International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 24(1) March, 2024, pp 115-132
DOI: https://doi.org/10.3329/ijss.v24i1.72030

On the Sample Size Determination based on the Randomized Response Surveys

Kajal Dihidar
Sampling and Official Statistics Unit,
Indian Statistical Institute, Kolkata, India
(Email: kajaldihidar@gmail.com)

[Received November 7, 2023; Accepted February 27, 2024]

Abstract

In general, the well known Chebyshev's inequality is used to determine the sample size in order to conduct a survey using direct responses. The same technique intending to cover for sensitive variables are attempted recently by many statisticians. However it has been observed that in many cases the acceptable sample sizes are hard to be obtained, mainly because of appearance of some easily non-controllable part. Chaudhuri and Sen (2020), Chaudhuri and Patra (2023) and others have illustrated different situations and solutions are proposed therein. In this paper, following Chaudhuri, Bose and Dihidar (2011), we have made an attempt to deterimine the sample size corresponding to the estimators of sensitive population proportion using multiple randomized responses from distinct persons sampled. Along with the theoretical derivations, some numerical illustrations are presented. Based on the important extractions of our numerical illustration results, the recommendable sample size in practical real survey situations are observed.

Keywords and Phrases: Randomized response survey; Sample-size determination; Multiple randomized responses.

AMS Classification: 62D05.

1 Introduction

Surveys to collect information on sensitive or stigmatizing attributes face the problem of untruthful responses or non-cooperation by respondents, both of which lead to biased estimates. To avoid this evasive answer bias and to preserve the privacy of the respondent, Warner (1965) introduced an innovative technique commonly referred to as randomized response (RR) technique. In his model, a sampled respondent answers 'Yes' or 'No' about the matching or non-matching of his/her own characteristic to either the sensitive question of interest or the complementary question, the question being chosen by randomly drawn cards, unnoticed by the inteviewer. Since then, many contributors have enriched the randomized response literature by providing alternative models and proving their efficiencies in comparison to the existing techniques. Important contributions are available for quantitative sensitive variables also.

However, prior to conducting a survey, it is mandatory to have the number of the units from whom we need to collect the data, called as the sample size. For direct surveys, Cochran (1953, 1963, 1977) and several others have prescribed the solutions mainly demanding normality in the distribution of the standardized difference between the estimate and the estimand parameter it seeks to estimate. Chaudhuri (2010, 2014, 2018, 2020) have shown the use of the Chebyshev's inequality in tackling this issue, where the assumption of normality is avoided. Chaudhuri & Dutta (2019) have considered a different approach while discussing the sample-size problem. In case a closed form expression of the variance formula is difficult to obtain, but an unbiased estimator for the variance is available at our hand, Chaudhuri & Dutta (2019) proposed to use that unbiased estimator for the variance to determine the sample size.

The situation is much more hard when the survey is about gathering data on the sensitive variables. Contributory researches are available in lierature in this regard, such as in Chaudhuri and Sen (2020), Chaudhuri and Patra (2023) and others. However it has been observed that in many cases the reasonably acceptable sample sizes are hard to be obtained. The difficulty arires to the point is that there is one part that may be controlled by an appropriately chosen sampling design and there is another part that is very hard to get suitably and naturally controlled. In this regard, various situations are illustrated and solutions are proposed in Chaudhuri and Sen (2020), Chaudhuri and Patra (2023) and others.

Chaudhuri, Bose and Dihidar (2011) have considered the estimation of the proportion of persons bearing a sensitive characteristic unbiasedly by Warner (1965)'s device using the multiple responses from distinct persons sampled. The problem studied in Chaudhuri et al. (2011) stems from the observation that in direct response surveys Basu (1958), Raj and Khamis (1958), Pathak (1962), Korwar and Serfling (1970) and others have shown that if one uses the responses from the distinct units sampled by simple random sampling with replacement (SRSWR) method, alternative unbiased estimators performng better than the classical estimator i.e. the sample mean, are available. Chaudhuri et al. (2011) investigated the similar things for the randomized responses. Mangat et al. (1995) gave an unbiased estimator based on only one randomized response for every distnct unit sampled and studied the relative efficiencies.

Being motivated by above researches, in this paper we have made an attempt to deterimine the sample size corresponding to the estimators of sensitive population proportion using multiple randomized responses from distinct persons obtained in an SRSWR sample. Details are given in the following sections.

2 Preliminaries on the sample size determination based on the direct surveys

Let us consider a finite population $U = (1, 2, \ldots, N)$ of N units, N being known and a real valued variable y with unknown values Y_i for $i \in U$. Suppose our objective is to estimate the population total $Y = \sum_{i=1}^{N} Y_i$ or the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ by a suitably drawn probability sample s of size $n(2 \leq n < N)$ with the selection probability $p(s)$ by a design p. Suppose we consider an estimator t to unbiasedly estimate Y which has a variance $V_p(t)$. Also suppose we need the estimator t to be so accurate that for a given positive proper fraction f, say, $f = 0.1$, or $f = 0.2$, etc. we want to maintain the absolute difference in between the estimate values and the unknown fixed Y value to be less than or equal to fY for many of the all possible samples, i.e. choosing a small positive value of α close to 0, say as 0.1, 0.05, 0.01,

$$P(|t - Y| \leq fY) \geq 1 - \alpha. \tag{1}$$

As per the definition of the mean squared error $M_p(t) = E_p(t - Y)^2$ of the estimator t, where $E_p()$ is the expectation operator with respect of the sampling design p, for a positive number K, we can have the result as

$$P\left(t - K \leq Y \leq t + K\right) \geq 1 - \frac{M_p(t)}{K^2}. \tag{2}$$

Taking $K = \lambda\sqrt{V_p(t)}$ for a positive number λ, and on noting that $M_p(t) = V_p(t) + B_p^2(t)$, where $B_p(t)$ denotes the bias of the estimator t, we have the result as

$$P\left(t - \lambda\sqrt{V_p(t)} \leq Y \leq t + \lambda\sqrt{V_p(t)}\right) \geq \left(1 - \frac{1}{\lambda^2}\right) - \frac{B_p^2(t)}{\lambda^2 V_p(t)}. \tag{3}$$

For t being an unbiased estimator, $B_p(t) = 0$, and the Chebyshev's inequality tells us

$$P\left(t - \lambda\sqrt{V_p(t)} \leq Y \leq t + \lambda\sqrt{V_p(t)}\right) \geq \left(1 - \frac{1}{\lambda^2}\right) \Longrightarrow P\left(|t - Y| \leq \lambda\sqrt{V_p(t)}\right) \geq \left(1 - \frac{1}{\lambda^2}\right). \tag{4}$$

Now combining the Eqns. (1) and (4) of above, let us take $\alpha = \frac{1}{\lambda^2}$ and $fY = \lambda\sqrt{V_p(t)}$, and on writing the coefficient of variation CV of t as $CV(t) = \frac{\sqrt{V_p(t)}}{Y} \times 100$, we have the relation as

$$\alpha = \frac{V_p(t)}{f^2 Y^2} \Longrightarrow \alpha = \frac{(CV(t))^2}{100^2 f^2}. \tag{5}$$

In case a simple random sampling, with or without replacement i.e. SRSWR or SRSWOR, $N\bar{y}$ with $\bar{y}$ as the sample mean, in either case is used to unbiasedly estimate Y.

For SRSWR, $V(N\bar{y}) = N^2\frac{\sigma_y^2}{n} = \frac{N(N-1)}{n}S_y^2$, where $\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ are two forms of the population variance of y, and $CV(\bar{y}) = 100 \times \frac{\sqrt{V(\bar{y})}}{\bar{Y}} = 100 \times \sqrt{\frac{N-1}{Nn}}\frac{S_y}{\bar{Y}}$.
Now on denoting CV as the $CV = \frac{S_y}{\bar{Y}}100$, the cooefficient of variation of N population values Y_i's, utilizing the given values of α, f, CV and N, the sample size may be determined by solving the above equation for n. For example, for SRSWR case, the sample size determination formula takes the following form.

$$\text{SRSWR case} : n = \left(\frac{N-1}{N}\right)\frac{1}{\alpha f^2}\left(\frac{CV}{100}\right)^2. \tag{6}$$

For SRSWOR case, on noting that $V(\bar{y}) = (\frac{1}{n} - \frac{1}{N})S_y^2$, the sample size determination formula takes the following form.

$$\text{SRSWOR case} : n = \frac{n_0}{1+\frac{n_0}{N}}, \text{ where } n_0 = \frac{1}{\alpha f^2}\left(\frac{CV}{100}\right)^2. \tag{7}$$

In Chaudhuri (2010, 2014, 2018) this is illustrated in details.
It is to be noted that in above formulae, the population CV of y values or the values of S_y and $\bar{Y}$ separately are required in advance from any reliable source, for example, from the results of recently conducted similar surveys or from some admininstrative sources. Many times this information are unavailable. In that situation, some important binary variable as per the objective of the survey is required to be identified and based on that binary variable, the sample size is determined. In both the SRSWR and SRSWOR cases, the sample proportion p is used to unbiasedly estimate the population proportion P associated to the considered binary variable. The variance of p are :

$$\text{for SRSWR case:} V(p) = \frac{P(1-P)}{n} \text{ and for SRSWOR case} : V(p) = \frac{N-n}{N-1}\frac{P(1-P)}{n}. \tag{8}$$

Using these, the Chebyshev's sample size determination relation $\alpha = \dfrac{V(p)}{f^2P^2}$ gives the formulae as:

$$\text{SRSWR case} : n = \frac{1}{\alpha f^2}\left(\frac{1-P}{P}\right). \tag{9}$$

$$\text{SRSWOR case} : n = \frac{n_0}{1+\frac{n_0}{N}}, \text{ where } n_0 = \left(\frac{N}{N-1}\right)\frac{1}{\alpha f^2}\left(\frac{1-P}{P}\right). \tag{10}$$

However, in community level surveys in health, nutrition etc., it is most common to use the following Cochran's formula for determining the sample size for estimating a population proportion parameter,

$$n = \frac{n_0}{1+\frac{n_0}{N}}, \text{ where } n_0 = \frac{z_\alpha^2 \times (p \times (1-p))}{d^2}, \tag{11}$$

where p is a reasonable guess value of the population proportion P, d is the margin of error and z_α is the value of the normal deviate corresponding to level of significance α. The margin of error d is usually taken as $d = e \times p$, where e is the relative permissible margin in error. If a reasonable guess value p about the population proportion P is not available beforehand, the middle value 0.5 is used in place of p.

It is to be noted that our present work is intended to avoid the usual normality assumption and to study in general the sample size determination problem corresponding to the estimators of sensitive population proportion using multiple randomized responses from distinct persons sampled.

For general sampling schemes and designs with suitable estimators, especially for unequal probability sampling designs, appealing to the concept of the coefficient of variation does not yield a useful solution to the problem of finding an appropriate value for n. For this reason, Chaudhuri & Dutta (2018) have recommended to solve this problem by postulating a regression model in between y and some well related auxiliary variable x as :

$$y_i = \beta x_i + \epsilon_i, i \in U, \tag{12}$$

where x_i are known for all $i \in U$ and β is an unknown constant and ϵ_i's are independent random variables with mean $E_m(\epsilon_i) = 0 \ \forall i \in U$ and variances $V_m(\epsilon_i) = \sigma^2 x_i^g$ with an unknown $\sigma > 0$ and an unknown constant $g(0 \leq g \leq 2)$, where $E_m()$ and $V_m()$ denote respectively the model-based expectation and variance operators. Based on this model, they recommended to solve for n from :

$$\alpha = \frac{E_m(V_p(t))}{f^2 E_m(Y^2)}. \tag{13}$$

It is to be noted that to implement this formula, we need to use the given values of N, f, α, σ^2, β, g, and the known values of x_i's for all $i \in U$. Chaudhuri and Dutta (2018) have presented the illustrations of sample size determinations for several unequal probability sampling designs. Every time they have noted that the sampling fractions obtained are quite reasonable in each case with numerical illustrations.
However, in the current research, we consider the equal probability and with replacement (SRSWR) design and study the details of finding the sample size based on the randomized responses from the distinct units obtained to cover sensitive issues. The details of our research are disclosed in the following sections. Unequal probability sampling designs will be considered in some consequent research work.

3 Preliminaries on the sample size determination based on the randomized response (RR) surveys

Let us consider the problem of estimation of the population proportion of a sensitive qualitative characteristic, say, A and the use of Warner (1965)'s RR technique. Let us define the variable y taking the value for unit i as $Y_i = 1/0$ according to bearing/non-bearing of the characteristic A, for $i \in U$. Our objective is to estimate the sensitive population proportion $\theta = \bar{Y}$. Let a sampled person i be approached with a box of cards in p ($\neq 0.5$) proportion with marked A and the rest $(1-p)$ proportion marked with A^c , with a request to draw randomly one card from the box and to truthfully answer if he/she gets a match or non-match of his/her own characteristic with the marks on the card randomly drawn. The generation of such randomized response data is done unnoticed by the interviewer, thus protecting the privacy of the respondent. Following Chaudhuri (2011) and Chaudhuri & Christofides (2013) let us define the random variable I denoting the numerical values of the RRs as :

$$\begin{aligned} I_i &= 1 \text{ if the person i gets a match} \\ &= 0 \text{ if the person i gets a non-match .} \end{aligned}$$

Let E_R, V_R denote the expectation and variance operators with respect to the RR collection from the sampled respondents, and E_P, V_P denote the expectation and variance operators with respect to the sampling of the respondents. Also let us suppose E and V denote the overall expectation and variance operators. Then $E_R(I_i) = py_i + (1-p)(1-y_i) = (1-p) + (2p-1)y_i$. This implies that on defining $r_i = \frac{I_i-(1-p)}{(2p-1)}$, we have $E_R(r_i) = y_i$. On noting that $y_i = 1/0$ and $I_i = 1/0$, and hence $I_i^2 = I_i$ and $y_i^2 = y_i$, we have $V_R(I_i) = p(1-p)$, and hence $V_R(r_i) = \frac{p(1-p)}{(2p-1)^2} = \phi_W$, say.

In order to estimate $\theta = \bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$, assuming that the population size N is known, the problem reduces to the estimation of the total $Y = \sum_{i=1}^{N} Y_i$. So utilizing the notation t for direct survey based unbiased estimator for Y, but for the RR data, let us call the estimator as e and e is defined as

$$e = t|_{y_i=r_i, \forall i \in s}. \tag{14}$$

This estimator e has the property that

$$E_R(e) = t \text{ and } E(e) = E_P E_R(e) = E_P(t) = Y. \tag{15}$$

Thus e is an unbiased estimator for Y, for RR-based survey.

$$V(e) = V_P E_R(e) + E_P V_R(e) = V_P(t) + E_P V_R(e). \tag{16}$$

It has been observed earlier by Chaudhuri and Sen (2020), Chaudhuri and Patra (2023) and others, that the $E_P V_R(e)$ part has shown much larger numerical values in comparison to the $V_P(t)$ part, resulting in absurd values for the sample sizes n using Chebyshev's inequality for the RR-based surveys. For example, for Warner (1965)'s RR technique, for both cases of SRSWR and SRSWOR designs, an unbiased estimator of θ, say, $\widehat{\theta}_W$ is :

$$\widehat{\theta}_W = \frac{e}{N} = \frac{1}{n}\sum_{i=1}^{n} r_i. \tag{17}$$

And about the variance of the estimator, on noting that $\sigma_y^2 = \theta(1-\theta)$,

$$\text{SRSWR case}: V(\widehat{\theta}_W) = \frac{\sigma_y^2}{n} + \frac{\phi_W}{n} = \frac{\theta(1-\theta)}{n} + \frac{\phi_W}{n}, \tag{18}$$

and

$$\text{SRSWOR case}: V(\widehat{\theta}_W) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 + \frac{\phi_W}{n} = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{N\sigma_y^2}{N-1} + \frac{\phi_W}{n}. \tag{19}$$

Taking advantage from Chebyshev's inequality, for SRSWR design, the sample size can be determined by solving the equation

$$\text{SRSWR case}: \alpha = \frac{V(\widehat{\theta}_W)}{f^2\theta^2} = \frac{\frac{\theta(1-\theta)}{n} + \frac{\phi_W}{n}}{f^2\theta^2} \Longrightarrow n = \frac{\theta(1-\theta) + \phi_W}{\alpha f^2\theta^2}. \tag{20}$$

Similarly for SRSWOR design, the sample size is determined from the relation :

$$\text{SRSWOR case}: \alpha = \frac{V(\widehat{\theta}_W)}{f^2\theta^2} = \frac{\left(\frac{1}{n} - \frac{1}{N}\right)\frac{N\sigma_y^2}{N-1} + \frac{\phi_W}{n}}{f^2\theta^2} \Longrightarrow n = \frac{\frac{N}{N-1}\theta(1-\theta) + \phi_W}{\alpha f^2\theta^2 + \frac{\theta(1-\theta)}{N-1}}. \tag{21}$$

Some numerical illustrations are presented below in Table 1.

Table 1: Sample size for direct and RR surveys for various parameters (Warner's RRT)

N	f	α	CV	$p(\neq 0.5)$	θ	n(DR:SRSWR)	n(RR:SRSWR)	n(DR:SRSWOR)	n(RR:SRSWOR)
80	0.10	0.05	10.00	0.45	0.25	20	798000	16	10372
60	0.10	0.05	8.00	0.45	0.25	13	798000	11	7772
100	0.10	0.05	10.00	0.45	0.25	20	798000	17	12955
50	0.10	0.05	5.00	0.45	0.25	5	798000	5	6466

From the outputs of this table, we see that the determinations of n for direct response (DR) surveys seem reasonable for both the designs SRSWR and SRSWOR, while the determinations of n for randomized response (RR) surveys seem quite absurd. A natural question arises, what happens about the sample size determination problem if we concentrate on the randomized responses obtained from the distinct units appearing in an SRSWR sample. Later sections are devoted to this direction.

4 Sample size determination based on the distinct units RR data in randomized response (RR) surveys

Let us consider Warner (1965)'s randomized response device with device parameter $p(\neq 0.5)$ and SRSWR sampling for choosing the respondents. For an SRSWR sample s of size $n(2 \leq n \leq N)$ drawn from U, let $\nu(1 \leq \nu \leq n)$ be the number of distinct persons appeared in s. We consider various estimators based on the distinct units' RRs and study the sample size determination problem for each of them. The details are given in the following subsections.

4.1 Sample size determination for Mangat et al. (1995)'s model

Mangat et al. (1995) studied the case where in an SRSWR sample chosen in n draws, the $\nu(1 \leq \nu \leq n)$ distinct persons found are requested to perform Warner's RR trial only once each. They proposed an unbiased estimator for θ based on these ν RR's and gave its variance. Let us denote this estimator by $\widehat{\theta}_{W1}$ and is given by the following.

$$\widehat{\theta}_{W1} = \frac{(\nu'/\nu) - (1-p)}{2p-1}, \tag{22}$$

where ν' is the number of persons out of the ν distinct persons, who find match with attribute A/A^c in Warner's RRD. Mangat et al. (1995) have given its variance as :

$$V(\widehat{\theta}_{W1}) = \phi_W E_P\left(\frac{1}{\nu}\right) + \left[N E_P\left(\frac{1}{\nu}\right) - 1\right]\frac{\theta(1-\theta)}{N-1}. \tag{23}$$

From Pathak (1962) and others, it is known to us that :

$$E_P\left(\frac{1}{\nu}\right) = \frac{1}{N^n}\sum_{j=1}^{N} j^{n-1}. \tag{24}$$

Utilizing this, the above variance expression takes the following form as :

$$V(\widehat{\theta}_{W1}) = \frac{\phi_W}{N}\sum_{j=1}^{N}\left(\frac{j}{N}\right)^{n-1} + \left[\sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}\right]\frac{\theta(1-\theta)}{N-1}. \tag{25}$$

Mangat et al. (1995) noted that $\widehat{\theta}_{W1}$ outperforms $\widehat{\theta}_W$, in the sense that $V(\widehat{\theta}_{W1}) < V(\widehat{\theta}_W)$ if N, n, p and θ happen to be such that :

$$\theta(1-\theta) > \left[\frac{n(N-1)(6N+n-1)}{N\{6Nn-12N-n(n-1)\}}\right] \times \phi_W. \tag{26}$$

In particular, they remarked that when $N = 100$, $n = 10$ and $p = 0.9$, $\widehat{\theta}_{W1} \succ \widehat{\theta}_W$ for $0.236 \leq \theta \leq 0.764$ and $\widehat{\theta}_W \succ \widehat{\theta}_{W1}$ otherwise. However, it is to be noted that a value of

p acceptable to a respondent should be away from 0 and 1 and most preferably near 0.5 on either side, say, $0.45 \leq p < 0.5$ or $0.5 < p \leq 0.55$. Now let us investigate about the sample sizes if some realistic sample size can be obtained for some acceptable ranges of p values.

From Chebyshev's inequality, the sample size n can be determined by solving the equation :

$$\alpha = \frac{V(\widehat{\theta}_{W1})}{f^2\theta^2} = \frac{\frac{\phi_W}{N}\sum_{j=1}^{N}\left(\frac{j}{N}\right)^{n-1} + \left[\sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}\right]\frac{\theta(1-\theta)}{N-1}}{f^2\theta^2}. \quad (27)$$

This shows that n can be obtained by solving the following equation as :

$$\frac{f^2\theta^2\alpha - \frac{\phi_W}{N}}{\frac{\phi_W}{N} + \frac{\theta(1-\theta)}{N-1}} = \sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}. \quad (28)$$

Some numerical illustrations are given below.

Table 2: Sample size for RR surveys for various parameters (Mangat et al. (1995) model), SRSWR

N	f	α	$p(\neq 0.5)$	θ	Soln. of n within N	Soln. of n beyond N
80	0.10	0.05	0.45	0.25	No solns within N	Even no solns within $\leq 20N$
60	0.10	0.05	0.45	0.25	No solns within N	Even no solns within $\leq 20N$
100	0.10	0.05	0.45	0.25	No solns within N	Even no solns within $\leq 20N$
50	0.10	0.05	0.45	0.25	No solns within N	Even no solns within $\leq 20N$

From the outputs of the above table, we see that it is very hard to find out the sample size for the parameters mentioned in the table. At this moment, let us perform our attempt by altering the device parameter, i.e. increasing p, as control of device parameter is at the hand of the statistician, and keeping the θ value as fixed because it is by nature. The following table shows some numerical results for $N = 100$.

Table 3: Sample size for RR surveys for various parameters (Mangat et al. (1995) model), SRSWR, attempt by increasing p

N	f	α	$p(\neq 0.5)$	θ	Soln. of n within N	Soln. of n beyond N
100	0.10	0.05	$0.65 \leq p \leq 0.99$	0.25	No solns within N	Even no solns within $\leq 20N$
100	0.10	0.05	0.995	0.25	No solns within N	Even no solns within $\leq 20N$
100	0.10	0.05	0.996	0.25	No solns within N	Even no solns within $\leq 20N$
100	0.10	0.05	0.997	0.25	No solns within N	775

From the outputs of the above table, we see that for many of the commomly used device parameter p values, even for large p values, neither any solution of n is obtained within $\leq N$, nor within $\leq 20N$. However, a very large value of n is obtained for extraordinarily large value $p = 0.997$, which is very hard for a respondent to be agreed to participate in the survey because of not protecting his/her privacy enough. Let us continue our attempts to investigate for n for various values of θ and p. Below are the results of our further attempts.

Table 4: Sample size for RR surveys for various parameters (Mangat et al. (1995) model), SRSWR, attempt for various values of θ and p (p values not mentioned means no solution.)

N	f	α	$p(\neq 0.5)$	θ	Soln. of n within N	Soln. of n beyond N
100	0.1	0.05	Any $p \in [0.01, 0.99]$	≤ 0.45	No solns within N	Even no solns within $\leq 20N$
100	0.1	0.05	$p = 0.01$ or 0.99	0.5	No solns within N	478
100	0.1	0.05	$p = 0.01$ or 0.99	0.55	No solns within N	399
100	0.1	0.05	$p = 0.01$ or 0.99	0.6	No solns within N	351
100	0.1	0.05	$p = 0.01$ or 0.99	0.65	No solns within N	313
100	0.1	0.05	$p = 0.01$ or 0.99	0.7	No solns within N	280
100	0.1	0.05	$p = 0.02$ or 0.98	0.7	No solns within N	428
100	0.1	0.05	$p = 0.01$ or 0.99	0.75	No solns within N	249
100	0.1	0.05	$p = 0.02$ or 0.98	0.75	No solns within N	344
100	0.1	0.05	$p = 0.01$ or 0.99	0.8	No solns within N	218
100	0.1	0.05	$p = 0.02$ or 0.98	0.8	No solns within N	288
100	0.1	0.05	$p = 0.01$ or 0.99	0.85	No solns within N	185
100	0.1	0.05	$p = 0.02$ or 0.98	0.85	No solns within N	240
100	0.1	0.05	$p = 0.03$ or 0.97	0.85	No solns within N	393
100	0.1	0.05	$p = 0.01$ or 0.99	0.9	No solns within N	146
100	0.1	0.05	$p = 0.02$ or 0.98	0.9	No solns within N	191
100	0.1	0.05	$p = 0.03$ or 0.97	0.9	No solns within N	284
100	0.1	0.05	$p = 0.01$ or 0.99	0.95	**98** (n/N is very high)	—
100	0.1	0.05	$p = 0.02$ or 0.98	0.95	No solns within N	136
100	0.1	0.05	$p = 0.03$ or 0.97	0.95	No solns within N	202

From above outputs, we see that for very large value of the device parameter p we may have some solution for n. However, only within $\leq N$ solution for n with very high sampling fraction is observed for $\theta = 0.95$ and $p = 0.01$ or 0.99. In other cases, for very high p values, some solution of n is observed, but not within $\leq N$, much exceeding N. Now as so large value of p is not acceptable for a respondent because of harming their privacy and at the same time the value of n exceeding much of N together conclude the impractical result for conducting a RR survey. So, at this monemt, let us study some other possible estimators and investigate if some practicable solutions of n can be obtained for those.

4.2 Sample size determination for Chaudhuri et al. (2011)'s model

Let for a sample s drawn by SRSWR scheme in n draws, f_i be the number of times unit i appears in the sample s. Let $I_{ij} = 1/0$ according as the i^{th} person in his j^{th} apparance in s gets a 'match' or 'mis-match' of his/her true attribute A or A^c and the statement written on the card drawn randomly from Warner's device with device parameter $p(\neq 0.5)$. Then clearly $\sum_{i=1}^{N} f_i = n$.

Let u denote the the set of distinct persons in s, and as $\nu(1 \leq \nu \leq n)$ is defined as the number of distinct persons appeared in s, ν is equal to the cardinality of u. Now for $i \in s$, $f_i > 0$. Chaudhuri et al. (2011) defined the following transformed randomized responses for $i \in s$ as :

$$m_i = \frac{1}{f_i}\sum_{j=1}^{f_i} I_{ij}, \quad g_i = \frac{m_i - (1-p)}{2p-1}. \tag{29}$$

Then

$$E_R(I_{ij}) = (1-p) + (2p-1)y_i = E_R(m_i), \tag{30}$$

$$V_R(I_{ij}) = E_R(I_{ij})(1 - E_R(I_{ij})) = p(1-p) \text{ and } V_R(m_i) = \frac{p(1-p)}{f_i}, \tag{31}$$

$$E_R(g_i) = y_i \text{ and } V_R(g_i) = \frac{\phi_W}{f_i}. \tag{32}$$

Based on these transformed randomized responses, Chaudhuri et al. (2011) proposed two alternative estimators for θ. In the two following subsections, we present Chaudhuri et al. (2011)'s alternative estimators and their properties. Additionally for each of them, we make an attempt to solve the sample size determination problem.

4.2.1 Sample size determination for Chaudhuri et al. (2011)'s first estimator of θ

Chaudhuri et al. (2011) proposed an estimator of θ, say, $\widehat{\theta}_{W2}$ defined as :

$$\widehat{\theta}_{W2} = \frac{1}{\nu}\sum_{i \in u} g_i. \tag{33}$$

They have shown that it is unbiased for θ, since $E(\widehat{\theta}_{W2}) = E_P E_R(\widehat{\theta}_{W2}) = E_P(\frac{1}{\nu}\sum_{i\in u} y_i) = \bar{Y} = \theta$.

Utilizing some results on direct surveys from Pathak (1962) and others, Chaudhuri et al. (2011) have obtained the variance of $\widehat{\theta}_{W2}$ as follows.

$$V(\widehat{\theta}_{W2}) = \left[\frac{1}{N^{n-1}(N-1)}\sum_{j=1}^{N-1} j^{n-1}\right](\theta - \theta^2) + \phi_W E_P\left(\frac{1}{\nu^2}\sum_{i\in u}\frac{1}{f_i}\right). \tag{34}$$

On noting that $f_i \geq 1$ and $f_i \leq n$ for all $i \in u$, and hence $\sum_{i\in u}\frac{1}{f_i} \leq \nu$ and $\sum_{i\in u}\frac{1}{f_i} \geq \frac{\nu}{n}$, Chaudhuri et al. (2011) observed that $V(\widehat{\theta}_{W2}) \leq V(\widehat{\theta}_{W1})$. Hence they concluded that $\widehat{\theta}_{W2}$ uniformly outperforms $\widehat{\theta}_{W1}$.

Here since it is hard for $V(\widehat{\theta}_{W2})$ to have a closed form expression, to take advantage of Chaudhuri and Dutta (2019)'s approach for sample size determination by solving $\alpha = \frac{(Estimate(CV(t)))^2}{100^2 f^2}$, we consider the variance estimator. In this regard, following Chaudhuri et al. (2011), let

$$C_{1a} = (N-1)\left[(1-\frac{1}{N})^n - (1-\frac{2}{N})^n\right] \text{ and } \tag{35}$$

$$C_{1b} = N(1-\frac{1}{N})^n - N^2(1-\frac{1}{N})^{2n} + N(N-1)(1-\frac{2}{N})^n. \tag{36}$$

An unbiased estimator for θ^2 is

$$\tilde{\theta^2} = \frac{1}{C_{1b} + N^2\pi_i^2 - C_{1a}\frac{N}{N-1}}\left[\sum\sum_{i\neq j, i,j\in u} g_i g_j - \widehat{\theta}_{W2}(C_{1a}\frac{N}{N-1} - N\pi_i)\right], \tag{37}$$

where π_i is the first order inclusion probability for $i \in s$. For SRSWR sampling design, $\pi_i = 1-(1-\frac{1}{N})^n$ for every i.

An unbiased estimator for $V(\widehat{\theta}_{W2})$ is given by

$$\widehat{V}(\widehat{\theta}_{W2}) = \left[\frac{1}{N^{n-1}(N-1)}\sum_{j=1}^{N-1} j^{n-1}\right](\widehat{\theta}_{W2} - \tilde{\theta^2}) + \phi_W(\frac{1}{\nu^2}\sum_{i\in u}\frac{1}{f_i}). \tag{38}$$

However, in practice the estimates of the coefficient of variation are usually higher for randomized response surveys than the direct response surveys, while the quantity $100f\sqrt{\alpha} = T$, say, ranges from 0.1 to 6.3245 for commonly taken measures of f ranging from 1% to 20% and the measures of α as $0.1, 0.05, 0.01$, it is very unlikely to expect the estimates of CV in percentages to be near about T. This point needs much more investigation to follow Chaudhuri and Dutta (2019)'s approach in our case. Rather we try to get some range of the sample size using the comparison critertia and Chebyshev's inequality.

Utilizing the comparison criteria $V(\widehat{\theta}_{W2}) \leq V(\widehat{\theta}_{W1})$, from Chebyshev's inequality, the sample size n in this case can be determined by solving the equation

$$\alpha = \frac{V(\widehat{\theta}_{W2})}{f^2\theta^2} \leq \frac{V(\widehat{\theta}_{W1})}{f^2\theta^2} = \frac{\frac{\phi_W}{N}\sum_{j=1}^{N}\left(\frac{j}{N}\right)^{n-1} + \left[\sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}\right]\frac{\theta(1-\theta)}{N-1}}{f^2\theta^2}. \tag{39}$$

This shows that n can be obtained by solving the following inequality as :

$$\frac{f^2\theta^2\alpha - \frac{\phi_W}{N}}{\frac{\phi_W}{N} + \frac{\theta(1-\theta)}{N-1}} \leq \sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}. \tag{40}$$

Some numerical illustrations are given below. Since the sample size determining equation is an inequality, we concentrate on the range of n. It is to be noted that the function $\frac{f^2\theta^2\alpha - \frac{\phi_W}{N}}{\frac{\phi_W}{N} + \frac{\theta(1-\theta)}{N-1}} - \sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}$ is an increasing function of n, keeping the other parameters as fixed. So in this case, we try to find the maximum value of n that satisfies the above mentioned inequality. Once these are obtained, we will be free to take any sample size within that limit. Our observations are presented in the following table.

Table 5: Maximum value of n determined for Chauduri et al. (2011)'s first estimator $\widehat{\theta}_{W2}$ for various values of p and θ

N	f	α	$p(\neq 0.5)$	θ	Maximum value of n(RR:SRSWR)
100	0.1	0.05	Any $p \in [0.01, 0.99]$	≤ 0.9	True for any n within N
100	0.1	0.05	$p = 0.01$ or 0.99	0.95	97
100	0.1	0.05	Other p values	0.95	True for any n within N

From above table, we see that many admissible choices of p values are available to perform the RR survey with any sample size within N and to use Chaudhuri et al. (2011)'s first estimator of θ based on the distinct units' multiple randomized responses.

4.2.2 Sample size determination for Chaudhuri et al. (2011)'s second estimator of θ

Chaudhuri et al. (2011) defined another estimator of θ, namely, the Horvitz Thompson estimator in this case, say, $\widehat{\theta}_{W3}$ defined as :

$$\widehat{\theta}_{W3} = \frac{1}{N}\sum_{i\in u}\frac{g_i}{\pi_i}, \tag{41}$$

where $\pi_i = 1 - (1 - \frac{1}{N})^n$, as defined earlier, is the first order inclusion probability for $i \in s$ for SRSWR. This estimator $\widehat{\theta}_{W3}$ is unbiased for θ.

The second order inclusion probability π_{ij} for every pair of units i and $j(\neq i)$ for SRSWR design is $\pi_{ij} = 1 - 2(1 - \frac{1}{N})^n + (1 - \frac{2}{N})^n$. Utilizing this and some results on direct surveys from Pathak (1962) and others, Chaudhuri et al. (2011) obtained the variance of $\widehat{\theta}_{W3}$ as :

$$\begin{aligned} V(\widehat{\theta}_{W3}) &= \frac{\phi_W}{N^2\pi_i^2}E_P\left[\sum_{i\in u}\frac{1}{f_i}\right] + \frac{\theta(1-\theta)}{N(N-1)\pi_i^2}\left[(N-1)\left\{(1-\frac{1}{N})^n - (1-\frac{2}{N})^n\right\}\right] \\ &+ \frac{\theta^2}{N^2\pi_i^2}\left[N(1-\frac{1}{N})^n - N^2(1-\frac{1}{N})^{2n} + N(N-1)(1-\frac{2}{N})^n\right]. \end{aligned} \tag{42}$$

On writing two constants, namely $A_1$and A_2 defined repectively as :

$$A_1 = \frac{1}{N^n}\sum_{j=1}^{N-1} j^{n-1} - \frac{N-1}{N^2\pi_i^2}\left[(1-\frac{1}{N})^n - (1-\frac{2}{N})^n\right], \tag{43}$$

and

$$A_2 = \frac{1}{N\pi_i^2}\left[(1-\frac{1}{N})^n - N(1-\frac{1}{N})^{2n} + (N-1)(1-\frac{2}{N})^n\right], \tag{44}$$

and also utilizing a relation from Korwar and Serfling (1970) which states that

$$Q - \frac{1}{720N} < E_P\left(\frac{1}{\nu}\right) \leq Q, \text{ where } Q = \frac{1}{n} + \frac{1}{2N} + \frac{n-1}{12N^2}, \tag{45}$$

Chaudhuri et al. (2011) obtained the conditions for comparison in between $\widehat{\theta}_{W3}$ and $\widehat{\theta}_{W1}$. These are as follows.

$$V(\widehat{\theta}_{W1}) \leq V(\widehat{\theta}_{W3}) \quad \text{if} \quad \frac{N\theta(1-\theta)}{N-1}A_1 - \theta^2 A_2 \leq \phi_W\left(\frac{1}{nN\pi_i} - Q\right), \tag{46}$$

and

$$V(\widehat{\theta}_{W3}) < V(\widehat{\theta}_{W1}) \quad \text{if} \quad \frac{N\theta(1-\theta)}{N-1}A_1 - \theta^2 A_2 \geq \phi_W\left(\frac{1}{N\pi_i} - Q + \frac{1}{720N}\right). \tag{47}$$

In practice, since it usually happens that $\frac{1}{720N} \approx 0$, Chaudhuri et al. (2011) concluded that

$$V(\widehat{\theta}_{W3}) < V(\widehat{\theta}_{W1}) \quad \text{if} \quad \frac{N\theta(1-\theta)}{N-1}A_1 - \theta^2 A_2 \geq \phi_W\left(\frac{1}{N\pi_i} - Q\right). \tag{48}$$

Following Pathak (1962), for large N, A_1 and A_2 can be approximated respectively by $\tilde{A}_1$ and $\tilde{A}_2$ shown as below :

$$\tilde{A}_1 = \frac{1}{2nN} + \frac{5(n-1)}{12nN^2} \text{ and } \tilde{A}_2 = \frac{n-1}{2nN} - \frac{(n-1)(n-2)}{3nN^2}. \tag{49}$$

Utilizing these approximations, for some given N, n and p values, Chaudhuri et al. (2011) have computed the range of θ for which $\widehat{\theta}_{W3} \succ \widehat{\theta}_{W1}$. Their results of illustrative computaions show that for most small θ values which usually may arise in reality for many of the cases and where we decide to apply the RR technique, the $\widehat{\theta}_{W3}$ will outperform $\widehat{\theta}_{W1}$.

Now we note that here also it is hard for $V(\widehat{\theta}_{W3})$ to have a closed form expression. So to take advantage of Chaudhuri and Dutta (2019)'s approach for sample size determination, we consider the variance estimator. In this regard, following Chaudhuri et al.

(2011), let

$$v_{HT}(g) = \frac{1}{N^2}\left[\sum_{i\in u} g_i^2(\frac{1-\pi_i}{\pi_i^2}) + \sum\sum_{i\neq i',\in u} g_i g_{i'}(\frac{\pi_{ii'}-\pi_i\pi_{i'}}{\pi_{ii'}\pi_i\pi_{i'}})\right]. \quad (50)$$

Then an unbiased estimator for $V(\widehat{\theta}_{W3})$ is given by

$$\widehat{V}(\widehat{\theta}_{W3}) = v_{HT}(g) + \frac{\phi_W}{N^2}\sum_{i\in u}\frac{1}{\pi_i f_i}. \quad (51)$$

Due to the same reasons i.e. of high value estimates of CV's which are far away from the T values as mentioned earlier in case of $\widehat{\theta}_{W2}$, instead of following Chaudhuri and Dutta (2019)'s approach, here also we try to get some range of the sample size using the comparison critertia and Chebyshev's inequality.

Utilizing the comparison criteria $V(\widehat{\theta}_{W3}) \leq V(\widehat{\theta}_{W1})$, from Chebyshev's inequality, the sample size n in this case can be determined by solving the two equations simultaneously :

$$\alpha = \frac{V(\widehat{\theta}_{W3})}{f^2\theta^2} \leq \frac{V(\widehat{\theta}_{W1})}{f^2\theta^2} = \frac{\frac{\phi_W}{N}\sum_{j=1}^{N}\left(\frac{j}{N}\right)^{n-1} + \left[\sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1}\right]\frac{\theta(1-\theta)}{N-1}}{f^2\theta^2}. \quad (52)$$

and

$$\frac{N\theta(1-\theta)}{N-1}A_1 - \theta^2 A_2 \geq \phi_W\left(\frac{1}{N\pi_i} - Q\right). \quad (53)$$

This shows that n can be obtained by solving the following two inequalities simultaneously as :

$$\frac{f^2\theta^2\alpha - \frac{\phi_W}{N}}{\frac{\phi_W}{N} + \frac{\theta(1-\theta)}{N-1}} \leq \sum_{j=1}^{N-1}\left(\frac{j}{N}\right)^{n-1} \quad \text{and} \quad \frac{N\theta(1-\theta)}{N-1}A_1 - \theta^2 A_2 \geq \phi_W\left(\frac{1}{N\pi_i} - Q\right). \quad (54)$$

Some numerical illustrations utilizing the approximations of A_1 and A_2 are given below. Since the sample size determining equations here are two inequalities, we concentrate on the range of n, for which two inequalities are satisfied together. Once these are obtained, we will be free to take any sample size within that range. Our observations are shown in the following table.

Here the word 'Success' means the above two inequalities together are true, otherwise 'No success'.

Table 6: Results for solution of $n < N$ for Chaudhuri et al. (2011)'s second estimator $\widehat{\theta}_{W3}$ for various values of p and θ

N	f	α	$p(\neq 0.5)$	θ	Remarks on n for Success
100	0.1	0.05	Any $p \in [0.02, 0.98]$	0.05	Success for any n within N
100	0.1	0.05	Any $p \in [0.15, 0.85]$	0.1	Success for any n within N
100	0.1	0.05	Any $p \in [0.25, 0.75]$	0.15	Success for any n within N
100	0.1	0.05	Any $p \in [0.31, 0.69]$	0.2	Success for any n within N
100	0.1	0.05	Any $p \in [0.35, 0.65]$	0.25	Success for any n within N
100	0.1	0.05	Any $p \in [0.37, 0.63]$	0.3	Success for any n within N
100	0.1	0.05	Any $p \in [0.39, 0.61]$	0.35	Success for any n within N
100	0.1	0.05	Any $p \in [0.40, 0.60]$	0.40	Success for any n within N
100	0.1	0.05	Any $p \in [0.42, 0.58]$	0.45	Success for any n within N
100	0.1	0.05	Any $p \in [0.43, 0.57]$	0.5	Success for any n within N
100	0.1	0.05	Any $p \in [0.43, 0.57]$	0.55	Success for any n within N
100	0.1	0.05	Any $p \in [0.44, 0.56]$	0.6	Success for any n within N
100	0.1	0.05	Any $p \in [0.45, 0.55]$	0.65	Success for any n within N
100	0.1	0.05	Any $p \in [0.45, 0.55]$	0.7	Success for any n within N
100	0.1	0.05	Any $p \in [0.45, 0.55]$	0.75	Success for any n within N
100	0.1	0.05	Any $p \in [0.46, 0.54]$	0.8	Success for any n within N
100	0.1	0.05	Any $p \in [0.46, 0.54]$	0.85	Success for any n within N
100	0.1	0.05	Any $p \in [0.46, 0.54]$	0.90	Success for any n within N
100	0.1	0.05	Any $p \in [0.46, 0.54]$	0.95	Success for any n within N

From above table, we see that many admissible choices of p values are available to perform the RR survey with any sample size within N and to use Chaudhuri et al. (2011)'s second estimator of θ based on the distinct units' multiple randomized responses.

5 Conclusion

The well known Chebyshev's inequality to determine the sample size prior to conducting a randomized response based survey for sensitive variables are attempted recently by many statisticians. It has been observed that in many cases the acceptable sample sizes are hard to be obtained. The reason is that there is one part that may be controlled by an appropriately chosen sampling design and there is another part that is not easy to get it suitably and naturally controlled. Various situations are illustrated and solutions are proposed in Chaudhuri and Sen (2020), Chaudhuri and Patra (2023) and others. In this paper we have made an attempt to deterimine the sample size corresponding to the estimators of sensitive population proportion using multiple randomized responses from distinct persons sampled as in Mangat et al. (1995), Chaudhuri, Bose and Dihidar (2011). We have derived the necessary theoretical equations to find the sample sizes based on our commitment as our estimator t to be so accurate that for a given positive proper fraction f, say, $f = 0.1$, or $f = 0.2$, etc., we want to maintain the absolute difference in between the estimate values and the unknown fixed θ value to be less than or equal to $f\theta$ for many of the all possible samples, i.e. for $(1-\alpha)\times 100$ cases, choosing a small positive value of α close to 0, say as 0.1, 0.01, 0.05, etc. We also tried to follow Chaudhuri and Dutta (2019)'s approach to use the unbiased variance estimator in case it is difficult for the variance expression to have the closed form expression. But due to the frequently obtained high value estimates of CV's which are far away from the T values for commonly taken values of f and α, instead of using that approach, we decide to find the range of n following the Chebyshev's inequality utilizing some comparison criteria. We have presented the results of our numerical illustrations.

Based on the results of our numerical illustrations, it can be concluded that though in general the sample sizes obtained for RR surveys are seen to be absurd in many cases, the results for distinct units RRs based estimators are rather impressive. However we see that for Mangat et al (1995)'s estimator based on the distinct units' randomized responses once, very large value of the device parameter p has some solution for n, but also not within N, much exceeding N. But so large value of p may not be acceptable for a respondent because of harming their privacy and at the same time the value of n exceeding much of N together conclude the impractical result for conducting a RR survey and to use Mangat et al (1995)'s estimator. Instead we see that many admissible choices of p values are available to perform the RR survey with any sample size within N and to use Chaudhuri et al. (2011)'s both the estimators of θ based on the distinct units' multiple randomized responses. So this result may be used to choose the sample size in the real practical RR surveys scenario as per the available budget of conducting the survey. This is the justification of this research.

Acknowledgment

The author is grateful to an anonymous referee for his/her fruitful comments and suggestions which led to the improvement of the earlier manuscript.

References

[1] Basu, D. (1958): On sampling with and without replacement. *Sankhya.* **20**. 287-294.

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

[3] Chaudhuri, A. (2011): Randomized response and indirect questioning techniques in surveys. Chapman & Hall, CRC Press, Taylor & Francis Group, Boca Raton, FL. USA.

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

[5] Chaudhuri, A. (2018): Survey Sampling. Taylor & Francis, CRC, Boca Raton, Florida, USA.

[6] Chaudhuri, A. (2020): A Review on Issues of Settling the sample-size in surveys: Two approaches: Equal and varying probability sampling - Crises in sensitive Issues. *Calcutta Statistical Association Bulletin.* SAGE. https://doi.org/10.1177/000806832095238

[7] Chaudhuri, A., Bose, M. and Dihidar, K. (2011): Estimating sensitive proportions by Warner's randomized response technique using multiple randomized responses from distinct persons sampled. *Statistical Papers.* **52**, 111-124.

[8] Chaudhuri, A. and Christofides, T. C. (2013): Indirect questioning in Sample Surveys, Springer-Verlag, Berlin.

[9] Chaudhuri, A. and Dutta, T. (2018): Determining the size of a sample to take from a finite population. *Statistics and Applications.* (ISSN 2452-7395 (on line)), **16(1)**. New Series, 37-44.

[10] Chaudhuri, A. and Dutta, T. (2019): Pursuing Further with an Innovative Approach the Issue to Settle the Size of a Sample to Draw from a Finite Survey Population. *Statistics and Applications.* (ISSN 2452- 7395 (on line)), **17(1)**. New Series, 63-72.

[11] Chaudhuri, A. and Patra, D. (2023): Fixing the size of a sample to draw in a randomized response survey. *Pakistan Journal of Statistics.* **39(3)**. 289-298.

[12] Chaudhuri, A. and Sen, A. (2020): Fixing the Sample-Size in Direct and Randomized Response Surveys. *Journal of The Indian Society of Agricultural Statistics.* **74(3)**. 201-208.

[13] Cochran, W. G. (1953, 1963, 1977): Sampling Techniques. 1st, 2nd, 3rd ED., John Wiley & Sons, N.Y. USA.

[14] Korwar, R. M. and Serfling, R. J. (1970): On averaging over distinct units in sampling with replacement. *Annals of Mathematical Statistics.* **41**, 2132-2134.

[15] Mangat N. S., Singh, R., Singh, S., Bellhouse, D. R. and Kashani, H. B. (1995): On efficiency of estimator using distinct respondents in randomized response survey. *Survey Methodology.* **21**, 21-23.

[16] Pathak, P. K. (1962): On simple random sampling with replacement. *Sankhya A.* **24**, 287-302.

[17] Raj, D. and Khamis, S. H. (1958): Some remarks on sampling with replacement. *Annals of Mathematical Statistics.* **39**, 550-557.

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