Pak. J. Statist.
2023 Vol. 39(3), 289-298

FIXING THE SIZE OF A SAMPLE TO DRAW IN A RANDOMIZED RESPONSE SURVEY

Arijit Chaudhuri[1§] and **Dipika Patra**[2]
[1] Indian Statistical Institute (ISI), Kolkata, India.
Email: arijitchaudhuri1@rediffmail.com
ORCID: https://orcid.org/0000-0002-4305-7686
[2] Seth Anandram Jaipuria College, Kolkata, India.
Email: dipika.patra1988@gmail.com
ORCID: https://orcid.org/0000-0003-4318-1123
[§] Corresponding author

ABSTRACT

A usual unbiased estimator for a population total, mean or a proportion based on randomized response data from a probability sample has a variance as a 'sum' of a term containing the values of a variable of interest and another term involving the variances of unbiased estimators of the variate values obtained by randomized responses. The first term may be controlled by an appropriately chosen sampling design and a suitably specified sample-size. But it is not easy to get the second term suitably and naturally controlled. Thus it remains a problem to suitably fix a sample-size to control the magnitude of this 'sum'. An exercise is presented to implement this task.

KEYWORDS

Equal Probability Sampling; Randomized Response Survey; Sample-size fixation

AMS subject classification: 62D05

1. INTRODUCTION

Suppose on a finite survey population $U = (1, \dots, i, .., N)$ is defined a real variable y taking on it the values y_i for i in U. Let us need to unbiasedly estimate the population total $Y = \sum_{i=1}^{N} y_i$ employing an unbiased estimator t for it based on a probability sample s with n $(2 \leq n < N)$ as its size suitably drawn from U. Suppose we need the estimator t to be so accurate that

$$Prob[|t - Y| \leq fY] \geq 1 - \alpha \tag{1.1}$$

Suitably choosing f and α as positive proper fractions, say, for example $f = 0.1$ or 0.2 etc. and α as close to 0 as $0.1, 0.01, 0.05$, say.

Chebyshev's inequality tells us

$$Prob\left[|t - Y| \leq \lambda\sqrt{V(t}\right] \geq 1 - \frac{1}{\lambda^2} \tag{1.2}$$

for a positive number λ.

Relating (1.1) and (1.2) on taking $fY = \lambda\sqrt{V(t)}$ and $\alpha = \frac{1}{\lambda^2}$ and noting $CV(t) = 100\frac{\sqrt{V(t)}}{Y}$ (taking $Y > 0$) which is the coefficient of variation of t, we may write

$$100f = \frac{CV(t)}{\sqrt{\alpha}} \tag{1.3}$$

This (1.3) may help us in recommending an appropriate size n of a sample to choose in a specific sample selection situation, we may illustrate below.

2.1 Sample-Size to Choose in Direct Surveys with SRSWR and SRSWOR

For SRSWR with $N\bar{y}$ to estimate Y or $\bar{y}$ to estimate $\bar{Y} = \frac{Y}{N}$, $V(\bar{y}) = \frac{\sigma^2}{n} = \frac{N-1}{Nn}S^2$ writing $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2 = \frac{N-1}{N}S^2$ implying $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$.

The coefficient of variation of $\bar{y}$ is $CV(\bar{y}) = 100\sqrt{\frac{N-1}{Nn}}\frac{S}{\bar{Y}}$.

Writing $CV = 100\frac{S}{\bar{Y}}$, the coefficient of variation of all the N values of y_i's in the population. A rule to choose the sample-size is

$$n = \frac{(N-1)(CV)^2}{N\alpha f^2} \quad \text{(using (1.3))} \tag{2.1}$$

For SRSWOR in n draws, if we estimate $\bar{Y}$ by the sample mean $\bar{y}$, then $V(\bar{y}) = \frac{N-n}{Nn}S^2$ and so, using (1.3) an appropriate 'sample-size fixing rule' gives

$$n = \frac{N}{1 + N\alpha f^2\left(\frac{100}{CV}\right)^2} \tag{2.2}$$

as both (2.1) and (2.2) may be checked from Chaudhuri and Dutta (2018) and also Chaudhuri (2020).

Choosing N, f, α and values of CV and following (2.1) the values of n for an SRSWR may be worked out 'rounding it up' to the nearest positive integer. Similarly, for SRSWOR, specifying N, f, α and CV the appropriate sample size n may be chosen using (2.2) above rounding it up to the least positive integer.

Thus one may construct the Table 1 below.

Table 1
Sample Sizes for SRSWR and SRSWOR

N	f	α	CV	n by eq(2.1)	n by eq(2.2)
80	0.1	0.05	0.1	20	16
60	0.1	0.05	0.08	13	11
100	0.1	0.05	0.1	20	17
50	0.1	0.05	0.05	5	5

For both SRSWR and SRSWOR the ratios $\frac{n}{N}$ look reasonable. These results relate to Direct Response (DR) surveys. Now let us turn to RR surveys.

2.2 Sample-Size Determination in case of SRSWR and SRSWOR sampling when (i) Warner's RR method with qualitative stigmatizing characteristics is employed and (ii) when an RR method with quantitative stigmatizing features is employed as given by Chaudhuri (2011)

(i) Warner's (1965) RR survey method prescribes an interviewer to approach a sampled person i $(in\ U)$ with a box of identical cards with a proportion p $(0 < p < 1, p \neq \frac{1}{2})$ marked A and the rest A^c, requesting him/her to randomly choose one and to respond

$I_i = 1$ if i's feature 'matches' the card type
$= 0$ if it 'does not match'.

Then, writing E_R, V_R as expectation, variance operators for the RR device, one gets

$$E_R(I_i) = py_i + (1-p)(1-y_i)$$

with $y_i = 1$ if i bears A
$= 0$ if i bears A^c.

Then, $r_i = \frac{I_i-(1-p)}{(2p-1)}$ has $E_R(r_i) = y_i$ and $V_R(r_i) = \frac{p(1-p)}{(2p-1)^2}$ $\forall$ i in U.

If, from U an SRSWR is taken in n draws, then $\bar{r}$, the sample mean of the r_i's has

$E_R(\bar{r}) = \bar{y}$, the sample mean of the y_i's and $V_R(\bar{r}) = \frac{p(1-p)}{n(2p-1)^2}$.

Writing E_P, V_P as design based expectation, variance and the overall expectation, variance operators as $E = E_P E_R = E_R E_P$ and $V = E_P V_R + V_P E_R = E_R V_P + V_R E_P$, we get

$E(\bar{r}) = \bar{Y} = \theta$, say, and

$$V(\bar{r}) = \frac{1}{nN}\sum_{i=1}^{N}(y_i - \bar{Y})^2 + \frac{p(1-p)}{n(2p-1)^2} = \frac{1}{n}\left[\frac{p(1-p)}{(2p-1)^2} + \frac{\theta(1-\theta)}{N}\right] \qquad (2.2.1)$$

Similarly, if RR data by Warner's technique (RRT) are gathered by SRSWOR in n draws, then $\bar{r}$ has

$E(\bar{r}) = \bar{Y} = \theta$, say and

$$V(\bar{r}) = \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(y_i - \bar{Y})^2 + \frac{p(1-p)}{n(2p-1)^2}$$
$$= \frac{1}{n}\left[\frac{p(1-p)}{(2p-1)^2} + \frac{N}{N-1}\theta(1-\theta)\right] - \frac{\theta(1-\theta)}{N-1} \qquad (2.2.2)$$

on noting $y_i^2 = y_i$ because $y_i = 1\ or\ 0$ and $\sum_{i=1}^{N} y_i^2 - N\bar{Y}^2 = N\theta(1-\theta)$.

The equation (1.3) gives in these two cases $100f = \frac{1}{\sqrt{\alpha}}\frac{\sqrt{V(\bar{r})}}{\bar{Y}} = \frac{1}{\sqrt{\alpha}}\frac{\sqrt{V(\bar{r})}}{\theta}$.

Then, (2.2.1) gives

$$100f = \frac{1}{\sqrt{\alpha}}\frac{1}{\theta}\left[\frac{1}{n}\left\{\frac{p(1-p)}{(2p-1)^2} + \frac{\theta(1-\theta)}{N}\right\}\right]^{1/2}$$

for SRSWR and

$$100f = \frac{1}{\sqrt{\alpha}}\frac{1}{\theta}\left[\frac{1}{n}\left\{\frac{p(1-p)}{(2p-1)^2} + \frac{N}{N-1}\theta(1-\theta)\right\}\right]^{1/2} \tag{2.2.3}$$

for SRSWOR

So, for SRSWR the rule is

$$n = \frac{\left[\frac{p(1-p)}{(2p-1)^2} + \frac{\theta(1-\theta)}{N}\right]}{(100f)^2\theta^2\alpha} \tag{2.2.4}$$

and for SRSWOR the rule is

$$n = \frac{\left[\frac{p(1-p)}{(2p-1)^2} + \frac{N}{N-1}\theta(1-\theta)\right]}{(100f)^2\theta^2\alpha + \frac{\theta(1-\theta)}{N-1}} \tag{2.2.5}$$

Though $\theta = \bar{Y}$ is the estimand parameter and can never be known, to get an insight into the possibilities of making a rational choice of n, one may construct a table choosing N, f, α, θ and p as the Table 2 below.

Table 2

Choosing n in SRSWR and SRSWOR in an RR Survey by Warner's RRT

N	f	α	θ	p	n by (2.2.4)	n by (2.2.5)
80	0.1	0.05	0.2	0.55	124	123
60	0.1	0.05	0.2	0.55	124	123
100	0.1	0.05	0.2	0.55	124	123
50	0.1	0.05	0.2	0.55	124	122

The results for n are awful.

(ii) Turning to the case of quantitative stigmatizing feature like duration of a stay behind the bar for a criminal punishment let us consider Chaudhuri's (2011) RR device as follows.

Let an investigator approach a sampled respondent i in U with two boxes with the 1st containing similar cards bearing numbers $a_1, a_2, \ldots a_j, \ldots, a_T$ with mean $\mu_a = \frac{1}{T}\sum_{j=1}^{T} a_j \neq 0$ and variance $\sigma_a^2 = \frac{1}{T-1}\sum_{j=1}^{T}(a_i - \mu_a)^2$ and the other with cards bearing numbers $b_1, \ldots, b_k, \ldots, b_M$ with mean $\mu_b = \frac{1}{M}\sum_{k=1}^{M} b_k$ and variance $\sigma_b^2 = \frac{1}{M-1}\sum_{k=1}^{M}(b_k - \mu_b)^2$.

The person i on request is to randomly draw a card from box 1, say found as a_j and independently draw randomly from the other box a card, found, say, as labelled b_k. Then, if he/she bears the value y_i, say, of the variable of interest as say y, then the RR from i is to be recorded as $z_i = a_j y_i + b_k$.

Then, $E_R(z_i) = \mu_a y_i + \mu_b$ and $V_R(z_i) = \sigma_a^2 y_i^2 + \sigma_b^2$.

Then, $r_i = \frac{z_i - \mu_b}{\mu_a}$ has $E_R(r_i) = y_i$ and $V_R(r_i) = y_i^2 \frac{\sigma_a^2}{\mu_a^2} + \frac{\sigma_b^2}{\mu_a^2} = V_i$, say.

Then, for an SRSWR in n draws for the sample mean $\bar{r}$, we have

$$E(\bar{r}) = \bar{Y} \text{ and } V(\bar{r}) = \frac{1}{n}\left[\frac{\sum_{i=1}^{N} V_i}{N} + \frac{(N-1)S^2}{N}\right]$$

So,

$$CV(\bar{r}) = \frac{100}{\sqrt{n}} \frac{\left[\frac{\sum_{i=1}^{N} V_i}{N} + \frac{(N-1)S^2}{N}\right]^{1/2}}{\bar{Y}}$$

So, (1.3) gives $100f = \frac{CV(\bar{r})}{\sqrt{\alpha}}$.

So,

$$(100f)^2 = \frac{100^2}{\alpha} \frac{\left[\frac{\sum_{i=1}^{N} V_i}{N} + \frac{(N-1)S^2}{N}\right]}{n\bar{Y}^2}$$

So, the rule for n is

$$n = \frac{\left[\left(\frac{100}{\bar{Y}}\right)^2 \frac{\sum_{i=1}^{N} V_i}{N} + \frac{N-1}{N}(CV)^2\right]}{(100f)^2 \alpha} \tag{2.2.6}$$

If, again the RR survey data is gathered by SRSWOR in n draws, then

$$V(\bar{r}) = \frac{N-n}{Nn} S^2 + \frac{1}{n} \frac{\sum_{i=1}^{N} V_i}{N}.$$

So,

$$[CV(\bar{r})]^2 = \left(\frac{100}{\bar{Y}}\right)^2 \left[\frac{N-n}{Nn} S^2 + \frac{1}{n} \frac{\sum_{i=1}^{N} V_i}{N}\right]$$

So, (1.3) gives the rule for n as

$$\begin{aligned}
(100f)^2 \alpha &= [CV(\bar{r})]^2 \\
&= \left(\frac{100}{\bar{Y}}\right)^2 \left[\frac{N-n}{Nn} S^2 + \frac{1}{n} \frac{\sum_{i=1}^{N} V_i}{N}\right] \\
&= \frac{1}{n}\left[\left(\frac{100}{\bar{Y}}\right)^2 \frac{\sum_{i=1}^{N} V_i}{N} + \left(\frac{N-n}{N}\right)(CV)^2\right] \\
&= \frac{1}{n}\left[\left(\frac{100}{\bar{Y}}\right)^2 \frac{\sum_{i=1}^{N} V_i}{N} + (CV)^2\right] - \frac{(CV)^2}{N}
\end{aligned}$$

or

$$n = \frac{\left[\left(\frac{100}{\bar{Y}}\right)^2 \frac{\sum_{i=1}^{N} V_i}{N} + (CV)^2\right]}{(100f)^2\alpha + \frac{(CV)^2}{N}} \tag{2.2.7}$$

So, to choose n for SRSWR and for SRSWOR the table to use is

Table 3
Choosing n in SRSWR and SRSWOR in RR Survey by Chaudhuri (2011) Device using (2.2.6) and (2.2.7)

N	f	α	$\bar{Y}$	CV	n by (2.2.6)	n by (2.2.6)
80	0.1	0.05	36	10	861	689
60	0.1	0.05	36	10	844	634
100	0.1	0.05	36	10	829	691
50	0.1	0.05	36	10	851	608

Looking at Tables 1, 2 and 3 we may observe that we have constructed Table 1 very easily employing only some arbitrary but reasonable values of CV to get quite appropriate choice of n which naturally turns out slightly larger for SRSWR than for SRSWOR. In Tables 2 and 3 also n for SRSWR exceeds that for SRSWOR. But their construction is different because we need values of $\theta = \bar{Y}$ and in addition we need CV for the quantitative case.

The sample-sizes for RR come out as absurd for the present approach based on Chebyshev's procedure though they are reasonable for DR. To circumvent this anomaly we may offer the following remedy.

For the RR's each variance of the estimator for the estimand parameter like the proportion, total or mean there are two terms: I for variance of RR's and II for variance of DR based values. In magnitude, I far exceeds II, vide Chaudhuri and Sen (2020). But I has a little relation to the sample selection procedure. So, in practice we should choose the sample-size directly to control the magnitude of II by the procedure based on Chebyshev's ideas and then examine how reasonable the magnitude of I turns out vis-a-vis the choice of sampling design and the sample-size.

In the present paper let us see how some other RR devices fare keeping the design fixed only as SRSWR and SRSWOR. In a separate paper we shall pursue with other sampling methods with several RR devices.

i) Simmons's URL procedure of RRT

The interviewer approaches a sampled person with two boxes, one containing cards marked A and B in proportions $p_1 : (1 - p_1) (0 < p_1 < 1)$ and the other the same in proportions $p_2 : (1 - p_2), (0 < p_2 < 1$ with $p_1 \neq p_2)$. Then, the person i drawing independently one card at random from each box is to respond as

$I_i = 1$ if the card type 'matches' his/her feature A or B from the 1st box
$= 0$ if it 'does not match'

and

$J_i = 1$ if 'match' for the 2nd box

$= 0$ if 'no match'.

Then, $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ has $E_R(r_i) = y_i$ and
$V_R(r_i) = \frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}(y_i - x_i)^2 = V_i$, say.

Here,

$x_i = 1$ if i bears B

$= 0$ if i bears B^c (complement of B).

B is an innocuous feature unrelated to A, the sensitive feature.

If an SRSWR in n draws is taken and RR's are observed, then $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$ has $E_R(\bar{r}) = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $E(\bar{r}) = \bar{Y} = \theta$ and $V(\bar{r}) = \frac{1}{n}[\frac{\theta(1-\theta)}{N} + \frac{\sum_{i=1}^{N} V_i}{N}]$ with V_i as above.

So, applying (1.3) one may derive n as in (2.2.6).

If an SRSWOR in n draws is taken and RR survey data are gathered, $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$ may be used to unbiasedly estimate $\bar{Y} = \theta$ and derived $V(\bar{r}) = \frac{N-n}{Nn}S^2 + \frac{\sum_{i=1}^{N} V_i}{nN}$ with V_i as above.

Then, by (1.3) a rule for n is as in (2.2.7).

So,

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \frac{\sum_{i=1}^{N} V_i}{N} + \frac{N-1}{N}(CV)^2}{(100f)^2\alpha} \qquad \text{for SRSWR and}$$

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \frac{\sum_{i=1}^{N} V_i}{N} + (CV)^2}{(100f)^2\alpha + \frac{(CV)^2}{N}} \qquad \text{for SRSWOR .}$$

Since V_i is difficult to anticipate in both the above formulae for n, V_i should be replaced by its unbiased estimator $v_i = r_i(r_i - 1)$ and $\sum_{i=1}^{N} V_i$ by $\frac{1}{n_0}\sum_{1}^{n_0} r_i(r_i - 1)$ in both the above formulae for n. Here by n_0 we mean an arbitrary size of an SRSWR taken to get the RR data as above to find r_i and hence use $\frac{1}{n_0}\sum_{1}^{n_0} r_i(r_i - 1)$, $\sum_{1}^{n_0}$ denoting sum over the n_0 values of $r_i(r_i - 1)$. For SRSWR and SRSWOR we use the same notation and in the above two formulae for n for SRSWR and n for SRSWOR. We simply replace $\frac{\sum_{i=1}^{N} V_i}{N}$ by $\frac{1}{n_0}\sum_{1}^{n_0} r_i(r_i - 1)$ which is calculated respectively for an SRSWR of size n_0 and SRSWOR of size n_0, using the realized values of r_i for them.

So, finally,

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \frac{\sum_1^{n_0} r_i(r_i - 1)}{n_0} + \frac{N-1}{N}(CV)^2}{(100f)^2\alpha} \qquad (i)$$

and

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \frac{\sum_1^{n_0} r_i(r_i - 1)}{n_0} + (CV)^2}{(100f)^2\alpha + \frac{(CV)^2}{N}} \qquad (ii).$$

Then we construct Table 4.

Table 4
Sample Size for URL by SRSWR and SRSWOR

N	f	α	θ	CV	p_1	p_2	n_0	n by (i)	n by (ii)
80	0.1	0.05	0.2	10	0.44	0.52	10	630019	504016
60	0.1	0.05	0.2	10	0.44	0.52	8	969019	675015
100	0.1	0.05	0.2	10	0.44	0.52	12	840019	583350
50	0.1	0.05	0.2	10	0.44	0.52	6	700019	250014

ii) Kuk's (1990) RRT

The interviewer approaches a sampled person i with two boxes, one with a proportion $\theta_1 (0 < \theta_1 < 1)$ of cards marked 'Red' and the rest 'Non-red' and the other with the same in proportion $\theta_2 (0 < \theta_2 < 1, \theta_1 \neq \theta_2)$ 'Red' and the rest 'Non-red'. The i^{th} person is asked to draw $k (> 1)$ cards by SRSWR from the 1st box if he/she bears the characteristics A or from the 2nd box if he/she bears A^c. The RR he/she is to give out is $'f_i'$, which is the number of 'Red' cards drawn.

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

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

Then, $r_i(k) = \frac{\frac{f_i}{k} - \theta_2}{\theta_1 - \theta_2}$ has $E_R\big(r_i(k)\big) = y_i$ and

$V_R\big(r_i(k)\big) = V_i(k)$, say

$= b_i(k)y_i + c_i(k)$, writing $b_i(k) = \frac{1 - \theta_1 - \theta_2}{k^2(\theta_1 - \theta_2)^2}$ and $c_i(k) = \frac{\theta_2(1 - \theta_2)}{k^2(\theta_1 - \theta_2)^2}$.

If an SRSWR in n draws is drawn producing such RR's as $r_i(k)$, then $\bar{r}(k) = \frac{1}{n}\sum_{i=1}^n r_i(k)$ unbiasedly estimates $\theta = \bar{Y} = \frac{1}{N}\sum_{i=1}^N y_i$ have the variance $V\big(\bar{r}(k)\big) = \frac{1}{n}\left[\frac{\theta(1-\theta)}{N} + \frac{\sum_{i=1}^N V_i(k)}{N}\right]$.

If, instead, an SRSWOR in n draws is taken producing the RR's as $r_i(k)$'s, then $\bar{r}(k) = \frac{1}{n}\sum_{i=1}^n r_i(k)$ would unbiasedly estimate $\theta = \bar{Y}$ with a variance as $V\big(\bar{r}(k)\big) = \frac{N-n}{Nn}S^2 + \frac{1}{nN}\sum_{i=1}^N V_i(k)$.

Then, applying (1.3), for SRSWR appropriate sample size is

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \left(\frac{\sum_{i=1}^{N} V_i(k)}{N}\right) + \left(\frac{N-1}{N}\right)(CV)^2}{(100f)^2\alpha} \qquad \ldots (iii)$$

Similarly, for SRSWOR it is

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \left(\frac{\sum_{i=1}^{N} V_i(k)}{N}\right) + (CV)^2}{(100f)^2\alpha + \frac{(CV)^2}{N}} \qquad \ldots (iv)$$

A usable Table then is, taking $k = 3$

Table 5
Sample-Size for Kuk by SRSWR and SRSWOR

N	f	α	θ	CV	p_1	p_2	n_0	n by (iii)	n by (iv)
80	0.1	0.05	0.2	10	0.44	0.52	10	665519	521127
60	0.1	0.05	0.2	10	0.44	0.52	8	643569	505483
100	0.1	0.05	0.2	10	0.44	0.52	12	636319	535124
50	0.1	0.05	0.2	10	0.44	0.52	6	636359	450371

In (iii) and (iv) also $\frac{\sum_{i=1}^{N} V_i(k)}{N}$ should be replaced by $\frac{1}{n_0}\sum_1^{n_0} V_i(k)$ just as in URL.

iii) Forced Response RRT

In applying this device the interviewer approaches a sampled person i in $U = (1, \ldots, i \ldots, N)$ with a box of a large number of cards marked 'Yes', 'No' and 'Genuine' in proportions p_1, p_2 and $(1 - p_1 - p_2)$ respectively $(0 < p_1, p_2 < 1$, $p_1 \neq p_2$ and $1 - p_1 - p_2 > 0$. The sampled person i is to give the RR I_i as 'Yes', 'No' or genuinely as 1 or 0 if his/her feature actually is stigmatizing or not. A 'Yes' or 'No' response is recorded respectively as 1 or 0.

Then, $r_i = \frac{I_i - p_1}{1 - p_1 - p_2}$ has $E_R(r_i) = y_i$ and $V_R(r_i) = \frac{p_1(1-p_1) + y_i(1-p_1-p_2)(p_2-p_1)}{(1-p_1-p_2)^2} = V_i$, say, which equals $\frac{p_1(1-p_1)}{(1-p_1-p_2)^2}$ if $y_i = 0$ and $\frac{p_2(1-p_2)}{(1-p_1-p_2)^2}$ if $y_i = 1$.

If an SRSWR is taken in n draws, then $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$ is unbiased for $\theta = \bar{Y}$. Also, $V(\bar{r}) = \frac{1}{n}\left[\frac{\theta(1-\theta)}{N} + \frac{1}{N}\sum_{i=1}^{N} V_i\right]$ with V_i as above.

If, instead, an SRSWOR is taken and RR's are derived as above, the same $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$ is an unbiased estimator for $\theta = \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. But its variance is $V(\bar{r}) = \frac{N-n}{Nn}S^2 + \frac{1}{Nn}\sum_{i=1}^{N} V_i = \frac{N-n}{n(N-1)}\theta(1-\theta) + \frac{1}{Nn}\sum_{i=1}^{N} V_i$.

Applying (1.3) appropriate n's for SRSWR and SRSWOR respectively

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \left(\frac{\sum_{i=1}^{N} V_i}{N}\right) + \left(\frac{N-1}{N}\right)(CV)^2}{(100f)^2\alpha} \qquad \ldots (v)$$

$$n = \frac{\left(\frac{100}{\theta}\right)^2 \left(\frac{\sum_{i=1}^{N} V_i}{N}\right) + (CV)^2}{(100f)^2\alpha + \frac{(CV)^2}{N}} \qquad \ldots(vi)$$

Here also as in the case of URL, the term $\frac{\sum_{i=1}^{N} V_i}{N}$ may be estimated by $\frac{1}{n_0}\sum_1^{n_0} v_i$ taking $v_i = p_1(1-p_1) + r_i(1-p_1-p_2)(p_2-p_1)$ as an unbiased estimator for V_i on taking an initial SRSWR and another SRSWOR of a sample size n_0 and with RR's found as above in that.

Then (v) and (vi) may be replaced by (v') and (vi') respectively with $\frac{\sum_{i=1}^{N} V_i}{N}$ replaced by $\frac{1}{n_0}\sum_1^{n_0} v_i$ therein.

Table 6
Sample size for Forced Response by SRSWR and SRSWOR

N	f	α	θ	CV	p_1	p_2	n_0	n by (v)	n by (vi)
80	0.1	0.05	0.2	10	0.44	0.52	10	7850020	6880016
60	0.1	0.05	0.2	10	0.44	0.52	8	7850020	6121890
100	0.1	0.05	0.2	10	0.44	0.52	12	9100020	6367933
50	0.1	0.05	0.2	10	0.44	0.52	6	8266520	5904657

3. CONCLUSION

Tables 4, 5 and 6 clearly reveal that sample sizes cannot be fixed in case of RR surveys by our approach utilizing Chebyshev's theorem as we could ingeniously achieve good results in DR surveys by SRSWR and SRSWOR. In a separate ensuing paper we shall reveal analogous results with varying probability sampling as well.

REFERENCES

1. Chaudhuri, A. (2020). A Review on Issues of Settling the Sample-size in Surveys: Two Approaches—Equal and Varying Probability Sampling—Crises in Sensitive Cases. *CSA Bulletin*, 72(1), 7-16.
2. Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: CRC Press.
3. Chaudhuri, A. and Dutta, T. (2018). Determining the size of a sample to take from a finite population. *Statistics and Applications,* 16(1), 37-44.
4. Chaudhuri, A. and Sen, A. (2020). Fixing the Sample-Size in Direct and Randomized Response Surveys. *Journal of Indian Society of Agricultural Statistics,* 74(3), 201-208.
5. Kuk, A.Y. (1990). Asking sensitive questions indirectly. *Biometrika,* 77(2), 436-438.
6. Warner, S.L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of American Statistical Association*, 60, 63-69.