Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 77(1) 2023 19–26

Fixing Size of a Varying Probability Sample in a Direct and a Randomized Response Survey

Arijit Chaudhuri[1] and Dipika Patra[2]

[1]*Indian Statistical Institute, Kolkata*
[2]*Seth Anandram Jaipuria College, Kolkata*

Received 06 March 2023; Accepted 27 March 2023

SUMMARY

For a Direct Survey on innocuous characteristics Chebyshev's inequality is helpful in prescribing the size of a sample in a survey. An extension of the same to cover stigmatizing features in Randomized Response (RR) survey is not smooth enough. Different situations are illustrated and solutions proposed.

AMS subject classification: 62D05.

Keywords: Randomized response survey; Sample-size fixation; Varying probability sampling.

1. INTRODUCTION

Our main concern here is to unbiasedly estimate the proportion of people in a community bearing a specific stigmatizing characteristic A, say, like criminal propensities, alcoholism, intoxicating drug habits and similar qualitative features or to estimate total or average expenses incurred because of such sensitive experiences like costs of treatment of AIDS, loss in gambling, paying fines for fraudulent conviction, income loss due to confinement in jail etc. A stigmatizing variable y will take real value y_i which may be simply 1 or 0 for a person i in a population $U = (1, 2, .., i, ..N)$ bearing a sensitive feature A or its complement A^c. The total $Y = \sum_{i=1}^{N} y_i$ or mean $\bar{Y} = \frac{Y}{N}$ is our estimated parameter of interest. A sample s from U of a 'suitable size n' is to be chosen according to a design p assigning a value $p(s)$ to s. It is to be surveyed gathering directly (called a Direct Response or DR survey) or by a Randomized Response (RR) Technique (RRT). Simplest design is SRSWR (Simple Random Sampling With Replacement) with its variant SRSWOR (Simple Random Sampling Without Replacement). Here we shall deal with more complex sampling designs, namely, PPSWR (Probability Proportional to Size With Replacement), IPPS (Inclusion Probability Proportional to Size) and RHC (Rao, Hartley and Cochran's) sampling scheme. Corresponding estimation procedures given by Hansen and Hurwitz (HH), Horvitz and Thompson (HT) and by RHC themselves will be described in Section 3 below. In Section 2, we describe a few RRTs we choose to deal with in this paper. Our main concern is of course to discuss how to prescribe sample size in respective sampling designs to be followed in DR and RR surveys.

2. A FEW ILLUSTRATIVE RR DEVICES

2.1 Warner's RR Device

Warner (1965) as the pioneer concerning RRT's prescribed essentially that an interviewer is to obtain an RR from a sampled person i of U as

$I_i = 1$ if a 'match' results in his/her feature A or A^c when he/she draws randomly from a pack of cards

Corresponding author: Arijit Chaudhuri
E-mail address: arijitchaudhuri1@rediffmail.com

offered containing a large number of cards marked A or A^c in proportions $p:(1-p)$, $(0<p<1, p\neq\frac{1}{2})$

$=0$ if there is 'no' match.

Writing E_R, V_R generically as expectation, variance operators,

$$E_R(I_i)=py_i+(1-p)(1-y_i),\ i\in U$$

$$V_R(I_i)=E_R(I_i^2)-E_R^2(I_i)=E_R(I_i)(1-E_R(I_i))=p(1-p),$$

since $I_i^2=I_i$ and $y_i^2=y_i$.

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.

2.2 Simmons's Unrelated Model or URL RRT

Here an RR emerges from a sampled person i of U as

$I_i=1$ if there is a 'match' in i's true characteristics namely, the stigmatizing A or an unrelated innocuous feature B when he/she on request randomly draws a card from a pack of cards marked A or B in proportions $p_1:(1-p_1)$, $(0<p_1<1,)$

$=0$ if there is 'no' match.

Another independent RR from i emerges as

$J_i=1$ if there is a 'match' when he/she on request draws similarly a card from second box with cards A and B in proportions $p_2:(1-p_2)$, $0<p_2<1$ but $p_1\neq p_2$.= 0 if there is no 'match'

Then, $r_i=\frac{p_2I_i-p_1J_i}{p_1-p_2}$ has $E_R(r_i)=y_i$ and

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

$x_i=1$ if i bears B

$=0$ if i bears B^c, the complement of B.

2.3 Kuk's RRT

Here the interviewer derives the RR from a sampled person i from U as f_i which is the number of red cards drawn from either a box with red and non-red cards in proportions $\theta_1:(1-\theta_1)$, $0<\theta_1<1$ if i bears A or he/she bears A^c, then from another similar box with the red: non-red in proportions $\theta_2:(1-\theta_2), \theta_1\neq\theta_2$ on choosing $k(>1)$ cards from either box by SRSWR.

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 $E_R(r_i(k))=y_i$ and $V_R(r_i(k))=V_i(k), say$

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

2.4 Forced Response RRT

Here the interviewer approaches a sampled person i from U with a box of large number of cards respectively marked 'Yes', 'No' and 'Genuine' in respective proportions p_1, p_2 and $(1-p_1-p_2)$, $0<p_1, p_2<1, p_1+p_2<1, p_1\neq p_2$ and on request he/she is to respond.

$I_i=1$ if he/she randomly draws a card marked 'Genuine' and his/her feature is A or he/she randomly chooses a card marked 'Yes'

$=0$ if he/she draws a card marked 'No' or he/she draws a card marked 'Genuine' and he/she bears A^c.

Then, $r_i=\frac{I_i-p_1}{1-p_1-p_2}$ has $E_R(r_i)=y_i$ and

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

2.5 Eriksson's RRT

Here the interviewer approaches a sampled person i from U with a proportion, $say, C\ (0<C<1)$ of cards marked 'Correct' and the remaining cards bear a real number $z_1, z_2, \ldots, z_m$ with known proportions $q_1, q_2, \ldots, q_m$ respectively such that $\sum_{j=1}^{m}q_j=1-C\ (0<q_j<1\forall j)$.

$I_i=y_i$ with probability C

$=z_j$ with probability q_j

Then, $r_i=\frac{I_i-\sum_{j=1}^{m}q_jz_j}{C}$ has $E_R(r_i)=y_i$ and $V_R(r_i)=\frac{1}{C^2}V_R(I_i)=ay_i^2+by_i+L$, where a, b, L are known constants.

3. A FEW ILLUSTRATIVE VARYING PROBABILITY SAMPLING SCHEMES

3.1 PPSWR (Probability proportional to size with replacement) sampling

Let x_i $(>0\,\forall i)$ denote size-measures of the units i of U, supposed to be well and positively correlated with the y_i values and $X=\sum_{i=1}^{N}x_i$, $p_i=\frac{x_i}{X}$, called the normed size-measures of the units i and be all known to the investigator. Then, in n independent draws from U the units i are selected with probabilities $p_i, i=1,2\ldots N$. Then, for a Direct survey we have the Hansen and Hurwitz (1943) unbiased estimator for $Y=\sum_{i=1}^{N}y_i$ as $t_{HH}=\frac{1}{n}\sum_{k=1}^{n}\frac{y_k}{p_k}$, denoting by y_k, p_k the values of y_i, p_i for the unit chosen on the k^{th} draw, $k=1,2\ldots n$.

Then, $V(t_{HH})=\frac{1}{n}\left(\sum_{i=1}^{N}\frac{y_i^2}{p_i}-Y^2\right)$ and

$v(t_{HH})=\frac{1}{2n^2(n-1)}\sum_{k\neq}^{n}\sum_{k'}^{n}(\frac{y_k}{p_k}-\frac{y_{k'}}{p_{k'}})^2$ has $E_p v(t_{HH})=V(t_{HH})$.

By E_p, V_p we shall denote generically the sampling based expectations, variance operators and by $E=E_PE_R=E_RE_P$ and $V=E_PV_R+V_PE_R=E_RV_P+V_RE_P$, the overall expectation, variance operators.

Using RR-survey data corresponding to t_{HH} an unbiased estimator for Y is

$e_{HH}=\frac{1}{n}\sum_{k=1}^{n}\frac{r_k}{p_k}$, denoted by r_k, the quantity generically, the value of r_i for the unit i chosen on the k^{th} draw.

Then, for Warner's RRT

$$V(e_{HH})=V(t_{HH})+\frac{p(1-p)}{n(2p-1)^2}\sum_{i=1}^{N}\frac{1}{p_i},$$

writing V_k for V_i for the unit i chosen on the k^{th} draw.

3.2 IPPS (Inclusion Probability Proportional to Size) sampling

Brewer and Hanif (1983) and Chaudhuri and Vos (1988) have narrated numerous IPPS sampling schemes. Here we shall consider only the following IPPS scheme in particular. Let z_i be certain known positive numbers and a unit i of U on the 1st draw be selected with a probability proportional to z_i, on the second draw a unit j of U other than i be selected with a probability proportional to z_j and out of the remaining $(N-2)$ units an SRSWOR in $(n-2)$ draws be chosen. Then the selection-probability of such a sample s of size n is

$$p(s)=\frac{z_i}{Z}\frac{z_j}{Z-z_i}\frac{1}{\binom{N-2}{n-2}}, \text{ where } Z=\sum_{i=1}^{N}z_i .$$

Then, the inclusion probability of i in such a sampling scheme is as follows.

First, in the 1st two draws inclusion probability of i is

$\pi_i(2)=\frac{z_i}{Z}+\sum_{i\neq j}^{N}\frac{z_j}{Z-z_j}\frac{z_i}{Z}$ and hence $\pi_i(n)=\pi_i(2)+(1-\pi_i(2))\frac{n-2}{N-2}$ is the inclusion probability in the entire sample of size n by this scheme.

Writing $Q_i=\frac{z_i}{Z}, Q_j=\frac{z_j}{Z-z_j}$, we have $\pi_{ij}(2)=\frac{Q_iQ_j}{1-Q_i}+\frac{Q_iQ_j}{1-Q_j}$ as the inclusion probability of i and j in the 1st two draws in this scheme and in the entire sample of size n in this scheme the inclusion probability of i and j both is

$$\pi_{ij}(n)=\pi_{ij}(2)+\left(\frac{n-2}{N-2}\right)(\pi_i(2)+\pi_j(2)-2\pi_{ij}(2))+\left(\frac{n-2}{N-2}\right)\left(\frac{n-3}{N-3}\right)(1-\pi_i(2)-\pi_j(2)+\pi_{ij}(2)).$$

If $\pi_i(n)$ here is equated to np_i, then the above scheme is an IPPS sampling scheme.

For any sampling scheme with the inclusion probability of i as π_i $(>0\, and \sum_{i=1}^{N}\pi_i=n)$ an unbiased estimator for Y is the Horvitz- Thompson (HT) estimator

$$t_{HT}=\sum_{i\in s}\frac{y_i}{\pi_i} .$$

Assuming every sample s has only distinct units and the number of units in s is a fixed number then

$$V_P(t_{HT})=\sum_{i<}^{N}\sum_{j=1}^{N}(\pi_i\pi_j-\pi_{ij})(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j})^2 .$$

For an RR survey, $e_{HT} = \sum_{i \in s} \frac{r_i}{\pi_i}$, with $E_R(r_i) = y_i$ is unbiased for Y in the sense

$$E(e_{HT}) = E_P(t_{HT}) = E_R(\sum_{i=1}^{N} r_i) = Y \text{ and}$$

$$V(e_{HT}) = V_P(t_{HT}) + \sum_{i=1}^{N} \frac{V_i}{\pi_i}.$$

3.3 Rao, Hartley and Cochran (RHC) sampling scheme

Here the population $U = (1, 2, \ldots i..N)$ is randomly split up into n disjoint parts by taking an SRSWOR of N_1 units forming the 1st group and then successively taking $(n-1)$ more SRSWOR's mutually exclusively of sizes $N_2, N_3, \ldots., N_n$ such that $\sum_n N_i = N$, $\sum_n$ denoting sum over the n disjoint groups thus formed. Then, p_{ij} values of p_i's for the respective groups $i = 1, 2 \ldots n$ are noted and from each of the n groups one unit j of the N_i units is chosen with the probability $\frac{p_{ij}}{Q_i}$, where $Q_i = \sum_{j=1}^{N_i} p_{ij}$ and this is independently repeated for all the n groups. Then,

$$t_{RHC} = \sum_n y_{ij} \frac{Q_i}{p_{ij}}$$

is taken as an unbiased estimator for $Y = \sum_n Y_i, Y_i = \sum_{j=1}^{N_i} y_{ij}$.

Then, it follows that

$$V(t_{RHC}) = \frac{\sum_n N_i^2 - N}{N(N-1)} \sum_n \sum_n p_i p_j (\frac{y_i}{p_i} - \frac{y_j}{p_j})^2$$

$$\text{and } v(t_{RHC}) = \frac{\sum_n N_i^2 - N}{N^2 - \sum_n N_i^2} \sum_n \sum_n Q_i Q_j (\frac{y_i}{p_i} - \frac{y_j}{p_j})^2$$

is an unbiased estimator of $V(t_{RHC})$.

A suitable choice of the N_i's is $N_i = \left[\frac{N}{n}\right]$, where $\left[\frac{N}{n}\right]$ is the integer part of N divided by n for $i = 1, \ldots m$ and $N_i = \left[\frac{N}{n}\right] + 1$ for $i = m+1, \ldots., n$ such that $\sum_{i=1}^{m} N_i + \sum_{i=m+1}^{n} N_i = N.$

For RR survey data r_i based on RHC an unbiased estimator for Y is

$$e_{RHC} = \sum_n r_{ij} \frac{Q_i}{p_{ij}} \text{ and}$$

$$v(e_{RHC}) = v(t_{RHC}) + \sum_n v_{ij} \frac{Q_i}{p_{ij}}.$$

4. FIXING SAMPLE-SIZE IN DR, RR SURVEY

Chaudhuri (2010, 2014, 2018, 2020), Chaudhuri and Dutta (2018), and Chaudhuri and Sen (2020) have proposed the following devices in sample-size specification.

Suppose t is an unbiased estimator for a finite population total Y and our intention is to choose t as so accurate that

$$Prob\left[|t - Y| < fY\right] \geq 1 - \alpha,$$

choosing f as proper fraction like $0.1, 0.2$ etc and α is a positive quantity so small as, say, $0.05, 0.01$ etc.

Chebyshev's inequality says

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

where λ is a positive number greater than 1. Combining these two inequalities we may take

$$fY = \lambda\sqrt{V(t)} \text{ and } \alpha = \frac{1}{\lambda^2}$$

giving us

$$100f = \frac{1}{\sqrt{\alpha}} CV(t) \qquad (I)$$

writing $CV(t) = 100\frac{\sqrt{V(t)}}{Y}$

which is the coefficient of variation of t.

In case an SRSWR or an SRSWOR is chosen and $N\bar{y}$ with $\bar{y}$ as the sample mean in n draws, in either case to unbiasedly estimate Y, then $V(N\bar{y})$ being equal to $N^2 \frac{\sigma^2}{n} = \frac{N^2}{n}\left(\frac{N-1}{N}\right) S^2$ for SRSWR, writing $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2, \bar{Y} = \frac{Y}{N}$, it is possible to use the formula (I) above to fix n vis-a-vis N, f and α on speculating magnitudes of $100\frac{S}{\bar{Y}}$, the co-efficient of variations from the N population values of y_i's $(i = 1, 2 \ldots N)$. But in case varying probability samples are surveyed to estimate Y, it is difficult to utilize such facilities to choose n. To circumvent this Chaudhuri and Dutta (2018) suggested postulating the following simple model connecting y variable with an auxiliary variable x, possibly well and positively correlated with y.

Let us introduce the model

$$y_i = \beta x_i + \epsilon_i, \quad i \in U \qquad (1)$$

with β as an unknown constant, ϵ_i's are independently distributed random variables with $E_m(\epsilon_i) = 0 \ \forall i$ and $V_m(\epsilon_i) = \sigma^2 x_i^g$ with $\sigma(>0)$, an unknown constant and g an unknown constant such that $0 \le g \le 2$.

Using (I) we note that we need

$$Prob\left[\left|t_{HH} - Y\right| \le fY\right] \ge 1 - \alpha = 1 - \frac{V(t_{HH})}{f^2 Y^2}.$$

Thus,

$$\alpha = \frac{V(t_{HH})}{f^2 Y^2}. \qquad (II)$$

From this we get no clue to fix n, the sample size. Chaudhuri and Dutta (2018), therefore, suggest taking

$$\alpha = \frac{E_m V(t_{HH})}{f^2 E_m(Y^2)} \qquad (III)$$

instead of (II), taking E_m as the expectation operator under the model we have postulated above as (1).

Similarly for t_{HT} and t_{RHC} the equation (III) above as an analogue to find a suitable clue for n.

Let us work out

$$E_m V(t_{HH}) = \frac{\sigma^2}{n}\left[X\sum_{i=1}^{N} x_i^{g-1} - \sum_{i=1}^{N} x_i^g\right]$$

$$E_m(Y^2) = \beta^2 X^2 + \sigma^2 \sum_{i=1}^{N} x_i^g.$$

Further restricting the Model (1) to suppose x has the density

$$f(x) = e^{-x}, x > 0,$$

it is easy to take a random sample of x_i values from this exponential density. Hence we may calculate the following table fixing n for PPSWR sampling to estimate Y in a Direct Survey

Table 1. Fixing n for PPSWR sampling in DR surveys

N	f	α	σ^2	β	g	n by (III)
80	0.1	0.05	2	15	1	18
60	0.1	0.05	2	15	1.5	19
100	0.1	0.05	2	15	2	17
50	0.1	0.05	2	15	0.5	19

Next we work out

$$E_m V(t_{HT}) = \frac{\sigma^2}{n}[X\sum_{i=1}^{N} x_i^{g-1} - n\sum_{i=1}^{N} x_i^g].$$

Hence analogously to (III) we tabulate Table 2, giving sample-size for estimating Y by HT estimator in a Direct Survey using (III) analogously as in HH estimator.

Table 2. Fixing n for HT estimator in DR surveys

N	f	α	σ^2	β	g	n by (III)
80	0.1	0.05	2	15	1	14
60	0.1	0.05	2	15	1.5	12
100	0.1	0.05	2	15	2	13
50	0.1	0.05	2	15	0.5	20

Next, for t_{RHC} we calculate

$$E_m V(t_{RHC}) = \sigma^2 \frac{\sum_n N_i^2 - N}{N(N-1)}\left[X\sum_{i=1}^{N} x_i^{g-1} - \sum_{i=1}^{N} x_i^g\right].$$

Hence, analogously to (III) we tabulate Table 3, giving sample size for estimating finite population total by RHC strategy in a Direct Survey.

Table 3. Fixing n for RHC strategy in DR surveys

N	f	α	σ^2	β	g	n by (III)
80	0.1	0.05	2	15	1	14
60	0.1	0.05	2	15	1.5	12
100	0.1	0.05	2	15	2	15
50	0.1	0.05	2	15	0.5	20

Table 1, 2 and 3 reveal that our approach of fixing sample-sizes in DR surveys employing varying probabilities sampling schemes is rather successful. The sampling fractions $\frac{n}{N}$'s are turning out quite elegant.

Let us now see what may happen in RR surveys.

Let us apply (III) to the situations

(i) PPSWR [Warner, URL, Kuk, Forced response, Eriksson's] with HH estimate

(ii) IPPS [Warner, URL, Kuk, Forced response, Eriksson's] with HT estimate

(iii) RHC[Warner, URL, Kuk, Forced response, Eriksson's] with RHC estimate

$$E_mV\left(e_{HH}|Warner\right)=E_mV\left(t_{HH}\right)+\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}\frac{p(1-p)}{(2p-1)^2}$$

$$=\frac{\sigma^2}{n}\left[X\sum_{i=1}^{N}x_i^{g-1}-\sum_{i=1}^{N}x_i^{g}\right]+\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}\frac{p(1-p)}{(2p-1)^2}$$

$$E_mV\left(e_{HH}|URL\right)=E_mV\left(t_{HH}\right)+\frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}E_m(y_i-x_i)^2$$

$$=\frac{\sigma^2}{n}\left[X\sum_{i=1}^{N}x_i^{g-1}-\sum_{i=1}^{N}x_i^{g}\right]+\frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}[(\beta-1)^2x_i^2+\sigma^2x_i^g]$$

$$E_mV\left(e_{HH}|Kuk\right)=E_mV\left(t_{HH}\right)+\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}[b_i(k)\beta x_i+c_i(k)]$$

$$E_mV\left(e_{HH}|Forced\ Response\right)=E_mV\left(t_{HH}\right)+\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}[\frac{p_1(1-p_1)+(1-p_1-p_2)(p_2-p_1)\beta x_i}{(1-p_1-p_2)^2}$$

$$E_mV\left(e_{HH}|Eriksson\right)=E_mV\left(t_{HH}\right)+E_m\left[\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}\left\{a\beta x_i^2+a\sigma^2x_i^g+b\beta x_i+L\right\}\right]$$

$$=\beta^2X^2+\sigma^2\sum_{i=1}^{N}x_i^g+\frac{1}{n}\left[a\left\{\beta\sum_{i=1}^{N}\frac{1}{p_i}x_i^2+\sigma^2\sum_{i=1}^{N}\frac{1}{p_i}x_i^g\right\}+b\sum_{i=1}^{N}\frac{1}{p_i}x_i+L\right]$$

$$E_m\left(Y^2\right)=\beta^2X^2+\sigma^2\sum_{i=1}^{N}x_i^g.$$

Fixing sample size in IPPS sampling and Horvitz-Thompson estimator in RR surveys

$$E_mV\left(e_{HT}|Warner\right)=E_mV\left(t_{HT}\right)+\sum_{i=1}^{N}\frac{1}{np_i}\frac{p(1-p)}{(2p-1)^2}$$

$$E_mV\left(e_{HT}|URL\right)$$

Table 4. Fixing sample-size in PPSWR sampling to estimate finite population total in RR surveys applying $(\boldsymbol{III})$

Warner's RR							
N	f	α	σ^2	β	g	p	n
80	0.1	0.05	2	15	1	0.35	66
60	0.1	0.05	2	15	1.5	0.35	96
100	0.1	0.05	2	15	2	0.35	128
50	0.1	0.05	2	15	0.5	0.35	87
URL							
N	f	α	σ^2	β	g	p_1, p_2	n
80	0.1	0.05	2	15	1	0.35,0.65	2442
60	0.1	0.05	2	15	1.5	0.35,0.65	2439
100	0.1	0.05	2	15	2	0.35,0.65	2441
50	0.1	0.05	2	15	0.5	0.35,0.65	2465
Kuk (Taking k=3)							
N	f	α	σ^2	β	g	θ_1,θ_2	n
80	0.1	0.05	2	5	1	0.35,0.65	24
60	0.1	0.05	2	5	1.5	0.35,0.65	21
100	0.1	0.05	2	5	2	0.35,0.65	39
50	0.1	0.05	2	5	0.5	0.35,0.65	44
Forced Response							
N.	f	α	σ^2	β	g	p_1, p_2	n
80	0.1	0.05	2	5	1	0.35,0.30	68
60	0.1	0.05	2	5	1.5	0.35,0.30	53
100	0.1	0.05	2	5	2	0.35,0.30	79
50	0.1	0.05	2	5	0.5	0.35,0.30	56
Eriksson's RRT							
N	f	α	σ^2	β	g	a,b,L	n
80	0.1	0.05	2	5	1	24,-5,33	1035
	0.1	0.05	2	5	1.5	24,-5,33	1699
100	0.1	0.05	2	5	2	24,-5,33	1536
50	0.1	0.05	2	5	0.5	24,-5,33	1344

$$=E_mV\left(t_{HT}\right)+\sum_{i=1}^{N}\frac{1}{np_i}\left\{\frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}\right\}\{(\beta-1)^2x_i^2+\sigma^2x_i^g\}$$

$$E_mV\left(e_{HT}|Kuk\right)=E_mV\left(t_{HT}\right)+\frac{1}{n}\sum_{i=1}^{N}\frac{1}{p_i}\left\{b_i(k)\beta x_i+c_i(k)\right\}$$

$$E_mV\left(e_{HT}|Forced\ Response\right)$$

$$=E_mV\left(t_{HT}\right)+\sum_{i=1}^{N}\frac{1}{np_i}\left\{\frac{p_1\left(1-p_1\right)+\left(1-p_1-p_2\right)\left(p_2-p_1\right)\beta x_i}{\left(1-p_1-p_2\right)^2}\right\}$$

$$E_mV\left(e_{HT}|Eriksson's\ RRT\right)=E_mV\left(t_{HT}\right)+\sum_{i=1}^{N}\frac{1}{np_i}\left\{a\beta x_i^2+a\sigma^2x_i^g+b\beta x_i+L\right\}.$$

Table 5. Fixing sample-size for Horvitz-Thompson estimate in RR surveys

Warner's RR							
N	f	α	σ^2	β	g	p	n
80	0.1	0.05	2	15	1	0.35	67
60	0.1	0.05	2	15	1.5	0.35	54
100	0.1	0.05	2	15	2	0.35	123
50	0.1	0.05	2	15	0.5	0.35	175
URL							
N	f	α	σ^2	β	g	p_1,p_2	n
80	0.1	0.05	2	5	1	0.35,0.65	1982
60	0.1	0.05	2	5	1.5	0.35,0.65	1758
100	0.1	0.05	2	5	2	0.35,0.65	1770
50	0.1	0.05	2	5	0.5	0.35,0.65	1530
Kuk (Taking k=3)							
N	f	α	σ^2	β	g	θ_1,θ_2	n
80	0.1	0.05	2	5	1	0.35,0.65	46
60	0.1	0.05	2	5	1.5	0.35,0.65	39
100	0.1	0.05	2	5	2	0.35,0.65	49
50	0.1	0.05	2	5	0.5	0.35,0.65	56
Forced Response							
N	f	α	σ^2	β	g	p_1,p_2	n
80	0.1	0.05	2	5	1	0.35,0.30	63
60	0.1	0.05	2	5	1.5	0.35,0.30	44
100	0.1	0.05	2	5	2	0.35,0.30	62
50	0.1	0.05	2	5	0.5	0.35,0.30	51
Eriksson's RRT							
N	f	α	σ^2	β	g	a,b,L	n
80	0.1	0.05	2	5	1	24,-5,33	409
60	0.1	0.05	2	5	1.5	24,-5,33	212
100	0.1	0.05	2	5	2	24,-5,33	508
50	0.1	0.05	2	5	0.5	24,-5,33	684

Fixing sample-size in estimating finite population total by RHC strategy in RR surveys

$$E_mV\left(e_{RHC}\right)=E_mV\left(t_{RHC}\right)+\sum_{i=1}^{N}\frac{Q_i}{p_i}V_R\left(r_i\right)$$

Table 6. Fixing sample-size for RHC strategy in RR surveys

Warner's RR							
N	f	α	σ^2	β	g	p	n
80	0.1	0.05	2	15	1	0.35	13
60	0.1	0.05	2	15	1.5	0.35	13
100	0.1	0.05	2	15	2	0.35	15
50	0.1	0.05	2	15	0.5	0.35	19
URL							
N	f	α	σ^2	β	g	p_1,p_2	n
80	0.1	0.05	2	5	1	0.35,0.65	14
60	0.1	0.05	2	5	1.5	0.35,0.65	12
100	0.1	0.05	2	5	2	0.35,0.65	15
50	0.1	0.05	2	5	0.5	0.35,0.65	20
Kuk (Taking k=3)							
N	f	α	σ^2	β	g	θ_1,θ_2	n
80	0.1	0.05	2	5	1	0.35,0.65	12
60	0.1	0.05	2	5	1.5	0.35,0.65	13
100	0.1	0.05	2	5	2	0.35,0.65	15
50	0.1	0.05	2	5	0.5	0.35,0.65	19
Forced Response							
N	f	α	σ^2	β	g	p_1,p_2	n
80	0.1	0.05	2	5	1	0.35,0.30	14
60	0.1	0.05	2	5	1.5	0.35,0.30	13
100	0.1	0.05	2	5	2	0.35,0.30	15
50	0.1	0.05	2	5	0.5	0.35,0.30	20
Eriksson's RRT							
N	f	α	σ^2	β	g	a,b,L	n
80	0.1	0.05	2	5	1	24,-5,33	14
60	0.1	0.05	2	5	1.5	24,-5,33	12
100	0.1	0.05	2	5	2	24,-5,33	15
50	0.1	0.05	2	5	0.5	24,-5,33	20

5. RECOMMENDATION AND CONCLUSION

We obtain a numerical confirmation that Chebyshev inequality-based sample-size fixing works well in DR surveys. But the same does not work for RR surveys with HH estimator and HT estimator. This is possibly because in the variance of an unbiased estimator based on RR data there is one term exclusively determined by sampling design and DR survey data but in the other term depending on RR-based variance which is too high irrespective of the sampling design specifics. So Chebyshev rule is not quite effective in controlling this term in the variance formula. So, we recommend to employ Chebyshev inequality in controlling the variance of the term in the variance which involves only DR-based features. The other variance-term cannot be controlled by our approach which aims at controlling only the DR-related materials.

REFERENCES

Brewer, K.R., and Hanif, M. (1983). *Sampling with unequal probabilities.* Springer.

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.

Chaudhuri, A. (2010). *Essentials of Survey Sampling.* Delhi, India: Prentice Hall of India.

Chaudhuri, A. (2014). *Modern Survey Sampling.* Chapman & Hall, CRC, Taylor & Francis, Boca Raton, Florida, USA.

Chaudhuri, A. (2018). *Survey Sampling.* Taylor & Francis, CRC, Boca Raton, Florida, USA.

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.

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.

Chaudhuri, A., and Vos, J.W. (1988). *Unified theory and strategies of survey sampling.* North- Holland, Amsterdam.

Hansen, M.M., and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of mathematical statistics,* **14**, 333-362.

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