A new unrelated question randomized response model†

Sanghamitra Pal* and Sarjinder Singhb

aDepartment of Statistics, West Bengal State University, Barasat, North 24 Pgs, West Bengal, India; bDepartment of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363, USA

(Received 11 December 2006; final version received 8 April 2010)


Keywords: estimation of population proportion; randomized response; unequal probability sampling; unrelated question model

2000 Mathematics Classification: 62 D05

1. Introduction

Politicians and social researchers are frequently interested in collecting information about stigmatizing issues, and the target parameter of many surveys turns out to be the population proportion \( \theta_y \) of individuals bearing a sensitive attribute. Respondents sometimes encounter sensitive questions, for example, their normatively charged values and their involvement in embarrassing behaviours such as visiting pornographic websites or illegal activities such as tax evasion as discussed in Elffers et al. [1]. In some surveys, respondents may be asked if they have abortion experience, about drug abuse history, homosexual activities and AIDS in some medical and epidemiological questionnaires. If respondents feel that they could be used to their advantage, they may choose not to answer or may intentionally provide false answers. The direct questioning methods may lead to refusals and biased answers which are likely to produce a severe underestimation of \( \theta_y \). The researchers commonly adopt the randomized response technique (RRT) which may strongly...
increase the interviewee’s cooperation as validated by many field studies. Indeed, the meta-analysis provided by Lensvelt-Mulders et al. [2] has emphasized that this survey technique yields more reliable estimation with respect to the direct questioning method. The RRT, which was initially introduced by Warner [3], utilizes a randomization device (such as spinner, dice, playing cards, random numbers, or computers) which enables the respondents to reply to sensitive questions without revealing their true status about the stigmatized attribute. The classical monograph by Chaudhuri and Mukerjee [4] as well as Chapter 13 by Singh and Mangat [5] and Chapter 11 by Singh [6] provide a useful introduction to the literature on RRT.

Marcheselli and Barabesi [7] have extended the Huang [8] randomized response (RR) procedure by considering the second stage as the Franklin [9] method. From a practical point of choice of randomization devices, the recent suggestions by Gjestvang and Singh (GS) [10] and Marcheselli and Barabesi [7] seem to be more effective and useful. Horvitz et al. [11] and Greenberg et al. [12] developed an unrelated question RR model \((U\)-Model) by introducing the use of a non-sensitive and unrelated question. The \(U\)-Model estimator of the population proportion \(\theta_x\) based on one sample makes use of known value of the population proportion \(\theta_x\) of the unrelated character, \(x\). To overcome the difficulty of the known value of \(\theta_x\), the investigators are supposed to take two independent random samples from the population by using simple random sampling (SRS) without replacement (SRSWR) sampling. On the same line, this \(U\)-model has been studied by Mangat et al. (MSS) [13].

In the present investigation, we follow GS [10] to modify the \(U\)-model due to MSS [13] and develop a decent one-sample-based estimator of the population proportion \(\theta_y\) which is free from the use of known value of \(\theta_x\) at the estimation stage. All the above-mentioned models are based on the SRSWR scheme. But in this paper, we propose an efficient unrelated question model which is based on any general unequal probability sampling scheme.

In Section 2, we introduce the problem. In Section 3, we discuss the recent model due to GS [10]. In Section 4, we propose a new unrelated question model. In Section 5, we consider the problem of estimation of different parameters involved in the proposed \(U\)-model. In Section 6, we consider the theoretical comparison of different RR models. In Section 7, simulation experiments are performed to compare the proposed RR model with the existing competitors, so that the magnitude of the gain in efficiency can be observed. The restrictions on the parameters have also been observed.

2. Problem

Let \(U = (1, \ldots, i, \ldots, N)\) denote a population of \(N\) persons, \(y_i = 1\) if the \(i\)th person bears a sensitive character \(A\), and \(y_i = 0\), if the \(i\)th person bears \(A^C\), \(i \in U\). In the same way, let \(x_i = 1\) if the \(i\)th person bears an unrelated character \(B\), and \(x_i = 0\), if the \(i\)th person bears \(B^C\), \(i \in U\).

Let \(\theta_y = (1/N) \sum_{i=1}^{N} y_i\) and \(\theta_x = (1/N) \sum_{i=1}^{N} x_i\). Our problem is to estimate: \(\theta_y = (1/N) \sum_{i=1}^{N} y_i\).

In practice, often the values of \(y_i, i \in U\), are non-ascertainable. In that situation, here we shall estimate the unknown \(y_i, i \in U\) using the following RR device. Let \(s\) be a sample of units drawn from \(U\) with a probability \(P(s)\) according to a sampling design, \(P\). The respondents of the sample \(s\) are requested to give out independent responses following the instructions below.

3. GS [10] model

We define two RR devices \(R_1\) and \(R_2\) as given by GS [10]. In \(R_1\), the respondents will report a scrambled response as \(1 + \beta_1 S_1\) with probability \(p = \alpha_1/(\alpha_1 + \beta_1)\) \((\alpha_1, \beta_1\) are two positive real
numbers, \( 0 < p < 1 \) or, with probability \( (1 - p) \), he/she will report \( 1 - \alpha_1 S_1 \). Here \( S_1 \) is the scrambling variable which can take positive, zero and negative values.

Following the instructions of \( R_2 \), the respondents will report \( \beta_2 S_2 \) with probability \( T = \alpha_2/(\alpha_2 + \beta_2) \) \( (\alpha_2, \beta_2 \) are two positive real numbers, \( 0 < T < 1 \) or, with probability \( (1 - T) \), he/she will report \( -\alpha_2 S_2 \). Here \( S_2 \) is the scrambling variable which can take positive, zero and negative values. The distributions of the variables \( S_1 \) and \( S_2 \) may or may not be known.

MSS [13] proposed a modification on the unrelated question model. Here we apply the GS [10] RR model in MSS’s [13] situation. Both of the above models are for the SRSWR sampling scheme. But here we present procedures applicable to general varying probability sampling, even without replacement.

4. Proposed unrelated question model

In this section, we consider a general unequal probability sampling scheme to select a sample of the respondents. Then each respondent selected in the sample is requested to use the proposed RR device as follows.

The respondents are requested to follow the instruction sheet, ST. In ST, two types of cards are there. They consist of the statements:

(i) I belong to the sensitive group \( A \) and (ii) I belong to the group \( A^C \).

The statements occur with probabilities \( \lambda (0 < \lambda < 1) \) and \( (1 - \lambda) \), respectively. If the answer of the question (i) is ‘yes’ they are further requested to follow the device \( ST(A) \) (will be discussed below). Otherwise the respondents will follow \( R_2 \).

Again, for (ii) the respondents will follow \( R_1 \) if their answer is ‘yes’ otherwise they are presented the device \( ST(A^C) \) (will be discussed below).

The device \( ST(A) \) also consists of the cards (a) I belong to \( A^C \) and (b) I belong to the group \( B \), which is totally unrelated to the character \( A \). The cards (a) and (b) are in proportions \( \lambda \) and \( (1 - \lambda) \), respectively. If the answer of the question of type (b) is ‘yes’ the respondents will follow GS’s [10] device \( R_1 \), otherwise he/she will follow \( R_2 \).

The statements of the device \( ST(A^C) \) are (c) ‘I belong to the group \( A \)’ and (d) ‘I belong to the group \( B^C \)’ which are in proportions \( (1 - \lambda) \) and \( \lambda \), respectively. If the answer of the question (d) is ‘yes’ they will follow \( R_1 \) otherwise they will follow \( R_2 \). In other words, for any yes response the respondents will follow \( R_1 \), and for any no answer they will follow \( R_2 \).

Denoting \( E_R \) and \( V_R \) as the expectation and variance operators with respect to the RR device and \( R_{ki} \) as the RR response of the \( i \)th individual operating the device \( R_1 \) and \( R_{2i} \) as the operating the device \( R_2 \), we may write

\[
E_R(R_{1i}) = p(1 + \beta_1 \theta_1) + (1 - p)(1 - \alpha_1 \theta_1) = 1; \quad (1)
\]
\[
E_R(R^2_{1i}) = 1 + (\theta^2_1 + v_1^2) \alpha_1 \beta_1, \quad (2)
\]
\[
V_R(R_{1i}) = (\theta^2_1 + v_1^2) \alpha_1 \beta_1, \quad (3)
\]
\[
E_R(R_{2i}) = T(\beta_2 \theta_2) + (1 - T)(-\alpha_2 \theta_2) = 0, \quad (4)
\]
\[
E_R(R^2_{2i}) = (\theta^2_2 + v_2^2) \alpha_2 \beta_2, \quad (5)
\]

and

\[
V_R(R_{2i}) = (\theta^2_2 + v_2^2) \alpha_2 \beta_2, \quad (6)
\]

where \( E_R(S_1) = \theta_1, V_R(S_1) = v_1^2 \) and \( E_R(S_2) = \theta_2, V_R(S_2) = v_2^2 \).
Now let us denote $I_i$ as the RR response from the $i$th individual using our proposed model, then using Equations (1)–(6) we can write

$$E_R(I_i) = \lambda [y_i (1 - \lambda) x_i E_R(R_{1i}) + (1 - x_i) E_R(R_{2i})] + (1 - y_i) E_R(R_{2i})]$$

$$+ (1 - \lambda) [y_i \lambda (1 - x_i) E_R(R_{1i}) + y_i x_i \lambda E_R(R_{2i})] + (1 - \lambda) (1 - y_i) E_R(R_{1i})$$

$$= \lambda [y_i (1 - \lambda) \{x_i \} + (1 - \lambda) [y_i \lambda (1 - x_i)] + (1 - \lambda) (1 - y_i)]$$

$$= \lambda (1 - \lambda) y_i + (1 - y_i) (1 - \lambda).$$

(7)

An unbiased estimator of $y_i$ is then

$$\hat{y}_i = r_i = \frac{(1 - \lambda) - I_i}{(1 - \lambda)^2}, \quad \lambda \neq 1.$$  

(8)

Also note that

$$E_R(I_i^2) = \lambda [y_i (1 - \lambda) x_i E_R(R_{1i}^2) + (1 - x_i) E_R(R_{2i}^2)] + (1 - y_i) E_R(R_{2i}^2)]$$

$$+ (1 - \lambda) [y_i \lambda (1 - x_i) E_R(R_{1i}^2) + y_i x_i \lambda E_R(R_{2i}^2)] + (1 - \lambda) (1 - y_i) E_R(R_{1i}^2)$$

$$= y_i \lambda (1 - \lambda) (R_{1i}^2 + R_{2i}^2) + (1 - y_i) [(1 - \lambda) R_{1i}^2 + \lambda R_{2i}^2],$$

(9)

where $R_{1i}^2 = E_R(R_{1i}^2) = 1 + (\theta_1^2 + \nu_1^2) \alpha_1 \beta_1$ and $R_{2i}^2 = E_R(R_{2i}^2) = (\theta_2^2 + \nu_2^2) \alpha_2 \beta_2$. Here $I_i$ and $I_j (i \neq j)$ are independent as they are independent responses. Then,

$$V_R(I_i) = E_R(I_i^2) - (E_R(I_i))^2$$

$$= y_i \lambda (1 - \lambda) (R_{1i}^2 + R_{2i}^2 - \lambda (1 - \lambda)] + (1 - y_i) [(1 - \lambda) R_{1i}^2 + \lambda R_{2i}^2 - (1 - \lambda)^2].$$

(10)

So,

$$V_R(r_i) = \frac{V_R(I_i)}{(1 - \lambda)^4}.$$  

(11)

An unbiased estimator of $V_R(r_i)$ is then

$$v_R(r_i) = \frac{r_i \lambda (1 - \lambda) (R_{1i}^2 + R_{2i}^2 - \lambda (1 - \lambda)] + (1 - r_i) [(1 - \lambda) R_{1i}^2 + \lambda R_{2i}^2 - (1 - \lambda)^2]}{(1 - \lambda)^4}.$$  

(12)

5. Unbiased estimation of parameters in the proposed model

We here first discuss a general method and then we will discuss different sampling schemes as a special case of it in the subsequent sections.

5.1. Sample is drawn by any general unequal probability sampling scheme

Using the Horvitz and Thompson [14] method of estimation, an unbiased estimator of the parameter $\theta_y = (1/N) \sum_{i=1}^{N} y_i$ is

$$\hat{\theta}_y = \frac{1}{N} \sum_{i \in s} \frac{r_i}{\pi_i}.$$  

(13)

(following Chaudhuri [15])
Here (i) $E_R(r_i) = y_i$, (ii) $V_R(r_i) = V_i(> 0)$, (iii) $r_i$s are independent over $i$ in $U$ and (iv) there exists $v_R(r_i)$ ascertainable from RRs such that $E_R(v_R(r_i)) = V_i(\forall i \in U)$.

Note that the estimator $\hat{\theta}_y$ in Equation (13) is free from the value of $\theta_x$ and makes use of only one sample.

Denoting the design-based expectation and variance as $E_P$ and $V_P$, the overall expectation and variance can be written as $E = E_P E_R$ and $V = E_P V_R + V_P E_R$.

We may write

$$E(\hat{\theta}_y) = E_P E_R(\hat{\theta}_y) = E_P \left( \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i} \right)$$  \hspace{1cm} (14)

and

$$V(\hat{\theta}_y) = E_P V_R(\hat{\theta}_y) + V_P E_R(\hat{\theta}_y)$$

$$= E_P \left[ \frac{1}{N^2} \sum_{i \in s} \frac{V_R(r_i)}{\pi_i^2} \right] + V_P \left( \frac{1}{N} \sum_{i \in s} \frac{y_i}{\pi_i} \right)$$

$$= E_P \left[ \frac{1}{N^2} \sum_{i \in s} \frac{V_R(r_i)}{\pi_i^2} \right] + \frac{1}{N^2} \sum_{i < j \in U} \sum (\pi_i \pi_j - \pi_{ij}) \left( \frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2, \hspace{1cm} (15)$$

where $\pi_{ij}$ is the joint inclusion probability of the units $i, j$.

Using Yates and Grundy’s [16] variance estimation formula, an unbiased estimator of the variance $V(\hat{\theta}_y)$ is

$$v(\hat{\theta}_y) = \frac{1}{N^2} \sum_{i \in s} \frac{v_R(r_i)}{\pi_i^2} + \frac{1}{N^2} \sum_{i < j \in s} \sum (\pi_i \pi_j - \pi_{ij}) \left( \frac{r_i}{\pi_i} - \frac{r_j}{\pi_j} \right)^2$$

$$- \frac{1}{N^2} \sum_{i < j \in s} \sum (\pi_i \pi_j - \pi_{ij}) \left( \frac{v_R(r_i)}{\pi_i^2} + \frac{v_R(r_j)}{\pi_j^2} \right). \hspace{1cm} (16)$$

5.2. Samples are drawn by SRSWR scheme

For the SRSWR scheme, an unbiased estimator of $\theta_y$ becomes

$$\hat{\theta}_y(SRS) = \frac{1}{n} \sum_{i \in s^*} r_i$$  \hspace{1cm} (17)

and with variance (following Chaudhuri and Pal [17]),

$$V(\hat{\theta}_y(SRS)) = E_P V_R(\hat{\theta}_y(SRS)) + V_P E_R(\hat{\theta}_y(SRS))$$

$$= E_P \left[ \frac{1}{n^2} \sum_{i \in s} V_R(r_i) \right] + V_P \left( \frac{1}{n} \sum_{i \in s^*} y_i \right)$$

$$= \frac{1}{nN} \sum_{i=1}^N V_i + \frac{\theta_y(1 - \theta_y)}{n} \text{ where } V_i = V_R(r_i). \hspace{1cm} (18)$$

Here $s^*$ is the simple random sample with $n$ draws.
It should be noted that the form of $V(\hat{\theta}, (SRS))$ (following Chaudhuri [15]) may not be reduced to the exact form of the existing variances for different RR models.

By following Chaudhuri [15], the result in Equation (18) is quite general.

For example, for Warner’s [3] scheme:

$$V_i = V_i(\text{Warner}) = \frac{k(1-k)}{(2k-1)^2}$$

(19)

defining the parameter of Warner’s [3] scheme as $k(0 < k < 1), k \neq .5$.

For Mangat and Singh’s [18] scheme (with parameters $k$ and $T$):

$$V_i = V_i(\text{Mangat}) = \frac{(1-T)(1-k)[1-(1-T)(1-k)]}{[(2k-1)+2T(1-k)]^2}. \quad (20)$$


For the Horvitz et al. [11] scheme:

$$V_i = V_i(\text{HSS}) = \frac{1-k}{k} [y_i(1-x_i) + x_i]. \quad (21)$$

For the MSS (Mangat-Singh-Singh) [13] improved unrelated question model:

$$V_i = V_i(\text{MSS}) = \frac{(1-y_i)x_i(1-k)}{1-x_i(1-k)}. \quad (22)$$

$V_i(\text{Warner}), V_i(\text{Mangat})$ and $V_i(\text{MSS})$ may be followed from Chaudhuri’s [15] result.

For the proposed model:

$$V_i = V_i(\text{Proposed})$$

$$V_R(ri) = \frac{y_i\lambda(1-\lambda)(R_1^2 + R_2^2 - 2\lambda(1-\lambda)) + (1-y_i)[(1-\lambda)R_1^2 + \lambda R_2^2 - (1-\lambda)^2]}{(1-\lambda)^4}. \quad (23)$$

In this section, we develop theoretical inequalities where the proposed model remains more efficient than its competitors.

6. Theoretical comparison among different models

To compare the efficiencies among different models, it is enough to compare the term $\sum_{i=1}^{N} (V_i/\pi_i)$ (from Equation (15)) for different RR models. But to simplify the comparison, we shall compare the simple random sampling (SRS) models under different RR schemes and use Equation (18). For the SRSWR scheme, it is enough to compare the term $\sum_{i=1}^{N} V_i(\cdot)$, where $V_i(\cdot)$ is the RR variance for any particular model.
6.1. Comparison between the proposed and the Horvitz–Shah–Simmons models

Using the results from Section 5.2, we have

\[
\sum_{i=1}^{N} V_i(\text{HSS}) = \frac{1 - k}{k} \left[ \sum_{i=1}^{N} y_i(1 - x_i) + \sum_{i=1}^{N} x_i \right]
\]

\[
= \frac{1 - k}{k} \left[ \sum_{i=1}^{N} y_i + \sum_{i=1}^{N} x_i(1 - y_i) \right].
\]

\[
\sum_{i=1}^{N} V_i(\text{Proposed}) = \sum_{i=1}^{N} \frac{\left[ y_i (1 - \lambda)(R_1^{s2} + R_2^{s2} - \lambda(1 - \lambda)) \right] + (1 - y_i)}{(1 - \lambda)^4}
\]

\[
\times \left[ \left( 1 - \lambda \right) R_1^{s2} + \lambda R_2^{s2} - (1 - \lambda)^2 \right]
\]

\[
= \frac{\left[ \lambda(1 - \lambda)(R_1^{s2} + R_2^{s2} - \lambda(1 - \lambda)) \right] \sum_{i=1}^{k} y_i}{(1 - \lambda)^4}
\]

\[
+ \frac{\left[ (1 - \lambda) R_1^{s2} + \lambda R_2^{s2} - (1 - \lambda)^2 \right] \sum_{i=1}^{k} (1 - y_i)}{(1 - \lambda)^4}
\]

\[
= m_1 \left( \sum_{i=1}^{N} y_i \right) + m_2 \left( \sum_{i=1}^{N} (1 - y_i) \right),
\]

where

\[ m_1 = \frac{\left( 1 - \lambda \right) R_1^{s2} + R_2^{s2} - \lambda(1 - \lambda))}{(1 - \lambda)^4} \quad \text{and} \quad m_2 = \frac{(1 - \lambda) R_1^{s2} + \lambda R_2^{s2} - (1 - \lambda)^2}{(1 - \lambda)^4}. \]

Note that \( \sum_{i=1}^{N} x_i(1 - y_i) \leq \sum_{i=1}^{N} (1 - y_i) \), which implies

\[
\sum_{i=1}^{N} x_i(1 - y_i) \geq \frac{1}{\sum_{i=1}^{N} (1 - y_i)}.
\]

Using Equations (24)–(26), we have

\[
\sum_{i=1}^{N} V_i(\text{Proposed}) < \sum_{i=1}^{N} V_i(\text{HSS})
\]

if

\[
(a)' \frac{1 - k}{k} \geq m_1
\]

\[
(b)' \left( \frac{1 - k}{k} \right) \geq m_2 \frac{\sum_{i=1}^{N} (1 - y_i)}{\sum_{i=1}^{N} x_i(1 - y_i)} \geq m_2
\]

i.e. if

\[
\left( \frac{1 - k}{k} \right) \geq m_2
\]

i.e. if

\[
\left( \frac{1 - k}{k} \right) \geq \max(m_1, m_2).
\]

Thus the inequality (28) could be used to decide the values of the parameters of the proposed randomization device so that it performs better than the Horvitz et al. [11] model.
6.2. Comparison between the proposed and MSS [13] model

The proposed model will be more efficient than MSS’s model [13] if

\[ \sum_{i=1}^{N} V_i(\text{Proposed}) < \sum_{i=1}^{N} V_i(\text{MSS}) \]

if \( \theta_y > \frac{1}{m_2 - m_1} \left[ m_2 - \frac{1}{N} \sum_{i=1}^{N} (1 - y_i)x_i(1 - k) \right] \) \( \left( 1 - x_i(1 - k) \right) \). \hspace{1cm} (29)

This will happen if

\[ m_2 < \frac{1}{N} \sum_{i=1}^{N} \frac{(1 - y_i)x_i(1 - k)}{1 - x_i(1 - k)} \] \hspace{1cm} (30)

or

\[ m_2 < m_1. \]

Choosing the parameters satisfying Equations (29) and (30), we always conclude that our proposed model is better than the MSS [13] model.

6.3. Comparison between the proposed and GS model [10]

For the GS [10] model \( V_i(\text{GS}) = y_i \alpha_1 \beta_1 (\theta_1^2 + v_1^2) + (1 - y_i) \alpha_2 \beta_2 (\theta_2^2 + v_2^2) \).

Then, we may write

\[ \frac{1}{N} \sum_{i=1}^{N} V_i(\text{GS}) = \theta_y \alpha_1 \beta_1 (\theta_1^2 + v_1^2) + (1 - \theta_y) \alpha_2 \beta_2 (\theta_2^2 + v_2^2). \]

\[ \sum_{i=1}^{N} V_i(\text{Proposed}) < \sum_{i=1}^{N} V_i(\text{GS}) \]

if \( m_1 \theta_y + m_2 (1 - \theta_y) < \theta_y \alpha_1 \beta_1 (\theta_1^2 + v_1^2) + (1 - \theta_y) \alpha_2 \beta_2 (\theta_2^2 + v_2^2) \)

or

\[ \theta_y > \frac{m_2 - \alpha_2 \beta_2 (\theta_2^2 + v_2^2)}{(m_2 - m_1) - \alpha_2 \beta_2 (\theta_2^2 + v_2^2) + \alpha_1 \beta_1 (\theta_1^2 + v_1^2)}. \]

This will happen if \( m_2 < \alpha_2 \beta_2 (\theta_2^2 + v_2^2) \) or

\[ (m_2 - m_1) < \alpha_2 \beta_2 (\theta_2^2 + v_2^2) - \alpha_1 \beta_1 (\theta_1^2 + v_1^2). \] \hspace{1cm} (31)

7. Simulation study

In order to see the magnitude of the gain in efficiency of the proposed RR model with respect to Warner [3], Mangat and Singh [18], Horvitz et al. [11], MSS [13] and GS [10] models, we considered a population of \( N = 24 \) units from a medical college of the city Calcutta with the known values \( l_i \) and the unknown values \( y_i \) for \( i = 1, 2, \ldots, N \).

Here \( y_i = 1 \) if the \( i \)th person (student) had the habit of using cocaine in any form, one or more times and \( y_i = 0 \) else.
Table 1. Data description of the numerical example.

<table>
<thead>
<tr>
<th>Serial no.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_i$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$l_i$</td>
<td>20.1</td>
<td>12.1</td>
<td>78.2</td>
<td>21.3</td>
<td>121.5</td>
<td>2.1</td>
<td>123.1</td>
<td>1.1</td>
<td>37.1</td>
<td>3.1</td>
<td>2.1</td>
<td>21.1</td>
<td>56.2</td>
<td>76.1</td>
<td>67.1</td>
<td>10.1</td>
<td>18.2</td>
<td>32.1</td>
<td>91.4</td>
<td>2.2</td>
<td>71.2</td>
<td>32.1</td>
<td>23.9</td>
<td>11.6</td>
</tr>
<tr>
<td>$x_i$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. PRE of the proposed model with respect to its competitors.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sampling scheme</th>
<th>SRSWR</th>
<th>HR [20]</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warner [3]</td>
<td>SRSWR</td>
<td>627.83</td>
<td>511.29</td>
<td>116.54</td>
</tr>
<tr>
<td>Mangat and Singh [18]</td>
<td>SRSWR</td>
<td>2447.54</td>
<td>1923.28</td>
<td>524.26</td>
</tr>
<tr>
<td>Horvitz et al. [11]</td>
<td>SRSWR</td>
<td>333.83</td>
<td>329.86</td>
<td>3.97</td>
</tr>
<tr>
<td>Mangat et al. [13] model</td>
<td>SRSWR</td>
<td>275.61</td>
<td>200.95</td>
<td>74.66</td>
</tr>
<tr>
<td>Gjestvang and Singh’s [10] model</td>
<td>SRSWR</td>
<td>136.94</td>
<td>105.99</td>
<td>30.95</td>
</tr>
</tbody>
</table>

\( l_i \): the per capita expenditure in Indian Rupees incurred in the household to which \( i \) belongs – this is the size measure used in choosing a sample \( s \) of the medical students.

Also, the unrelated character \( B \) is defined to implement Horvitz et al. [11] and MSS’s [13] unrelated question model. In this case, \( x_i = 1 \) if the date of birth of the \( i \)th person (student) falls between January and June and \( x_i = 0 \) if the date of birth of the \( i \)th person (student) falls between July and December.

The data are given in Table 1.

In the first case, we selected an SRSWR sample of \( n = 5 \) units and later we used the Hartley–Rao (HR [20]) scheme to select a sample of \( n = 5 \) units. In the HR [20] scheme, we use the size measure \( l_i \). Here the units in a population are first randomly permuted. Then, from the permuted vector of labelled units, one probability proportionate to size circular systematic sample is chosen in \( n = 5 \) draws provided \( nq_i < 1 \ \forall \ i \in U \), \( q_i = l_i / \sum_{i=1}^{N} l_i \).

We decided to make practicable choices of different parameters of the randomization devices as: \( k = 0.13 \), \( T = 0.35 \) and \( \lambda = 0.35 \).

Here \( \theta_y = 0.375 \), \( \theta_x = 0.542 \).

Also, \( \theta_1^2 = 0.006 \), \( v_1^2 = 0.05 \), \( \alpha_1 = 0.016 \), \( \beta_1 = 0.05 \) and \( \theta_2^2 = 0.75 \), \( v_2^2 = 0.65 \), \( \alpha_2 = 0.66 \), \( \beta_2 = 0.45 \). According to the GS [10] model \( \alpha_1 \), \( \alpha_2 \), \( \beta_1 \), \( \beta_2 \) are any positive real numbers. So \( p = \alpha_1 / (\alpha_1 + \beta_1) = 0.24 \) and \( T = \alpha_2 / (\alpha_2 + \beta_2) = 0.59 \) which are quite moderate. The percent relative efficiency (PRE) of the proposed model with respect to its competitor is computed by taking the ratio of the variances multiplied by 100. PRE of the proposed model with respect to Warner [3] is

\[
\text{PRE(}\text{Warner}) = 100 \frac{V(\theta_y|\text{Warner})}{V(\theta_y|\text{proposed})}.
\]

The results so obtained are presented in Table 2.

8. Discussion of results

The most practicable comparison of the proposed RR model is with respect to the MSS [13] unrelated question model. It is already shown that the MSS [13] model remains more efficient than the Horvitz et al. [11] model. It is interesting to note that the relative efficiency of the proposed model with respect to the MSS [13] model is 74.66% while using the SRSWR design for both models. The use of the HR scheme provides 3.97% extra gain in the relative efficiency. Note that this gain in the relative efficiency comes from the use of right choice of design, but has nothing to do with the proposed randomization device. We may conclude that the proposed RR model is better than others with some restrictions on the parameters. We can also see that the proposed estimator can be used for any design, thus it has more practical utility than the other estimators discussed in the present investigation.
Acknowledgement

The authors are grateful to the editor and the referees for several valuable comments and suggestions.

References
