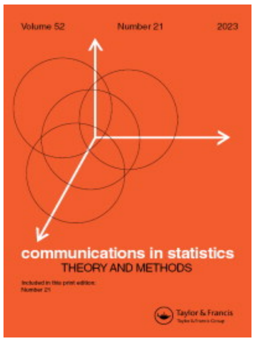

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

How to use randomized response survey data at hand by a specific procedure to judge its efficiency versus a possible rival

Arijit Chaudhuri & Dipika Patra

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2023.2250489

How to use randomized response survey data at hand by a specific procedure to judge its efficiency versus a possible rival

Arijit Chaudhuri[a] and Dipika Patra[b]

[a]Indian Statistical Institute (ISI) Kolkata, Kolkata, West Bengal, India; [b]Seth Anandram Jaipuria College, Kolkata, West Bengal, India

ABSTRACT
It is well-known that given survey data at hand by a given method, it is possible to examine how this method promises to fare vis- à -vis another one that might have been employed but not actually implemented. We illustrate how this may be extended to cover Randomized Response (RR) Techniques (RRT) by citing a few RR procedures in combination with a few sample selection methods generally in use. Developing the theory, simulated results are presented to demonstrate interesting numerical findings in this context.

ARTICLE HISTORY
Received 19 December 2022
Accepted 15 August 2023

KEYWORDS
Equal and varying probability sampling; RR procedure; stigmatizing issues; qualitative and quantitative features

AMS SUBJECT CLASSIFICATION:
62D05

1. Introduction

Cochran (1977) gave a formula to examine the gain in efficiency in stratified simple random sampling over a simpler procedure of unstratified simple random sampling both without replacement. Rao (1961) gave an interesting alternative with a novel approach for the same purpose. Chaudhuri and Pal (2022) mentioned a possible application of Rao (1961)'s approach in order to examine if the sample survey data gathered by a rather complex selection method and employed to estimate a finite population total or mean by an unbiased estimator, may yield an unbiased variance estimate for it less than that by a rival sample selection and unbiased estimation method that might have been employed for the sake of simplicity but not actually implemented. Their book referred to the work by Chaudhuri and Samaddar (2022) which used this Rao (1961) approach illustrating strategies of sample selection by probability proportional to size without replacement (PPSWOR) combined with Des (1956) estimator of total versus Simple Random Sampling Without Replacement (SRSWOR) and similarly two other pairs of survey sampling strategies.

In this article, we attempt an extension of Rao (1961) approach to cover rival procedures of sample selection, RR data gathering techniques (RRT) and use of several unbiased estimators of finite population totals and means. Section 2 presents briefly the theoretical procedures of estimation of totals and estimations of their variances when the survey data are gathered by Warner's device. Sections 3 and 4 develop the estimation procedure in general approach. Section 5 discusses how to simulate data to present numerical findings to compare relative efficiencies of rival estimation procedures. Section 6 presents tables showing the findings and the conclusions to draw.

CONTACT Dipika Patra dipika.patra1988@gmail.com Seth Anandram Jaipuria College, Kolkata, West Bengal, India.
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

2. Estimation of finite population totals from survey data gathered by warner's RR procedure by SRSWR and PPSWR sampling methods

In human surveys, sometimes situations arise demanding coverage of sensitive, stigmatizing and incriminating issues. Enquirers feel delicacies to directly ask sampled persons such questions relating, for examples, to habitual drunkenness, unlawfully speedy driving, conjugal misbehaviors, experiences in induced abortions, fraudulent tax-evasions, false claims for public grants-in aid, experiences in treatment, for AIDS and expenses incurred for treatment of venereal diseases and similar items. Even if asked, truthful responses are hard to come by.

Warner (1965) gave his Randomized Response Technique (RRT) as a globally admired clue. To implement Warner's RRT, an enquirer approaches a respondent with a box containing a number of identical cards marked A and its complement A^c and requests for a truthful "Yes" or "No" responses on randomly drawing a card from the box about whether the card type A or A^c drawn "matches" or "does not match" his/her own characteristic A or A^c and return the card to the box. The enquirer is not to see the card type drawn and hence, the respondent believes not to have divulged his/her true feature and thus confidentiality or secrecy is protected.

Denoting the finite survey-population $U = (1, 2, \ldots, i, \ldots N)$ of N individuals and the vector $\underline{Y} = (y_1 y_2, \ldots, y_i, \ldots y_N)$ of values y_i for the i^{th} $(i \in U)$ population unit on a real variable y so that

$$y_i = \begin{cases} 1 & \textit{if i bears A} \\ 0 & \textit{if i bears } A^c \end{cases} ; \; i = 1, 2, \ldots, N$$

Further, we shall denote $y_i = 1 \; or \; 0$ as the y-value for a unit chosen on the i^{th} draw and $I_i = 1 \; or \; 0$ as the response from a unit chosen on the i^{th} draw in case the units are chosen "with replacement."

Denoting by $\underline{X} = (x_1 x_2, \ldots, x_i \ldots, x_N)$, a vector of N co-ordinates x_i for the units i of U on another variable x such that $0 < x_i < 1$ for $i = 1, 2, \ldots, N$ and $X = \sum_{i=1}^{N} x_i$ and these x_i's are called "Size measures." Then $p_i = \frac{x_i}{X}$ for $i = 1, 2 \ldots N$, are called "Normed size-measures."

Now suppose our intention is to estimate $Y = \sum_{i=1}^{N} y_i$ the total or the mean $\overline{Y} = \frac{Y}{N}$ and for this a sample s is drawn from U in n $(2 \leq n \leq N)$ draws by the probability proportional to size with replacement (PPSWR) method using the values of $\underline{X}$ supposed to be all fixed and known-the values y_i of course are unknown and cannot be ascertained but I_i values are gathered with respective probabilities $p_i = \frac{x_i}{X}$ in n independent draws. Thus, let $E_R V_R C_R$ denote operators for expectation, variance, and covariance with respect to RR method not only due to Warner but generically any RRT. Then, we work out

$$E_R\,(I_i) = p y_i + \left(1 - p\right)(1 - y_i)$$

for Warner's RRT on taking a proportion p of cards marked A and $(1-p)$ of cards marked A^c so that $0 < p < 1$.

Then, $r_i = \frac{I_i - (1-p)}{(2p-1)}$, taking $p \neq \frac{1}{2}$ has $E_R\,(r_i) = y_i$.

Also, $V_R\,(r_i) = \frac{1}{(2p-1)^2} V_R\,(I_i) = \frac{p(1-p)}{(2p-1)^2}$ because $I_i^2 = I_i$ and $y_i^2 = y_i$.

Using the I_i values from a PPSWR sample s in n draws, we get an estimator $e_{HH} = \frac{1}{n} \sum_{i=1}^{N} \frac{r_i}{p_i} f_{si}$ for which $E_P E_R\,(e_{HH}) = Y$ and $E = E_P E_R = E_R E_P$ is obvious from Arnab (1990). Here E_P, E_R

are the operators for design-based and RR-based expectation, respectively, and f_{si} denotes the number of times the i^{th} unit is selected in the sample s.

Now, $V_R\left(e_{HH}\right)=\frac{1}{n}\frac{p(1-p)}{(2p-1)^2}$ and $V\left(e_{HH}\right)=E_R E_P(e_{HH}-Y)^2=E_P E_R(e_{HH}-Y)^2$ implies

$$V\left(e_{HH}\right)=\frac{1}{n}\sum_{i=1}^{N}p_i(\frac{y_i}{p_i}-Y)^2+\frac{1}{n}\left(\frac{p\left(1-p\right)}{(2p-1)^2}\right)\sum_{i=1}^{N}\frac{1}{p_i}+\frac{n-1}{n}N\left(\frac{p\left(1-p\right)}{(2p-1)^2}\right)$$

It follows that

$$\begin{aligned}v(e_{HH})&=\frac{1}{n(n-1)}\sum_{i=1}^{n}f_{si}\left(\frac{r_i}{p_i}-e_{HH}\right)^2+\frac{1}{n}\left(\frac{p(1-p)}{(2p-1)^2}\right)\sum_{i=1}^{N}\frac{f_{si}}{p_i}\text{ has }E\left(v(e_{HH})\right)\\&=E_P E_R v(e_{HH})\\&=E_R E_P v(e_{HH})\\&=V(e_{HH})\end{aligned}$$

that is, $v\left(e_{HH}\right)$ is an unbiased estimator of the variance of the Hansen Hurwitz (HH) (Hansen and Hurwitz, 1943) estimator e_{HH} for Y in case a PPSWR sample s is drawn from U in n draws and RR data are obtained then by Warner (1965) RR with the PPSWR sampling and combining this with Hansen Hurwitz estimator is more complicated than SRSWR sampling and combining that with the expansion estimator coupled with Warner's RRT would be a simpler alternative.

In this case, for the expansion estimator denoted by $e_{SRSWR,RR}=\frac{N}{n}\sum_{i=1}^{n}r_i$, then variance would have been $V\left(e_{SRSWR,RR}\right)=\frac{N}{n}\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{N^2}{n}(\frac{p(1-p)}{(2p-1)^2})$.

Given the Warner-based RR survey data at hand gathered by PPSWR sampling in n draws an unbiased estimate of this variance would be derived as follows:

$$\hat{Y^2}=e_{HH}^2-v\left(e_{HH}\right)=(\sum_{i=1}^{N}\frac{r_i f_{si}}{np_i})^2-v\left(e_{HH}\right)$$

So, an unbiased estimate of the above $V\left(e_{SRSWR,RR}\right)$ would be writing $e_{HH}\left(y^2\right)$ based on y_i^2's though in this qualitative case $y_i^2=y_i$ for all i,

$$v\left(e_{SRSWR,RR}\right)=\frac{N}{n}\left(e_{HH}\left(y^2\right)-\frac{(\sum_{i=1}^{N}\frac{r_i f_{si}}{np_i})^2-v\left(e_{HH}\right)}{N}\right)+\frac{N^2}{n}\left(\frac{p\left(1-p\right)}{(2p-1)^2}\right)$$

Thus gain in efficiency of PPSWR, Hansen Hurwitz estimate based on Warner's RR's over a possible alternative of SRSWR, expansion estimate based on Warner's RR is

$$\begin{aligned}G_{RR}&=v\left(e_{SRSWR,RR}\right)-v\left(e_{HH}\right)\\&=\frac{1}{n}(N\sum_{i=1}^{N}\frac{r_i f_{si}}{np_i}-\left(\sum_{i=1}^{N}\frac{r_i f_{si}}{np_i})^2\right)+\frac{N^2}{n}(\frac{p\left(1-p\right)}{(2p-1)^2})-\left(1-\frac{1}{n}\right)v\left(e_{HH}\right)\end{aligned}$$

Now, let us consider a more general approach of the above in the following section.

3. A general approach of RR estimation

To estimate $= \sum_{i=1}^{N} y_i$, let a linear homogeneous unbiased estimator for Y based on DR data be

$$t\left(s,\underline{Y}\right) = \sum_{i \in s} b_{si} y_i$$

where b_{si}'s are constants and free of any element of $\underline{Y} = (y_1, y_2 \ldots, y_N)$ and $b_{si} = 0$ for $i \notin s$. The estimator $t\left(s, \underline{Y}\right)$ will be unbiased for Y in the sense $E_P\left(t\left(s, \underline{Y}\right)\right) = Y$ if and only if $\sum_{s \ni i} p\left(s\right) b_{si} = 1\, \forall i$. Here $p(s)$ is the selection probability of a sample s.

Then the variance of $t\left(s, \underline{Y}\right)$ is

$$V\left(t\left(s,\underline{Y}\right)\right) = E_P(t\left(s,\underline{Y}\right) - Y)^2 = E_P(\sum_{i \in s} b_{si} y_i - Y)^2 = \sum_{i=1}^{N} y_i^2 a_{ii} + \sum_{i \neq}^{N} \sum_{j=1}^{N} a_{ij} y_i y_j$$

where $a_{ii} = E_P\left(b_{si}^2\right) - 1 = \sum_{s \ni i} b_{si}^2 p(s) - 1$ and $a_{ij} = E_P\left(b_{si} b_{sj}\right) - 1 = \sum_{s \ni i,j} b_{si} b_{sj} p(s) - 1$.

Since $\underline{Y}$ is unknowable, let us consider $\underline{R} = (r_1 r_{2,} \ldots r_N)$. These r_i's are obtained from a suitable RR device and assumed to follow the model R :

$$R: \quad \begin{array}{c} E_R\left(r_i\right) = y_i \\ V_R\left(r_i\right) = \varphi_i \\ C_R\left(r_i, r_j\right) = 0 \; for \; i \neq j = 1, 2 \ldots N \end{array} \quad \text{where } \varphi_i (> 0) \text{ is a function of } y_i \text{ alone.}$$

Arnab (1990) proposed the above general RR model and it can be shown that most of the qualitative and quantitative RR models satisfy the model R It was also assumed that a non negative unbiased estimator of φ_i is $\hat{\varphi}_i$.

Now, replacing y_i by r_i in $t\left(s, \underline{Y}\right)$ we get the estimator

$$e\left(s,\underline{R}\right) = t\left(s,\underline{Y}\right) |_{\underline{Y}=\underline{R}} = \sum_{i \in s} b_{si} r_i \, . \tag{1}$$

Here $b_{si}'s$ are the constants free from r_i's as well as y_i's and $b_{si} = 0$ for $i \notin s$. The estimator will be unbiased for Y if $E_P E_R\left(e\left(s, \underline{R}\right)\right) = E_R E_P\left(e\left(s, \underline{R}\right)\right) = Y$.

Now, the variance of $e\left(s, \underline{R}\right)$ is

$$V\left(e\left(s,\underline{R}\right)\right) = E_P E_R (e\left(s,\underline{R}\right) - Y)^2 = V\left(t\left(s,\underline{Y}\right)\right) + \sum_{i} \alpha_{ii} \varphi_i$$

where $\alpha_{ii} = E_P(b_{si}^2)$ and $\varphi_i = V_R(r_i)$.

Arnab (1994) proposed an unbiased estimator of $V\left(e\left(s, \underline{R}\right)\right)$ as

$$v(e\,(s,R)\,) = \hat{V}\left(t\left(s,\underline{R}\right)\right) + \sum_{i \in s} b_{si} \hat{\varphi}_i$$

where $\hat{V}\left(t\left(s,\underline{R}\right)\right)$ is obtained by writing r_i in place of y_i in $\hat{V}\left(t\left(s,\underline{Y}\right)\right) = \sum_{i \in s} c_{si} y_i^2 + \sum_{i \neq} \sum_{j \in s} c_{sij} y_i y_j$, as an unbiased estimator of $V\left(t\left(s,\underline{Y}\right)\right)$. Here, $\sum_{s \supset i} p\left(s\right) c_{si} = E_P\left(b_{si}^2\right) - 1$ and $\sum_{s \supset i,j} p(s) c_{sij} = E_P\left(b_{si} b_{sj}\right) - 1$ for $i \neq j = 1, 2, \ldots, N$. We refer to Arnab (2017) for details. Below we consider different forms of b_{si} to deduce the Equation (1) in different well-known estimators like Horvitz-Thompson (HT 1952) estimator and Rao-Hartley-Cochran (RHC 1962) estimator.

1. If we consider $b_{si} = \frac{1}{\pi_i}$, $\alpha_{ii} = \frac{1}{\pi_i}$ and $a_{ij} = \frac{\pi_{ij}}{\pi_i \pi_j} - 1$, the unbiased estimator of Y defined in Equation (1) reduces to the **Horvitz-Thompson estimator** (1952):

$$e\left(s,\underline{R}\right)|_{b_{si}=\frac{1}{\pi_i}} = \sum_s \frac{r_i}{\pi_i} = e_{HT}$$

The variance and unbiased variance estimator of e_{HT} are

$$V\left(e_{HT}\right) = \sum_{i<j=1}^{N}\sum\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{\varphi_i}{\pi_i}\ and$$

$$v\left(e_{HT}\right) = \sum_{i,j=1}^{n}\sum_{j>i}\frac{\left(\pi_i\pi_j - \pi_{ij}\right)}{\pi_{ij}}\left(\frac{r_i}{\pi_i} - \frac{r_j}{\pi_j}\right)^2 + \sum_{i=1}^{n}\frac{\hat{\varphi}_i}{\pi_i}$$

2. Taking $b_{si} = \frac{Q_i}{p_i}$, the Equation (1) reduces to **Rao-Hartley-Cochran estimator** (1962):

$$e\left(s,\underline{R}\right)|_{b_{si}=\frac{Q_i}{p_i}} = \sum_s \frac{r_i Q_i}{p_i} = e_{RHC}$$

Here, $Q_i = \sum_{j=1}^{N} {}_iP_{ij}$. The variance $V\left(e_{RHC}\right)$ and unbiased variance estimator $v\left(e_{RHC}\right)$ are as follows:

$$V\left(e_{RHC}\right) = V_P\left(\sum_{i=1}^{n} y_i\frac{Q_i}{p_i}\right) + E_P\left(\sum_{i=1}^{n} V_i\left(\frac{Q_i}{p_i}\right)^2\right)$$

$$= \left(\frac{\sum_{i=1}^{n} N_i^2 - N}{N\left(N-1\right)}\right)\sum_{i=1}^{N} p_i\left(\frac{y_i}{p_i} - Y\right)^2 + \sum_{i=1}^{N}\varphi_i\left(\frac{Q_i}{p_i}\right)$$

$$v\left(e_{RHC}\right) = \left(\frac{\sum_{i=1}^{n} N_i^2 - N}{N^2 - \sum_{i=1}^{n} N_i^2}\right)\sum_{i=1}^{n} Q_i\left(\frac{r_i}{p_i}\right)^2 - e_{RHC}^2 + \sum_{i=1}^{n}\hat{\varphi}_i\frac{Q_i}{p_i}.$$

3. For Lahiri–Midzuno-Sen (1951, 1952, 1953) sampling scheme, if $b_{si} = \frac{1}{p_s}$; $p_s = \sum_{i\epsilon s} p_i$, we get **the Ratio estimator** of Y as $e_{LMS} = \frac{\sum_{i=1}^{n} r_i}{p_s}$ and the unbiased variance estimator is

$$v\left(e_{LMS}\right) = \sum_{i<j\in s}\sum\left(\frac{N-1}{n-1}\frac{1}{p_s} - \frac{1}{p_s^2}\right)p_i p_j\left[\left(\frac{r_i}{p_i} - \frac{r_j}{p_j}\right)^2 - \left(\frac{\hat{\varphi}_i}{p_i^2} + \frac{\hat{\varphi}_j}{p_j^2}\right)\right] + \sum_s \frac{\hat{\varphi}_i}{p_s^2}.$$

For many aspects of this section, Adhikary (2016) is handy to consider.

4. For the PPSWOR sample in n draws selection-probabilities respective draw-wise are $\frac{p_{ir}}{1 - p_{i1} - p_{i2} - \ldots - p_{i(r-1)}}$, $r = 1, 2, \ldots, n$ and unbiased estimators for Y draw-wise are $e_1 = \frac{r_{i1}}{p_{i1}}$, $e_2 = r_{i1} + \frac{r_{i2}}{p_{i2}}\left(1 - p_{i1}\right), \ldots, e_n = r_{i1} + r_{i2} + \ldots + r_{i(n-1)} + \frac{r_{in}}{p_{in}}(1 - p_{i1} - p_{i2} - \ldots - p_{i(n-1)})$

Des Raj estimator of Y is $e_D = \frac{1}{n}\left(e_1 + e_2 + \ldots + e_n\right)$ and its unbiased variance estimate is $v\left(e_D\right) = v_P\left(e_D\right) + \sum_{i=1}^{n}\hat{\varphi}_i = v_P\left(e_D\right) + e_D(v)$ where $v_P\left(e_D\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(e_i - \bar{e}\right)^2 =$

Table 1. Different RR procedures with $\mathbf{r_i}$, φ_i and $\hat{\varphi}_i$.

RR	$\mathbf{r_i}$	φ_i	$\hat{\varphi}_i$
URL (Greenberg et al. (1969) and Horvitz, Shah, and Simmons (1967))	$\frac{(1-p_2)I_i-(1-p_1)J_i}{(p_1-p_2)}$	$\frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}(y_i-x_i)^2$	$r_i(r_i-1)$
Kuk (1990)	$\frac{f_i/k-\theta_2}{\theta_1-\theta_2}$	$\frac{(1-\theta_1-\theta_2)}{k(\theta_1-\theta_2)}y_i+\frac{\theta_2(1-\theta_2)}{k(\theta_1-\theta_2)^2}$	$\frac{(1-\theta_1-\theta_2)}{k(\theta_1-\theta_2)}r_i+\frac{\theta_2(1-\theta_2)}{k(\theta_1-\theta_2)^2}$
Christofides (2003)	$\frac{z_i-\mu}{M+1-2\mu}$	$\frac{\sum_{K=1}^{M}K^2p_K-\mu^2}{(M+1-2\mu)^2}$; $\mu=\sum_{K=1}^{M}Kp_K(known)$	–
Boruch (1972)	$\frac{I_i-p_1}{(1-p_1-p_2)}$	$\frac{(p_2-p_1)}{(1-p_1-p_2)}y_i+\frac{p_1(1-p_1)}{(1-p_1-p_2)^2}$	$\frac{(p_2-p_1)}{(1-p_1-p_2)}r_i+\frac{p_1(1-p_1)}{(1-p_1-p_2)^2}$
Mangat and Singh (1990)	$\frac{z_i-(1-T)(1-p)}{T+(1-T)(2p-1)}$	$\frac{(1-T)(1-p)[T+(1-T)p]}{(T+(1-T)(2p-1))^2}$ $(known)$	–
Chaudhuri (2011)'s Device I	$\frac{z_i-\gamma}{\mu}$	$\frac{\sigma^2}{\mu^2}y_i^2+\frac{\psi^2}{\mu^2}$	$\frac{\frac{\sigma^2}{\mu^2}r_i^2+\frac{\psi^2}{\mu^2}}{1+\frac{\sigma^2}{\mu^2}}$
Chaudhuri (2011)'s Device II	$\frac{1}{C}\left(z_i-\sum_{j=1}^{m}q_jx_j\right)$	$\alpha y_i^2+\beta y_i+\gamma$; where $\alpha=C(1-C)$, $\beta=-2C(\sum_{j=1}^{M}q_jx_j)$, $\gamma=\sum q_jx_j^2-(\sum q_jx_j)^2$	$\frac{\alpha r_i^2+\beta r_i+\gamma}{1+\alpha}$

$\frac{1}{2n^2(n-1)}\sum\sum_{i\neq j}(e_i-e_j)^2$ and $e_D(v)$ is the Des Raj estimator of $\sum_{i=1}^{N}V_i$, replacing r_{ij} term in (2) by $v_{ij}=r_{ij}\left(r_{ij}-1\right)$.

In the following table (Table 1), we mention a few well-known qualitative and quantitative RR devices other than Warner's RR device with its $\mathbf{r_i}$, φ_i and $\hat{\varphi}_i$. For detailed study of different RR models, we refer to Chaudhuri (2011) and Arnab (2017).

4. A general approach for gain in efficiency

We have already given our theory for evaluating gain in efficiency in employing PPSWR sampling combined with Hansen-Hurwitz estimator by Warner's RRT data over SRSWR plus expansion estimator with the same RRT data. For each of the other RRT's qualitative as well as quantitative RRT data the change is obvious, simply using the right $V_R(r_i)'s$ already all given in Table 1. Now let us present the same in general approach.

In this section, we intend to examine how much efficiency we have gain using the RR data at hand compared to a possibility that we might instead have gone for an SRSWOR of the same sample size.

For SRSWOR, unbiased estimator of Y is the expansion estimator which is $N\bar{r}$, with $\bar{r}=\frac{1}{n}\sum_{i\in s}r_i$. In this case, $b_{si}=\frac{1}{\pi_i}=\frac{N}{n}$ and $\pi_{ij}=\frac{n(n-1)}{N(N-1)}$

$$V\left(N\bar{r}\right)=V\left(N\bar{y}\right)+\frac{N}{n}\sum_{i=1}^{N}\varphi_i=N^2\left(\frac{N-n}{Nn}\right)\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{N}{n}\sum_{i=1}^{N}\varphi_i$$

$$=C_1\left[\sum_{i=1}^{N}y_i^2-\frac{Y^2}{N}\right]+\frac{N}{n}\sum_{i=1}^{N}\varphi_i,\ \text{writing } C_1=N(\frac{N-n}{n})\frac{1}{N-1}$$

Now, $V\left(e_{Design}\right) = E\left(e^2_{Design}\right) - Y^2$. It is true for any design.

So, from PPSWOR in n draws $\hat{Y^2} = e^2_{Design} - v(e_{Design})$.

Also, $\sum_{i=1}^{N} y_i^2$ is unbiasedly estimated by $e_{Design}(y^2)$ which is e_{Design} with each y in e_{Design} replaced by y^2, of course, for qualitative cases $y_i^2 = y_i$ because each y_i equals only 1 or 0. It should be noted that for quantitative cases $E_R\left(r_i^2 - v_R(r_i)\right) = y_i^2$. In other words, $r_i^2 - v_R(r_i)$ is an unbiased estimator of y_i^2 for quantitative cases.

So, unbiased estimate of $V(N\bar{r})$ based on PPSWOR is

$$v(N\bar{r}) = C_1\left[e_{Design}\left(y^2\right) - \frac{\hat{Y^2}}{N}\right] + \frac{N}{n}\widehat{\sum_{i=1}^{N}\varphi_i}.$$

So, the gain in efficiency is $G_{RR} = v(N\bar{r}) - v\left(e_{Design}\right)$ from above.

5. Numerical exercise and simulation

For both qualitative and quantitative characteristics, we utilize numerical data borrowing from the comprehensive paper by Chaudhuri, Christofides, and Saha (2009), Chaudhuri and Christofides (2013)'s book may be considered also. Here, $N = 116$, $Y = \begin{cases} 96\ (qualitative) \\ 105,336\ (quantitative) \end{cases}$. An innocuous character is also considered here for URL model. A size measure variable is taken to draw samples in varying probability sampling schemes. Different sample sizes with different parameter values are considered here for simulation. We

Table 2. φ_i and $\widehat{\sum_{i=1}^{N}\varphi_i}$, for different qualitativeand quantitative models.

RR	φ_i	$\widehat{\sum_{i=1}^{N}\varphi_i}$
Warner	$\frac{p(1-p)}{(2p-1)^2}$ (Known)	$\frac{Np(1-p)}{(2p-1)^2}$
URL	$\frac{(1-p_1)(1-p_2)(p_1+p_2-2p_1p_2)}{(p_1-p_2)^2}(y_i - x_i)^2$	$e_{Design}(V_i)$ E.g. $\sum_{i=1}^{n}\frac{v_i}{\pi_i}$ (Design: HT estimator)
Kuk	$\frac{(1-\theta_1-\theta_2)}{k(\theta_1-\theta_2)}y_i + \frac{\theta_2(1-\theta_2)}{k(\theta_1-\theta_2)^2}$	$\frac{(1-\theta_1-\theta_2)}{k(\theta_1-\theta_2)}e_{Design} + \frac{N\theta_2(1-\theta_2)}{k(\theta_1-\theta_2)^2}$
Christofides	$\frac{\sum_{K=1}^{M}K^2p_K-\mu^2}{(M+1-2\mu)^2}$; $\mu = \sum_{K=1}^{M}Kp_K(known)$	$\frac{N(\sum_{K=1}^{M}K^2p_K-\mu^2)}{(M+1-2\mu)^2}$
Boruch	$\frac{(p_2-p_1)}{(1-p_1-p_2)}y_i + \frac{p_1(1-p_1)}{(1-p_1-p_2)^2}$	$\frac{(p_2-p_1)}{(1-p_1-p_2)}e_{Design} + \frac{Np_1(1-p_1)}{(1-p_1-p_2)^2}$
Mangat and Singh	$\frac{(1-T)(1-p)[T+(1-T)p]}{(T+(1-T)(2p-1))^2}$ (known)	$\frac{N(1-T)(1-p)[T+(1-T)p]}{(T+(1-T)(2p-1))^2}$
Device I	$\frac{\sigma^2}{\mu^2}y_i^2 + \frac{\psi^2}{\mu^2}$	$\frac{\sigma^2}{\mu^2}e_{Design}(y^2) + N\frac{\psi^2}{\mu^2}$
Device II	$\alpha y_i^2 + \beta y_i + \gamma$; where $\alpha = C(1-C)$, $\beta = -2C(\sum_{j=1}^{M}q_jx_j), \gamma = \sum q_jx_j^2 - (\sum q_jx_j)^2$	$\alpha e_{Design}(y^2) + \beta e_{Design} + N\gamma$

Table 3. Numerical study for HH estimator.

RR	Sample size	Parameter values	e_{HH}	$v(e_{SRSWR,RR})$ $(I+\left(\frac{N}{n}\right)II)$	I	II $(\sum_{i=1}^{\hat{N}} \varphi_i)$	$v(e_{HH})$	G_{RR}	RelativeG_{RR}
Warner	30	$p = 0.3$	94.0054	672.1424	83.4423	152.25	435.6612	236.4812	35.1832
	35	$p = 0.4$	89.5121	2400.621	93.8783	696	914.7597	1485.862	61.8948
	40	$p = 0.7$	86.8784	513.2048	71.6797	152.25	337.1536	176.0512	34.3043
URL	30	$(p_1p_2) = (0.25,0.55)$	99.3856	618.5839	70.3993	141.7719	460.7506	157.8332	25.5152
	35	$(p_1p_2) = (0.25,0.65)$	104.7705	556.6212	46.7251	153.8479	458.8608	97.7604	17.5632
	40	$(p_1p_2) = (0.4,0.6)$	108.1509	979.7275	36.4904	325.2542	610.731	368.9965	37.6632
Kuk	30	$(\theta_1\theta_2k) = (0.4,0.7,2)$	90.9360	671.7671	89.8750	150.4893	417.0333	254.7339	37.9199
	35	$(\theta_1\theta_2k) = (0.35,0.55,2)$	85.2623	1209.026	90.2580	337.5594	538.2643	670.7621	55.4795
	40	$(\theta_1\theta_2k) = (0.4,0.65,2)$	85.6297	497.8537	73.1333	146.4553	324.7316	173.1221	34.7736
Christofides	30	$(p_1p_2p_3p_4k) = (0.2,0.4,0.3,0.1, 4)$	112.338	2317.737	47.0368	587.25	999.722	1318.015	56.8664
	35	$(p_1p_2p_3p_4k) = (0.3,0.4,0.2,0.1,4)$	102.2405	587.7682	53.1325	161.3125	452.863	134.9052	22.9521
	40	$(p_1p_2p_3p_4k) = (0.2,0.4,0.2,0.1,0.1,5)$	87.7579	558.5937	70.8136	168.2	354.0786	204.5151	36.6125
Boruch	30	$(p_1p_2) = (0.3,0.4)$	86.0889	1261.528	103.9908	299.363	544.709	716.8186	56.8215
	35	$(p_1p_2) = (0.2, 0.6)$	113.608	2328.503	37.6116	691.2171	1044.717	1283.786	55.1335
	40	$(p_1p_2) = (0.35,0.25)$	80.1432	499.5675	79.35257	144.9017	300.4238	199.1437	39.86322
Mangat and Singh	30	$(T, p) =(0.3,0.45)$	110.1709	2059.6	52.0110	519.2042	918.1363	1141.464	55.4216
	35	$(T, p) = (0.6,0.3)$	83.8506	486.3934	86.0496	120.7934	315.9933	170.4002	35.0334
	40	$(T, p) = (0.4,0.6)$	71.77607	311.4234	84.5027	78.24852	205.8897	105.5337	33.88753
Device I	30	$(\mu, \gamma) = (4.75,25)$	98030.44	402578858	213281007	48956341	359017542	43561316	10.82057
	35	$(\mu, \gamma) = (5.75,21.75)$	90004.03	312717957	206167369	32148884	257224983	55492974	17.74538
	40	$(\mu, \gamma) = (5.75,22.5)$	93493.97	250588292	173399887	26616692	239847949	10740343	4.286051
Device II	30	$C = 0.4$	97313.65	1286093324	931194002	91784307	401122565	884970759	68.81077
	35	$C = 0.5$	111307.8	706185262	482885261	67375000	409745412	296439849	41.97763
	40	$C = 0.6$	96540.03	437593831	298051560	48118025	276529576	161064255	36.80679

Table 4. Numerical study for DesRaj estimator.

RR	Sample size	Parameter value	e_D	$v(N\bar{r})$ $(I + \left(\frac{N}{n}\right) II)$	I	II $(\sum_{i=1}^{\hat{N}} \varphi_i)$	$v(e_D)$	G_{RR}	RelativeG_{RR}
Warner	30	$p = 0.3$	94.47683	653.297	64.59696	152.25	557.9488	95.34816	14.59492
	35	$p = 0.4$	90.46729	2394.628	87.8847	696	2057.235	337.3928	14.08957
	40	$p = 0.7$	88.98523	487.8514	46.32637	152.25	400.0489	87.8025	17.9978
URL	30	$(p_1 p_2) = (0.25,0.55)$	106.6275	713.8531	39.39523	174.4288	581.0268	132.8264	18.60696
	35	$(p_1 p_2) = (0.25,0.65)$	88.2961	399.2426	54.27368	104.0855	250.7867	148.4559	37.18439
	40	$(p_1 p_2) = (0.4,0.6)$	96.78497	850.9591	39.2042	279.9155	513.1597	337.7994	39.69632
Kuk	30	$(\theta_1 \theta_2 k) = (0.4,0.7,2)$	93.03297	651.3142	68.0707	150.8388	594.0522	57.26192	8.791752
	35	$({}_1\theta_2 k) = (0.35,0.55,2)$	75.05203	1209.889	82.66068	340.112	1034.293	175.5959	14.51339
	40	$(\theta_1 \theta_2 k) = (0.4,0.65,3)$	105.253	184.052	21.04642	56.20883	142.7088	41.3432	22.46278
Christofides	30	$(p_1 p_2 p_3 p_4 k) = (0.2,0.4,0.3,0.1,4)$	84.05548	2390.934	120.2343	587.25	2138.24	252.6947	10.56887
	35	$(p_1 p_2 p_3 p_4 k) = (0.3,0.4,0.2,0.1,4)$	84.99945	597.7536	63.11791	161.3125	501.3849	96.36875	16.12182
	40	$(p_1 p_2 p_3 p_4 k) = (0.2,0.4,0.2,0.1,0.1,5)$	84.31026	538.1238	50.34376	168.2	375.3521	162.7716	30.24799
Boruch	30	$(p_1 p_2) = (0.30,0.40)$	91.37146	1246.151	81.80583	301.1238	1031.4	214.7517	17.2332
	35	$(p_1 p_2) = (0.20,0.60)$	96.67454	2254.503	75.8608	657.3491	1901.346	353.1572	15.66452
	40	$(p_1 p_2) = (0.35,0.25)$	73.75693	483.0846	58.23967	146.4983	409.3133	73.7713	15.27088
Mangat and Singh	30	$(T, p) = (0.3,0.45)$	92.0819	2105.61	98.0206	519.2042	1729.798	375.8124	17.84815
	35	$(T, p) = (0.6,0.3)$	103.9994	432.8063	32.4625	120.7934	365.0539	67.75239	15.65421
	40	$(T, p) = (0.4,0.6)$	84.50195	274.7262	47.80553	78.24852	231.8459	42.88034	15.60839
Device I	30	$(\mu, \gamma) = (2)$	99619.41	285811931	113727547	44504582	260040924	25771007	9.016771
	35	$(\mu, \gamma) = (5.75,21.75)$	99449.4	233619788	122219732	33612086	218686101	14933687	6.392304
	40	$(\mu, \gamma) = (5.75,22.5)$	113532.2	213535136	98141319	39790971	190508283	23026853	10.78364
Device II	30	$C = 0.4$	103408.8	1014880428	664650935	90576593	956189735	58690694	5.783016
	35	$C = 0.5$	119220.5	358242473	195184912	49198402	334833881	23408592	6.534287
	40	$C = 0.6$	107130.6	328434691	191380841	47259948	298458283	29976408	9.127053

Table 5. Numerical illustration for RHC estimation.

RR	Sample size	Parameter values	e_{RHC}	$v_{SRSWR,RHC}$ $(I+\left(\frac{N}{n}\right)II)$	I	II $(\sum_{i=1}^{\hat{N}}\varphi_i)$	$v(e_{RHC})$	G_{RR}	RelativeG_{RR}
Warner	30	$p=0.7$	82.71556	674.0837	85.38366	152.25	672.1336	1.950078	0.289293
	35	$p=0.4$	83.76781	2407.58	100.8368	696	2310.697	96.8826	4.024066
	40	$p=0.3$	80.18881	497.1356	55.61064	152.25	494.2505	2.885125	0.580349
URL	30	$(p_1p_2)=(0.45,0.55)$	116.4334	2400.6	50.98201	607.6598	2095.671	304.9285	12.70218
	35	$(p_1p_2)=(0.25,0.75)$	91.15981	181.0729	49.06172	39.83095	173.5163	7.556588	4.17323
	40	$(p_1p_2)=(0.4,0.6)$	118.8408	628.1583	4.690581	214.9889	621.5088	6.649483	1.058568
Kuk	30	$(\theta_1\theta_2k)=(0.7,0.4,2)$	70.00256	648.1851	95.25346	142.9996	601.2756	46.90949	7.237052
	35	$(\theta_1\theta_2k)=(0.6,0.2,2)$	81.23329	322.0101	62.47394	78.30832	280.2007	41.80936	12.98387
	40	$(\theta_1\theta_2k)=(0.4,0.8,3)$	89.84693	197.5675	42.00815	53.64116	192.8254	4.742125	2.400255
Christofides	30	$(p_1p_2p_3p_4k)=(0.2,0.4,0.3,0.1,4)$	89.77156	2388.549	117.8493	587.25	2373.109	15.44068	0.646446
	35	$(p_1p_2p_3p_4k)=(0.3,0.4,0.2,0.1,4)$	110.7706	557.0138	22.37806	161.3125	532.7332	24.28057	4.359062
	40	$(p_1p_2p_3p_4k)=(0.23,0.2,0.23,0.23,0.11,5)$	114.4068	3384.114	54.63156	1148.098	3124.373	259.7409	7.6753
Boruch	30	$(p_1p_2)=(0.25,0.20)$	89.30057	313.7125	67.08646	63.78259	306.9851	6.727364	2.144436
	35	$(p_1p_2)=(0.2,0.5)$	85.26495	295.3854	58.56638	71.45402	289.6205	5.764874	1.951645
	40	$(p_1p_2)=(0.3,0.3)$	98.78468	477.4575	35.93247	152.25	474.2499	3.20754	0.671796
Mangat and Singh	30	$(T,p)=(0.3,0.6)$	84.28277	547.283	80.21525	120.7934	544.7212	2.561868	0.468106
	35	$(T,p)=(0.6,0.8)$	100.0099	73.71064	33.60853	12.09977	70.88166	2.828976	3.837949
	40	$(T,p)=(0.4,0.45)$	89.19868	693.9094	50.5006	221.8651	665.9802	27.92921	4.024907
Device I	30	$(\mu,\gamma)=(2)$	108137	360779229	150423079	54402453	337720983	23058246	6.391234
	35	$(\mu,\gamma)=(5.75,21.75)$	113583.7	267462795	131684044	40967727	254475090	12987705	4.855892
	40	$(\mu,\gamma)=(5.75,22.5)$	113343.3	189446736	80804565	37462818	188889954	556782.6	0.293899
Device II	30	$C=0.6$	103036.1	409662882	254197569	40206546	400406984	9255898	2.259394
	35	$C=0.5$	106056	476137721	292847416	55303110	472565929	3571793	0.750159
	40	$C=0.4$	112520.8	572255886	357660563	73998387	567898773	4357113	0.761392

Table 6. Numerical illustration for HT estimator.

RR	Sample size	Parameter value	e_{HT}	$v_{SRSWOR,HT}$ $(I+\left(\frac{N}{n}\right) II)$	I	II $(\sum_{i=1}^{\hat{N}} \varphi_i)$	$v(e_{HT})$	G_{RR}	RelativeG_{RR}
Warner	30	$p = 0.3$	86.9956	668.2478	79.54782	152.25	667.9071	0.3407225	0.050987
	35	$p = 0.4$	99.3096	2388.154	81.41084	696	2387.898	0.2557052	0.010707
	40	$p = 0.7$	108.7086	462.2528	20.7278	152.25	461.9396	0.3132072	0.067756
URL	30	$(p_1 p_2) = (0.25,0.55)$	96.31851	791.7799	62.82621	188.5225	624.6624	167.1175	21.10656
	35	$(p_1 p_2) = (0.25,0.65)$	93.08111	409.8009	49.3765	108.7487	320.2695	89.53132	21.84752
	40	$(p_1 p_2) = (0.4,0.7)$	114.219	395.8766	9.110023	133.3678	347.9689	47.90761	12.10165
Kuk	30	$(\theta_1 \theta_2 k) = (0.7,0.4,2)$	71.83497	547.1368	74.20754	142.6942	514.8798	32.25698	5.895596
	35	$(\theta_1 \theta_2 k) = (0.6,0.2,2)$	103.5487	310.018	31.99191	83.88718	300.4065	9.611459	3.100291
	40	$(\theta_1 \theta_2 k) = (0.4,0.8,3)$	99.06987	190.8004	30.78333	55.17831	185.9354	4.865005	2.549787
Christofides	30	$(p_1 p_2 p_3 p_4 k) = (0.2,0.4,0.3,0.1,4)$	77.28757	2401.606	130.9062	587.25	2259.482	142.1243	5.917883
	35	$(p_1 p_2 p_3 p_4 k) = (0.3,0.3,0.3,0.1,4)$	93.87059	1087.193	61.97369	309.3333	1002.257	84.93617	7.81243
	40	$(p_1 p_2 p_3 p_4 k) = (0.23,0.23,0.23,0.11,0.2,5)$	90.26807	5387.224	124.2248	1814.827	5196.097	191.1266	3.547776
Boruch	30	$(p_1 p_2) = (0.3,0.4)$	90.18269	1252.058	89.24488	300.7276	1251.898	0.1600751	0.012784
	35	$(p_1 p_2) = (0.2, 0.2)$	110.4581	1083.691	34.12206	316.6803	1083.422	0.268762	0.024801
	40	$(p_1 p_2) = (0.4,0.5)$	93.92581	216.9602	46.09033	51.55556	216.9548	0.005360967	0.002471
Mangat and Singh	30	$(T,p) = (0.3,0.45)$	108.4247	2079.861	72.27136	519.2042	2077.913	1.947572	0.093639
	35	$(T,p) = (0.6,0.3)$	99.4203	442.4133	42.0695	120.7934	442.1317	0.2815831	0.063647
	40	$(T,p) = (0.6,0.4)$	96.99773	261.6923	34.77154	78.24852	261.4161	0.276103	0.105506
Device I	30	$(\mu,\gamma) = (2)$	113244.2	357123287	138905786	56435560	356403504	719782.9	0.201550
	35	$(\mu,\gamma) = (5.75,21.75)$	113985.4	233150263	105705536	38453150	232870526	279736.9	0.119981
	40	$(\mu,\gamma) = (5.75,22.5)$	113557.1	251637815	119991650	39720826	251444882	192932.9	0.076671
Device II	30	$C = 0.4$	114436.5	957095095	633742424	83625691	955620004	1475090	0.154122
	35	$C = 0.5$	118392.7	612103074	390826077	66764611	611847676	255398.2	0.041725
	40	$C = 0.6$	104219.5	337668700	198358328	48038059	337269440	399259.5	0.11824

Table 7. Numerical illustration for ratio estimator.

RR	Sample size	Parameter value	e_{LMS}	$v_{SRSWOR,LMS}$ I $+ \left(\frac{N}{n}\right)$ II	I	II $(\sum_{i=1}^{\hat{N}} \varphi_i)$	$v(e_{LMS})$	G_{RR}	RelativeG_{RR}
Warner	30	$p = 0.3$	103.5386	636.5365	47.83646	152.25	628.7862	7.750236	1.217564
	35	$p = 0.4$	114.7051	2355.894	49.15164	696	2293.881	62.01332	2.632262
	40	$p = 0.7$	103.147	470.985	29.46004	152.25	457.3574	13.62765	2.893435
URL	30	$(p_1p_2) = (0.25,0.55)$	97.22797	733.7927	61.12204	173.9666	626.8229	106.9698	14.57767
	35	$(p_1p_2) = (0.25,0.65)$	101.1195	341.7829	37.06471	91.94083	337.084	4.698911	1.374823
	40	$(p_1p_2) = (0.4,0.7)$	95.62189	336.3876	37.4976	103.0655	320.9983	15.38932	4.574878
Kuk	30	$(\theta_1\theta_2k) = (0.7,0.4,2)$	72.43992	645.0038	93.64283	142.5933	601.1132	43.89059	6.804702
	35	$(\theta_1\theta_2k) = (0.6,0.2,2)$	88.25329	320.6011	55.2484	80.06332	296.6295	23.9716	7.477078
	40	$(\theta_1\theta_2k) = (0.4,0.8,3)$	103.3261	186.3604	24.28613	55.88769	160.4091	25.95132	13.92534
Christofides	30	$(p_1p_2p_3p_4k) = (0.2,0.4,0.3,0.1,4)$	106.3129	2352.27	81.56971	587.25	2242.409	109.8606	4.670406
	35	$(p_1p_2p_3p_4k) = (0.3,0.3,0.3,0.1,4)$	108.1223	1245.204	49.11519	309.3333	1118.562	126.6424	10.17041
	40	$(p_1p_2p_3p_4k) = (0.23,0.23,0.23,0.11,0.2,5)$	105.113	5357.818	94.81965	1814.827	4594.714	763.104	14.24281
Boruch	30	$(p_1p_2) = (0.3,0.3)$	88.41615	665.318	76.61801	152.25	634.7719	30.54614	4.591209
	35	$(p_1p_2) = (0.2, 0.2)$	88.88427	223.6556	52.78571	51.55556	212.8315	10.82404	4.839605
	40	$(p_1p_2) = (0.2,0.3)$	93.1991	309.4818	40.13036	92.87982	303.92	5.561812	1.797137
Mangat and Singh	30	$(T,p) = (0.3,0.45)$	90.07977	2069.813	62.22348	519.2042	161.2863	1908.527	92.20769
	35	$(T,p) = (0.6,0.3)$	92.49639	444.9065	44.56267	120.7934	40.38073	404.5257	90.92377
	40	$(T,p) = (0.6,0.4)$	85.23108	271.3561	44.43538	78.24852	67.04195	204.3141	75.29374
Device I	30	$(\mu,\gamma) = (2)$	94582.05	262942518	106053999	40574617	230212560	32729958	12.44757
	35	$(\mu,\gamma) = (5.75,21.75)$	114580.2	255704235	121639455	40450580	224948506	30755729	12.02785
	40	$(\mu,\gamma) = (5.75,22.5)$	101523.4	221444964	115614057	12555103	210169224	11275740	5.091893
Device II	30	$C = 0.4$	104776.4	900425711	581984336	82355528	866615869	33809842	3.754873
	35	$C = 0.5$	96227.52	482750864	287235380	58991741	381658791	101092074	20.94084
	40	$C = 0.6$	99933.06	364953389	218231694	50593688	324831407	40121982	10.99373

have shown relative gain in efficiency using $100\frac{G_{RR,Design}}{v(e_{SRSWR(WOR),\,Design})}$ along with the actual gain in efficiency$(G_{RR,Design})$.

6. Tables

In Table 2 below, we present the details for evaluating $\widehat{\sum_{i=1}^{N}\varphi_i}$ using RR survey data at hand, for various RRT's. For example, Warner's RR model has $\varphi_i = \frac{p(1-p)}{(2p-1)^2}$ which is known. Therefore, $\widehat{\sum_{i=1}^{N}\varphi_i}$ for this model is just the sum of all φ_i's. However, φ_i of Kuk's model depends on the unknown y-value. Thus, $\sum_{i=1}^{N}\varphi_i$ can be estimated here from the PPS sample at hand and the estimator is mentioned in the third column of Table 2. Utilizing the formulae for various sampling strategies combined with various RRT's, we present the numerical evaluations of the gain in efficiency in employing complex strategies versus simpler alternatives in the Tables 3–7. Different parameter values are mentioned in the third column of the mentioned tables. For example, the parameter p is taken as .3, 0.4 and 0.7 for Warner's RR model. For URL model, parameters $(p_1 p_2)$ are taken as $(0.25, 0.55)$, $(0.25, 0.65)$ and $(0.4, 0.6)$ in Table 3. Estimated population total (e_{Design}) and its unbiased variance estimate $(v\left(e_{Design}\right))$ are given for each sampling design in fourth and eighth columns of the tables. Using the RR survey data at hand, unbiased estimate of variance for simpler alternative is computed and shown in the fifth column of Tables 3–7. Here "I" stands for the design-based estimated variance from PPS sample and "II" for the estimate of $\sum_{i=1}^{N}\varphi_i$.

7. Conclusion

From the values of the estimated variances, it can be concluded that (PPSWR-HH estimator) is substantially more gainful in efficiency than (SRSWR-Expansion estimator) irrespective of the considered RR models. Also, (PPSWOR-Des Raj estimator) is more gainful compared to (SRSWOR-Expansion estimator). Compared to Lahiri-Midzuno-Sen (LMS) sampling design -Ratio estimator and Horvitz Thompson estimator with SRSWOR-Expansion estimator, LMS-Ratio estimator is more gainful for each and every RR model.

Acknowledgments

The authors gratefully acknowledge the second referee for approving our entire first write-up and the 1st referee for recommending substantial revision which is implemented for a possible improvement on our first draft.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Arijit Chaudhuri https://orcid.org/0000-0002-4305-7686
Dipika Patra https://orcid.org/0000-0003-4318-1123

References

Adhikary, A. K. 2016. Variance estimation in randomizsed response surveys, *Handbook of statistics*, 34, 191–207. North Holland, Amsterdam: Elsevier.

Arnab, R. 1994. Nonnegative variance estimation in randomized response surveys. *Communications in Statistics- Theory and Methods* 23 (6):1743–52. doi:10.1080/03610929408831351.

Arnab, R. 1990. On comutativity of design and model expectations in randomized response surveys. *Communications in Statistics- Theory and Methods* 19 (10):3751–8.

Arnab, R. 2017. *Survey sampling theory and applications*, London UK: Academic Press.

Boruch, R. F. 1972. Relations among statistical methods for assuring confidentiality of social research data. *Social Science Research* 1 (4):403–14. doi:10.1016/0049-089X(72)90085-3.

Chaudhuri, A. 2011. *Randomized response and indirect questioning techniques in surveys*, Boca Raton: CRC Press.

Chaudhuri, A. and T. C. Christofides, 2013. *Indirect questioning in sample surveys*, Berlin, Germany: Springer Verlag.

Chaudhuri, A. and S. Pal, *2022. A comprehensive textbook on sample surveys*, Singapore: Springer Nature.

Chaudhuri, A, and S. Samaddar. 2022. Estimating the population mean using a complex sampling design dependent on an auxiliary variable. *Statistics in Transition New Series* 23 (1):39–54. doi:10.2478/stattrans-2022-0003.

Chaudhuri, A., T. C. Christofides, and A. Saha. 2009. Protection of privacy in efficient application of randomized response techniques. *Statistical Methods and Applications* 18 (3):389–418. doi:10.1007/s10260-008-0097-5.

Christofides, T. C. 2003. A generalized randomized response technique. *Metrika* 57 (2):195–200. doi:10.1007/s001840200216.

Cochran, W. G. 1977. *Sampling techniques*, New York, NY: John Wiley & Sons.

Des, R. 1956. Some estimators in sampling with varying probabilities without replacement. *Journal of American Statistical Association* 51:269–84.

Greenberg, B. G., A. Abul-Ela, W. R. Simmons, and D. G. Horvitz. 1969. The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association* 64 (326):520–39. doi:10.1080/01621459.1969.10500991.

Hansen, M. H, and W. N. Hurwitz. 1943. On the Theory of Sampling from Finite Populations. *The Annals of Mathematical Statistics* 14 (4):333–62. doi:10.1214/aoms/1177731356.

Horvitz, D. G, and D. J. Thompson. 1952. A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association* 47 (260):663–85. doi:10.1080/01621459.1952.10483446.

Horvitz, D. G., B. V. Shah, and W. R. Simmons, 1967. The unrelated question randomized response model, *Proceedings of Social Statistics Section*, 65–72.: American Statistical Association.

Kuk, A. Y. C. 1990. Asking sensitive questions indirectly. *Biometrika* 77 (2):436–8. doi:10.1093/biomet/77.2.436.

Lahiri, D. B. 1951. A method of sample selection providing unbiased ratio estimates. *Bulletin of International Statistical Institute* 3 (2):133–73.

Mangat, N. S, and R. Singh. 1990. An alternative randomized response procedure. *Biometrika* 77 (2):439–42. doi:10.1093/biomet/77.2.439.

Midzuno, H. 1952. An outline of the theory of sampling systems. *Annals of the Institute of Statistical Mathematics* 1 (1):149–56. doi:10.1007/BF03029250.

Rao, J. N. K. 1961. On the estimate of the variance in unequal probability sampling. *Annals of the Institute of Statistical Mathematics* 13 (1):57–60. doi:10.1007/BF02868860.

Rao, J. N. K., H. O. Hartley, and W. G. Cochran. 1962. On a Simple Procedure of Unequal Probability Sampling Without Replacement. *Journal of the Royal Statistical Society: Series B (Methodological)* 24 (2):482–91. doi:10.1111/j.2517-6161.1962.tb00475.x.

Sen, A. R. 1953. Recent Advances in Sampling with Varying Probabilities. *Calcutta Statistical Association Bulletin* 5 (1):1–15. doi:10.1177/0008068319530101.

Warner, S. L. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60 (309):63–9. doi:10.1080/01621459.1965.10480775.