Statistics and Applications {ISSN 2454-7395 (online)}
Volume 19, No. 2, 2021 (New Series), pp 161-179

Estimating Sensitive Population Proportion Permitting Options for Various Respondents' Choices

Purnima Shaw[1] and Sanghamitra Pal[2]
[1]*Department of Statistics and Information Management, Reserve Bank of India*
[2]*Department of Statistics, West Bengal State University*

Received: 31 May 2020; Revised: 08 December 2020; Accepted: 11 December 2020

Abstract

Randomized Response (RR) Techniques (RRT's) are employed to reduce possible bias in gathering data related to sensitive characteristics. Alternatively, Item Count Techniques (ICT's) are also used for indirect questioning related to stigmatizing characteristics. Anticipating that a characteristic may be viewed as stigmatizing by some of the potential respondents but as innocuous by the rest who may not hesitate to give out direct responses (DR), literature is already developed as Optional Randomized Response (ORR) Techniques (ORRT's) permitting respondents to answer either an RR or a DR, exercising respective judgments. In this paper, two ORR techniques relating to qualitative sensitive characteristics are proposed allowing individuals selected by a general sampling scheme to choose DR, RR or ICT according to his/her own choice. Based on simulation results, estimates obtained from the proposed techniques are competitive to those obtained from an existing ORRT.

Key words: Item count technique; Optional randomized response technique; Stigmatizing characteristic; Unequal probability sampling.

AMS Subject Classification: 62 DO5

1. Introduction

Let $\boldsymbol{U} = (1,2,\ldots,i,\ldots,N)$ denote a finite population of a known number N of persons. Let y_i be a stigmatizing variate value such that $y_i = 1$ or 0, if the i^{th} $(i = 1,2,\ldots,N)$ person bears a stigmatizing characteristic A or its complement A^C, respectively. Our objective is to estimate a finite population proportion of individuals bearing A, *i.e.*,

$$\theta = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{1}$$

Warner's (1965) RRT is a well-known device for estimating θ. In this technique, a sampled person i is provided with a box containing similar cards marked A and A^C in proportions $p(\neq 0.5)$ and $(1-p)$, respectively. The individual's response is 1, if the card drawn randomly by him/her matches his/her characteristic and the response is 0, if there is no match. Warner's (1965) RR device was followed by several developments in which selection of sample was restricted to Simple Random Sampling with Replacement (SRSWR). Chaudhuri (2011) and Chaudhuri and Christofides (2013) recommended unequal probability sampling for selecting the units from the population and explained that RRT's are not conditioned by the sampling

Corresponding Author: Sanghamitra Pal
E-mail address: mitrapal2013@gmail.com

schemes. With this amendment, using RR's, y_i can be unbiasedly estimated for each i, followed by estimation of θ and estimate of its standard error.

As respondents may be suspicious of revelation of their privacy in RRT, the ICT, also known as the Block Total Response or the Unmatched Count Technique was introduced by Raghavarao and Federer (1979), Miller (1984) and Miller *et al*. (1986). Further developments in this area include those of Chaudhuri and Christofides (2007), in which two independent samples are required to be selected from U. A questionnaire is provided to the participants in the first sample in which there are G innocuous item statements and the $(G+1)^{th}$ item is —"I bear characteristic A or F", with F as an innocuous characteristic unrelated to A. Another questionnaire is provided to the participants in the second sample in which the same G innocuous item-statements along with the $(G+1)^{th}$ statement —"I do not bear characteristic A or I do not bear characteristic F", are present. Each of the participants responds the number of statements out of $(G+1)$ that are valid for him/her without revealing the answers to the individual statements. Using the responses from the two independent samples, θ and its standard error can be estimated. The developments in this area include those of Chaudhuri and Christofides (2013) and Shaw (2016) among others.

While some individuals may consider A as sensitive, others may prefer giving a direct response (DR). To tackle such situations, ORR devices were contributed by Arnab (2004), Chaudhuri and Saha (2005), Pal (2008), Mehta *et al.* (2012) and Sihm and Gupta (2015), among others. Several ORR devices are elaborated in Arnab and Rueda (2016). In the ORR device by Chaudhuri and Dihidar (2009) as explained in Chaudhuri (2011), each individual i, $i = 1,2,\dots,N$, in the population, bears an unknown probability C_i, $0 \le C_i \le 1$, to opt for giving a direct answer and a probability $(1 - C_i)$ for preferring an RR. Respondents in a sample s drawn from U, are requested to either answer directly about A or provide an RR. However, they are instructed not to reveal the option chosen by them. Another similar response, independent of the first response, is collected from the same set of individuals. Pal (2007) developed an optional method in which, a sampled individual i in s is given the option to either provide an RR or answer to an ICT questionnaire, without revealing the choice of response to the investigator.

It is observed that all the ORR devices existing in the literature provide only two response options to the sampled individuals, *i.e.*, either DR and RR or RR and ICT. It is anticipated that while a few individuals in the population may prefer DR, some may opt for RR and the rest may be comfortable in answering to an ICT questionnaire. Motivated to fill up this gap in the literature, a generalized version of ORR device providing all the three modes of responses *viz.*, DR, RR and ICT is proposed in Section 2. This device mandates selection of 2 independent samples from the population by using a general sampling scheme. An alternative ORR device proposed in Section 3 requires selection of three independent samples chosen by a general sampling scheme. Section 4 provides a derivation of optimum allocation of sample sizes for a given cost of survey. In Pal's (2007) ORRT, there are two response options, *viz.*, RR and ICT; however, option for DR is not provided. So, we have compared our new ORRT's with Pal (2007). Hence in Section 5, performances of the two proposed devices have been compared with the performance of Pal (2007) ORR device, on the basis of a simulated data. The concluding remarks are presented in Section 6.

2. Proposed ORR Device Using Two Independent Samples

A respondent $i (i = 1,2,\dots,N)$, in the population, bears an unknown probability $C_{1i} (0 \le C_{1i} \le 1)$, with preference for opting a DR, a probability $C_{2i} (0 \le C_{1i} + C_{2i} \le 1)$, for an RR and with the remaining probability $(1 - C_{1i} - C_{2i}), 0 \leq 1 - C_{1i} - C_{2i} \leq 1$, for an ICT. Consider a sample s_1 selected from $\boldsymbol{U}$ according to an unequal probability sampling design P admitting positive first order and second order inclusion probabilities $\pi_i = \sum_{s_1 \ni i} P(s_1)$, $\pi_{ij} = \sum_{s_1 \ni i,j} P(s_1)$, $i \neq j, (i,j = 1,2,\dots,N)$. A respondent i is provided with options to either give a DR after multiplying with a constant or answer as per an RR device or answer to an ICT questionnaire, without divulging the chosen option.

If a respondent opts for DR, then, he/ she has to multiply the direct answer y_i with 2 and then give the resulting number in his/her response. In the option for RR, the respondent is requested to multiply his/ her value y_i with 2 and then add it with a number, say a_{11i}, randomly chosen from $(1,2,3,\dots,G-1)$. The questionnaire for ICT consists of G innocuous item statements, the $(G+1)^{th}$ statement being "I have characteristic A or F", *i.e.*, $(A \cup F)$, where F is an innocuous characteristic unrelated to A. A respondent opting for ICT, has to answer the total number of statements holding true for him/ her, say t_{1i}. Considering $f_i = 1$ or 0, if the i^{th} person bears innocuous characteristic F or its complement F^C, respectively, t_{1i} can be expressed as,

$$t_{1i} = \sum_{h=1}^{G} u_{ih} + y_i + f_i - y_i f_i \tag{2}$$

where, u_{ih} takes value 1 if the individual i bears the h^{th} innocuous characteristic, $h = 1,2,\dots,G$. Consider, the i^{th} respondent's answer as z_{11i}, where,

$$z_{11i} = \begin{cases} 2y_i \; with\; probability\; C_{1i}, \quad for\; DR \\ (2y_i + a_{11i}) \; with\; probability\; C_{2i}, \quad for\; RR \\ t_{1i} \; with\; probability\; (1 - C_{1i} - C_{2i}), \quad for\; ICT \end{cases} \tag{3}$$

The respondent i is requested to provide another response, say z_{12i}, independent of z_{11i}, following the same procedure. Let a_{12i} be the random number chosen from $(1,2,3,\dots,G-1)$ by the respondent opting RR, independent of the selection of a_{11i},

$$z_{12i} = \begin{cases} 2y_i \; with\; probability\; C_{1i}, \quad for\; DR \\ (2y_i + a_{12i}) \; with\; probability\; C_{2i}, \quad for\; RR \\ t_{1i} \; with\; probability\; (1 - C_{1i} - C_{2i}), \quad for\; ICT \end{cases} \tag{4}$$

The set of all possible answers for the DR is $\{0,2\}$, for the RR is $\{1,2,3,\dots,G+1\}$ and that for the ICT questionnaire is $\{0,1,2,\dots,G+1\}$. Hence, the investigator remains unaware of the respondent's choice. Taking E_R and V_R as the RR-based expectation and variance operators, respectively,

$$\begin{aligned} & E_R(z_{11i}) = E_R(z_{12i}) \\ &= 2C_{1i}y_i + C_{2i}\left(2y_i + \frac{G}{2}\right) + (1 - C_{1i} - C_{2i})\left(\sum_{h=1}^{G} u_{ih} + y_i + f_i - y_i f_i\right) \end{aligned} \tag{5}$$

Taking the concept of inter-penetrating network of sub-samples developed by Mahalanobis in 1936, consider,

$$r_{1i} = \frac{z_{11i} + z_{12i}}{2}, \quad v_{1i} = \frac{(z_{11i} - z_{12i})^2}{4} \tag{6}$$

Then,

$$E_R(r_{1i}) = E_R(z_{11i}) = E_R(z_{12i}) \tag{7}$$

and,

$$E_R(v_{1i}) = V_R(r_{1i}) \tag{8}$$

Consider a second sample s_2, selected from $\boldsymbol{U}$ (independent of the selection of s_1) according to an unequal probability sampling design P admitting positive first order and second order inclusion-probabilities $\pi_k = \sum_{s_2 \ni k} P(s_2)$, $\pi_{kl} = \sum_{s_2 \ni k,l} P(s_2)$, $k \neq l (k, l = 1,2, \dots, N)$. The selected individuals in s_2 are provided the options for DR, RR and ICT, which are slightly different from the device used for the first sample. A respondent k is provided with options to either give DR added with a constant or answer as per an RR device or answer according to an ICT questionnaire, without revealing the response medium opted by him/ her.

If DR is chosen, then, y_k is to be added with 1. If RR is chosen, then y_k is to be added with $(1 + a_{21k})$, where a_{21k} is randomly chosen from $(0,1,2, \dots, G)$. The ICT questionnaire contains the same G innocuous statements as in the questionnaire used for the first sample, with the $(G+1)^{th}$ statement being "I do not bear characteristic A or I do not bear characteristic F", *i.e.*, $(A^C \cup F^C)$ and the $(G+2)^{th}$ statement being "I bear characteristic F". A respondent opting for ICT, answers t_{2k}, where,

$$t_{2k} = \sum_{h=1}^{G} u_{kh} + 1 - y_k f_k + f_k \tag{9}$$

where, u_{kh} takes value 1 if the individual k bears the h^{th} innocuous characteristic, $h = 1,2, \dots, G$. Consider, the k^{th} respondent's answer as z_{21k}, where,

$$z_{21k} = \begin{cases} y_k + 1 \; with \; probability \; C_{1k}, \quad for \; DR \\ (y_k + 1 + a_{21k}) \; with \; probability \; C_{2k}, \quad for \; RR \\ t_{2k} \; with \; probability \; (1 - C_{1k} - C_{2k}), \quad for \; ICT \end{cases} \tag{10}$$

The respondent k is requested to provide another response, say z_{22k}, independent of z_{21k}, following the same procedure. Let a_{22k} be the number randomly chosen from $(0,1,2, \dots, G)$ by the respondent, independent of the selection of a_{21k},

$$z_{22k} = \begin{cases} y_k + 1 \; with \; probability \; C_{1k}, \quad for \; DR \\ (y_k + 1 + a_{22k}) \; with \; probability \; C_{2k}, \quad for \; RR \\ t_{2k} \; with \; probability \; (1 - C_{1k} - C_{2k}), \quad for \; ICT \end{cases} \tag{11}$$

The set of all possible answers for DR being $\{1,2\}$, for RR being $\{1,2,3, \dots G+2\}$ and for the ICT being $\{1,2,3, \dots G+2\}$, the medium of response chosen by the respondent is not revealed. Then, similar to (6), taking,

$$r_{2k} = \frac{z_{21k} + z_{22k}}{2}, \quad v_{2k} = \frac{(z_{21k} - z_{22k})^2}{4} \tag{12}$$

gives,

$$E_R(r_{2k}) = E_R(z_{21k}) = E_R(z_{22k}) \tag{13}$$

and, $$E_R(v_{2k}) = V_R(r_{2k}) \tag{14}$$

We consider the Horvitz Thompson (1952) estimator e to estimate θ, where,

$$e = 1 + \frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i} - \frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k} \tag{15}$$

Then, assuming E_P and V_P as the design-based expectation and variance operators respectively,

$$E(e) = E_R E_P(e) = E_P E_R(e) = \frac{1}{N}\sum_{i=1}^{N} y_i = \theta \tag{16}$$

Hence, e is an unbiased estimator of θ. Now, taking clue from Chaudhuri and Pal (2002), variance of e can be expressed as,

$$V(e) = V\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) + V\left(\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}\right)$$

$$= E_P V_R\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) + V_P E_R\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) + E_P V_R\left(\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}\right) + V_P E_R\left(\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}\right)$$

$$= E_R V_P\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) + V_R E_P\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) + E_R V_P\left(\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}\right) + V_R E_P\left(\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}\right)$$

$$= E_R\left[\frac{1}{N^2}\left\{\sum_{i}^{N}\sum_{<j}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{r_{1i}}{\pi_i} - \frac{r_{1j}}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{\beta_i}{\pi_i}{r_{1i}}^2\right\}\right] + \frac{1}{N^2}\sum_{i=1}^{N}V_R(r_{1i})$$
$$+ E_R\left[\frac{1}{N^2}\left\{\sum_{k}^{N}\sum_{<l}^{N}(\pi_k\pi_l - \pi_{kl})\left(\frac{r_{2k}}{\pi_k} - \frac{r_{2l}}{\pi_l}\right)^2 + \sum_{k=1}^{N}\frac{\beta_k}{\pi_k}{r_{2k}}^2\right\}\right] + \frac{1}{N^2}\sum_{k=1}^{N}V_R(r_{2k}) \tag{17}$$

writing,

$$\beta_i = 1 + \frac{1}{\pi_i}\sum\nolimits_{j\neq i}^{N}\pi_{ij} - \sum\nolimits_{i=1}^{N}\pi_i, \qquad \beta_k = 1 + \frac{1}{\pi_k}\sum\nolimits_{k\neq l}^{N}\pi_{kl} - \sum\nolimits_{k=1}^{N}\pi_k, \tag{18}$$

If every sample s_1 and s_2 contains a common number of distinct units in it, then, $\beta_i = 0\ \forall$ i and $\beta_k = 0\ \forall$ k throughout in $V(e)$ above.
Then, taking clue from Chaudhuri and Pal (2002), an unbiased estimator for $V(e)$ is,

$$v(e) = \frac{1}{N^2}\left\{\sum_{i<}\sum_{j\in s_1}\left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{r_{1i}}{\pi_i} - \frac{r_{1j}}{\pi_j}\right)^2 + \sum_{i\in s_1}\frac{\beta_i}{{\pi_i}^2}{r_{1i}}^2\right\} + \frac{1}{N^2}\sum_{i\in s_1}\frac{v_{1i}}{\pi_i} \tag{19}$$

$$+\frac{1}{N^2}\left\{\sum_{k<}\sum_{l\in s_2}\left(\frac{\pi_k\pi_l-\pi_{kl}}{\pi_{kl}}\right)\left(\frac{r_{2k}}{\pi_k}-\frac{r_{2l}}{\pi_l}\right)^2+\sum_{k\in s_2}\frac{\beta_k}{{\pi_k}^2}{r_{2k}}^2\right\}+\frac{1}{N^2}\sum_{k\in s_2}\frac{v_{2k}}{\pi_k}$$

with $\beta_i = 0\ \forall$ i and $\beta_k = 0\ \forall$ k in $v(e)$ when applicable. Hence, $v(e)$ is an unbiased estimator of $V(e)$, such that $E\{v(e)\} = E_P E_R\{v(e)\} = E_R E_P\{v(e)\} = V(e)$. A $100(1-\alpha)\%$ Confidence Interval for θ is, $[L, U]$, where,

$$L = e - \left\{\tau_{\alpha/2}\sqrt{v(e)}\right\}, \qquad U = e + \left\{\tau_{\alpha/2}\sqrt{v(e)}\right\} \tag{20}$$

where, $\tau_{\alpha/2}$ is the upper $\alpha/2$ point of $N(0,1)$ distribution

This device, although provides three choices to the respondents, it has a limitation. A respondent may prefer to give direct value to the investigator, instead of multiplying or adding it with a constant. The authors have resolved this issue in the proposed device in Section 3.

3. Proposed ORR Device Using Three Independent Samples

In this device, a respondent i in the first sample s_1 opting DR, has to answer y_i directly. For giving an RR, y_i is to be multiplied with a number, say a'_{11i}, randomly chosen from $(0,1,2,\ldots,G+1)$. The ICT questionnaire contains G innocuous item statements and the $(G+1)^{th}$ statement is "I bear characteristic A or F". Let t'_{1i} be the total number of statements valid for respondent i who has chosen ICT. Then,

$$t'_{1i} = \sum_{h=1}^{G} u_{ih} + y_i + f_i - y_i f_i \tag{21}$$

where, u_{ih} takes value 1 if the individual i bears the h^{th} innocuous characteristic, $h = 1,2,\ldots,G$. Consider, the i^{th} respondent's answer as z'_{11i},

$$z'_{11i} = \begin{cases} y_i\ with\ probability\ C_{1i}, & for\ DR \\ a'_{11i}y_i\ with\ probability\ C_{2i}, & for\ RR \\ t'_{1i}\ with\ probability\ (1-C_{1i}-C_{2i}), & for\ ICT \end{cases} \tag{22}$$

The respondent is requested to provide another response, say z'_{12i}, independent of z'_{11i}, following the same procedure. Let a'_{12i} denote the number randomly chosen from $(0,1,2,\ldots,G+1)$ by the respondent, independent of the selection of a'_{11i}, then,

$$z'_{12i} = \begin{cases} y_i\ with\ probability\ C_{1i}, & for\ DR \\ a'_{12i}y_i\ with\ probability\ C_{2i}, & for\ RR \\ t'_{1i}\ with\ probability\ (1-C_{1i}-C_{2i}), & for\ ICT \end{cases} \tag{23}$$

The set of all possible answers for DR is $\{0,1\}$, for RR is $\{0,1,2,\ldots G+1\}$ and that for ICT questionnaire is $\{0,1,2,\ldots G+1\}$. Hence, the response option chosen by the respondent remains unknown. Then, taking

$$r'_{1i} = \frac{z'_{11i} + z'_{12i}}{2}, \quad v'_{1i} = \frac{(z'_{11i} - z'_{12i})^2}{4} \tag{24}$$

gives,

$$E_R(r'_{1i}) = E_R(z'_{11i}) = E_R(z'_{12i}) \tag{25}$$

and,

$$E_R(v'_{1i}) = V_R(r'_{1i}) \tag{26}$$

Respondents in the second sample s_2 are also provided with all the three response options but the ORR device in this case is differently designed as compared to the device used for the first sample. The RR is obtained by adding a number, say a'_{21k}, randomly chosen from $(0,1,2,\dots,G+1)$ with another number, say b'_{21k}, randomly chosen from $(0,1,2,\dots,H)$ and then multiplying this sum with the y −value. The ICT questionnaire contains the same G innocuous item statements as in the questionnaire used for the first sample, with an additional set of H innocuous item statements and the $(G+H+1)^{th}$ statement being "I do not bear characteristic A or I do not bear characteristic F". A sampled individual choosing ICT, has to answer, say t'_{2k}, where,

$$t'_{2k} = \sum_{h=1}^{G} u_{kh} + \sum_{o=1}^{H} w_{ko} + 1 - y_k f_k \tag{27}$$

where, u_{kh} takes value 1 if the individual k bears the h^{th} innocuous characteristic, $h = 1,2,\dots,G$ and w_{ko} takes value 1 if the individual k bears the o^{th} innocuous characteristic, $o = 1,2,\dots,H$. Let the k^{th} sampled individual's answer be z'_{21k}, where

$$z'_{21k} = \begin{cases} y_k \; with\, probability\; C_{1k}, \quad for\; DR \\ (a'_{21k} + b'_{21k}) y_k \; with\, probability\; C_{2k}, \quad for\; RR \\ t'_{2k} \; with\, probability\; (1 - C_{1k} - C_{2k}), \quad for\; ICT \end{cases} \tag{28}$$

The respondent k is requested to provide another response, say z'_{22k}, independent of z'_{21k}, following the same procedure. Let a'_{22k} denote the number randomly chosen from $(0,1,2,\dots,G+1)$ and another number, say b'_{22k}, randomly chosen from $(0,1,2,\dots,H)$ by the respondent, independent of the selection of b'_{21k},

$$z'_{22k} = \begin{cases} y_k \; with\, probability\; C_{1k}, \quad for\; DR \\ (a'_{22k} + b'_{22k}) y_k \; with\, probability\; C_{2k}, \quad for\; RR \\ t'_{2k} with\, probability\; (1 - C_{1k} - C_{2k}), \quad for\; ICT \end{cases} \tag{29}$$

The sets of all possible answers for DR, RR and ICT questionnaire being $\{0,1\}$, $\{0,1,2,\dots G+H+1\}$ and $\{0,1,2,\dots G+H+1\}$, respectively, the respondent's choice remains hidden. Then, taking

$$r'_{2k} = \frac{z'_{21k} + z'_{22k}}{2}, \quad v'_{2k} = \frac{(z'_{21k} - z'_{22k})^2}{4} \tag{30}$$

gives,

$$E_R(r'_{2k}) = E_R(z'_{21k}) = E_R(z'_{22k}) \tag{31}$$

and,

$$E_R(v'_{2k}) = V_R(r'_{2k}) \tag{32}$$

A third sample s_3 is chosen (independent of the selection of s_1 and s_2) from $\boldsymbol{U}$ according to an unequal probability sampling design P with positive first and second order inclusion-probabilities $\pi_d = \sum_{s_3 \ni d} P(s_3)$, $\pi_{dq} = \sum_{s_3 \ni d,q} P(s_3), d \neq q (d, q = 1,2, \ldots, N)$. The instructions for RR and ICT for the sampled individuals in this case differ from those in the first two samples. Respondents are free to choose any of the three response options and answer accordingly without revealing the chosen response option.

RR is generated by adding a number, say b'_{31d}, randomly chosen from $(0,1,2, \ldots, H)$, with 1 and then multiplying the resulting number with the respondent's $y -$ value. All the H innocuous item statements in the second questionnaire are repeated in the current ICT questionnaire along with the $(H+1)^{th}$ statement "I do not bear characteristic F". Let the response for ICT, if chosen by the d^{th} respondent, be t'_{3d} where,

$$t'_{3d} = \sum_{o=1}^{H} w_{do} + 1 - f_d \tag{33}$$

where, w_{do} takes value 1 if the individual d bears the o^{th} innocuous characteristic, $o = 1,2, \ldots, H$. Consider, the d^{th} respondent's answer as z'_{31d},

$$z'_{31d} = \begin{cases} y_d \; with \; probability \; C_{1d}, \quad for \; DR \\ (1 + b'_{31d}) y_d \; with \; probability \; C_{2d}, \quad for \; RR \\ t'_{3d} \; with \; probability \; (1 - C_{1d} - C_{2d}), \quad for \; ICT \end{cases} \tag{34}$$

The respondent d is requested to provide another response, say z'_{32d}, independent of z'_{31d}, following the same procedure. Then, taking b'_{32d} as the number randomly chosen from $(0,1,2, \ldots, H)$ by the respondent, independent of the selection of b'_{31d},

$$z'_{32d} = \begin{cases} y_d \; with \; probability \; C_{1d}, \quad for \; DR \\ (1 + b'_{32d}) y_d \; with \; probability \; C_{2d}, \quad for \; RR \\ t'_{3d} \; with \; probability \; (1 - C_{1d} - C_{2d}), \quad for \; ICT \end{cases} \tag{35}$$

The sets of all possible responses for DR, RR and ICT are $\{0,1\}$, $\{0,1,2, \ldots, H+1\}$ and $\{0,1,2, \ldots, H+1\}$, respectively, thus indicating that the investigator is unaware of the choice of the respondent. Taking,

$$r'_{3d} = \frac{z'_{31d} + z'_{32d}}{2}, \quad v'_{3d} = \frac{(z'_{31d} - z'_{32d})^2}{4} \tag{36}$$

gives,

$$E_R(r'_{3d}) = E_R(z'_{31d}) = E_R(z'_{32d}) \tag{37}$$

and,

$$E_R(v'_{3d}) = V_R(r'_{3d}) \tag{38}$$

We consider the Horvitz Thompson (1952) estimator e', where,

$$e' = \frac{1}{N} \sum_{i \in s_1} \frac{r'_{1i}}{\pi_i} - \frac{1}{N} \sum_{k \in s_2} \frac{r'_{2k}}{\pi_k} + \frac{1}{N} \sum_{d \in s_3} \frac{r'_{3d}}{\pi_d} \tag{39}$$

Then,

$$E(e') = E_R E_P(e') = E_P E_R(e') = \frac{1}{N}\sum_{i=1}^{N} y_i = \theta \tag{40}$$

Hence, e' is an unbiased estimator of θ. Now, to find out the variance of e',

$$V(e') = V\left(\frac{1}{N}\sum_{i\in s_1}\frac{r'_{1i}}{\pi_i}\right) + V\left(\frac{1}{N}\sum_{k\in s_2}\frac{r'_{2k}}{\pi_k}\right) + V\left(\frac{1}{N}\sum_{d\in s_3}\frac{r'_{3d}}{\pi_d}\right)$$

$$= E_P V_R\left(\frac{1}{N}\sum_{i\in s_1}\frac{r'_{1i}}{\pi_i}\right) + V_P E_R\left(\frac{1}{N}\sum_{i\in s_1}\frac{r'_{1i}}{\pi_i}\right) + E_P V_R\left(\frac{1}{N}\sum_{k\in s_2}\frac{r'_{2k}}{\pi_k}\right) + V_P E_R\left(\frac{1}{N}\sum_{k\in s_2}\frac{r'_{2k}}{\pi_k}\right)$$
$$+ E_P V_R\left(\frac{1}{N}\sum_{d\in s_3}\frac{r'_{3d}}{\pi_d}\right) + V_P E_R\left(\frac{1}{N}\sum_{d\in s_3}\frac{r'_{3d}}{\pi_d}\right)$$

$$= E_R V_P\left(\frac{1}{N}\sum_{i\in s_1}\frac{r'_{1i}}{\pi_i}\right) + V_R E_P\left(\frac{1}{N}\sum_{i\in s_1}\frac{r'_{1i}}{\pi_i}\right) + E_R V_P\left(\frac{1}{N}\sum_{k\in s_2}\frac{r'_{2k}}{\pi_k}\right) + V_R E_P\left(\frac{1}{N}\sum_{k\in s_2}\frac{r'_{2k}}{\pi_k}\right)$$
$$+ E_R V_P\left(\frac{1}{N}\sum_{d\in s_3}\frac{r'_{3d}}{\pi_d}\right) + V_R E_P\left(\frac{1}{N}\sum_{d\in s_3}\frac{r'_{3d}}{\pi_d}\right)$$

$$= E_R\left[\frac{1}{N^2}\left\{\sum_{i}^{N}\sum_{<j}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{r'_{1i}}{\pi_i} - \frac{r'_{1j}}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{\beta_i}{\pi_i}{r'_{1i}}^2\right\}\right] + \frac{1}{N^2}\sum_{i=1}^{N} V_R(r'_{1i})$$

$$+E_R\left[\frac{1}{N^2}\left\{\sum_{k}^{N}\sum_{<l}^{N}(\pi_k\pi_l - \pi_{kl})\left(\frac{r'_{2k}}{\pi_k} - \frac{r'_{2l}}{\pi_l}\right)^2 + \sum_{k=1}^{N}\frac{\beta_k}{\pi_k}{r'_{2k}}^2\right\}\right] + \frac{1}{N^2}\sum_{k=1}^{N} V_R(r'_{2k}) \tag{41}$$

$$+E_R\left[\frac{1}{N^2}\left\{\sum_{d}^{N}\sum_{<q}^{N}(\pi_d\pi_q - \pi_{dq})\left(\frac{r'_{3d}}{\pi_d} - \frac{r'_{3q}}{\pi_q}\right)^2 + \sum_{d=1}^{N}\frac{\beta_d}{\pi_d}{r'_{3d}}^2\right\}\right] + \frac{1}{N^2}\sum_{d=1}^{N} V_R(r'_{3d})$$

where, $\beta_i = 1 + \frac{1}{\pi_i}\sum_{j\neq i}^{N}\pi_{ij} - \sum_{i=1}^{N}\pi_i$, $\quad \beta_k = 1 + \frac{1}{\pi_k}\sum_{k\neq l}^{N}\pi_{kl} - \sum_{k=1}^{N}\pi_k$

and,

$$\beta_d = 1 + \frac{1}{\pi_d}\sum_{d\neq q}^{N}\pi_{dq} - \sum_{d=1}^{N}\pi_d \tag{42}$$

If every sample s_1, s_2 and s_3 contains a common number of distinct units in it, then, $\beta_i = 0\ \forall$ i, $\beta_k = 0\ \forall$ k and $\beta_d = 0\ \forall$ d throughout in $V(e)$ above, using Chaudhuri and Pal (2002). Then, taking clue from Chaudhuri and Pal (2002), an unbiased estimator of $V(e')$ is,

$$v(e') = \frac{1}{N^2}\left\{\sum_{i<}\sum_{j\in s_1}\left(\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right)\left(\frac{r'_{1i}}{\pi_i} - \frac{r'_{1j}}{\pi_j}\right)^2 + \sum_{i\in s_1}\frac{\beta_i}{{\pi_i}^2}{r'_{1i}}^2\right\} + \frac{1}{N^2}\sum_{i\in s_1}\frac{v'_{1i}}{\pi_i} \tag{43}$$

$$+\frac{1}{N^2}\left\{\sum_{k<}\sum_{l\in s_2}\left(\frac{\pi_k\pi_l-\pi_{kl}}{\pi_{kl}}\right)\left(\frac{r'_{2k}}{\pi_k}-\frac{r'_{2l}}{\pi_l}\right)^2+\sum_{k\in s_2}\frac{\beta_k}{\pi_k{}^2}r'_{2k}{}^2\right\}+\frac{1}{N^2}\sum_{k\in s_2}\frac{v'_{2k}}{\pi_k}$$

$$+\frac{1}{N^2}\left\{\sum_{d<}\sum_{q\in s_3}\left(\frac{\pi_d\pi_q-\pi_{dq}}{\pi_{dq}}\right)\left(\frac{r'_{3d}}{\pi_d}-\frac{r'_{3q}}{\pi_q}\right)^2+\sum_{d\in s_3}\frac{\beta_d}{\pi_d{}^2}r'_{3d}{}^2\right\}+\frac{1}{N^2}\sum_{d\in s_3}\frac{v'_{3d}}{\pi_d}$$

with $\beta_i = 0\ \forall$ i, $\beta_k = 0\ \forall$ k and $\beta_d = 0\ \forall$ d in $v(e')$ when applicable. Hence, $v(e')$ is an unbiased estimator of $V(e')$, such that, $E\{v(e')\} = E_P E_R\{v(e')\} = E_R E_P\{v(e')\} = V(e')$. A $100(1-\alpha)\%$ Confidence Interval for θ is, $[L', U']$, where,

$$L' = e' - \left\{\tau_{\alpha/2}\sqrt{v(e')}\right\}, \qquad U' = e' + \left\{\tau_{\alpha/2}\sqrt{v(e')}\right\} \tag{44}$$

4. Optimum Sample Size Allocation for Fixed Survey Cost

As the two proposed ORR devices with options for DR, RR and ICT, mandate selection of multiple samples from the population, this section demonstrates a procedure to minimize the variance of the estimate of $\bar{Y}$ by assuming a fixed cost of the survey. Consider the ORR device in Section 2 and assume that both the samples (of sizes n_1 and n_2, say) are chosen independently from the population by following the Hartley and Rao (1962) sampling scheme and using variable x as the size measure for sample selection, with population total $X = \sum_{i=1}^{N} x_i$. Then, putting the expressions for first and second order inclusion probabilities and using (17), variance of the estimate $\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}$ obtained from the first sample can be written as,

$$V\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) = E_P V_R\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) + V_P E_R\left(\frac{1}{N}\sum_{i\in s_1}\frac{r_{1i}}{\pi_i}\right) = \frac{V_{11}}{n_1} + V_{12} \tag{45}$$

where,

$$V_{11} = \frac{1}{N^2}\sum_{i=1}^{N}\frac{V_R(r_{1i})}{p_i} + \frac{1}{N^2}\left[\sum_{i}^{N}\sum_{<j}^{N} p_{ij}\left\{\frac{E_R(r_{1i})}{p_i} - \frac{E_R(r_{1j})}{p_j}\right\}^2\right], \tag{46}$$

$$V_{12} = \frac{1}{N^2}\left[\sum_{i}^{N}\sum_{<j}^{N}(p_i p_j - p_{ij})\left\{\frac{E_R(r_{1i})}{p_i} - \frac{E_R(r_{1j})}{p_j}\right\}^2\right], \tag{47}$$

$$p_i = \frac{x_i}{X}, \qquad p_j = \frac{x_j}{X}, \tag{48}$$

$$E_R(r_{1i}) = 2C_{1i}y_i + C_{2i}\left(2y_i + \frac{G}{2}\right) + (1 - C_{1i} - C_{2i})\left(\sum_{h=1}^{G} u_{ih} + y_i + f_i - y_i f_i\right), \tag{49}$$

$$V_R(r_{1i}) = \frac{E_R(r_{1i}^2) - E_R^2(r_{1i})}{2}, \tag{50}$$

with,

$$E_R(r_{1i}^2) = 4C_{1i}y_i + C_{2i}\left(4y_i + \frac{G^2}{3} - \frac{G}{6} + 4y_i\frac{G}{2}\right) + (1 - C_{1i} - C_{2i})\left(\sum_{h=1}^{G} u_{ih} + y_i + f_i - y_i f_i\right)^2 \tag{51}$$

Similarly, variance of the estimate $\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}$ obtained from the second sample can be expressed as,

$$V\left(\frac{1}{N}\sum_{k\in s_2}\frac{r_{2k}}{\pi_k}\right) = \frac{V_{21}}{n_2} + V_{22} \tag{52}$$

where,

$$V_{21} = \frac{1}{N^2}\sum_{k=1}^{N}\frac{V_R(r_{2k})}{p_k} + \frac{1}{N^2}\left[\sum_{k}^{N}\sum_{<l}^{N} p_{kl}\left\{\frac{E_R(r_{2k})}{p_k} - \frac{E_R(r_{2l})}{p_l}\right\}^2\right], \tag{53}$$

$$V_{22} = \frac{1}{N^2}\left[\sum_{k}^{N}\sum_{<l}^{N}(p_k p_l - p_{kl})\left\{\frac{E_R(r_{2k})}{p_k} - \frac{E_R(r_{2l})}{p_l}\right\}^2\right], \tag{54}$$

$$p_k = \frac{x_k}{X}, \qquad p_l = \frac{x_l}{X}, \tag{55}$$

$$E_R(r_{2k}) = C_{1k}(y_k + 1) + C_{2k}\left(y_k + 1 + \frac{G}{2}\right) + (1 - C_{1k} - C_{2k})\left(\sum_{h=1}^{G} u_{kh} + 1 - y_k f_k + f_k\right), \tag{56}$$

$$V_R(r_{2k}) = \frac{E_R(r_{2k}^2) - E_R^2(r_{2k})}{2}, \tag{57}$$

with,

$$E_R(r_{2k}^2) = C_{1k}(y_k + 1)^2 + C_{2k}\left\{(y_k + 1)^2 + \frac{G^2}{3} + \frac{G}{6} + 2(y_k + 1)\frac{G}{2}\right\} + (1 - C_{1k} - C_{2k})\left(\sum_{h=1}^{G} u_{kh} + 1 - y_k f_k + f_k\right)^2 \tag{58}$$

Hence,

$$V(e) = \frac{V_{11}}{n_1} + \frac{V_{21}}{n_2} + V_{12} + V_{22} \tag{59}$$

Let C be the total cost of the survey, C_0 be the overhead cost and C' be the cost per unit in the samples selected from the population. Then,

$$C = C_0 + C'(n_1 + n_2) \tag{60}$$

In order to find n_1 and n_2 under the above cost function, consider the following Lagrangian function with λ as the Lagrange multiplier,

$$\Phi = V(e) + \lambda(C - C_0) = \frac{V_{11}}{n_1} + \frac{V_{21}}{n_2} + \lambda C'(n_1 + n_2) + V_{12} + V_{22}$$
$$= \left(\sqrt{\frac{V_{11}}{n_1}} - \sqrt{\lambda C' n_1}\right)^2 + \left(\sqrt{\frac{V_{21}}{n_2}} - \sqrt{\lambda C' n_2}\right)^2 + 2\sqrt{\lambda C'}\left(\sqrt{V_{11}} + \sqrt{V_{21}}\right) + V_{12} + V_{22} \tag{61}$$

Thus, Φ is minimum when,

$$n_1 = \frac{\sqrt{V_{11}}}{\sqrt{\lambda C'}} \qquad n_2 = \frac{\sqrt{V_{21}}}{\sqrt{\lambda C'}} \tag{62}$$

Now, considering C as a pre-specified fixed survey cost, then, $C'(n_1 + n_2) = C - C_0$ gives,

$$\sqrt{\lambda} = \frac{\sqrt{C'}\left(\sqrt{V_{11}} + \sqrt{V_{21}}\right)}{(C - C_0)} \tag{63}$$

Hence,

$$n_1 = \frac{\sqrt{V_{11}}(C - C_0)}{C'\left(\sqrt{V_{11}} + \sqrt{V_{21}}\right)} \qquad n_2 = \frac{\sqrt{V_{21}}(C - C_0)}{C'\left(\sqrt{V_{11}} + \sqrt{V_{21}}\right)} \tag{64}$$

Thus, for fixed survey cost, the variance $V(e)$ is minimum if the two independent samples chosen from the population are of sizes as specified above.

For the ORR device proposed in Section 3, sizes of the three independent samples, say n_1, n_2 and n_3 for which variance of estimate e' is minimum for fixed survey cost, can be similarly worked out, as,

$$n_1 = \frac{\sqrt{V'_{11}}(C - C_0)}{C'\left(\sqrt{V'_{11}} + \sqrt{V'_{21}} + \sqrt{V'_{31}}\right)} \qquad n_2 = \frac{\sqrt{V'_{21}}(C - C_0)}{C'\left(\sqrt{V'_{11}} + \sqrt{V'_{21}} + \sqrt{V'_{31}}\right)}$$
$$n_3 = \frac{\sqrt{V'_{31}}(C - C_0)}{C'\left(\sqrt{V'_{11}} + \sqrt{V'_{21}} + \sqrt{V'_{31}}\right)} \tag{65}$$

where,

$$V'_{11} = \frac{1}{N^2}\sum_{i=1}^{N}\frac{V_R(r'_{1i})}{p_i} + \frac{1}{N^2}\left[\sum_{i}^{N}\sum_{<j}^{N} p_{ij}\left\{\frac{E_R(r'_{1i})}{p_i} - \frac{E_R(r'_{1j})}{p_j}\right\}^2\right] \tag{66}$$

$$V'_{21} = \frac{1}{N^2}\sum_{k=1}^{N}\frac{V_R(r'_{2k})}{p_k} + \frac{1}{N^2}\left[\sum_{k}^{N}\sum_{<l}^{N} p_{kl}\left\{\frac{E_R(r'_{2k})}{p_k} - \frac{E_R(r'_{2l})}{p_l}\right\}^2\right] \tag{67}$$

$$V'_{31} = \frac{1}{N^2}\sum_{k=1}^{N}\frac{V_R(r'_{3d})}{p_k} + \frac{1}{N^2}\left[\sum_{k}^{N}\sum_{<l}^{N} p_{kl}\left\{\frac{E_R(r'_{3d})}{p_k} - \frac{E_R(r'_{3q})}{p_l}\right\}^2\right] \tag{68}$$

It may be noted that for both the proposed ORR devices, the probabilities of choosing DR, RR and ICT are different and unknown for each individual.

5. Numerical Illustration

To examine the performances of the Generalized ORR devices proposed in Sections 2 and 3, a simulated population of $N = 117$ individuals has been considered, wherein, A indicates consumption of alcohol, F indicates the individual's preference in playing football and z is the number of family members of the respondent. It is desired to estimate the population proportion of individuals consuming alcohol. A sample of size $n = 11$ individuals is drawn from the population by following Hartley and Rao's (1962) sampling scheme. In this scheme, a systematic sample is drawn by Probability Proportional to Size (PPS) method following a random arrangement of the population units. The size measure used for the sampling purpose is z.

For the sake of simplicity in notations, the device by Pal (2007), the generalized ORR device proposed with two independent samples (Section 2) and the alternative ORR device with three independent samples (Section 3) are denoted as Device-I, Device-II and Device-III, respectively. For individuals opting to answer an RR, the specifications of the RR device are stated below:

(i) **Device-I:** An individual in the first or second sample, opting for RR has to choose a random number from $(0,1,2,\dots,G)$.

(ii) **Device-II:** For the first sample, RR in this device is generated by choosing a random number from $(1,2,3,\dots,G-1)$. For the second sample, a random number is to be chosen from $(0,1,2,\dots,G)$.

(iii) **Device-III:** This device for the first sample mandates selection of a random number from $(0,1,2,\dots,G+1)$. In the second sample, RR is computed by choosing two numbers randomly from $(0,1,2,\dots,G+1)$ and $(0,1,2,\dots,H)$, respectively. A random number is to be chosen from $(0,1,2,\dots,H)$ for generating RR in the third sample.

As per requirement of ICT, the two sets of $G=5$ and $H=4$ innocuous items statements denoted by B_1, B_2, B_3, B_4, B_5 and E_1, E_2, E_3, E_4, considered here are described below:

Set-1

B_1 : I like listening to music.
B_2 : I am diagnosed with liver disease.
B_3 : I am married.
B_4 : I am planning to buy a house.
B_5 : I love painting.

Set-2

E_1 : I like watching movies.
E_2 : I prefer cricket test matches over one day matches.
E_3 : I am currently employed.
E_4 : My birthday is in December.

The specifications of the ICT questionnaire followed are stated below:

(i) **Device-I:** For both the samples, G innocuous item statements in the questionnaire are as given in Set-1. The $(G+1)^{th}$ item in the questionnaire for the first sample is "I consume alcohol or I love playing football". On the other hand, the $(G+1)^{th}$ item in the questionnaire for the second sample is "I do not consume alcohol or I don't love playing football".

(ii) **Device-II:** G innocuous item statements from Set-1 are used for both the samples. In the questionnaire for the first sample, the $(G+1)^{th}$ statement is "I consume alcohol or I love playing football". In the questionnaire to be used for the second sample, the $(G+1)^{th}$ statement is "I do not consume alcohol or I don't love playing football" and the $(G+2)^{th}$ statement is "I love playing football".

(iii) **Device-III:** For the first sample, G innocuous item statements in the questionnaire are as given in Set-1. The $(G+1)^{th}$ statement is "I consume alcohol or I love playing football". For the second sample, $(G+H)$ innocuous item statements in the questionnaire are provided in Set-1 and Set-2. The $(G+H+1)^{th}$ statement is “I do not consume alcohol or I don’t love playing football”. The H innocuous item statements in Set-2 are used in the questionnaire for the third sample. The $(H+1)^{th}$ statement is “I do not love playing football”.

For each of the three devices, independent samples are drawn each of size $n = 11$. Various scenarios on different proportion of individuals in the sample opting for DR, RR and ICT for Devices I, II and III are identified. For each of these scenarios, e, $v(e)$, L and U for Device-II and e', $v(e')$, L' and U' for Device-III are calculated. Similarly, the proportion estimate e'', its variance estimate $v(e'')$ and confidence interval (L'', U'') are also computed for Device-I. The estimates are derived each time for $D = 1000$ re-samples drawn from the population and then to compare Devices II and III with Device-I, the following are calculated:

Average Estimates: $\frac{1}{D}\sum_{d=1}^{1000} e_d$, $\frac{1}{D}\sum_{d=1}^{1000} e'_d$ and and $\frac{1}{D}\sum_{d=1}^{1000} e''_d$,

Average Relative Efficiency (Device-II relative to Device-I): $\frac{\frac{1}{D}\sum_{d=1}^{1000} v(e''_d)}{\frac{1}{D}\sum_{d=1}^{1000} v(e_d)}100$,

Average Relative Efficiency (Device-III relative to Device-I): $\frac{\frac{1}{D}\sum_{d=1}^{1000} v(e''_d)}{\frac{1}{D}\sum_{d=1}^{1000} v(e'_d)}100$,

Average Relative Bias: $\left|\frac{\frac{1}{D}\sum_{d=1}^{1000} e_d-\theta}{\theta}\right|$, $\left|\frac{\frac{1}{D}\sum_{d=1}^{1000} e'_d-\theta}{\theta}\right|$ and $\left|\frac{\frac{1}{D}\sum_{d=1}^{1000} e''_d-\theta}{\theta}\right|$,

Actual Coverage Percentage for Devices I, II and III *viz.*, percentage of cases out of 1,000 re-samples, in which (L, U), (L', U') and (L'', U'') covers θ and Average Length of the 1,000 replicates of Confidence Intervals for θ for Devices I, II and III are also computed.

If Average Relative Efficiency of a proposed device relative to Device-I is more than 100, then the proposed device is more efficient than Device-I. On the other hand, lower the Average Relative Bias, better the device. Further, closer the Actual Coverage Percentage to 95% and smaller the Average Length, better is the performance of that device.

The Average Estimates (AE), Average Relative Bias (ARB), Actual Coverage Percentage (ACP) and Average Length (AL) obtained from Devices II and III are compared to those obtained from Device-I for various scenarios of individuals’ choices on the medium of response. Further, the Average Relative Efficiency (ARE) of the estimates obtained from each of Device-II and Device-III relative to those calculated using Device-I are also derived. Few such comparisons are displayed in Tables 1 and 2. Table 3 provides a similar comparison of the performances of the proposed Devices II and III.

From Table 1, it is observed that Device-II outperforms Device-I marginally in all aspects *viz*., ARE, ARB and ACP and AL. From Table 2, it is observed that Device-III is better than Device-I in respect of ARE, ARB, ACP as well as AL. Amongst Device-II and Device-III, Device-II shows better performance in terms of ARE, ARB and AL (Table 3). Hence, it can be safely concluded that the proposed Devices II and III are competitive with Device-I. The very purpose of proposing Devices II and III is to accommodate a variety of responses, *viz.*, DR, RR and ICT. The proposed devices not only fulfil this purpose but also perform efficiently in comparison to the existing device Pal (2007) with two response options.

Table 1: Comparison of performance of proposed ORR Device-II with Device-I by Pal (2007)

Sample proportion with a chosen response option					AE ($\theta = 0.67$)		ARE of Device II relative to Device I	ARB		ACP		AL	
Device-II			Device-I		Device			Device		Device		Device	
DR	RR	ICT	RR	ICT	II	I		II	I	II	I	II	I
0.4	0.4	0.2	0.8	0.2	0.67	0.58	103.6	0.005	0.131	97.9	94.8	4.57	4.57
0.6	0.1	0.3	0.7	0.3	0.64	0.60	131.0	0.052	0.098	97.7	95.4	3.80	4.42
0.7	0.1	0.2	0.8	0.2	0.73	0.76	152.4	0.096	0.130	97.8	95.4	3.51	4.47
0.8	0.1	0.1	0.9	0.1	0.68	0.70	210.6	0.021	0.050	96.7	95.4	3.11	4.57
0.6	0.2	0.2	0.8	0.2	0.66	0.71	148.2	0.008	0.055	98.0	96.3	3.76	4.64
0.7	0.2	0.1	0.9	0.1	0.67	0.68	166.4	0.003	0.016	97.8	96.3	3.56	4.57
0.4	0.4	0.2	0.4	0.6	0.67	0.75	113.4	0.005	0.125	97.9	94.8	4.57	4.66
0.4	0.2	0.4	0.2	0.8	0.71	0.62	112.9	0.062	0.071	98.3	95.8	4.32	4.58
0.6	0.2	0.2	0.2	0.8	0.66	0.62	162.1	0.008	0.069	98.0	96.6	3.76	4.71

Table 2: Comparison of performance of proposed ORR Device-III with Device-I by Pal (2007)

Sample proportion with a chosen response option					AE ($\theta = 0.67$)		ARE of Device III relative to Device I	ARB		ACP		AL	
Device-III			Device-I		Device			Device		Device		Device	
DR	RR	ICT	RR	ICT	III	I		III	I	III	I	III	I
0.7	0.1	0.2	0.8	0.2	0.66	0.76	115.4	0.008	0.130	100.0	95.4	4.16	4.47
0.6	0.2	0.2	0.8	0.2	0.67	0.71	111.7	0.000	0.055	100.0	96.3	4.49	4.64
0.7	0.2	0.1	0.9	0.1	0.67	0.68	158.5	0.003	0.016	99.8	96.3	3.74	4.57
0.6	0.3	0.1	0.9	0.1	0.67	0.64	117.8	0.007	0.047	99.5	94.9	4.27	4.62
0.8	0.1	0.1	0.9	0.1	0.70	0.63	181.2	0.052	0.059	99.8	94.7	3.30	4.41
0.6	0.1	0.3	0.7	0.3	0.70	0.60	120.5	0.041	0.104	100.0	95.0	4.63	4.76
0.7	0.1	0.2	0.1	0.9	0.66	0.66	127.0	0.008	0.015	100.0	95.5	4.16	4.50
0.8	0.1	0.1	0.1	0.9	0.65	0.72	173.0	0.025	0.076	99.7	92.5	3.38	4.33
0.6	0.2	0.2	0.2	0.8	0.67	0.62	122.2	0.000	0.069	100.0	96.6	4.49	4.71
0.7	0.2	0.1	0.2	0.8	0.67	0.65	160.8	0.003	0.026	99.8	93.6	3.74	4.46
0.6	0.3	0.1	0.3	0.7	0.67	0.64	113.5	0.007	0.050	99.5	94.8	4.27	4.40

Table 3: Comparison of performances of proposed ORR devices *viz.*, Device-II and Device-III

Sample proportion with a chosen response option			AE ($\theta = 0.67$)		ARE of Device II relative to Device III	ARB		ACP		AL	
DR	RR	ICT	Device			Device		Device		Device	
			II	III		II	III	II	III	II	III
0.1	0.3	0.6	0.68	0.73	139.9	0.018	0.088	96.0	99.4	5.00	6.26
0.1	0.8	0.1	0.63	0.74	104.6	0.060	0.109	95.1	98.9	5.25	5.74
0.2	0.3	0.5	0.69	0.61	155.9	0.028	0.082	97.1	99.5	4.82	6.38
0.2	0.6	0.2	0.72	0.73	125.0	0.075	0.094	96.9	99.2	5.18	5.89
0.3	0.5	0.2	0.65	0.64	128.8	0.035	0.046	97.5	99.3	4.91	5.71

0.4	0.4	0.2	0.67	0.58	134.4	0.005	0.140	97.9	99.5	4.57	5.38
0.4	0.2	0.4	0.68	0.64	159.9	0.021	0.045	97.9	99.8	4.46	5.87
0.6	0.1	0.3	0.64	0.73	140.8	0.052	0.096	97.7	99.9	3.80	4.77
0.8	0.1	0.1	0.68	0.71	103.5	0.012	0.055	97.0	100.0	3.16	3.31
0.6	0.3	0.1	0.65	0.60	114.7	0.023	0.099	97.9	99.9	3.91	4.27

Next, as the two proposed devices mandate selection of multiple samples from the population, an attempt has been made here to compute the optimum samples sizes, based on the discussion in Section 4. The probabilities of choosing DR, RR and ICT are different and unknown for each individual. However, for conducting the numerical computations, it is assumed that these probabilities are same for each individual in the population. Taking this assumption, the optimum sample sizes are calculated for various scenarios of proportion of individuals opting for DR, RR and ICT. Tables 4 and 5 illustrate the optimum sample sizes of independent samples required to be drawn from the population for the two proposed devices. Subsequently, the resulting variances of the estimated population proportions given a fixed survey cost are also displayed. For both the devices, it is observed that for increase in survey costs, gain in efficiency of estimates is achieved with increasing sample sizes.

Table 4: Population variance for fixed survey cost in ORR device with options for RR and ICT using two independent samples (Device-II)

Proportion of individuals with a chosen response option			C (Rs.)	C_0 (Rs.)	C' (Rs.)	n_1	n_2	$V(e)$
DR	**RR**	**ICT**						
0.1	0.2	0.7	300	12	22	6	7	4.65
			600	23	29	9	11	3.36
			900	32	35	12	13	2.87
			1200	47	43	13	14	2.72
			1900	73	64	13	15	2.61
0.2	0.4	0.4	300	12	22	6	7	4.30
			600	23	29	9	10	3.09
			900	32	35	12	13	2.63
			1200	47	43	13	14	2.49
			1900	73	64	14	15	2.39
0.2	0.3	0.5	300	12	22	6	7	4.25
			600	23	29	9	10	3.06
			900	32	35	12	13	2.61
			1200	47	43	13	14	2.47
			1900	73	64	14	15	2.37
0.2	0.6	0.2	300	12	22	6	7	4.52
			600	23	29	9	10	3.25
			900	32	35	12	13	2.76
			1200	47	43	13	14	2.62
			1900	73	64	14	15	2.50
0.4	0.2	0.4	300	12	22	6	7	3.42
			600	23	29	9	10	2.45
			900	32	35	12	13	2.09
			1200	47	43	13	14	1.97
			1900	73	64	14	15	1.89

0.6	0.2	0.2	300	12	22	6	7	2.38
			600	23	29	10	10	1.70
			900	32	35	12	13	1.44
			1200	47	43	13	14	1.36
			1900	73	64	14	15	1.30

Table 5: Population variance for fixed survey cost in ORR device with options for DR, RR and ICT using three independent samples (Device-III)

Proportion of individuals with a chosen response option			C (Rs.)	C_0 (Rs.)	C' (Rs.)	n_1	n_2	n_3	$V(e')$
DR	RR	ICT							
0.1	0.2	0.7	300	12	22	4	6	3	8.87
			600	23	29	6	9	5	6.22
			900	32	35	8	11	6	5.21
			1200	47	43	8	12	6	4.90
			1900	73	64	9	13	7	4.67
0.2	0.4	0.4	300	12	22	4	6	3	7.37
			600	23	29	6	9	5	5.09
			900	32	35	7	11	6	4.23
			1200	47	43	8	12	7	3.96
			1900	73	64	9	13	7	3.77
0.2	0.3	0.5	300	12	22	4	6	3	7.60
			600	23	29	6	9	5	5.27
			900	32	35	8	11	6	4.39
			1200	47	43	8	12	7	4.12
			1900	73	64	9	13	7	3.92
0.2	0.6	0.2	300	12	22	4	6	3	7.00
			600	23	29	6	9	5	4.85
			900	32	35	7	11	7	4.03
			1200	47	43	8	12	7	3.78
			1900	73	64	8	13	8	3.59
0.4	0.2	0.4	300	12	22	4	6	3	6.02
			600	23	29	6	9	5	4.15
			900	32	35	8	11	6	3.44
			1200	47	43	8	12	7	3.23
			1900	73	64	9	13	7	3.06
0.6	0.2	0.2	300	12	22	4	6	3	3.41
			600	23	29	6	9	5	2.32
			900	32	35	8	11	6	1.91
			1200	47	43	8	12	7	1.78
			1900	73	64	9	13	7	1.69

6. Conclusion

To estimate a proportion of individuals bearing a sensitive characteristic in the population, ORR devices present in the literature provide only two types of response options to the survey participants. It is anticipated that in reality, the population is heterogeneous enough to contain individuals out of which a few may opt DR, a few may choose RR and the rest may opt for answering an ICT questionnaire. In such a case, using any ORR device existing in the literature which provides only two response options (DR and RR or RR and ICT), would result in plausible non-responses. Hence, to avoid this issue, two ORR devices are proposed here, both of which provide all the three response options (DR, RR and ICT) to each sampled individual who may choose any one option at his/her discretion without disclosing the choice to the investigator. The first proposed device requires selection of two independent samples from the population and the second device mandates selection of three independent samples. Based on a simulation exercise with different scenarios of respondents' choices for DR, RR and ICT, it is concluded that both the proposed devices are competitive to the existing ORR device.

Acknowledgements

The authors are thankful to the referee and the editor for providing useful comments which have helped in improving the manuscript.

References:

Arnab, R. (2004). Optional randomized response techniques for complex designs. *Biometrical Journal*, **46**, 114-124.

Arnab, R., and Rueda, M. (2016). Optional Randomized Response: A Critical Review. *Handbook of Statistics*, **34**. On *Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits*, 253-271.

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys*. Boca Raton: Chapman and Hall, CRC Press, Taylor & Francis Group (ISBN: 9781138115422).

Chaudhuri, A., and Christofides, T. C. (2007). Item count technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference*, **137**, 589-593.

Chaudhuri, A., and Christofides, T. C. (2013). *Indirect Questioning in Sample Surveys*. Springer.

Chaudhuri, A., and Dihidar, K. (2009). Estimating means of stigmatizing qualitative and quantitative variables from discretionary responses randomized or direct. *Sankhya*, **B71**, 123-136.

Chaudhuri, A., and Pal, S. (2002). On certain alternative mean square error estimators in complex survey sampling. *Journal of Statistical Planning and Inference*, **104**, 363-375.

Chaudhuri, A., and Saha, A. (2005a). On relative efficiencies of optional versus compulsory randomization in responses: A simulation-based numerical study covering three RR schemes. *Pakistan Journal of Statistics*, **21**(**1**), 87-98.

Chaudhuri, A., and Saha, A. (2005b). Optional versus compulsory randomized response techniques in complex surveys. *Journal of Statistical Planning and Inference*, **135**, 516-527.

Hartley, H. O., and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *Annals of Mathematical Statistics*, **33**(**2**), 350-374.

Horvitz, D. G., and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **77**, 89-96.

Mehta, S., Dass, B. K., Shabbir, J., and Gupta, S. N. (2012). A three-stage optional randomized response model. *Journal of Statistical Theory and Practice*, **6**(**3**), 417–427.

Miller, J. D. (1984). *A New Survey Technique for Studying Deviant Behaviour*. Ph.D. Thesis, The George Washington University.

Miller, J. D., Cisin, I. H., and Harrel, A. V. (1986). A new technique for surveying deviant behavior: item count estimates of marijuana, cocaine and heroin. Paper presented at the annual meeting of the *American Association for Public Opinion Research, St. Petersburg, Florida*.

Pal, S. (2007). Estimating the proportion of people bearing a sensitive issue with an option to item count lists and randomized response. *Statistics in Transition*, **8**(**2**), 301-310.

Pal, S. (2008). Unbiasedly estimating the total of a stigmatizing variable from a complex survey on permitting options for direct or randomized responses. *Metrika*, **49**, 157-164.

Raghavarao, D., and Federer, W. F. (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society*, **B41**, 40-45.

Shaw, P. (2016). Estimating a finite population proportion bearing a sensitive attribute from a single probability sample by Item Count Technique. *Handbook of Statistics*, **34**. On *Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits*, 387-403.

Sihm, J. S., and Gupta, S. (2015). A two-stage binary optional response model. *Communications in Statistics – Simulation and Computation*, **44**(**9**), 2278–2296.

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