Sankhyā : The Indian Journal of Statistics
2016, Volume 78-B, Part 1, pp. 66-77

On Comparisons of With and Without Replacement Sampling Strategies for Estimating Finite Population Mean in Randomized Response Surveys

S. Sengupta
University of Calcutta, Kolkata, India

Abstract

We consider the problem of unbiased estimation of a finite population mean (or proportion) related to a sensitive character under a randomized response model and present results on the comparisons of some with and without replacement sampling strategies based on equal and unequal probability sampling designs paralleling those for a direct survey.

AMS (2000) subject classification. 62D05.
Keywords and phrases. Linear unbiased strategy, Multiple responses, Population mean, Population proportion, Randomized response, Sampling with and without replacement, Unequal probability sampling.

1 Introduction

Consider a finite population of labeled units and suppose that the problem is to estimate certain population parameters on surveying a random sample of units. In an open set-up it is assumed that an exact response can be obtained from each sampled unit through a direct survey. However, if the character of interest is sensitive or stigmatizing such as drinking alcohol or gambling habit, drug addiction, tax evasion, history of induced abortions etc., a direct survey is likely to yield unreliable responses and an alternative technique, introduced by Warner (1965), is to obtain responses through a *randomized response* (RR) survey wherein every sampled unit is asked to give a response through an RR device as per instructions from the investigator. We refer to Chaudhuri and Mukerjee (1988), Chaudhuri (2011) and Chaudhuri and Christofides (2013) for a comprehensive review of such RR procedures.

Several researchers compared different linear unbiased sampling strategies for estimating a finite population mean (or proportion) based on with replacement (WR) sampling designs with equal and unequal selection probabilities in an open set-up with some comparable sampling strategies based on without replacement (WOR) sampling designs. It was found that most often the WOR sampling strategies are better than the WR sampling strategies in the sense of having smaller variances. In this paper it is shown that all these results relating to comparisons of with and without replacement sampling strategies in an open set-up also remain true under an RR model covering various RR plans when the WR strategies are based on a single randomized response obtained from every sampled unit even if selected more than once in the sample. However, for with replacement sampling, a population unit may be selected more than once in the sample and independent randomized responses may be obtained from it as many times as it is selected. In fact, as shown in Arnab (1999) and Sengupta (2015), any linear unbiased estimator for estimating the population mean (or proportion) for an WR sampling design based on a single randomized response can be improved upon by an unbiased estimator based on independent multiple responses. It is demonstrated that the results on the comparisons of with and without replacement sampling strategies in the open set-up are not, however, generally true under the RR model when the WR strategies are based on such independent multiple responses although some of the results still remain true under the general RR model while some other hold good for a special case of the model.

2 Notations and Preliminaries

Let $U = \{1, 2, \ldots, i, \ldots, N\}$ be a finite population of N labeled units and Y be a real variable with unknown value y_i for the population unit $i, 1 \leq i \leq N$. For a dichotomous population, Y is considered to be an indicator variable where y_i is 1 or 0 according as the unit i does or does not possess a certain attribute. The problem of interest is to estimate unbiasedly the unknown population mean (or proportion) $\theta = \frac{1}{N}\sum_{i=1}^{N} y_i$ on surveying a sample (a sequence of units of U with or without repetitions) s selected from a set of samples S with a given probability $p(s)(> 0)$ i.e. according to a given sampling design p. It is assumed that $p(s)$ is independent of $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$. Any p is said to be an WOR sampling design if $f_{si} = 1$ for each $i \in s$ and for each s and an WR sampling design otherwise, where f_{si} is the number of times the population unit i is selected in s.

For a given p, an estimator $e(s, y)$ of θ is called a linear unbiased estimator (lue) if it is of the form

$$e(s, y) = \sum_{i \in s} b_{si} y_i, \sum_{s \supset i} b_{si} p(s) = 1/N \forall i \tag{2.1}$$

for known coefficients b_{si}'s, where $\{i : i \in s\}$ is the set of distinct population units in s. The variance of an lue is given by

$$V_p(e(s,y)) = \sum_{i=1}^{N} y_i^2 \sum_{s \supset i} b_{si}^2 p(s) + \sum_{i \neq} \sum_{j=1}^{N} y_i y_j \sum_{s \supset i,j} b_{si} b_{sj} p(s) - \theta^2 \qquad (2.2)$$

where the suffix p on E(or V) is used to denote the expectation (or variance) with respect to p. A sampling design p together with an lue $e(s,y)$ is called a linear unbiased sampling strategy for estimating θ and is denoted by $(p, e(s,y))$.

We consider the character Y to be sensitive and suppose that some RR device R is employed to produce a randomized response z_i on the population unit i when included in s. We assume that under R, z_i's are independently distributed with

$$E_R(z_i) = ay_i + b \text{ and } V_R(z_i) = \phi(y_i), \text{ say, } 1 \leq i \leq N \qquad (2.3)$$

for some known constants $a(\neq 0)$ and b and for some known function ϕ. The suffixes R and both p and R on E(or V) will be used to denote the expectations (or variances) with respect to R and both p and R. The model (2.3) can be equivalently written as

$$E_R(r_i) = y_i \text{ and } V_R(r_i) = \frac{\phi(y_i)}{a^2} = \psi(y_i), \text{ say, } 1 \leq i \leq N \qquad (2.4)$$

in terms of transformed randomized response $r_i = (z_i - b)/a$. The model (2.4) holds for the RR device due to Warner (1965) for estimating a population proportion which consists of asking a population unit i to report y_i or $1 - y_i$ with probability q or $1 - q(0 < q < 1, q \neq \frac{1}{2})$ known to the investigator. The RR device due to Eriksson (1973) for estimating population proportions is also a special case of (2.4) in which a population unit i is asked to report $y_i, 1$ or 0 with given probabilities q_0, q_1 or $1 - q_0 - q_1 (q_0, q_1 > 0, q_0 + q_1 < 1)$. As illustrated in Chaudhuri and Christofides (2013, Chapter 4), the model (2.4) is also true for several other RR plans for a qualitative character suggested in the literature e.g. for those suggested in Boruch (1972), Kuk (1990), Mangat and Singh (1990), Mangat (1994) and Christofides (2003). The model (2.4) holds for several RR plans for a quantitative character as well (see Chaudhuri and Christofides 2013, Chapter 5). As for example, this is true for a scrambled RR plan discussed in

Pollock and Bek (1976) and Eichhorn and Hayre (1983) wherein a population unit i is asked to report $Ay_i + B$, where A and B are two independent random variables with known probability distributions with $E(A) \neq 0$. The special case of the model (2.4) for which the function $\psi(y_i)$ is a constant independent of y_i, say ψ_0, will be denoted by R_0. This special model holds e.g. for the RR devices due to Warner (1965), Mangat and Singh (1990) and Christofides (2003) and is also true for the scrambled RR plan when the probability distribution of A is degenerate (see Chaudhuri and Christofides 2013, Chapters 4, 5).

If a single response is obtained from each distinct population unit in s, one obtains the data $\{r_i : i \in s\}$ and for a given p under an RR plan R, an lue $e(s, \boldsymbol{r})$ of θ is defined as

$$e(s,\boldsymbol{r}) = \sum_{i\in s} b_{si} r_i, \sum_{s\ni i} b_{si} p(s) = 1/N \forall i \tag{2.5}$$

for known coefficients b_{si}'s, where $\boldsymbol{r} = (r_1, r_2, \ldots, r_N)$. It can be readily verified from (2.4) that for an lue $e(s,\boldsymbol{r})$, the estimator $e(s,\boldsymbol{y})$ obtained by replacing r_i by y_i is an lue in the open set-up and vice versa. We shall call $e(s,\boldsymbol{r})$ to be the derived lue from $e(s,\boldsymbol{y})$ and the linear unbiased strategy $(p, e(s,\boldsymbol{r}))$ to be the derived strategy from $(p, e(s,\boldsymbol{y}))$ in the open set-up. The variance of a derived linear unbiased strategy $(p, e(s,\boldsymbol{r}))$ is given by

$$V_{p^R}(e(s,\boldsymbol{r})) = V_p E_R(e(s,\boldsymbol{r})) + E_p V_R(e(s,\boldsymbol{r})) = V_p(e(s,\boldsymbol{y})) + \sum_{i=1}^{N} \psi(y_i) \sum_{s\ni i} b_{si}^2 p(s)$$

$$= V_p(e(s,\boldsymbol{y})) + \sum_{i=1}^{N} \psi(y_i)[V_{pi}(e(s,\boldsymbol{y})) + N^{-2}] \tag{2.6}$$

where $V_{pi}(e(s,\boldsymbol{y}))$ is $V_p(e(s,\boldsymbol{y}))$ at an $\boldsymbol{y}$ with $y_i = 1$ and $y_j = 0 \forall j (\neq i)$.

If, for an WR sampling design, independent randomized responses are obtained from each population unit as many times as it is selected in the sample, one obtains the data $\{r_{ij}, j = 1, \ldots, f_{si}, i \in s\}$, where r_{ij} is the transformed response from unit i in its jth selection. As shown in Arnab (1999) and Sengupta (2015), a derived lue $e(s,\boldsymbol{r})$ for an WR sampling design p based on a single response can then be improved upon by the unbiased estimator $e^*(s,\boldsymbol{r})$, where $e^*(s,\boldsymbol{r}) = \sum_{i\in s}\frac{b_{si}}{f_{si}}\sum_{j=1}^{f_{si}} r_{ij}$. Also since $E_R(e^*(s,\boldsymbol{r})) = E_R(e(s,\boldsymbol{r}))$, we have

$$V_{p^R}(e^*(s,\boldsymbol{r})) = V_p(e(s,\boldsymbol{y})) + E_p V_R(e^*(s,\boldsymbol{r})) = V_p(e(s,\boldsymbol{y})) + \sum_{i=1}^{N}\psi(y_i)\sum_{s\supset i}\frac{b_{si}^2}{f_{si}}p(s). \tag{2.7}$$

The unbiased estimator $e^*(s,\boldsymbol{r})$ obtained from an lue $e(s,\boldsymbol{y})$ based on an WR sampling design p in the open set-up will be called a derived estimator and the strategy $(p, e^*(s,\boldsymbol{r}))$ a derived strategy based on multiple responses.

The theorem stated below follows readily from (2.6).

Theorem 2.1. *Let for a given R satisfying (2.4), $(p, e(s,\boldsymbol{r}))$ and $(p', e'(s,\boldsymbol{r}))$ be two derived linear unbiased strategies based on single response. Then $V_p(e(s,\boldsymbol{y})) \leq V_{p'}(e'(s,\boldsymbol{y}))\forall\boldsymbol{y}$ implies $V_{p^R}(e(s,\boldsymbol{r})) \leq V_{p'^R}(e'(s,\boldsymbol{r}))\forall\boldsymbol{y}$.*

Consequently, for any given R satisfying (2.4), all well-known results on the comparisons of with and without replacement sampling strategies in the open set-up get extended through the above theorem for the derived strategies based on single response. Thus, for example, the results relating to comparisons of strategies based on simple random sampling with and without replacement and the comparison of Hansen and Hurwitz (1943) strategy with Des Raj (1956) strategy, Murthy's (1957) strategy, Rao et al. (1962) strategy and Horvitz and Thompson (1952) strategies based on inclusion probability proportional to size (IPPS) sampling designs in the open set-up hold as well for the derived strategies when the WR strategies are based on a single response from every sampled unit.

Theorem 2.1 is not, however, true for the derived strategies based on multiple responses. This may be demonstrated through the following example in Sengupta (2015).

Example 2.1. *Consider Warner's (1965) RR plan with $q = 0.4$ and let p be the SRSWR design involving n draws. Let $e_1^*(s,\boldsymbol{r})$ and $e_2^*(s,\boldsymbol{r})$ be the derived estimators based on multiple responses from the sample means in the open set-up based, respectively, on all n units and on v distinct population units in the sample. It can be shown that for $N = 4, n = 3, V_{p^R}(e_1^*(s,\boldsymbol{r})) < V_{p^R}(e_2^*(s,\boldsymbol{r}))\forall\boldsymbol{y}$ although $V_p(e_2(s,\boldsymbol{y})) \leq V_p(e_1(s,\boldsymbol{y}))\forall\boldsymbol{y}$ (see Basu 1958; Raj and Khamis 1958).*

As such the results on the comparisons of with and without replacement sampling strategies in the open set-up do not generally get extended for the derived strategies based on multiple responses (see Example 3.1). However, in what follows we show that some of the results in the open set-up are still true under the general RR model (2.4) while some other hold good for the RR model R_0.

3 Comparisons of RR Strategies

We consider the following unbiased RR strategies derived from some well-known sampling strategies in an open set-up with equal and unequal selection probabilities based on known normed size measure w_i for unit $i, 1 \leq i \leq N, w_i > 0 \forall i, \sum_{i=1}^{N} w_i = 1$.

$(p_{1n}, e_{1n}(s, \boldsymbol{r}))$: p_{1n} is the simple random sampling (SRS) WOR sampling design of size n and $e_{1n}(s, \boldsymbol{r}) = \frac{1}{n}\sum_{i \in s} r_i$ is the derived estimator from the sample mean in the open set-up.

$(p_2, e^*_{21}(s, \boldsymbol{r}))$ and $(p_2, e^*_{22}(s, \boldsymbol{r}))$: p_2 is the SRSWR sampling design involving n draws and $e^*_{21}(s, \boldsymbol{r}) = \frac{1}{n}\sum_{i \in s}\sum_{j=1}^{f_{si}} r_{ij}$ and $e^*_{22}(s, \boldsymbol{r}) = \frac{1}{v(s)}\sum_{i \in s}\frac{1}{f_{si}}\sum_{j=1}^{f_{si}} r_{ij}$ are, respectively, the derived estimators based on multiple responses obtained from the sample means in the open set-up based on all n units and $v(s)$ distinct population units in the sample.

$(p_3, e^*_{31}(s, \boldsymbol{r}))$ and $(p_3, e^*_{32}(s, \boldsymbol{r}))$: p_3 is the inverse SRSWR sampling design (see Raj and Khamis 1958) in which units are selected independently WR with equal selection probabilities until a fixed number of v distinct population units get selected and $e^*_{31}(s, \boldsymbol{r}) = \frac{1}{n(s)}\sum_{i \in s}\sum_{j=1}^{f_{si}} r_{ij}$ and $e^*_{32}(s, \boldsymbol{r}) = \frac{1}{v}\sum_{i \in s}\frac{1}{f_{si}}\sum_{j=1}^{f_{si}} r_{ij}$ are, respectively, the derived estimators based on multiple responses obtained from the sample means in the open set-up based on all $n(s)$ units and v distinct population units in the sample.

$(p_4, e^*_4(s, \boldsymbol{r}))$: p_4 is the probability proportional to size (PPS) WR sampling design involving n draws and $e^*_4(s, \boldsymbol{r}) = \frac{1}{nN}\sum_{i \in s}\frac{1}{w_i}\sum_{j=1}^{f_{si}} r_{ij}$ is the derived estimators based on multiple responses obtained from the Hansen and Hurwitz (1943) estimator in the open set-up.

$(p_5, e_5(s, \boldsymbol{r}))$: p_5 is an IPPS sampling design in which n units are selected WOR such that the inclusion probability of the population unit i in the sample is nw_i and $e_5(s, \boldsymbol{r}) = \frac{1}{nN}\sum_{i \in s} \frac{r_i}{w_i}$ is the derived estimator from the Horvitz and Thompson (1952) estimator in the open set-up.

$(p_6, e_6(s, \boldsymbol{r}))$: p_6 is the (Rao et al., 1962) sampling design in which U is divided at random into n groups $G_t, t = 1, \ldots, n$ each of size N/n (assumed to be an integer) and one unit i_t is selected from G_t with probability w_{i_t}/W_t independently for each t, where $W_t = \sum_{i \in G_t} w_i$ and $e_6(s, \boldsymbol{r}) = \frac{1}{N}\sum_{t=1}^{n} r_{i_t} \frac{W_t}{w_{i_t}}$ is the derived estimator from the Rao et al. (1962) estimator in the open set-up.

$(p_7, e_7(s, \boldsymbol{r}))$: p_7 is the PPSWOR sampling design of size n and $e_7(s, \boldsymbol{r})$ is the derived estimator from Murthy's (1957) estimator defined as $e_7(s, \boldsymbol{r}) = \frac{\sum_{i \in s} r_i p_7(s|i)}{N p_7(s)}$ where $p_7(s)$ is the probability of selecting an unordered sample (sub-set of U) s and $p_7(s \mid i)$ is the conditional probability of selecting s given that the first unit selected is the population unit i.

In the following theorem we present the results on some comparisons of these RR strategies based on different with and without replacement sampling designs.

Theorem 3.1. (a) *For a given R satisfying (2.4),*

(i) $V_{p_{1nR}}(e_{1n}(s, \boldsymbol{r})) \leq V_{p_2^R}(e_{2k}^*(s, \boldsymbol{r})) \forall \boldsymbol{y}, k = 1, 2$

(ii) $V_{p_{1n^*R}}(e_{1n^*}(s, \boldsymbol{r})) \leq V_{p_3^R}(e_{3k}^*(s, \boldsymbol{r})) \forall \boldsymbol{y}, k = 1, 2$ *where* $n^* = E_{p_3}(n(s))$ *(assumed to be an integer)*

(iii) $V_{p_5^R}(e_5(s, \boldsymbol{r})) \leq V_{p_4^R}(e_4^*(s, \boldsymbol{r})) \forall \boldsymbol{y}$ *if* $V_{p_5}(e_5(s, \boldsymbol{y})) \leq V_{p_4}(e_4(s, \boldsymbol{y})) \forall \boldsymbol{y}$ *for the IPPS sampling design p_5 e.g. for those due to Sampford (1967) and Brewer (1975) (see Gabler 1981, Sengupta 1986).*

(b) *For a given R following the RR model R_0,*

(i) $V_{p_6^R}(e_6(s, \boldsymbol{r})) \leq V_{p_4^R}(e_4^*(s, \boldsymbol{r})) \forall \boldsymbol{y}$

(ii) $V_{p_7^R}(e_7(s, \boldsymbol{r})) \leq V_{p_4^R}(e_4^*(s, \boldsymbol{r})) \forall \boldsymbol{y}$.

PROOF.

(a) It can be readily verified that for an R satisfying (2.4),

$$\begin{aligned}
E_{p_{1n}} V_R(e_{1n}(s,\boldsymbol{r})) &= E_{p_2} V_R(e^*_{21}(s,\boldsymbol{r})) = \frac{1}{nN}\sum_{i=1}^{N}\psi(y_i) \\
E_{p_2} V_R(e^*_{22}(s,\boldsymbol{r})) &= E_{p_2}\left(\frac{1}{v^2(s)f_{si}}\right)\sum_{i=1}^{N}\psi(y_i) \\
E_{p_3} V_R(e^*_{31}(s,\boldsymbol{r})) &= \frac{1}{N}E_{p_3}\left(\frac{1}{n(s)}\right)\sum_{i=1}^{N}\psi(y_i) \\
E_{p_3} V_R(e^*_{32}(s,\boldsymbol{r})) &= \frac{1}{v^2}E_{p_3}\left(\frac{1}{f_{si}}\right)\sum_{i=1}^{N}\psi(y_i) \\
E_{p_4} V_R(e^*_{4}(s,\boldsymbol{r})) &= E_{p_5} V_R(e_5(s,\boldsymbol{r})) = \frac{1}{nN^2}\sum_{i=1}^{N}\frac{\psi(y_i)}{w_i}.
\end{aligned}$$

$$\text{Since } E_{p_2}\left(\frac{1}{v^2(s)f_{si}}\right) \geq \frac{E^2_{p_2}\left(\frac{1}{v(s)}\right)}{E_{p_2}(f_{si})} = \frac{NE^2_{p_2}\left(\frac{1}{v(s)}\right)}{n} \geq \frac{1}{nN},$$

$$E_{p_3}\left(\frac{1}{n(s)}\right) \geq \frac{1}{n^*} \text{ and } \frac{1}{v^2}E_{p_3}\left(\frac{1}{f_{si}}\right) \geq \frac{1}{v^2E_{p_3}(f_{si})} = \frac{N}{v^2n^*} \geq \frac{1}{n^*N},$$

it follows that

$$\begin{aligned}
E_{p_{1n}} V_R(e_{1n}(s,\boldsymbol{r})) &\leq E_{p_2} V_R(e^*_{22}(s,\boldsymbol{r})) \\
E_{p_{1n^*}} V_R(e_{1n^*}(s,\boldsymbol{r})) &\leq E_{p_3} V_R(e^*_{3k}(s,\boldsymbol{r})), k = 1,2.
\end{aligned}$$

Also it is well known that (see Basu 1958; Raj and Khamis 1958; Chikkagoudar 1966)

$$\begin{aligned}
V_{p_{1n}}(e_{1n}(s,\boldsymbol{y})) \leq V_{p_2}(e_{22}(s,\boldsymbol{y})) \leq V_{p_2}(e_{21}(s,\boldsymbol{y}))\forall\, \boldsymbol{y} \\
V_{p_{1n^*}}(e_{1n^*}(s,\boldsymbol{y})) \leq V_{p_{1v}}(e_{1v}(s,\boldsymbol{y})) = V_{p_3}(e_{32}(s,\boldsymbol{y})) \leq V_{p_3}(e_{31}(s,\boldsymbol{y}))\forall\, \boldsymbol{y}.
\end{aligned}$$

Hence, the proof follows from (2.6) and (2.7).

(b) The variance of the Rao et al. (1962) strategy in the open set-up is given by

$$V_{p_6}(e_6(s,\boldsymbol{y})) = \frac{N-n}{N-1}V_{p_4}(e_4(s,\boldsymbol{y})) = \frac{N-n}{nN^2(N-1)}\left[\sum_{i=1}^{N}\frac{y_i^2}{w_i} - N^2\theta^2\right]$$

whence it can be seen using (2.6) that for an R following the RR model R_0,

$$\begin{aligned} V_{p_6^R}(e_6(s,\boldsymbol{r})) &= \frac{N-n}{N-1}V_{p_4}(e_4(s,\boldsymbol{y})) \\ &\quad + \frac{\psi_0}{nN^2(N-1)}\sum_{i=1}^{N}\left[\frac{N-n}{w_i} + N(n-1)\right] \\ &\leq V_{p_4}(e_4(s,\boldsymbol{y})) + \frac{\psi_0}{nN^2}\sum_{i=1}^{N}\frac{1}{w_i} = V_{p_4}(e_4(s,\boldsymbol{y})) \\ &\quad + E_{p_4}V_R(e_4^*(s,\boldsymbol{r})) = V_{p_4^R}(e_4^*(s,\boldsymbol{r}))\forall\,\boldsymbol{y} \end{aligned}$$

since $\sum_{i=1}^{N}\frac{1}{w_i} \geq N^2$.

It similarly follows that for an R following the RR model R_0,

$$\begin{aligned} V_{p_7^R}(e_7(s,\boldsymbol{r})) &= V_{p_7}(e_7(s,\boldsymbol{y})) + \frac{\psi_0}{N^2}\sum_{i=1}^{N}\sum_{s\supset i}\frac{p_7^2(s\mid i)}{p_7(s)} \\ &\leq V_{p_4}(e_4(s,\boldsymbol{y})) + \frac{\psi_0}{nN^2}\sum_{i=1}^{N}\frac{1}{w_i} = V_{p_4}(e_4(s,\boldsymbol{y})) \\ &\quad + E_{p_4}V_R(e_4^*(s,\boldsymbol{r})) = V_{p_4^R}(e_4^*(s,\boldsymbol{r}))\forall\,\boldsymbol{y} \end{aligned}$$

since it is well known that $V_{p_7}(e_7(s,\boldsymbol{y})) \leq V_{p_4}(e_4(s,\boldsymbol{y}))\forall\,\boldsymbol{y}$ and as proved in the Appendix,

$$\sum_{i=1}^{N}\left[\frac{1}{nw_i} - \sum_{s\supset i}\frac{p_7^2(s\mid i)}{p_7(s)}\right] \geq 0 \qquad (3.1)$$

We remark that (b) of Theorem 3.1 is not necessarily true under the general RR model R if $\psi(y_i)$ is not independent of y_i. This may be demonstrated through the following example.

Example 3.1. *Consider Eriksson's (1973) RR plan for estimation of population proportion for which* $\psi(y_i) = c_1 y_i + c_2$ *with* $c_1 = (1 - q_0 - 2q_1)/q_0, c_2 = q_1(1 - q_1)/q_0^2$. *As in the proof of Theorem 3.1, it can be verified that*

$$\begin{aligned} V_{p_4^R}(e_4^*(s, \boldsymbol{r})) - V_{p_6^R}(e_6(s, \boldsymbol{r})) &= \frac{n-1}{N-1} V_{p_4}(e_4(s, \boldsymbol{y})) \\ &\quad + \frac{n-1}{nN^2(N-1)} \sum_{i=1}^{N} \psi(y_i) \left[\frac{1}{w_i} - N \right] \\ &= \frac{n-1}{nN^2(N-1)} \left\{ \sum_{i=1}^{N} \frac{y_i^2}{w_i} - N^2\theta^2 + \sum_{i=1}^{N} \psi(y_i) \left[\frac{1}{w_i} - N \right] \right\}. \end{aligned} \tag{3.2}$$

For $N = 4, n = 2, w_1 = 0.4, w_2 = w_3 = w_4 = 0.2, y_1 = 1, y_2 = y_3 = y_4 = 0$, *the RHS of (3.2) is* $\frac{1+c_2-c_1}{64}$ *which is negative if* $c_1 - c_2 > 1$ *e.g. for* $q_0 = 0.2, q_1 = 0.05$.

Similar example can be provided to demonstrate that (ii) of (b) is not necessarily true under the general RR model (2.4).

Acknowledgments. We acknowledge with thanks some constructive suggestions from the Associate Editor and the referee which helped a lot to improve an earlier version of the paper.

Appendix

The proof of (3.1) is based on the following lemma.

Lemma A.1. *If for* $i \neq j \in s, w_i \geq w_j$ *ihen*
(*i*) $p_7(s \mid i) \leq p_7(s \mid j)$ (*ii*) $w_i p_7(s \mid i) \geq w_j p_7(s \mid j)$.

PROOF. For $i \in s$, one can write (see Andreatta and Kaufman 1986)

$$\begin{aligned} p_7(s \mid i) &= (1 - \sum_{i \in s} w_i) \int_0^\infty e^{-\lambda} \frac{e^{\lambda w_i}}{e^{\lambda w_i} - 1} \prod_{j \in s} \{e^{\lambda w_j} - 1\} d\lambda \\ &= (1 - \sum_{i \in s} w_i) \int_0^1 \frac{1}{1 - t^{w_i}} \prod_{j \in s} \{t^{-w_j} - 1\} dt. \end{aligned}$$

It is easy to verify that for $t \in (0, 1)$,

$$\frac{1}{1-t^x} \text{ and } \frac{x}{1-t^x}$$

are, respectively, decreasing and increasing in $x \in (0,1)$ and this proves the lemma.

Now as $p_7(s) = \sum_{i \in s} w_i p_7(s \mid i)$, $\sum_{s \supset i} p_7(s \mid i) = 1$

$$\sum_{i=1}^{N} [\frac{1}{nw_i} - \sum_{s \supset i} \frac{p_7^2(s \mid i)}{p_7(s)}] = \frac{1}{n} \sum_{i=1}^{N} \frac{1}{w_i} \sum_{s \supset i} \frac{p_7(s \mid i)\{p_7(s) - nw_i p_7(s \mid i)\}}{p_7(s)}.$$

$$= \frac{1}{n} \sum_{i=1}^{N} \frac{1}{w_i} \sum_{s \supset i} \frac{p_7(s \mid i) \sum_{j(\neq i) \in s} \{w_j p_7(s \mid j) - w_i p_7(s \mid i)\}}{p_7(s)}.$$

$$= \frac{1}{n} \sum_{s \in S} \frac{1}{p_7(s)} \sum_{i \neq} \sum_{j \in s} \frac{p_7(s \mid i)\{w_j p_7(s \mid j) - w_i p_7(s \mid i)\}}{w_i}$$

$$= \frac{1}{n} \sum_{s \in S} \frac{1}{p_7(s)} \sum_{i <} \sum_{j \in s} \{\frac{p_7(s \mid i)}{w_i} - \frac{p_7(s \mid j)}{w_j}\}\{w_j p_7(s \mid j) - w_i p_7(s \mid i)\}$$

$$= \frac{1}{n} \sum_{s \in S} \frac{1}{p_7(s)} \sum_{i <} \sum_{j \in s} \frac{\{w_j p_7(s \mid i) - w_i p_7(s \mid j)\}}{w_i w_j} \{w_j p_7(s \mid j) - w_i p_7(s \mid i)\}$$

$$= \frac{1}{n} \sum_{s \in S} \frac{1}{p_7(s)} \sum_{i <} \sum_{j \in s} \frac{[\{w_j p_7(s \mid j) - w_i p_7(s \mid i)\}^2 + (w_i + w_j)\{p_7(s \mid i) - p_7(s \mid j)\}\{w_j p_7(s \mid j) - w_i p_7(\mid i)\}]}{w_i w_j}.$$

Hence, (3.1) follows by Lemma A.1.

References

ANDREATTA, G. and KAUFMAN, G. M. (1986). Estimation of finite population properties when sampling is without replacement and proportional to magnitude. *J. Amer. Statist. Assoc.* **81**, 657–666.

ARNAB, R. (1999). On use of distinct respondents in RR surveys. *Biom. J.* **41**, 507–513.

BASU, D. (1958). On sampling with and without replacement. *Sankhya* **20**, 287–294.

BORUCH, R. F. (1972). Relations among statistical methods for assuring confidentiality of social research data. *Soc. Sci. Res.* **1**, 403–414.

BREWER, K. R. W. (1975). A simple procedure for sampling ΠPSWOR. *Austral. J. Statist.* **17**, 166–172.

CHAUDHURI, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys.* CRC Press, Chapman and Hall, Taylor & Francis Group, Boca Raton, Florida, USA.

CHAUDHURI, A. and CHRISTOFIDES, T. C. (2013). *Indirect Questioning in Sample Surveys.* Springer- Verlag, Berlin, Heidelberg, Germany.

CHAUDHURI, A. and MUKERJEE, R. (1988). *Randomized Response: Theory and Techniques.* Marcel Dekker, New York.

CHIKKAGOUDAR, M. S. (1966). A note on inverse sampling with equal probabilities. *Sankhya* **28**, 93–96.

CHRISTOFIDES, T. C. (2003). A generalized randomized response technique. *Metrika* **57**, 195–200.

EICHHORN, B. H. and HAYRE, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *J. Statist. Plann. Inference* **7**, 307–316.

ERIKSSON, S. A. (1973). A new model for randomized response. *Rev. Internat. Stat. Inst.* **41**, 101–113.

GABLER, S. (1981). A comparison of Sampford's sampling procedure versus unequal probability sampling with replacement. *Biometrika* **68**, 725–727.

HANSEN, M. H. and HURWITZ, W. N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.* **14**, 333–362.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement. *J. Amer. Statist. Assoc.* **47**, 663–685.

KUK, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika* **77**, 436–438.

MANGAT, N. S. (1994). An improved randomized response strategy. *J. Roy. Statist. Soc.* **56**, 93–95.

MANGAT, N. S. and SINGH, R. (1990). An alternative randomized response procedure. *Biometrika* **77**, 439–442.

MURTHY, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhya* **18**, 379–390.

POLLOCK, K. H. and BEK, Y. (1976). A comparison of three randomized response models for quantitative data. *J. Amer. Statist. Assoc.* **71**, 884–886.

RAJ, D. (1956). Some estimators in sampling with varying probabilities without replacement. *J. Amer. Statist. Assoc.* **51**, 269–284.

RAJ, D. and KHAMIS, S. H. (1958). Some remarks on sampling with replacement. *Ann. Math. Statist.* **39**, 550–557.

RAO, J. N. K., HARTLEY, H. O. and COCHRAN, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. Ser B* **24**, 482–491.

SAMPFORD, M. R. (1967). On sampling without replacement with unequal probability of selection. *Biometrika* **54**, 499–513.

SENGUPTA, S. (1986). A comparison between PPSWR and Brewer's π PSWOR procedures. *Cal. Statist. Assoc. Bull.* **35**, 207–210.

SENGUPTA, S. (2015). Estimation of finite population proportion in randomized response surveys using multiple responses. *Sankhya* **77**, 75–83.

WARNER, S. L. (1965). Randomized response - A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60**, 63–69.

S. SENGUPTA
DEPARTMENT OF STATISTICS,
UNIVERSITY OF CALCUTTA,
35, BALLYGUNGE CIRCULAR ROAD,
KOLKATA 700019, INDIA
E-mail: samindras@gmail.com

Paper received: 30 June 2014.