51.20	80.00	116.9	164.1	224.9	305.0
	0.4	2.599	9.890	21.63	38.03	59.72	87.88	124.4	172.6	238.0
	0.5	1.821	6.919	15.17	26.80	42.37	62.90	90.13	127.1	179.4
	0.6	1.186	4.481	9.846	17.49	27.88	41.86	60.92	87.80	128.1
	0.7	0.692	2.576	5.659	10.10	16.26	24.75	36.75	54.53	83.34
	0.8	0.340	1.208	2.624	4.689	7.609	11.77	17.91	27.64	45.24
	0.9	0.131	0.384	0.773	1.334	2.138	3.319	5.164	8.369	15.25
0.8	0.1	15.18	53.65	111.5	189.6	292.1	427.1	608.7	862.3	1238
	0.2	12.02	42.48	88.33	150.4	232.0	340.0	486.0	691.9	1002
	0.3	9.232	32.62	67.92	115.8	179.2	263.5	378.6	543.3	797.3
	0.4	6.820	24.08	50.21	85.85	133.3	197.0	285.1	413.7	618.6
	0.5	4.780	16.84	35.18	60.34	94.13	140.0	204.6	301.2	461.7
	0.6	3.112	10.91	22.80	39.26	61.58	92.35	136.5	204.7	324.5
	0.7	1.816	6.268	13.09	22.60	35.67	54.02	81.03	124.4	206.1
	0.8	0.893	2.940	6.061	10.45	16.57	25.34	38.68	61.18	107.6
	0.9	0.344	0.935	1.784	2.962	4.617	7.042	10.87	17.77	33.99
0.9	0.1	35.17	107.9	208.1	340.8	519.0	767.3	1135	1731	2866
	0.2	27.85	85.37	164.7	269.9	411.3	608.9	902.2	1382	2310
	0.3	21.39	65.54	126.5	207.4	316.5	469.5	697.9	1076	1823
	0.4	15.80	48.35	93.36	153.3	234.3	348.4	520.2	808.1	1394
	0.5	11.08	33.80	65.27	107.3	164.3	245.2	368.1	577.6	1019
	0.6	7.212	21.86	42.20	69.44	106.6	159.7	241.4	383.3	694.9
	0.7	4.209	12.56	24.15	39.74	61.15	92.00	140.1	225.6	422.7
	0.8	2.070	5.884	11.15	18.25	28.06	42.35	64.98	106.3	207.5
	0.9	0.797	1.870	3.270	5.133	7.712	11.50	17.62	29.18	59.66

Table 3. The relative efficiency of the proposed estimator $\hat{\pi}$ with respect to direct estimator $\hat{\pi}_D$ for $n = 2000$.

P_1	T	π								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.7	0.1	11.56	44.04	95.86	167.2	259.8	376.8	523.7	708.4	943.5
	0.2	9.144	34.85	75.95	132.7	206.6	300.5	419.2	569.9	764.1
	0.3	7.015	26.75	58.41	102.3	159.9	233.7	328.1	449.7	609.9
	0.4	5.170	19.73	43.19	75.97	119.3	175.6	248.7	345.0	475.8
	0.5	3.608	13.78	30.26	53.49	84.61	125.7	180.1	254.0	358.6
	0.6	2.331	8.889	19.59	34.85	55.61	83.55	121.6	175.4	255.9
	0.7	1.337	5.069	11.20	20.07	32.35	49.32	73.29	108.8	166.4
	0.8	0.627	2.323	5.122	9.224	15.04	23.33	35.59	55.03	90.20
	0.9	0.203	0.665	1.408	2.500	4.083	6.424	10.09	16.48	30.21
0.8	0.1	30.33	107.3	223.0	379.2	584.2	854.2	1217	1724	2476
	0.2	24.00	84.89	176.6	300.6	463.9	679.8	971.8	1384	2003
	0.3	18.41	65.15	135.7	231.5	358.2	526.8	757.0	1086	1594
	0.4	13.57	48.04	100.3	171.5	266.4	393.7	569.9	827.0	1237
	0.5	9.471	33.54	70.16	120.4	188.0	279.7	408.8	601.9	922.9
	0.6	6.118	21.63	45.37	78.23	122.8	184.3	272.6	409.0	648.4
	0.7	3.509	12.33	25.91	44.89	70.99	107.6	161.6	248.3	411.5
	0.8	1.647	5.652	11.83	20.55	32.75	50.25	76.87	121.8	214.6
	0.9	0.533	1.618	3.247	5.551	8.818	13.63	21.25	35.00	67.35
0.9	0.1	70.29	215.7	416.1	681.5	1038	1534	2269	3462	5732
	0.2	55.61	170.6	329.3	539.6	822.4	1217	1804	2763	4619
	0.3	42.66	130.9	252.8	414.6	632.7	938.5	1395	2150	3645
	0.4	31.44	96.46	186.4	306.2	468.1	696.3	1040	1616	2788
	0.5	21.95	67.29	130.2	214.1	328.2	489.9	735.6	1154	2037
	0.6	14.18	43.37	83.97	138.4	212.7	318.8	482.0	765.6	1388
	0.7	8.135	24.71	47.81	78.92	121.7	183.3	279.3	450.2	844.1
	0.8	3.818	11.31	21.76	35.89	55.47	83.98	129.1	211.7	413.8
	0.9	1.235	3.236	5.952	9.620	14.73	22.26	34.44	57.46	118.2

mating the truthful response probability T. In addition, since the values of P_1 and n are determined by the researcher, and the guessed values of π and T are available, with the help of the expression for RE_2, there is a clear suggestion that the RR method is apt to outperform the regular DR scheme in a variety of situations. Actually, this is an extra utility of the proposed survey strategy and has significantly increased the value and acceptance of survey sampling approach.

In the end, we recommend that the investigator should adopt the proposed survey technique to obtain the estimated values $\hat{\pi}$ and $\hat{T}$. Then he or she may check if a RR strategy is suitable to apply by replacing the unidentified π and T in (4.2) by the estimated values $\hat{\pi}$ and $\hat{T}$ respectively. If the resulting value of estimated RE_2 is less than unity, then one may go for the common DR method thereafter, otherwise the proposed strategy or other improved RR technique may be better in use.

References

Chaudhuri A, Mukherjee R (1987) Randomized response techniques: a review. Statistica Neerlandica 41:27–44

Chaudhuri A, Mukherjee R (1988) Randomized response: Theory and techniques. Marcel Dekker, New York

Fox J, Tracy P (1986) Randomized response: A method for sensitive surveys. Sage Publication, Beverley Hills

Hedayat AS, Sinha BK (1991) Design and inference in finite population sampling. Wiley, New York

Kuk AYC (1990) Asking sensitive questions indirectly. Biometrika 77:2, 436–438

Mahmood M, Singh S, Horn S (1998) On the confidentiality guaranteed under randomized response sampling: a comparison with several new techniques. Biom. J. 40:2, 237–242

Mangat NS (1994) An improved randomized response strategy. J. R. Statist. Soc. B. 56:93–95

Mangat NS, Singh R (1990) An alternative randomized response procedure. Biometrika 77:2, 439–442

Mangat NS, Singh R, Singh S (1997) Violation of respondent's privacy in Moor's model – its rectification through a random group strategy. Commun. Statist. Theory Meth. 26:3, 743–754

Singh S, Singh R, Mangat NS (2000) Some alternative strategies to Moor's model in randomized response sampling. J. Statist. Plann. Infere. 83:243–255

Warner SL (1965) Randomized response: a survey technique for eliminating evasive answer bias. J. Amer. Statist. Assoc. 60:63–69