STATISTICS IN TRANSITION new series, June 2020
Vol. 21, No. 2, pp. 61–87, DOI 10.21307/stattrans-2020-014
Submitted – 16.04.2019; accepted for publishing – 14.04.2020

How privacy may be protected in optional randomized response surveys

Sanghamitra Pal[1]**, Arijit Chaudhuri**[2]**, Dipika Patra**[3]

ABSTRACT

There are materials in literature about how privacy on stigmatizing features like alcoholism, history of tax-evasion, or testing positive in AIDS-related testing may be partially protected by a proper application of randomized response techniques (RRT). The paper demonstrates what amendments are necessary for this approach while applying optional RRTs covering qualitative characteristics, permitting a sampled respondent either to directly reveal sensitive data or choose a randomized response respectively with complementary probabilities. Only a few standard RRTs are illustrated in the text.

AMS subject classification: 62D05

Key words: protection of privacy, randomized response, sensitive issues, Warner and other techniques.

1. Introduction

Chaudhuri (2011) and Chaudhuri and Christofides (2013) in their books and Chaudhuri and Dihidar (2009), Chaudhuri and Saha (2005) and Chaudhuri, Christofides and Saha (2009) in their published papers have recounted details about how to protect privacy in randomized responses (RR) given out by the respondents following various RR devices.

We have reservations about only a few RR techniques because in a couple of text books and several authentic published review papers, only a few RR techniques are illustrated as we have done with no prejudice against the ones we omit to save space.

Here, we intend to investigate possibilities of protecting privacy in generating optional RR's covering qualitative stigmatizing issues. The optional RR (ORR)

[1] Department of Statistics, West Bengal State University, India. Corresponding author. E-mail : mitrapal2013@gmail.com. ORCID: https://orcid.org/0000-0002-5752-8282.

[2] Applied Statistics Unit, Indian Statistical Institute, Kolkata, India. E-mail : arijitchaudhuri1@rediffmail.com. ORCID: https://orcid.org/0000-0002-4305-7686.

[3] Department of Statistics, West Bengal State University, India. E-mail : dipika.patra1988@gmail.com. ORCID: https://orcid.org/0000-0003-4318-1123.

technique was introduced by Chaudhuri and Mukerjee (1985). A large number of developments following Chaudhuri and Mukerjee (1985) approach were proposed by Gupta (2001), Gupta et al. (2002), Pal (2008) and many others. Subsequent developments are due to Arnab (2004), Chaudhuri and Saha (2005), Saha (2007), Huang (2008), Arnab and Rueda (2016) among others with slight differences in approaches. As we see, in ORR a sampled person is offered an option either to (i) report directly whether he/she bears a stigmatizing feature, say A (which may mean alcoholism or testing HIV positive, etc.) or (ii) give out an RR adopting a device offered and explained to him/her. The option (i) may be implemented with an unknowable probability and (ii) with the complementary probability. How to implement (i) or (ii) may be clearly explained to the respondent who may or may not divulge which of these options is actually applied. Different ORR techniques are described in this article.

In the cases of RR's, it is observed that privacy is protected only for specific parametric combinations in the RR devices and protection leads to loss of control in achieving accuracy in estimation of the population proportion of people bearing sensitive features. Such features will be seen in what follows with optional RR situations as well. But certain other striking possibilities are revealed below with optional RR's (ORR) rather than with compulsory RR's (CRR). Details are shown in Sections 2 and 3 below. Section 4 presents some numerical findings, through simulation.

2. Certain basics for protection of privacy in general sampling

Let $U = (1, 2, ..., N)$ denote a finite population of units. On drawing a sample according to a general sampling design P, the selected units are approached with a request to provide ORR's in order to estimate the proportion of the population units bearing a sensitive characteristic A, say.

Let, for a person labelled i, L_i be the unknowable prior probability that i bears *A* and $L_i(R)$ denote the posterior probability that given the RR or DR denoted R, the respondent bears *A*. Following Chaudhuri, Christofides and Saha (2009), the literature considers for a measure of jeopardy inherent in the response R the quantity

$$J_i(R) = \frac{L_i(R)/L_i}{(1-L_i(R))/(1-L_i)}$$

assuming the denominator is non-zero, with the RR device parameters rightly chosen.

Let, for a general ORR device, $c_i (0 < c_i < 1 \forall i \in U)$ be an unknowable probability that the i^{th} person chooses to answer directly without divulging this secret to the enquirer. Further, let for i,

I_i = the DR, with probability c_i

= the RR for a specified device with probability (1- c_i) of course.

The investigator is to explain to a respondent a formal way to implement choosing such an undisclosed c_i and 1- c_i to be the probability of giving a DR and respectively an RR with no option to change it for an alternative RR device. For example, c_i may be fixed (without disclosing to the investigator) as $\frac{13}{100}$ on choosing a 2-digited random number from 01, ..., 13 leaving the rest namely 14, ...,99,00 for giving out an RR.

Warner's RRT demands from a chosen person i a response

$$R_i = y_i \text{ with probability } p\,(0 < p < 1,\ p \neq \frac{1}{2}),$$

$$= 1 - y_i \text{ with probability } 1\text{-}p$$

if i chooses a card marked A and bears the stigmatizing feature A or chooses a card marked the complement of A (A^c) from a pack of cards with a proportion marked A and the rest marked the complement of $A(A^c)$ and

$$y_i = 1 \text{ if } i \text{ bears } A$$

$$= 0, \text{if } i \text{ bears } A^c.$$

Then, for this RRT due to Warner (1965) the expected value of R_i is

$$E(R_i) = py_i + (1-p)(1-y_i) = (1-p) + (2p-1)y_i \tag{I}$$

for every i in U .

For the ORR technique instead for Warner's RRT the ORR is

$$OR_i = y_\mathrm{i} \text{ with probability } c_i \text{ in the closed interval from 0 to 1}$$

$$= R_i \text{ with probability } (1-c_i)$$

$$\text{Then, } E(OR_i) = c_i y_i + (1-c_i)[(1-p) + (2p-1)y_i]$$

$$= (1-c_i)(1-p) + [c_i + (1-c_i)(2p-1)]\,\mathrm{y_i} \tag{II}$$

Clearly, if c_i equals zero, (II) matches (I) and if c_i differs from 0, (II) differs from (I) as well.

For simplicity let the response be either 'Yes' or 'No' only. We may write, applying Bayes' theorem, writing A^c as the complement of A ,

$$\text{Prob}(A \mid Yes) = \frac{L_i \Pr\text{ob}(Yes \mid A)}{L_i \Pr\text{ob}(Yes \mid A) + (1-L_i)\Pr\text{ob}(Yes \mid A^c)}$$

on supposing that Warner's RR device in Chaudhuri's (2001) form is employed.

Defining $y_i = 1$ if i bears A and 0 if i does not bear A, we may work out

$$\Pr ob(Yes \mid A) = c_i y_i + (1-c_i) p y_i$$

$$= p + c_i(1-p) \text{ since } y_i = 1$$

and $\Pr ob(Yes \mid A^c) = (1-c_i)(1-p)(1-y_i)$

$$= (1-p)(1-c_i) \text{ since } y_i = 0$$

Hence, it follows that

$$\Pr ob(A \mid Yes) = \frac{L_i[p + c_i(1-p)]}{L_i[p + c_i(1-p)] + (1-L_i)(1-p)(1-c_i)}$$

$$= \frac{L_i[p + (1-p)c_i]}{pL_i + (1-p)[1 - L_i(1-c_i)]}$$

and $J_i(1) = \dfrac{L_i(1)/L_i}{(1-L_i(1))/(1-L_i)}$.

With a little algebra,

$$1 - L_i(1) = \frac{(1-p)(1-L_i)}{pL_i + (1-p)[1 - L_i(1-c_i)]};$$

so,

$$J_i(1) = \frac{L_i(1)}{1-L_i(1)} \frac{1-L_i}{L_i}$$

$$= \frac{p + c_i(1-p)}{(1-p)(1-c_i)}. \qquad (1)$$

Again,

$$J_i(0) = \frac{L_i(0)/L_i}{(1-L_i(0))/(1-L_i)}.$$

Now,

$$\Pr ob(A \mid No) = \frac{L_i \Pr ob(No \mid A)}{L_i \Pr ob(No \mid A) + (1-L_i)\Pr ob(No \mid A^c)}$$

$$\Pr ob(No \mid A) = c_i(1-y_i) + (1-c_i)(1-p)$$

$$= (1-c_i)(1-p) \qquad \text{since } y_i = 1;$$

$$\Pr ob(No \mid A^c) = c_i(1-y_i) + (1-c_i)p$$

$$= c_i + p(1-c_i) \qquad \text{since } y_i = 0.$$

So,

$$L_i(0) = \frac{L_i(1-c_i)(1-p)}{L_i(1-c_i)(1-p) + (1-L_i)[c_i + (1-c_i)p]}$$

$$1 - L_i(0) = \frac{(1-L_i)[c_i + (1-c_i)p]}{(1-L_i)[c_i + (1-c_i)p] + L_i(1-c_i)(1-p)};$$

since $L_i(0) = \dfrac{L_i \operatorname{Pr ob}(No/\mathrm{A})}{L_i \operatorname{Pr ob}(No/\mathrm{A}) + (1-L_i)\operatorname{Pr ob}(No/\mathrm{A}^c)}$,

so,

$$J_i(0) = \frac{L_i(0)/L_i}{(1-L_i(0))/(1-L_i)}$$

$$= \frac{(1-c_i)(1-p)}{c_i + (1-c_i)p} = \frac{(1-c_i)(1-p)}{p + c_i(1-p)} . \tag{2}$$

Hence,

$$J_i(1) \times J_i(0) = (1) \times (2) = 1$$

Thus, our proposed measure of jeopardy is $\overline{J}_i \equiv$ the G.M. of $J_i(1)$ and $J_i(0)$ and this carries over for every *i* as $\overline{J}_i = 1$.

Ensuring privacy protection is not enough. The estimation of the variance of the estimator employed is also a crucial requirement. So, adjustments in the RRT's are needed. Thus, in employing Warner's RRT not just one RR is adequate; two independent RR's are needed when the ORR technique is to be employed by Warner's RRT allowing options for DR's. This is elaborated in Section 3.

3. Optional Randomized Response Technique with two independent randomized responses

The person labelled i is requested to give out two ORR's independently with different known RR device probabilities. Denoting the responses as R and R', the posterior probability and the measure of jeopardy may be written as $L_i(R,R')$ and $J_i(R,R')$ respectively, corresponding to the i^{th} person's response (R,R').

Now, applying Bayes' theorem,

$$\Pr ob(A|(R,R')) = \frac{L_i \Pr ob(R|A)\Pr ob(R'|A)}{L_i \Pr ob(R|A)\Pr ob(R'|A) + (1-L_i)\Pr ob(R|A^c)\Pr ob(R'|A^c)} \tag{3}$$

as the responses are independent for every person, and the response specific jeopardy measure for the ith person

$$J_i(R,R') = \frac{L_i(R,R')/L_i}{(1-L_i(R,R'))/(1-L_i)} \tag{4}$$

indicates the risk of divulging the respondent's status due to his/her specific response (R,R'). Chaudhuri et al. (2009) preferred an average measure. Here, we propose geometric mean as an average measure instead of arithmetic mean, earlier suggested by Chaudhuri Christofides and Saha (2009). A Geometric Mean (GM) in lieu of Arithmetic Mean (AM) is proposed to achieve an algebraic simplicity. Thus, the measure of jeopardy for the ith person is

$$\overline{J}_i = \text{G.M of } J_i(R,R') \ \forall R,R' \ . \tag{5}$$

In this section, we discuss the response specific measure of jeopardy in ORR technique and our proposed measure combining all the response specific jeopardy measures for qualitative characteristics.

Although $J_i(R,R')$ depends on unknown probability c_i, it can be shown in the later sections that the measure of jeopardy $\overline{J}_i$ is free from c_i.

3.1. ORR using Warner's (1965) RR model

Suppose the sampled person labelled i is directed to respond his/her true value of the specific stigmatizing attribute or by Warner's RR device. A box with identical cards with A or A^c in proportions $p_1 : 1-p_1 \ (0<p_1<1,\ p_1 \neq \frac{1}{2})$ is given to the respondent. He/she is requested to draw a card and without divulging the card-type drawn he/she is to truthfully say his/her outcome if the card type drawn matches or not his/her characteristic. The whole process is repeated one more time independently but with different Warner's RR device with another similar box - cards marked by A or A^c, which are in proportions $p_2 : 1-p_2 \ (0<p_2<1,\ p_2 \neq \frac{1}{2})$.

Thus, the independent Optional randomized responses for i th $(i = 1,2,...,N)$ person are Z_i and Z_i'.

Here, $Z_i = y_i$, with unknown probability c_i

= the Warner's RR, with unknown probability $1-c_i$, using first box

and

$Z_i' = y_i$, with unknown probability c_i

= the Warner's RR, with unknown probability $1-c_i$, using another box.

Here,

$y_i = 1$ if the person bears the sensitive characteristic

= 0, else.

Note that the investigator's instruction is to keep the same c_i for both z and z'.

Then, denoting RR based expectations and variances as E_R and V_R we may write

$$E_R(Z_i) = c_i y_i + (1-c_i)[p_1 y_i + (1-p_1)(1-y_i)]$$

$$E_R(Z_i') = c_i y_i + (1-c_i)[p_2 y_i + (1-p_2)(1-y_i)]$$

Thus, an unbiased estimator of y_i is $r_i = \dfrac{(1-p_2)Z_i - (1-p_1)Z_i'}{p_1 - p_2}$, $p_1 \neq p_2$ and an unbiased estimator of the variance $V_R(r_i)$ is $v_i = \dfrac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(Z_i - Z_i')^2$. The details of the proof is given in Appendix 1.This variance estimator form is slightly different from Chaudhuri and Dihidar's (2009).We prefer this form as it is a function of two independent responses.

The possible responses for each individual in the above method are (1, 1) (0, 0) (1, 0) and (0, 1).

Note that a different response 1 or 0 may come from the same person for the first and second trials; of course it does not matter because it may reveal that a person may have opted for an RR rather than a DR; this does not reveal the person's sensitive feature.

Suppose the response of i^{th} labelled person is (1, 1). Then, using the equation (3) we get

$$\frac{L_i(1,1)}{L_i} = \frac{P(Z_i = 1 \mid y_i = 1)P(Z_i' = 1 \mid y_i = 1)}{L_i P(Z_i = 1 \mid y_i = 1)P(Z_i' = 1 \mid y_i = 1) + (1-L_i)P(Z_i = 1 \mid y_i = 0)P(Z_i' = 1 \mid y_i = 0)}$$

$$= \frac{\{c_i + (1-c_i)p_1\}\{c_i + (1-c_i)p_2\}}{L_i\{c_i + (1-c_i)p_1\}\{c_i + (1-c_i)p_2\} + (1-L_i)\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}}$$

$$\frac{1-L_i(1,1)}{1-L_i} = \frac{\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}}{L_i\{c_i + (1-c_i)p_1\}\{c_i + (1-c_i)p_2\} + (1-L_i)\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}}$$

So, from equation (4) the response (1,1) - specific jeopardy measure is

$$J_i(1,1) = \frac{L_i(1,1)/L_i}{(1-L_i(1,1))/(1-L_i)} = \frac{\{c_i + (1-c_i)p_1\}\{c_i + (1-c_i)p_2\}}{\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}} \tag{6}$$

If the i^{th} person's response is (0,0) then we may write

$$\frac{L_i(0,0)}{L_i}=\frac{P(Z_i=0\,|\,y_i=1)P(Z_i=0\,|\,y_i=1)}{L_iP(Z_i=0\,|\,y_i=1)P(Z_i=0\,|\,y_i=1)+(1-L_i)P(Z_i=0\,|\,y_i=0)P(Z_i=0\,|\,y_i=0)}$$

$$=\frac{\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}}{L_i\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}+(1-L_i)\{c_i+(1-c_i)p_1\}\{c_i+(1-c_i)p_2\}}$$

$$\frac{1-L_i(0,0)}{1-L_i}=\frac{\{c_i+(1-c_i)p_1\}\{c_i+(1-c_i)p_2\}}{L_i\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}+(1-L_i)\{c_i+(1-c_i)p_1\}\{c_i+(1-c_i)p_2\}}.$$

So, the response (0,0) - specific jeopardy measure is

$$J_i(0,0)=\frac{L_i(0,0)/L_i}{(1-L_i(0,0))/(1-L_i)}=\frac{\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}}{\{c_i+(1-c_i)p_1\}\{c_i+(1-c_i)p_2\}}. \tag{7}$$

For the response (1,0), the corresponding posterior probability $L_i(1,0)$ and Jeopardy measure $J_i(1,0)$ may be expressed as

$$L_i(1,0)=\frac{L_i\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}}{L_i\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}+(1-L_i)\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}}$$

$$J_i(1,0)=\frac{\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}}{\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}}. \tag{8}$$

Similarly, for the response (0,1), the posterior probability is

$$L_i(0,1)=\frac{L_i\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}}{L_i\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}+(1-L_i)\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}}$$

and the response specific measure of jeopardy is

$$J_i(0,1)=\frac{\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}}{\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}}. \tag{9}$$

Thus, our proposed measure of jeopardy is the geometric mean of all response specific jeopardy measures (6), (7), (8) and (9), which is exactly 1 for each and every individual. If $p_1 \rightarrow p_2$, responses of every individual are well protected but estimate of the variance tends to be infinite. Yet the overall measure does not reveal the status of the respondent.

3.2. ORR using Greenberg et al.'s (1969) unrelated question RR model

The ORR technique with an unrelated question model is same as the above discussed technique except the RR device. Here, the RR device is Greenberg et al.'s unrelated question model (1969) instead of Warner's model. In this RR, two boxes contain cards marked as A, the stigmatizing attribute or B, the innocuous attribute. The attribute A, is unrelated to the attribute B . The two types of cards are mixed with different known proportions say p_1 and $(1-p_1)$ and p_2 and $(1-p_2)$ in box 1 and box 2 respectively. Each respondent is requested to draw two cards independently from box 1 and box 2 respectively and report according to the above device.

So, the Optional randomized response for i^{th} person is

$Z_i = y_i$, with unknown probability c_i

= the Greenberg et al.'s RR, with unknown probability $1-c_i$, using box 1

and

$Z_i' = y_i$, with unknown probability c_i

= the Greenberg et al.'s RR, with unknown probability $1-c_i$, using box 2.

Defining

$x_i = 1$ if the person bears the innocuous character B

$= 0$ if the person bears the innocuous character B^{C}, the complement of B,

we may write,

$$P(Z_i = 1) = c_i y_i + (1-c_i)[p_1 y_i + (1-p_1)x_i] \text{ and}$$

$$P(Z_i = 0) = c_i(1-y_i) + (1-c_i)[p_1(1-y_i) + (1-p_1)(1-x_i)].$$

Hence, it follows that

$$E_R(Z_i) = c_i y_i + (1-c_i)[p_1 y_i + (1-p_1)x_i].$$

Similarly,

$$E_R(Z_i') = c_i y_i + (1-c_i)[p_2 y_i + (1-p_2)x_i].$$

Thus, an unbiased estimator of y_i under the above model is $r_i = \frac{(1-p_2)Z_i - (1-p_1)Z_i'}{p_1 - p_2}$ taking $p_1 \neq p_2$ and an unbiased estimator of the variance $V_R(r_i)$ is $v_i = \frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(Z_i - Z_i')^2$ since $p_1 \neq p_2$. The proof is given in Appendix 2.

The necessary conditional probabilities are shown below to calculate posterior probabilities and response specific jeopardy measures defined as in the equation (3).

Now, $P(Z_i=1|y_i=0)=(1-c_i)(1-p_1)x_i=(1-c_i)(1-p_1)$, as the situation arises if the response of the i^{th} individual is 1 but the true value of the sensitive characteristic is zero. This is possible only if the respondent chooses RR device and responds to the question regarding the innocuous attribute *B* due to the assumption that the respondents provide true response. So, $x_i=1$ is obvious.

With the same line of reasoning, we get $P(Z_i=0|y_i=1)=(1-c_i)(1-p_1)$.

As we know, $P(A|B)+P(A^C|B)=1$.

Clearly, $P(Z_i=1|y_i=1)=1-P(Z_i=0|y_i=1)=c_i+(1-c_i)p_1$.

Similarly, $P(Z_i=0|y_i=0)=1-P(Z_i=1|y_i=0)=c_i+(1-c_i)p_1$.

Proceeding as described in 3.1, their response specific jeopardy measures are

$$J_i(1,1)=\frac{L_i(1,1)/L_i}{(1-L_i(1,1))/(1-L_i)}=\frac{\{c_i+(1-c_i)p_1\}\{c_i+(1-c_i)p_2\}}{\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}} \tag{10}$$

$$J_i(0,0)=\frac{\{(1-c_i)(1-p_1)\}\{(1-c_i)(1-p_2)\}}{\{c_i+(1-c_i)p_1\}\{c_i+(1-c_i)p_2\}} \tag{11}$$

$$J_i(1,0)=\frac{\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}}{\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}} \tag{12}$$

$$J_i(0,1)=\frac{\{(1-c_i)(1-p_1)\}\{c_i+(1-c_i)p_2\}}{\{c_i+(1-c_i)p_1\}\{(1-c_i)(1-p_2)\}}. \tag{13}$$

Now, our proposed measure of jeopardy by the equation (5) is the geometric mean (G.M) of the above response specific jeopardy measures. Here, the GM is $\bar{J}_i=\{J_i(1,1)\times J_i(0,0)\times J_i(1,0)\times J_i(0,1)\}^{1/4}=\{(10)\times(11)\times(12)\times(13)\}^{1/4}=1$, whatever be the value of the selection probabilities of a card from RR devices. Here p_1 cannot tend to p_2, otherwise variance estimate will be infinite.

3.3. ORR using Forced response model

In ORR with forced response model the sampled person labelled i is requested to give out the truthful response y_i with unknown probability c_i or the forced RR response with probability $1-c_i$. In forced RR device, the person is offered two boxes with three types of cards marked as "Yes", "No" and "Honest Response" but they are in different proportions. For the first box, "Yes", "No" and "Honest Response" are in proportions p_1, p_2 and $1-p_1-p_2$ $(0<p_1,\ p_2<1)$ respectively. For the second box, they are in proportions p_3, p_4 and $1-p_3-p_4$ $(0<p_3,\ p_4<1)$ respectively. But we should add the restriction $p_1p_4=p_2p_3$ on the known probabilities $p_1, p_2, p_{3,} p_4$ to derive an unbiased estimator for the proportion with stigmatizing attribute A.

So, the Optional randomized response for i^{th} person is

$Z_i = y_i$, with unknown probability c_i

= the Forced RR, with unknown probability $1-c_i$, using the first box

and

$Z_i' = y_i$, with unknown probability c_i

= the Forced RR, with unknown probability $1-c_i$, using the second box.

Then,

$$P(Z_i=1)=c_iy_i+(1-c_i)[(1-p_1-p_2)y_i+p_1]$$

$$P(Z_i=0)=c_i(1-y_i)+(1-c_i)[(1-p_1-p_2)(1-y_i)+p_2]$$

$$P(Z_i'=1)=c_iy_i+(1-c_i)[(1-p_3-p_4)y_i+p_3]$$

$$P(Z_i'=0)=c_i(1-y_i)+(1-c_i)[(1-p_3-p_4)(1-y_i)+p_4].$$

The unbiased estimator of y_i is $r_i=\dfrac{p_3Z_i-p_1Z_i'}{p_3-p_1}$, $(p_3\neq p_1)$ and the unbiased estimator of the variance $V_R(r_i)$ is $v_i=\dfrac{p_1p_3(Z_i-Z_i')^2}{(p_3-p_1)^2}$. It is proved in Appendix 2.

Then, the posterior probabilities and the response specific jeopardy measures for different responses are shown below.

$$L_i(1,1)=\frac{L_i\{c_i+(1-c_i)(1-p_2)\}\{c_i+(1-c_i)(1-p_4)\}}{L_i\{c_i+(1-c_i)(1-p_2)\}\{c_i+(1-c_i)(1-p_4)\}+(1-L_i)\{(1-c_i)p_1\}\{(1-c_i)p_3\}}$$

$$L_i(0,0)=\frac{L_i\{(1-c_i)p_2\}\{(1-c_i)p_4\}}{L_i\{(1-c_i)p_2\}\{(1-c_i)p_4\}+(1-L_i)\{c_i+(1-c_i)(1-p_1)\}\{c_i+(1-c_i)(1-p_3)\}}$$

$$L_i(1,0)=\frac{L_i\{c_i+(1-c_i)(1-p_2)\}\{(1-c_i)p_4\}}{L_i\{c_i+(1-c_i)(1-p_2)\}\{(1-c_i)p_4\}+(1-L_i)\{(1-c_i)p_1\}\{c_i+(1-c_i)(1-p_3)\}}$$

$$L_i(0,1)=\frac{L_i\{(1-c_i)p_2\}\{c_i+(1-c_i)(1-p_4)\}}{L_i\{(1-c_i)p_2\}\{c_i+(1-c_i)(1-p_4)\}+(1-L_i)\{c_i+(1-c_i)(1-p_1)\}\{(1-c_i)p_3\}}$$

$$J_i(1,1)=\frac{\{c_i+(1-c_i)(1-p_2)\}\{c_i+(1-c_i)(1-p_4)\}}{\{(1-c_i)p_1\}\{(1-c_i)p_3\}} \quad (14)$$

$$J_i(0,0)=\frac{\{(1-c_i)p_2\}\{(1-c_i)p_4\}}{\{c_i+(1-c_i)(1-p_1)\}\{c_i+(1-c_i)(1-p_3)\}} \quad (15)$$

$$J_i(1,0)=\frac{\{c_i+(1-c_i)(1-p_2)\}\{(1-c_i)p_4\}}{\{(1-c_i)p_1\}\{c_i+(1-c_i)(1-p_3)\}} \quad (16)$$

$$J_i(0,1)=\frac{\{(1-c_i)p_2\}\{c_i+(1-c_i)(1-p_4)\}}{\{c_i+(1-c_i)(1-p_1)\}\{(1-c_i)p_3\}}. \quad (17)$$

Here, the proposed jeopardy measure for the i th person is the G.M of $J_i(1,1), J_i(0,0), J_i(1,0), J_i(0,1)$.

It is $$\bar{J}_i=[\frac{p_2^2p_4^2\{c_i+(1-c_i)(1-p_2)\}^2\{c_i+(1-c_i)(1-p_4)\}^2}{p_1^2p_3^2\{c_i+(1-c_i)(1-p_1)\}^2\{c_i+(1-c_i)(1-p_3)\}^2}]^{1/4}$$

$$=\frac{p_2}{p_1}[\frac{\{c_i+(1-c_i)(1-p_2)\}\{c_i+(1-c_i)(1-p_4)\}}{\{c_i+(1-c_i)(1-p_1)\}\{c_i+(1-c_i)(1-p_3)\}}]^{1/2} \quad (17.1)$$

Thus, the GM need not always be unity as is also the case in (18.1) and later and also in (19) below.

Thus, the measure of jeopardy depends on the selection of p_1, p_2, p_3 and p_4 Here $\bar{J}_i \to 1$ if $p_1 \to p_2$ and $p_3 \to p_4$.

3.4. ORR using Kuk's (1990) RR model

Let the sampled person be instructed to record his/her true value of bearing the sensitive attribute A using the ORR device adopting the RR device or direct response. The respondent is directed to draw k (with replacement) number of cards from one of two boxes having red and black cards in different proportions $(\theta_1 : 1-\theta_1$ and $\theta_2 : 1-\theta_2)$ with $0<\theta_1,\theta_2<1$ and requested to report the number $\frac{(\frac{f_i}{k}-\theta_2)}{\theta_1-\theta_2}$, f_i being the number of red cards out of k cards if the sampled person i decides to adopt Kuk's RR device. Cards should be drawn from the first box if the respondent bears the sensitive attribute, otherwise the cards are drawn from the second box having the proportion of red and black cards in proportions $(\theta_2 : 1-\theta_2)$ without disclosing which box is used to draw the cards.

So, the ORR response for i^{th} person is

$$Z_i = y_i \text{ with the unknown probability } c_i$$

$$= \frac{\frac{f_i}{k} - \theta_2}{\theta_1 - \theta_2} \text{ with the unknown probability } 1-c_i \text{, and}$$

$$E_R(f_i) = k[y_i\theta_1 + (1-y_i)\theta_2]$$

leading to $E_R(Z_i) = c_i y_i + (1-c_i)E_R(\frac{\frac{f_i}{k} - \theta_2}{\theta_1 - \theta_2}) = y_i$.

To estimate the variance, the process is repeated one more time and the response variable Z_i' is the same as above but the number of red cards is denoted by f_i' . So, the final unbiased estimator of y_i is $\frac{Z_i + Z_i'}{2}$ and the related unbiased variance estimator is $v_i = \frac{1}{4}(Z_i - Z_i')^2$ following Chaudhuri et al. (2013, 2016).

The posterior probability can be defined as

$$L_i(f_i, f_i') = \frac{L_i P(Z_i = f_i \mid y_i = 1) P(Z_i' = f_i' \mid y_i = 1)}{L_i P(Z_i = f_i \mid y_i = 1) P(Z_i' = f_i' \mid y_i = 1) + (1-L_i) P(Z_i = f_i \mid y_i = 0) P(Z_i' = f_i' \mid y_i = 0)}$$

$$= \frac{L_i \psi_{1i} \psi_{1i}'}{L_i \psi_{1i} \psi_{1i}' + (1-L_i)\psi_{2i}\psi_{2i}'}$$

where $\psi_{1i} = c_i I_i + (1-c_i)\theta_1^{f_i}(1-\theta_1)^{k-f_i}$ with the indicator function I_i defining as $I_i = 1$ if $f_i = 1$ and 0 otherwise and $\psi_{2i} = c_i I_i' + (1-c_i)\theta_2^{f_i}(1-\theta_2)^{k-f_i}$ with another indicator function I_i' defined as $I_i' = 1$ if $f_i = 1$ and 0 otherwise.

Similarly, $\psi_{1i}' = c_i I_i + (1-c_i)\theta_1^{f_i'}(1-\theta_1)^{k-f_i'}$ and $\psi_{2i}' = c_i I_i' + (1-c_i)\theta_2^{f_i'}(1-\theta_2)^{k-f_i'}$ with two indicator functions defined as just above, and the response specific jeopardy measure is

$$J_i(f_i, f_i') = \frac{\psi_{1i}\psi_{1i}'}{\psi_{2i}\psi_{2i}'} = J_i(f_i).J_i(f_i') \tag{18}$$

where $J_i(f_i) = \dfrac{\psi_{1i}}{\psi_{2i}}$; $J_i(f_i') = \dfrac{\psi'_{1i}}{\psi'_{2i}}$ for all f_i, f_i' = 0,1,2,...k

and

$$\bar{J}_i = (\prod_{\forall f_i, f_i'} J_i(f_i, f_i'))^{1/k+1} = (\prod_{\forall f_i, f_i'} J_i(f_i) J_i(f_i'))^{1/k+1} = (\prod_{\forall f_i} J_i(f_i))^{1/(k+1)} . \tag{18.1}$$

Consequently, $J_i(0) = \dfrac{(1-c_i)(1-\theta_1)^k}{c_i + (1-c_i)(1-\theta_2)^k}$ and

$J_i(1) = \dfrac{c_i + (1-c_i)\theta_1(1-\theta_1)^{k-1}}{(1-c_i)\theta_2(1-\theta_2)^{k-1}}$ do not tend to 1 whatever the choice of θ_1, θ_2

But $J_i(f_i) = \dfrac{(1-c_i)\theta_1^{f_i}(1-\theta_1)^{k-f_i}}{(1-c_i)\theta_2^{f_i}(1-\theta_2)^{k-f_i}} = (\dfrac{\theta_1}{\theta_2})^{f_i}(\dfrac{1-\theta_1}{1-\theta_2})^{k-f_i}$, for all

$f_i, f_i' = 2,3..k$.

And it tends to 1 if $\theta_1 \rightarrow \theta_2$

$$\bar{J}_i = [J_i(0).J_i(1).J_i(2).....J_i(k-1).J_i(k)]^{1/k+1}$$

$$= [\frac{(1-c_i)(1-\theta_1)^k}{c_i + (1-c_i)(1-\theta_2)^k} \frac{c_i + (1-c_i)\theta_1(1-\theta_1)^{k-1}}{(1-c_i)\theta_2(1-\theta_2)^{k-1}} . \frac{\theta_1^2(1-\theta_1)^{k-2}}{\theta_2^2(1-\theta_2)^{k-2}} \frac{\theta_1^3(1-\theta_1)^{k-3}}{\theta_2^3(1-\theta_2)^{k-3}} \cdots \frac{\theta_1^k}{\theta_2^2}]^{1/k+1}$$

$$= [\frac{c_i + (1-c_i)\theta_1(1-\theta_1)^{k-1}}{c_i + (1-c_i)(1-\theta_2)^k} \frac{(1-c_i)(1-\theta_1)^k}{(1-c_i)\theta_2(1-\theta_2)^{k-1}} . \frac{\theta_1^2(1-\theta_1)^{k-2}}{\theta_2^2(1-\theta_2)^{k-2}} \frac{\theta_1^3(1-\theta_1)^{k-3}}{\theta_2^3(1-\theta_2)^{k-3}} \cdots \frac{\theta_1^k}{\theta_2^2}]^{1/k+1} . \tag{19}$$

It is observed that $\bar{J}_i$ tends to 1 if $\theta_1, \theta_2 \rightarrow \dfrac{1}{2}$.

4. Simulation study

In this section, we present some numerical illustrations. The tables along the figures provide how our proposed method works for different prior probabilities L_i with the probability of direct response c_i , which is actually unknown, but here, for calculating posterior probabilities with their response specific jeopardy measures, we assume c_i artificially, say, 0.06,0.12,0.63,0.57,0.91, etc. In Figures 1 , 2 , 3 , taking L_i along horizontal axis (in graph "L$_i$") and c_i along vertical directions (in graph "C$_i$"), the plotted points represent the geometric mean of response specific jeopardy measures along with relevant *"p"* values or "θ" values (denoted by (**GM_J;** p$_1$, p$_2$) or (**GM_J;** p$_1$, p$_2$, p$_3$, p$_4$) or (**GM_J;** θ$_1$, θ$_2$)in graphs). Table 1 shows the calculations for ORR using

Warner's model and Greenberg et al.'s unrelated question model with the overall measure of jeopardy $\overline{J}_i$ in the last column, which is exactly 1, whatever the values of L_i, c_i, p_1 and p_2 as discussed in Section 3.1. Here, it is slightly different from 1 due to approximations in posterior probabilities and response specific jeopardy measures. Figure 1 is a representation of the measure of jeopardy $\overline{J}_i$ for all the combinations of (L_i, c_i) as mentioned in Table 1. Table 2 represents the calculation for ORR using Forced response model imposing the restriction $p_1 p_4 = p_2 p_3$ as pointed out in Section 3.3. Figure 2 is a diagrammatic representation of the measure of jeopardy $\overline{J}_i$ while the ORR survey is performed by using the forced model. The numerical study for ORR using Kuk's model is shown in Table 3 along with Figure 3. If the number of cards (k) drawn for RR devices is 2, an artificial data set is used for the simulation study and the results are shown in Table 4. The data consist of an imaginary set of 116 undergraduate students aged below 20 and their reckless driving with weekly expenditures. We are interested to estimate the proportion of the students who broke the traffic rules last year. An unrelated auxiliary variate, whether they are interested in painting, takes for the numerical illustration of optional randomized techniques with Greenberg et al.'s (1969) RR device, as mentioned in Section 3.2. Let U= (1,2,...*i*...,N) be a finite labelled population with N units and the proportion π may be defined as $\pi = \frac{1}{N}\sum_{i=1}^{N} y_i$ treating y as a "study qualitative stigmatizing variable", as mentioned in Section 3.

Samples are taken from the population with unequal probability sampling scheme of Lahiri (1951) – Midzuno (1952) – Sen (1953) used for the selection of a sample of 39 units to estimate the population proportion. Here, the first unit is selected with the probability $p_i^* = \frac{z_i}{Z}$ (where $Z = \sum_{1}^{N} z_i$), the normed size measure and the remaining ones are selected by simple random sampling without replacement (SRSWOR) from the remaining units in the population after the first draw. The variable "Have you ever been fined for breaking traffic rules" is our study qualitative characteristic with "Weekly expenditure" as the size measure. In this design, the inclusion probability π_i of the i^{th} unit in the sample of size n from the population of size N is $p_i^* + (1 - p_i^*)\frac{n-1}{N-1}$ as the i^{th} unit may be selected in first position with probability p_i^* or in any other position with probability $(1 - p_i^*)$ through SRSWOR with probability $\frac{n-1}{N-1}$. Clearly, the second order inclusion probability of the unit (i, j) may be obtained by the following formula $\pi_{ij} = \frac{(n-1)(N-n)(p_i^* + p_j^*) + (n-1)(n-2)}{(N-1)(N-2)}$. We employ Horvitz-Thompson estimator (HTE) to estimate the proportion $\pi = \frac{1}{N}\sum_{i \in s} \frac{y_i}{\pi_i}$ in the case of qualitative character. Since y_i is not directly assessable, an unbiased estimator r_i of y_i is assigned here. Hence, $e = \frac{1}{N}\sum_{i \in s} \frac{r_i}{\pi_i}$ is our the final unbiased estimator of the population proportion with an unbiased estimator of variance $\mathrm{v}(e) = \frac{1}{N^2}[\sum\sum_{i<j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}(\frac{r_i}{\pi_i} - \frac{r_j}{\pi_j})^2 + \sum_{i \in s} \frac{v_i}{\pi_i}]$ where v_i is an unbiased estimator of variance of r_i. The HT estimator e for the proportion need not be a proper fraction and this anomaly arises not because of ORR as it is natural even for DR's. Proportion estimation is a big challenge in statistics. In Randomized response surveys with unequal probabilities, we usually do not face the problem of getting e values outside the range [0,1].

To judge the efficacy of our results, average coverage probabilities (ACP), average coefficient of variation (ACV) and the average Length (AL) of the 95% confidence intervals based on $e \pm 1.96\sqrt{v(e)}$ have been used. To calculate, we draw $T = 1000$ samples from the population by Lahiri (1951) – Midzuno (1952) – Sen (1953) sampling scheme. For each sample we perform ORR methods to calculate the estimates and variance estimates.

The point estimator will be judged good if the estimated coefficient of variation, namely $\mathrm{CV} = 100\frac{\sqrt{v(e)}}{e}$, has a small magnitude, preferably less than 10% or at most 30%. A confidence interval (CI) will be judged good if on drawing a large number of simulated samples, say B in the number taken as 1000, from a population at hand, the (1) CI's happen to cover the known value of the parameter, a percentage of times close to 95% -this percentage is called the ACP, the Average Coverage Percentage and (2) if the average value of the length, AL, say, of a CI is small enough. Between two CI's the one with a lower value of AL will be preferred unless its ACP is too far from 95% compared to that for the other. Tables 4.1., 4.2., 4.3. and 4.4. represent the ACV (in %), ACP (in %) and AL for four different ORR techniques. Figures 4.1., 4.2., 4.3. represent the ACV and ACP values denoted as (ACV, ACP) taking paired "p" ("θ" for optional Kuk model) values along horizontal and vertical axes.

Table 1. ORR with Warner and Unrelated - measure of jeopardy.

L_i	c_i	p_1	p_2	$J_i(0,1)$	$J_i(1,0)$	$J_i(0,0)$	$J_i(1,1)$	**Warner** $\overline{J}_i$	**Unrelated** $\overline{J}_i$
0.1	0.06	0.44	0.49	1.2216	0.8186	1.0409	0.9607	1	1
0.3	0.63	0.3	0.73	3.1622	0.3162	0.039	25.6154	0.9989	0.9989
0.4	0.42	0.95	0.11	0.0285	35.028	0.0335	29.8462	0.9981	0.9981
0.5	0.91	0.42	0.28	0.8246	1.2128	0.0034	297.6667	1.0121	1.0121
0.6	0.37	0.07	0.98	142.46	0.007	0.0145	68.7966	0.9948	0.9948
0.8	0.55	0.43	0.62	1.7154	0.5829	0.072	13.8959	1.0004	1.0004
0.9	0.53	0.57	0.73	1.6731	0.5977	0.0374	26.7692	1.0012	1.0012

Table 2. ORR with Forced model - measure of jeopardy.

L_i	c_i	p_1	p_2	p_3	p_4	$J_i(1,0)$	$J_i(0,1)$	$J_i(0,0)$	$J_i(1,1)$	$\overline{J}_i$
0.1	0.42	0.64	0.23	0.24	0.0863	0.1367	1.4002	0.012	15.9556	0.4375
0.2	0.43	0.45	0.4	0.52	0.4622	1.1	0.7667	0.1154	7.3051	0.9183
0.6	0.18	0.61	0.25	0.47	0.1926	0.4195	0.8615	0.1049	3.4467	0.6013
0.6	0.26	0.39	0.31	0.43	0.3418	0.9762	0.7592	0.1191	6.2231	0.8609
0.7	0.3	0.25	0.4	0.37	0.5920	2.2162	0.7749	0.1892	9.0769	1.3105
0.9	0.1	0.69	0.21	0.6	0.1826	0.4544	0.7778	0.1739	2.0323	0.5945
0.9	0.12	0.58	0.35	0.37	0.2233	0.4039	1.5337	0.1889	3.2799	0.7871

Table 3. ORR with Kuk model (k=2) - measure of jeopardy.

L_i	c_i	θ_1	θ_2	$J_i(0)$	$J_i(1)$	$J_i(2)$	$\overline{J}_i$
0.1	0.18	0.72	0.43	0.1333	1.75	2.867	0.874
0.2	0.36	0.78	0.13	0.0357	6.7143	39	2.107
0.3	0.55	0.54	0.34	0.1333	6.6	2.6	1.318
0.4	0.72	0.49	0.5	0.0886	11.286	1	1
0.5	0.8	0.58	0.53	0.0476	17	1.167	0.981
0.6	0.74	0.55	0.2	0.0549	20	8	2.063
0.7	0.78	0.75	0.76	0.0127	20.5	0.923	0.622
0.9	0.49	0.77	0.81	0.0588	7.25	0.909	0.729

Table 4.1. ACV, ACP, AL for Optional Warner Model

p_1	p_2	ACV	ACP	AL
0.28	0.19	48.1865	94.6	3.8623
0.49	0.36	37.5586	99.1	2.0169
0.54	0.29	26.3733	98	1.0650
0.63	0.56	42.3268	97.3	2.6278
0.66	0.45	24.7465	97.9	0.9712
0.77	0.65	26.8792	99.4	1.0953
0.81	0.63	20.2835	92.5	0.7143

Table 4.2. ACV, ACP, AL for Optional Unrelated Model

p_1	p_2	ACV	ACP	AL
0.36	0.23	42.2166	95.2	2.4704
0.51	0.39	38.2907	86.8	2.0682
0.56	0.49	45.1583	86.5	3.0992
0.69	0.54	28.0311	98.3	1.1740
0.72	0.55	24.9143	93.6	0.9746
0.88	0.34	11.7061	96.8	0.3454
0.92	0.61	12.1766	93.5	0.3634

Table 4.3. ACV, ACP, AL for Optional Forced Model

p_1	p_2	p_3	p_4	ACV	ACP	AL
0.64	0.23	0.24	0.0863	33.7349	90.1	0.5066
0.45	0.4	0.52	0.4622	46.4117	79.4	3.1305
0.39	0.31	0.43	0.3418	55.3157	76.4	4.6584
0.25	0.4	0.37	0.5920	27.8805	91.2	1.1824
0.32	0.38	0.4	0.4750	37.776	84.4	2.0412
0.35	0.23	0.47	0.3089	32.5996	87.7	1.5525
0.15	0.13	0.22	0.1907	28.2077	98.1	1.1993

Table 4.4. ACV, ACP, AL for Optional Kuk Model if k=2

θ_1	θ_2	ACV	ACP	AL
0.6	0.2	5.4740	94.5	0.1749
0.8	0.6	11.3493	95.4	0.3933
0.56	0.4	12.1329	96	0.3856

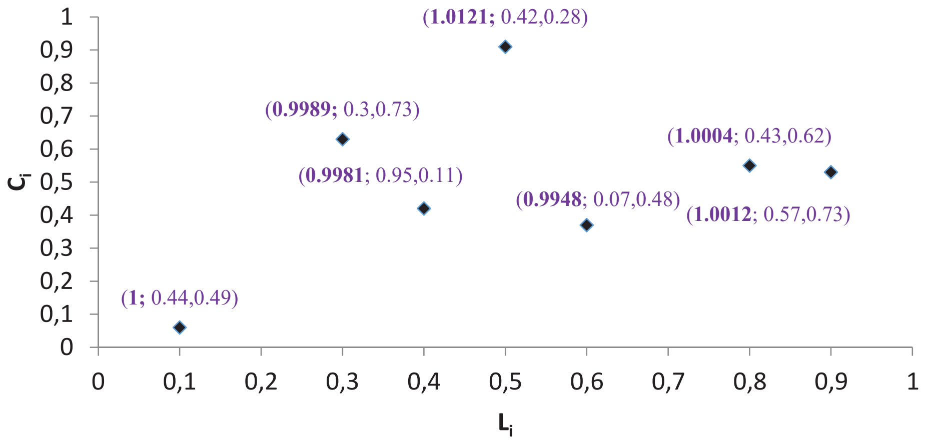

Figure 1. Measure of Jeopardy for Warner and Unrelated ORR

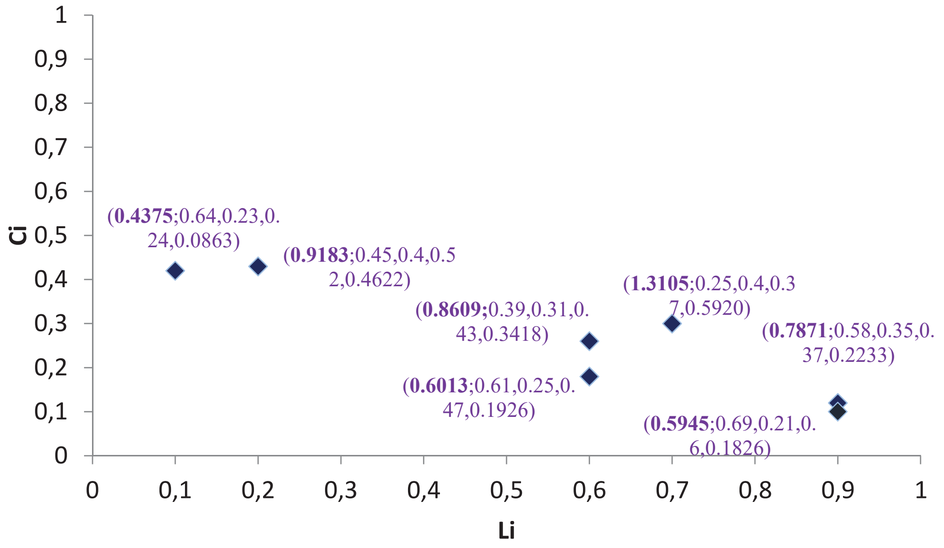

Figure 2. Measure of Jeopardy for Forced ORR

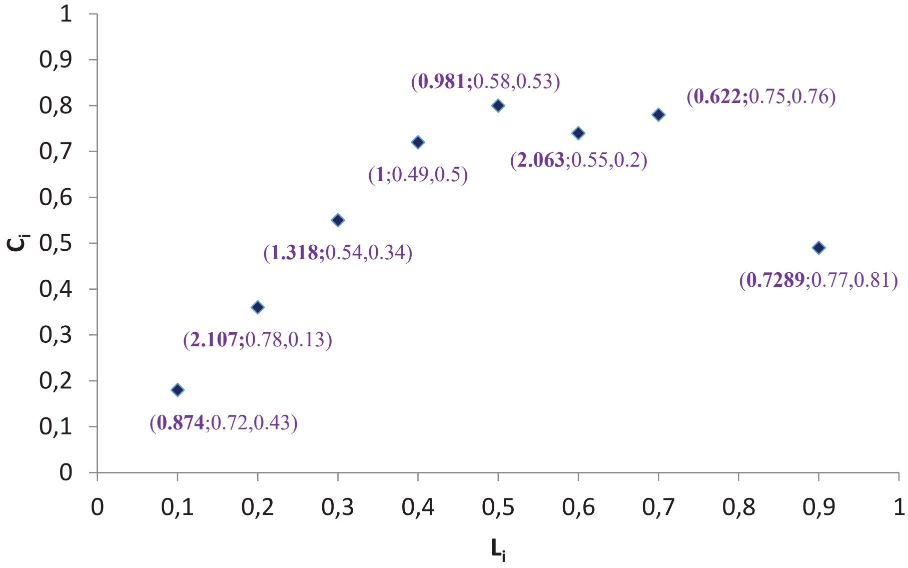

Figure 3. Measure of Jeopardy for Kuk ORR

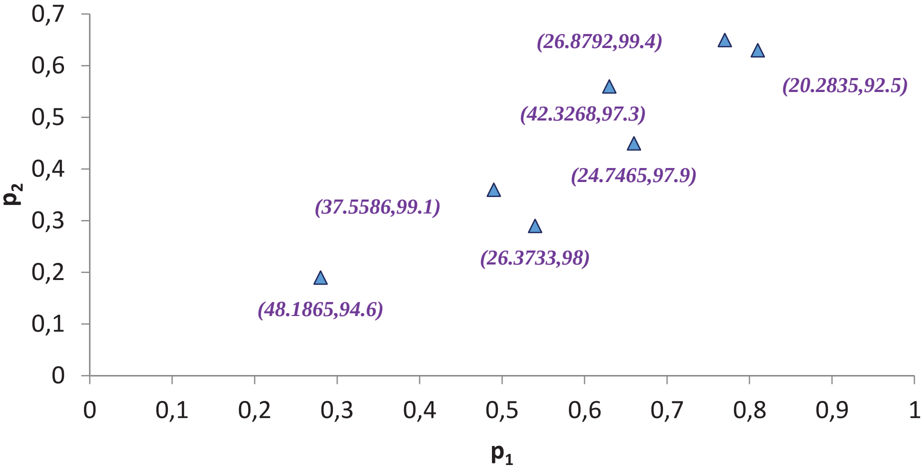

Figure 4.1. Representation of ACP , ACV for Warner ORR

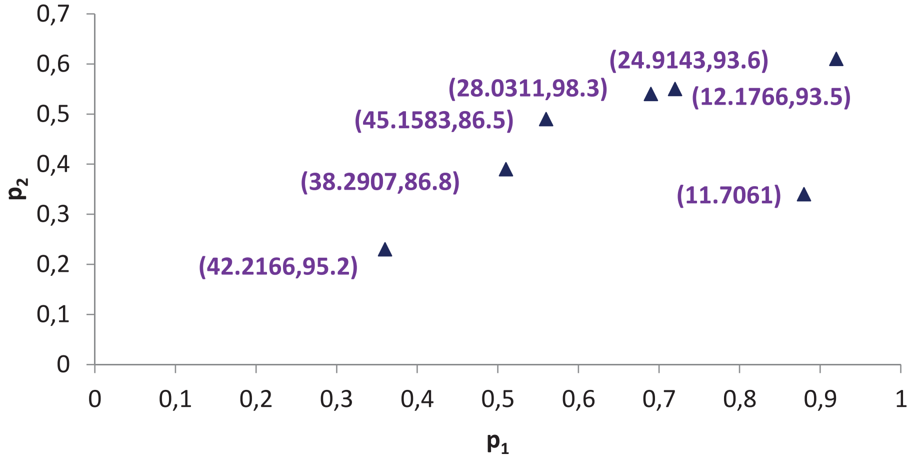

Figure 4.2. Representation of ACP , ACV for Unrelated ORR

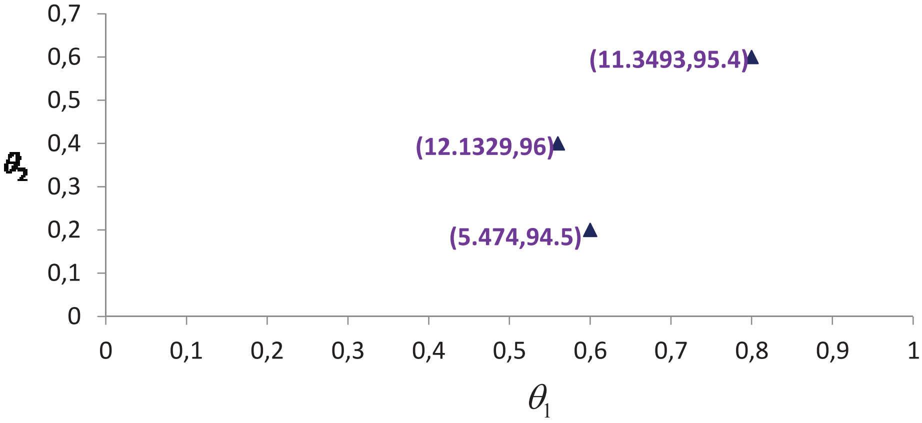

Figure 4.3. Representation of ACP , ACV for Kuk ORR

5. Concluding remarks

Most of the literature on the theory of RR is restricted to simple random sampling (SRS) with replacement (SRSWR). We strongly believe that extension of the theory of RR to varying probability sampling is necessary.

In our proposed ORR method, the probability of choosing between a 'direct' and an 'RR' should vary across individuals rather than be a constant and that is unknown. To get an unbiased variance estimator, two responses from each individual are necessarily required. Regarding the privacy protection of each individual, we can proceed with the ORR method. As a measure of jeopardy, the average jeopardy measure with geometric mean is successfully carried out.

From our results we observe that all of the competing ORR methods show satisfactory results in terms of ACP and ACV values.

Acknowledgement

We are greatly appreciative of the kind comments from the referees as a means to our fruitful efforts in implementation of a possible improvement on our initial draft.

REFERENCES

ARNAB, R., (2004). Optional randomized response techniques for complex survey designs. Biom, J. 46, pp. 114–124.

ARNAB, R., RUEDA, M., (2016). Optional randomized response: A critical review. Handbook of statistics, Elsevier, 34, pp. 253–271

CHAUDHURI, A., (2001). Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population, *Journal of Statistical Planning and Inference*, 94, pp. 37–42.

CHAUDHURI, A., (2011). Randomized response and indirect questioning techniques in surveys. CRC Press, Fl. USA.

CHAUDHURI, A., CHRISTOFIDES, T. C., (2013). Indirect questioning in sample surveys. Springer-Verlag, Berlin, Heidelberg.

CHAUDHURI, A., DIHIDAR, K., (2009). Estimating means of stigmatizing qualitative and quantitative variables from discretionary responses randomized or direct. *Sankhya B*, 71, pp. 123–136.

CHAUDHURI, A., MUKERJEE, R., (1985). Optionally randomized responses techniques. *Calcutta Statistical Association Bulletin,* 34, pp. 225–229.

CHAUDHURI, A., SAHA, A., (2005). Optional versus compulsory randomized response techniques in complex surveys, *Journal of Statistical Planning and Inference*, 135, pp. 516–527.

CHAUDHURI, A., CHRISTOFIDES, T. C., RAO, C. R., (2016). Handbook of statistics, Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits. Elsevier, NL, 34, pp. 2–525.

CHAUDHURI, A., CHRISTOFIDES, T. C., SAHA, A., (2009). Protection of privacy in efficient application of randomized response techniques, *Statistical Methods and Applications,* 18, pp. 389–418.

GREENBERG, B. G., ABUL-ELA, A.-L., SIMMONS, W. R., HORVITZ, D. G., (1969). The unrelated question RR model: Theoretical framework, *Journal of American Statistical Association,* 64, pp. 520–539.

GUPTA, S., (2001). Qualifying the sensitivity level of binary response personal interview survey questions. *Journal of Combinatorics, Information and System Sciences,* 26 (1- 4), pp. 101−109.

GUPTA, S., GUPTA, B., SINGH, S., (2002). Estimation of sensitivity level of personal interview survey question, *Journal of Statistical Planning and Inference,* 100, pp. 239−247.

HORVITZ, D. G., THOMPSON, D. J., (1952). A generalization of sampling without replacement from a finite universe, *Journal of American Statistical Association,* 47, pp. 663−685.

HUANG, K. C., (2008). Estimation of sensitive characteristics using optional randomized techniques, Qual. Quant. 42, pp. 679−686.

KUK, A. Y. C., (1990). Asking sensitive questions indirectly. *Biometrika*, 77(2), pp. 436–438.

LAHIRI, D. B., (1951). A method of sample selection providing unbiased ratio estimates, *Bulletin of International Statistical Institute,* 3, pp. 133−140

MIDZUNO, H., (1952). On the sampling system with probability proportional to the sum of the sizes, *Annals of the Institute of Statistical Mathematics*, 3, pp. 99−107

PAL, S., (2008). Unbiasedly estimating the total of a stigmatizing variable from a complex survey on permitting options for direct or randomized responses, *Statistical Papers,* 49, pp. 157−164

SAHA, A., (2007). Optional randomized response in stratified unequal probability sampling. A simulation based numerical study with Kuk's method, Test 16, pp. 346−354.

SEN, A. R., (1953). On the estimator of the variance in sampling with varying probabilities, *Journal of Indian Society of Agricultural Statistics,* 5, pp. 119−127.

WARNER, S. L., (1965). Randomized response: a survey technique for eliminating evasive answer bias, *Journal of American Statistical Association*, 60, pp. 63−69.

APPENDICES

Appendix 1. variance estimation in ORR using Warner's (1965) RR model (under Section 3.1.)

Estimator $r_i = \dfrac{(1-p_2)Z_i - (1-p_1)Z_i'}{p_1 - p_2}$ (from Section 3.1)

$V_R(r_i) = \dfrac{(1-p_2)^2 V_R(Z_i) + (1-p_1)^2 V_R(Z'_i)}{(p_1 - p_2)^2}$ and

$$
\begin{aligned}
V_R(Z_i) &= E_R(Z_i^2) - [E_R(Z_i)]^2 \\
&= E_R(Z_i) - [E_R(Z_i)]^2 \\
&= E_R(Z_i)[1 - E_R(Z_i)] \\
&= [c_i y_i + (1-c_i)\{(1-p_1) + (2p_1 - 1)y_i)\}][1 - c_i y_i - (1-c_i)\{(1-p_1) + (2p_1 - 1)y_i)\}] \\
&= c_i y_i + (1-c_i)(1-p_1) + (1-c_i)(2p_1 - 1)y_i - c_i^2 y_i - c_i(1-c_i)(1-p_1)y_i - c_i(1-c_i)(2p_1 - 1)y_i \\
&\quad - c_i(1-c_i)(1-p_1)y_i - (1-c_i)^2(1-p_1)^2 - (1-c_i)^2(1-p_1)(2p_1 - 1)y_i \\
&\quad - c_i(1-c_i)(2p_1 - 1)y_i - (1-c_i)^2(2p_1 - 1)^2 y_i \\
&= c_i(1-c_i)y_i + (1-c_i)(1-p_1) + (1-c_i)(2p_1 - 1)y_i - c_i(1-c_i)(1-p_1 + 2p_1 - 1)y_i \\
&\quad - c_i(1-c_i)p_1 y_i - (1-c_i)^2\{(1-p_1)^2 + 2(1-p_1)(2p_1 - 1)y_i + (2p_1 - 1)^2 y_i\} \\
&= c_i(1-c_i)y_i + (1-c_i)(1-p_1) + (1-c_i)(2p_1 - 1)y_i - 2c_i(1-c_i)p_1 y_i \\
&\quad - (1-c_i)^2(1-p_1)^2 - (1-c_i)^2(2p_1 - 1)(2 - 2p_1 + 2p_1 - 1)y_i \\
&= (1-c_i)[c_i y_i + (1-p_1) + (2p_1 - 1)y_i - 2c_i p_1 y_i - (1-c_i)(1-p_1)^2 - (1-c_i)(2p_1 - 1)y_i] \\
&= (1-c_i)[(1-p_1) + c_i y_i + (2p_1 - 1)(1 - 1 + c_i)y_i - 2c_i p_1 y_i - (1-c_i)(1-p_1)^2] \\
&= (1-c_i)[(1-p_1) + c_i y_i + (2p_1 - 1)c_i y_i - 2p_1 c_i y_i - (1-c_i)(1-p_1)^2] \\
&= (1-c_i)(1-p_1)[1 - (1-c_i)(1-p_1)] \\
&= (1-c_i)(1-p_1)[c_i + (1-c_i)p_1]
\end{aligned}
$$

Similarly, $V_R(Z_i') = (1-c_i)(1-p_2)[c_i + (1-c_i)p_2]$.

So,

$$V_R(r_i)=\frac{(1-p_2)^2(1-c_i)(1-p_1)(c_i+(1-c_i)p_1)+(1-p_1)^2(1-c_i)(1-p_2)(c_i+(1-c_i)p_2)}{(p_1-p_2)^2}$$

$$=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(1-p_2)(c_i+(1-c_i)p_1)+(1-p_1)(c_i+(1-c_i)p_2)]$$

$$=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(2-p_1-p_2)c_i+(1-c_i)\{p_1(1-p_2)+p_2(1-p_1)\}]$$

$$=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(2-p_1-p_2)c_i+(1-c_i)\{p_1+p_2-2p_1p_2\}]$$

$$=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(2-p_1-p_2)\{1-(1-c_i)\}+(1-c_i)\{p_1+p_2-2p_1p_2\}]$$

$$=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(2-p_1-p_2)-(1-c_i)\{2-2p_1-2p_2+2p_1p_2\}]$$

$$=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(2-p_1-p_2)-2(1-c_i)(1-p_1)(1-p_2)]$$

Now,

$$E(Z_i-Z_i')^2=E(Z_i+Z_i'-2Z_iZ_i')$$

$$=E(Z_i)+E(Z_i')-2E(Z_iZ_i')$$

$$=2c_iy_i+(1-c_i)[(p_1+p_2)y_i+(2-p_1-p_2)(1-y_i)]-2\{c_iy_i+(1-c_i)(1-p_1+(2p_1-1)y_i\}$$

$$\{c_iy_i+(1-c_i)(1-p_2+(2p_2-1)y_i\}$$

$$=2c_iy_i+(1-c_i)[(2-p_1-p_2)+2(p_1+p_2-1)y_i]-2[c_i^2y_i+c_i(1-c_i)(1-p_1+2p_1-1+1-p_2+2p_2-1)y_i$$

$$+(1-c_i)^2\{(1-p_1)(1-p_2)+(1-p_2)(2p_1-1)y_i+(1-p_1)(2p_2-1)y_i+(2p_1-1)(2p_2-1)y_i\}]$$

$$=2c_i(1-c_i)y_i+(1-c_i)[(2-p_1-p_2)+2(p_1+p_2-1)y_i]-2c_i(1-c_i)(p_1+p_2)y_i$$

$$-2(1-c_i)^2\{(1-p_1)(1-p_2)+(2p_1-4p_1p_2-1+p_2+2p_2-1+p_1+4p_1p_2+1-2p_1-2p_2)y_i\}$$

$$=2c_i(1-c_i)(1-p_1-p_2)y_i+(1-c_i)[(2-p_1-p_2)+2(p_1+p_2-1)y_i]$$

$$-2(1-c_i)^2[(1-p_1)(1-p_2)+(p_1+p_2-1)y_i]$$

$$=(1-c_i)(2-p_1-p_2)+2(1-c_i)^2(p_1+p_2-1)y_i-2(1-c_i)^2(1-p_1)(1-p_2)-2(1-c_i)^2(p_1+p_2-1)y_i$$

$$=(1-c_i)(2-p_1-p_2)-2(1-c_i)^2(1-p_1)(1-p_2)$$

Thus,

$$\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}E(Z_i-Z_i')^2=\frac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(1-c_i)[(2-p_1-p_2)-2(1-c_i)(1-p_1)(1-p_2)]=V_R(r_i)$$

i.e. $v_i=\dfrac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(Z_i-Z_i')^2$ is an unbiased estimator of $V_R(r_i)$.

Appendix 2. ORR using Greenberg et al.'s (1969) unrelated question RR model (under Section 3.2)

Estimator $r_i = \dfrac{(1-p_2)Z_i - (1-p_1)Z_i'}{p_1 - p_2}$ (from Section 3.2)

$$V_R(r_i) = \frac{(1-p_2)^2 V_R(Z_i) + (1-p_1)^2 V_R(Z_i')}{(p_1-p_2)^2} \text{ and}$$

$$V_R(Z_i) = E_R(Z_i)(1-E_R(Z_i)) = (y_i - x_i)^2(1-p_1)(1-c_i)(p_1 + (1-p_1)c_i)$$

Similarly, $V_R(Z_i') = (y_i - x_i)^2(1-p_2)(1-c_i)(p_2 + (1-p_2)c_i)$

So, $V_R(r_i)$ can be written as,

$$V_R(r_i) = \frac{(y_i - x_i)^2(1-c_i)(1-p_1)(1-p_2)}{(p_1-p_2)^2}[2c_i(1-p_1)(1-p_2) + p_1 + p_2 - 2p_1p_2].$$

Now,

$$E_R(Z_i - Z_i')^2 = E_R(Z_i) + E_R(Z_i') - 2E_R(Z_i)E_R(Z_i')$$

$$= (y_i - x_i)^2(1-c_i)[(p_1+p_2) - 2p_1p_2 + 2c_i(1-p_1)(1-p_2)]$$

Thus, $v_i = \dfrac{(1-p_1)(1-p_2)}{(p_1-p_2)^2}(Z_i - Z_i')^2$ is an unbiased estimator of $V_R(r_i)$.

ORR using Forced response model (under Section 3.3.)

Estimator $r_i = \dfrac{p_3 Z_i - p_1 Z_i'}{p_3 - p_1}$ (from Section 3.3) ,

$$V_R(r_i) = \frac{p_3^2}{(p_3-p_1)^2}V_R(Z_i) + \frac{p_1^2}{(p_3-p_1)^2}V_R(Z_i') \text{ and}$$

$$V_R(Z_i) = (2y_i - 1)(1-c_i)\{(p_1+p_2)y_i - p_1\} - (1-c_i)^2\{(p_1+p_2)y_i - p_1\}^2$$

$$= p_1^2(2y_i-1)(1-c_i)(\frac{p_1+p_2}{p_1}y_i - 1)\{\frac{1}{p_1} - (1-c_i)(\frac{p_1+p_2}{p_1}y_i - 1)(2y_i - 1)\} \quad \text{as } (2y_i-1)^2 = 1$$

Similarly,

$$V_R(Z_i') = (2y_i - 1)(1-c_i)\{(p_3+p_4)y_i - p_3\} - (1-c_i)^2\{(p_3+p_4)y_i - p_3\}^2$$

$$= p_3^2(2y_i-1)(1-c_i)(\frac{p_3+p_4}{p_3}y_i - 1)\{\frac{1}{p_3} - (1-c_i)(\frac{p_3+p_4}{p_3}y_i - 1)(2y_i - 1)\}$$

Using the condition $p_1 p_4 = p_2 p_3$, (see Section 3.3)

$$V_R(r_i) = \frac{p_1^2 p_3^2}{(p_3 - p_1)^2}(2y_i - 1)(1 - c_i)(\frac{p_1 + p_2}{p_1} y_i - 1)(\frac{1}{p_1} + \frac{1}{p_3} - 2(1 - c_i)(\frac{p_1 + p_2}{p_1} y_i - 1)(2y_i - 1)) \cdot$$

Now,

$$E(Z_i - Z_i')^2 = p_1 p_3 (2y_i - 1)(1 - c_i)(\frac{p_1 + p_2}{p_1} y_i - 1)(\frac{1}{p_1} + \frac{1}{p_3} - 2(1 - c_i)(\frac{p_1 + p_2}{p_1} y_i - 1)(2y_i - 1)) \cdot$$

So, $v_i = \frac{p_1 p_3}{(p_3 - p_1)^2}(Z_i - Z_i')^2$ is an unbiased estimator for $V_R(r_i)$.