Journal of Statistical Planning and Inference 135 (2005) 516–527

journal of statistical planning and inference

Optional versus compulsory randomized response techniques in complex surveys

Arijit Chaudhuri*[,1], Amitava Saha[2]

Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

Received 8 July 2003; accepted 19 April 2004
Available online 21 July 2004

Abstract

In estimating the proportion of people bearing a sensitive attribute in a community, to mitigate possible evasive answer biases, Warner (J. Amer. Statist. Assoc. 60 (1965) 63) introduced a technique of randomized response (RR) in human surveys, by way of protecting individual privacy. Chaudhuri and Mukerjee (Calcutta Statist. Assoc. Bull. 34 (1985) 225; Randomized Response: Theory and Techniques, Marcel Dekker, New York) presented a modification allowing a direct response (DR) option to whom the attribute does not appear to be stigmatizing enough. Warner himself and many of his followers restrict the application of their RR devices to surveys with selection exclusively by 'simple random sampling with replacement'. Chaudhuri (J. Statist. Plann. Inference 34 (2001a) 37; Pakistan J. Statist. 17 (3) (2001b) 259; Calcutta Statist. Assoc. Bull. 52 (205–208) (2002) 315) showed the efficacy of some of these devices when sample selection is by general unequal probabilities possibly even without replacement. Here, we present theories for unbiased estimation of the proportion along with unbiased estimation of the variances of the estimators when 'compulsory' or 'optional' RR's are gathered from persons sampled with varying probabilities. Gains in efficiency by allowing DR option rather than RR compulsion are illustrated numerically through simulation from data.

MSC: 62 D05

Keywords: Optional randomized response; Sensitive proportion estimation; Simulation; Varying probability sampling

* Corresponding author.
E-mail addresses: achau@isical.ac.in (A. Chaudhuri), saha_amitava@hotmail.com (A. Saha).

[1] His research is partially supported by CSIR Grant no. 21(0539)/02/EMR-II.

[2] Amitava Saha is working as Deputy Director in the Directorate General of Mines Safety, Dhanbad, Jharkhand 826001, India.

1. Introduction

Warner (1965) is the pioneer introducing the idea of randomized response (RR) when encountering the social problem of estimation for a given community the proportion of people bearing a sensitive attribute like tax evasion, drunken driving, gambling, drug abuse, bribe taking for examples. Anticipating direct questioning to be embarrassing for both the interviewer and the interviewee, in order to protect the latter's privacy and expect truthful answers he devised a technique known as randomized response or RR technique.

Chaudhuri and Mukerjee (1985, 1988) relaxed a compulsion in RR and permitted an option for a direct response (DR) to those who volunteer to divulge the truths viewing the attribute not stigmatizing enough. Gupta (2001) provides an example of a practical application of compulsory RR's combined with optional RR's. As it is an unpublished piece of work and we have an access to only a short announcement of this through the internet, we have no comments on the possibilities of his approach to ORR.

Warner (1965) and most of his followers developed their estimation theories demanding the sample to be chosen by the simple random sampling with replacement (SRSWR) method alone. Chaudhuri (1987) gave a general theory covering qualitative as well as quantitative characteristics suspected to be socially stigmatizing provided the RR-device employed admits an unbiased estimator based on the RR from a sampled person for the person's true characteristic value such that the variance of this estimator is a quadratic function of the true values with known coefficients.

However, Chaudhuri (2001a, b, 2002) had to present additional procedures covering RR's based on unequal probability samples in estimating population proportions. The present work consolidates some of the scattered ideas covering optional randomized response (ORR) techniques in unequal probability sampling illustrating a few RR-devices in estimating sensitive proportions. Possible gains in efficiency by ORR vis-à-vis a compulsory RR (CRR) are illustrated numerically through simulations from certain data with reference to a few RR-devices and sampling designs. For further activities in RR a reader may refer to Chaudhuri (1999); Greenberg et al. (1977); Horvitz et al. (1976); Mangat (1991); Saha (2003) cited at the end of this work.

Section 2 presents the related theories and Section 3, the numerical findings.

2. RR generation, sample selection and estimation methods

2.1. Examples of RR procedures

We illustrate here for application only three RR techniques, namely those given by Warner (1965), Kuk (1990) and the unrelated question RR techniques of Horvitz et al. (1967) further strengthened by Greenberg et al. (1969). These are described with necessary alterations below to suit unequal probability sample selection.

Let $U = (1, \ldots, i, \ldots, N)$ denote a survey population with y_i as the value for its unit labeled i on a variable y such that

$$y_i = 1 \text{ if } i \text{ bears a sensitive attribute } A$$
$$= 0 \text{ if } i \text{ bears the complementary attribute } A^C.$$

Letting $\sum$ denote sum over i in U, $Y = \sum y_i$, $\theta = Y/N$ the problem is to estimate θ, equivalently Y, with N as known. We shall use generic notations E_R, V_R for expectation, variance operators with respect to any RR device employed.

2.1.1. *Warner's (1965) RR device*

Let a box contain similar cards marked A and A^C, respectively, in proportions $p : (1-p), 0 < p < 1$. A sampled person is requested to draw randomly from the box one card, unnoticed by the interviewer, and to report if the card-type drawn 'matches' or 'mismatches' his/her true y-value. This reporting is independent across the persons.

Letting $I_i = 1$ if i reports a 'match', $= 0$ if i reports a 'mismatch', we have $E_R(I_i) = py_i + (1-p)(1-y_i), i \in U$. Taking $p \neq \frac{1}{2}$ it follows that

$$V_R(I_i) = E_R(I_i)(1 - E_R(I_i)) = p(1-p), \quad \text{noting } y_i^2 = y_i, \ i \in U,$$
$$\text{for } r_i = \frac{I_i - (1-p)}{(2p-1)}, \quad E_R(r_i) = y_i$$

and

$$V_R(r_i) = \frac{p(1-p)}{(2p-1)^2} = V_i, \quad \text{say}, \ i \in U.$$

2.1.2. *Kuk's (1990) RR device*

Two boxes marked, respectively, A with Red and Black cards in proportions $p_1 : (1-p_1), 0 < p_1 < 1$ and A^C with proportions $p_2 : (1-p_2), 0 < p_2 < 1$ are presented to a sampled person. The person is requested, unnoticed by the interviewer, to independently draw with replacement a card k times from the box marked matching his/her A/A^C feature and to report the number f_i, of Red cards drawn. Then,

$$E_R(f_i) = k[p_1 y_i + p_2(1-y_i)],$$
$$V_R(f_i) = k[p_1(1-p_1)y_i + p_2(1-p_2)(1-y_i)]$$

and taking $p_1 \neq p_2$, it follows that

$$r_i = \frac{(f_i/k - p_2)}{(p_1 - p_2)} \quad \text{satisfies } E_R(r_i) = y_i$$

and

$$V_R(r_i) = \frac{V_R(f_i)}{k^2(p_1-p_2)^2} = V_i = \beta_i y_i + \theta_i, \ \text{say},$$

where

$$\beta_i = \frac{(1-p_1-p_2)}{k^2(p_1-p_2)^2}, \quad \theta_i = \frac{p_2(1-p_2)}{k^2(p_1-p_2)^2}, \quad i \in U.$$

An unbiased estimator for V_i is then $v_i = \beta_i r_i + \theta_i, i \in U$.

2.1.3. Unrelated question model' RR device

Let B be an innocuous characteristics unrelated to A and for a variable x, x_i's be the values in respect of the characteristic B, such that $x_i = 1$ if i bears B, $=0$ if i bears B^C, the complement of B. A sampled person, say, i is presented two boxes marked I and II containing cards marked A and B in proportions $p_1 : (1-p_1)$ and $p_2 : (1-p_2)$, respectively. Then, he is requested to independently draw two cards with replacement from the box marked I and repeat this independently twice from the box marked II and to report in each case as either '1' or '0' according as the card type 'matches' or 'does not match' the characteristic A or B, respectively. Thus, writing

$$\begin{aligned} I_i &= 1 \text{ if the draw from I matches for } i \\ &= 0 \text{ if the draw from I mismatches for } i, \\ J_i &= 1 \text{ if the draw from II matches for } i \\ &= 0 \text{ if the draw from II mismatches for } i \end{aligned}$$

and I_i' as defined similarly to I_i and J_i' to J_i we get

$$E_R(I_i) = p_1 y_i + (1-p_1)x_i = E_R(I_i'),$$
$$E_R(J_i) = p_2 y_i + (1-p_2)x_i = E_R(J_i').$$

Taking $p_1 \neq p_2$, letting

$$r_i' = \frac{(1-p_2)I_i - (1-p_1)J_i}{(p_1-p_2)}, \quad r_i'' = \frac{(1-p_2)I_i' - (1-p_1)J_i'}{(p_1-p_2)},$$

we get $E_R(r_i') = y_i = E_R(r_i''), i \in U$.

So, letting $r_i = \frac{1}{2}(r_i' + r_i'')$ we get $E_R(r_i) = y_i$ and writing $V_i = V_R(r_i)$ we get $v_i = \frac{1}{4}(r_i' - r_i'')^2$ satisfying $E_R(v_i) = V_i, i \in U$.

2.2. Sample selection and estimation

We shall illustrate only two schemes of sampling without replacement namely SRSWOR and Rao, Hartley and Cochran's (RHC, 1962) scheme. The former needs no elaboration. The RHC scheme, in vogue over decades, is well known. RHC is illustrated as a specimen of a varying probability sampling scheme because of its inherent properties of simple and universal application in yielding efficient estimator for a population total with non-negative unbiased variance estimator. For completeness and clarification of notations, however, we need to briefly describe it. In it certain positive integers N_i are first fixed as the numbers allotted to n non-overlapping groups into which U is to be randomly divided such that $\sum_n N_i = N$, writing $\sum_n$ as the sum over the n groups formed. Certain normed size-measures P_i $(0 < P_i < 1, \sum P_i = 1)$ are supposed to be known. Writing $P_{i_1}, \ldots, P_{i_{N_i}}$ as the P_i-values for the units assigned randomly into the ith group and $Q_i = P_{i_1} + \cdots + P_{i_{N_i}}$, one unit, say, i_k is chosen from the ith group with probability P_{i_k}/Q_i and this is independently repeated for all the n groups. If y_i's were ascertainable then

$$t = \sum_n y_i Q_i / P_i,$$

writing (y_i, P_i) for the y_i and P_i-values for the unit chosen from the ith group, is an unbiased estimator for Y based on RHC scheme.

Also,

$$v_{pR}(t) = \frac{\sum_n N_i^2 - N}{N^2 - \sum_n N_i^2}\left[\sum_n Q_i\left(\frac{y_i}{P_i}\right)^2 - t^2\right]$$

is given by RHC as an unbiased estimator for $V_{pR}(t)$, the variance of t. We shall write E_p, V_p as the generic notations for operators for expectation and variance, respectively, with respect to any sampling scheme. Also, we shall write E, V as the over-all expectation, variance operators such that $E = E_pE_R = E_RE_p$, $V = E_pV_R + V_pE_R = E_RV_p + V_RE_p$. Note that when y_i's are not directly ascertained if r_i's are gathered for i in a sample s chosen according to any design p with a probability $p(s)$

$$e = \sum_n r_iQ_i/P_i,$$

when s is chosen by RHC scheme satisfies the following:

$$E(e) = E_pE_R(e) = E_p\left(\sum_n y_iQ_i/P_i\right) = Y$$

and also

$$E(e) = E_RE_p(e) = E_R\left(\sum r_i\right) = Y.$$

Further, from Chaudhuri et al. (2000) we know that using

$$V(e) = E_RV_p(e) + V_RE_p(e) = E_RE_pv_p(e) + V_R\left(\sum r_i\right)$$
$$= E_RE_pv_{pR}(e) + \sum V_i = E_RE_pv_{pR}(e) + E_p\left(\sum_n w_iQ_i/P_i\right),$$
$$v(e) = v_{pR}(e) + \sum_n w_iQ_i/P_i$$

is an unbiased estimator for $V(e)$, writing $v_{pR}(e) = v_{pR}(t)|_{\underline{Y}=\underline{R}}$, where $\underline{Y} = (y_1, \ldots, y_i, \ldots, y_N)$, $\underline{R} = (r_1, \ldots, r_i, \ldots, r_N)$ and w_i is V_i if known or is v_i if V_i is unknown but unbiasedly estimated by v_i.

For a general sampling design we shall write

$$t_b = \sum y_ib_{si}I_{si}$$

writing b_{si} as constants free of $\underline{Y}$, $\underline{R}$, $I_{si} = 1$ if $i \in s$ and $=0$ if $i \notin s$ such that $E_p(b_{si}I_{si}) = 1 \forall i$. Then

$$e_b = \sum r_ib_{si}I_{si}$$

will satisfy

$$E(e_b) = E_pE_R(e_b) = E_p(t_b) = Y$$

and also

$$E(e_b) = E_R E_p(e_b) = E_R\left(\sum r_i\right) = Y.$$

Writing $V_p(t_b)=\sum y_i^2 C_i+\sum_{i\neq j}\sum y_i y_j C_{ij}$, where $C_i=E_p(b_{si}^2 I_{si})-1$, $C_{ij}=E_p(b_{si}I_{si}-1)(b_{sj}I_{sj}-1)$ if c_{si}, c_{sij} are available free of $\underline{Y}$, $\underline{R}$, $I_{sij}=I_{si}I_{sj}$ such that $E_p(c_{si}I_{si})=C_i$ and $E_p(c_{sij}I_{sij})=C_{ij}$, then

$$v_p(t_b) = \sum y_i^2 c_{si} I_{si} + \sum_{i\neq j}\sum y_i y_j c_{sij}$$

satisfies $E_p v_p(t_b) = V_p(t_b)$. We shall write $v_p(e_b) = v_p(t_b)|_{\underline{Y}=\underline{R}}$. The literature on survey sampling abounds with examples of such p, b_{si}, c_{si}, c_{sij}'s as one may check from Chaudhuri and Stenger (1992).

It is easy to check that two unbiased estimators as follows are available for $V(e_b) = E_R V_p(e_b) + V_R E_p(e_b)$ as

$$\hat{V}_1(e_b) = v_p(e_b) + \sum w_i b_{si} I_{si} \text{ with } w_i\text{'s as before and for}$$

$$V(e_b) = E_p V_R(e_b) + V_p E_R(e_b) = E_p\left(\sum V_i b_{si}^2 I_{si}\right) + V_p(t_b)$$

as

$$\hat{V}_2(e_b) = v_p(e_b) + \sum w_i (b_{si}^2 - c_{si}) I_{si}.$$

2.3. ORR

If the people in a sub-sample s_1 of s feel the attribute not sensitive enough and divulge their true y_i-values then since knowing these values the interviewer himself/herself may generate r_i for $i \in s_1$ and hence get the option to employ two estimators—one using r_i for $i \in s$ and the other using y_i for $i \in s_1$ and r_i in $s_2 = s - s_1$, namely $e_b = \sum r_i b_{si} I_{si}$ as before and

$$e_b^* = \sum_{i\in s_1} y_i b_{si} I_{si} + \sum_{i\in s_2} r_i b_{si} I_{si}.$$

Writing E_{DR} as the operator for the conditional expectation over RR-device employed only for the units opting for DR keeping the RR's given as fixed it follows that $E_{\mathrm{DR}}(e_b)=e_b^*$. Then, we have the

Theorem. *$E_R(e_b^* = t_b = E_R(e_b)$ and $\hat{V}(e_b^*) = \hat{V}(e_b) - (e_b - e_b^*)^2$.*

Proof. $E_R(e_b^*) = E_R(\sum_{i\in s_1} y_i b_{si} I_{si} + \sum_{i\in s_2} r_i b_{si} I_{si}) = \sum y_i b_{si} I_{si} = E_R(e_b) = t_b$.

Noting that

$$E_R(e_b - e_b^*)^2 = E_R[(e_b - t_b) - (e_b^* - t_b)]^2 = V_R(e_b) - V_R(e_b^*)$$

because

$$E_R(e_b - t_b)(e_b^* - t_b) = E_R(e_b^* - t_b)E_{\mathrm{DR}}(e_b - t_b) = E_R(e_b^* - t_b)^2$$

it follows that

$$V(e_b^*) = E_p V_R(e_b^*) + V_p E_R(e_b^*) = E_p V_R(e_b) - E_p E_R(e_b - e_b^*)^2 + V_p E_R(e_b)$$
$$= V(e_b) - E_p E_R(e_b - e_b^*)^2.$$

Hence the theorem. □

Thus, given any unbiased estimator $\hat{V}(e_b)$, say for $V(e_b)$ we can take

$$\hat{V}(e_b^*) = \hat{V}(e_b) - (e_b - e_b^*)^2$$

as an unbiased estimator for $V(e_b^*)$.

Since

$$E_R(e_b - e_b^*)^2 = E_R\left[\sum_{i \in s_1} (r_i - y_i) b_{si} I_{si}\right]^2 = \sum_{i \in s_1} V_i b_{si}^2 I_{si},$$

an alternative unbiased estimator for $V(e_b^*)$ may also be taken as

$$\hat{V}^*(e_b^*) = \hat{V}(e_b) - \sum_{i \in s_1} w_i b_{si}^2 I_{si},$$

with w_i as before.

In Section 3, we illustrate the use of $\hat{V}(e_b^*)$ rather than $\hat{V}^*(e_b^*)$. For e based on RHC scheme also a similar theory follows with e^* likewise defined.

3. Numerical findings in efficiency gains by optional rather than compulsory RR's

We use artificial data comprising 113 households for which the last month's expenses (Indian Rupees) denoted by z_i for them are used as the size-measures to draw samples by RHC scheme. For one representative member of these households denoted by $i = 1, \ldots, N$ we assigned values of y and x, where

$y_i = 1(0)$ is to be interpreted as the ith person is (not) a habitual gambler, and similarly,

$x_i = 1(0)$ is to be interpreted as the ith person prefers (does not prefer) cricket to football.

From $U = (1, \ldots, N = 113)$ we draw samples by (1) SRSWR, (2) SRSWOR and (3) RHC schemes of sizes $n = 33$ and suppose $n_1 = 24$ randomly selected persons in the sample opt for RR and the remaining $n_2 = 9$ opt for DR's. For SRSWR and SRSWOR we employ the sample mean in estimating θ and use $\hat{V}$ in estimating the variance of the estimate of the total in both SRS and RHC sampling. For an estimator $\hat{\mu}$ for a parameter μ employed with v as the estimate of the variance of $\hat{\mu}$ treating the pivotal $\delta = (\mu - \hat{\mu})/\sqrt{v}$ as a standard normal deviate we take $(\hat{\mu} - 1.96\sqrt{v}, \hat{\mu} + 1.96\sqrt{v})$ as a 95% confidence interval (CI) for μ. Since the population is completely specified we repeatedly draw a sample $\tau = 1000$ times by each method and use the three criteria for comparison namely, (1) ACP, actual coverage percentage which is the percent of the replicated samples for which CI covers μ—the closer it is to 95 the better, (2) ACV, the average coefficient of variation which is the average, over the replicated samples, of the values $100 \times \sqrt{v}/\hat{\mu}$ and (3) AL, the average length of the CI's over the replicated samples. It is worth mentioning that with increasing ACP, the ACV and AL also may undesirably go on increasing. So, an observed value of ACP nearer 96 or 97 may not be more desirable than one nearer 95. For various choices of p, p_1, p_2 for the RR devices illustrated, the table below gives the relative performances of alternative procedures based on repeated samplingof 1000 times each.

For the data about which the results are presented in the Table 1 below $\frac{\sum y_i}{N} = 0.8230$ and $\frac{\sum x_i}{N} = 0.7345$.

In presenting this table we have tried to emphasize that though the RR procedures illustrated involve parameters like p, p_1, p_2, k permitted to be assigned several values, their performances though vary with these values, continue to remain well in terms of the coefficients of variation (CV) and the length and coverage properties of the confidence intervals. So, the results are displayed over variation in parametric values showing that they retain coverage probabilities closer to 95% and yet with CV's desirably low enough.

4. Concluding remarks

As expected, SRSWOR yields less average coefficient of variation than SRSWR though the coverage percentages for the confidence intervals do not often increase. RHC scheme in the present example does not turn out to be an appropriate selection procedure possibly because the size-measures used in sampling are not quite well correlated with the RR's. This correlation is required to be high to control the variance of the estimator. However, optional RR technique necessarily shows improvement compared to the compulsory RR technique, as it should. The main purpose of this paper is to illustrate that if (1) a sample is chosen with unequal selection-probabilities and (2) our objective is to unbiasedly estimate an unknown finite population proportion covering a supposedly sensitive characteristic then a solution is readily available through the uses of RR devices (3) allowing a possible improvement using DR's opted for by the volunteers who do not see it sensitive enough.

In an actual RR survey on addiction possibly practised by some university students many announced 'no inhibition to divulge facts' and we used the revealed facts in employing RR-based analysis without utilizing these DR's in the way explained here thus incurring a loss in efficacy which could be avoided.

Table 1
Comparative performances of alternative procedures

(A) Warner's method

	CRR			ORR		
	ACV	ACP	AL	ACV	ACP	AL
(p)						
SRSWR						
0.08	11.5	95.0	0.36	10.8	95.2	0.34
0.09	11.9	95.4	0.37	11.2	95.7	0.35
0.25	23.9	95.6	0.73	21.3	98.1	0.66
0.30	32.3	97.0	0.94	27.8	98.2	0.84
0.37	56.3	96.6	1.50	48.3	97.5	1.34
0.38	80.9	96.4	1.60	59.2	96.9	1.43
0.66	37.4	97.2	1.15	35.8	97.0	1.03
0.78	20.2	96.2	0.62	18.1	96.4	0.57
0.88	13.4	93.2	0.42	12.4	96.1	0.39
SRSWOR						
0.08	10.0	91.0	0.32	9.2	94.4	0.29
0.09	10.4	91.0	0.33	9.6	94.4	0.31
0.25	20.1	95.6	0.62	17.4	96.0	0.54
0.30	27.0	93.7	0.81	22.8	95.3	0.70
0.37	48.4	95.1	1.34	44.7	94.4	1.15
0.38	65.4	93.6	1.43	43.1	95.1	1.23
0.66	35.9	92.4	1.04	35.5	95.4	0.90
0.78	18.3	93.5	0.56	15.9	96.0	0.50
0.88	11.7	93.5	0.37	10.6	95.3	0.33
RHC						
0.08	21.1	99.4	0.70	20.7	99.5	0.68
0.09	21.2	99.2	0.70	20.8	99.1	0.69
0.25	32.3	98.0	1.02	30.2	96.8	0.95
0.30	42.8	97.5	1.22	38.5	96.5	1.14
0.66	54.9	97.4	1.48	51.7	96.5	1.37
0.78	28.4	98.5	0.91	26.8	98.4	0.86
0.88	22.2	98.6	0.72	21.6	98.6	0.70

(B) Kuk's method

(k, p_1, p_2)						
SRSWR						
(3, 0.13, 0.41)	10.9	92.7	0.34	8.7	86.0	0.28
(3, 0.08, 0.72)	9.0	92.1	0.28	8.0	86.5	0.25
(3, 0.76, 0.75)	9.0	93.9	0.28	8.3	89.3	0.25
(4, 0.29, 0.80)	8.5	93.1	0.27	8.1	90.1	0.25
(4, 0.25, 0.98)	8.2	93.4	0.26	8.2	92.7	0.26
(4, 0.27, 0.71)	8.9	93.1	0.28	8.4	89.2	0.25
(4, 0.44, 0.90)	8.4	92.8	0.26	8.3	91.2	0.25

Table 1 (*continued*)

(B) Kuk's method

	CRR			ORR		
	ACV	ACP	AL	ACV	ACP	AL
(5, 0.17, 0.66)	8.9	93.9	0.28	8.5	90.2	0.25
(5, 0.56, 0.68)	8.9	92.2	0.28	8.6	87.6	0.25
(5, 0.59, 0.84)	8.5	91.6	0.27	8.4	89.6	0.25
(5, 0.25, 0.84)	8.5	92.7	0.26	8.3	90.6	0.24
(6, 0.67, 0.56)	9.1	93.6	0.28	8.2	88.4	0.25
(6,0.84, 0.72)	8.7	93.7	0.27	8.5	91.3	0.25
(6, 0.57, 0.63)	8.7	92.7	0.27	8.2	89.9	0.25
(6, 0.42, 0.68)	8.7	94.5	0.27	8.2	90.8	0.24
SRSWOR						
(3, 0.64, 0.94)	7.0	95.1	0.22	6.6	92.0	0.21
(3, 0.43, 0.81)	7.4	94.4	0.23	6.5	88.6	0.21
(3, 0.39, 0.77)	7.6	95.5	0.24	6.6	89.4	0.21
(3, 0.79, 0.87)	7.3	93.1	0.23	6.6	90.2	0.20
(4, 0.84, 0.97)	7.1	95.0	0.22	6.9	93.4	0.22
(4, 0.20, 0.72)	7.5	94.7	0.24	6.6	89.2	0.21
(4, 0.49, 0.65)	7.7	94.5	0.24	6.5	88.3	0.21
(4, 0.38, 0.77)	7.4	94.7	0.23	6.5	90.6	0.21
(5, 0.43, 0.95)	7.0	94.5	0.22	6.8	92.2	0.22
(5, 0.86, 0.90)	7.1	93.4	0.22	6.7	91.0	0.21
(5, 0.64, 0.73)	7.2	94.2	0.23	6.4	88.4	0.20
(5, 0.88, 0.74)	7.3	94.2	0.23	6.5	90.4	0.21
(6, 0.98, 0.82)	7.0	93.8	0.22	6.5	90.2	0.21
(6, 0.25, 0.67)	7.4	95.0	0.23	6.5	90.2	0.22
(6, 0.29, 0.59)	7.6	94.7	0.24	6.6	89.5	0.21
(6, 0.52, 0.88)	7.1	95.0	0.22	6.7	93.2	0.21
RHC						
(3, 0.57, 0.99)	11.5	90.2	0.39	11.2	89.1	0.38
(3, 0.11, 0.99)	12.0	89.3	0.41	11.8	88.8	0.40
(3, 0.34, 0.77)	13.1	94.0	0.44	11.9	89.2	0.40
(3, 0.46, 0.48)	15.5	96.0	0.51	13.2	91.6	0.44
(4, 0.07, 0.99)	11.9	88.6	0.40	11.8	88.6	0.40
(4, 0.43, 0.98)	11.8	90.7	0.40	11.5	89.5	0.39
(4, 0.25, 0.98)	11.7	90.9	0.40	11.5	90.4	0.39
(4, 0.80, 0.45)	15.6	95.1	0.51	13.2	90.4	0.44
(5, 0.40, 0.75)	13.0	93.1	0.44	12.0	90.6	0.41
(5, 0.25, 0.84)	12.5	92.5	0.43	11.9	90.3	0.41
(5, 0.28, 0.57)	13.8	94.8	0.46	12.3	90.6	0.42
(5, 0.46, 0.50)	14.1	94.8	0.47	12.5	90.8	0.42
(6, 0.67, 0.56)	13.0	94.0	0.44	11.6	89.2	0.40
(6, 0.84, 0.72)	13.0	92.7	0.44	11.9	88.7	0.41
(6, 0.57, 0.63)	13.0	93.4	0.44	11.9	89.7	0.40
(6, 0.78, 0.35)	15.5	94.3	0.50	13.1	90.1	0.44

Table 1 (*continued*)

(C) Unrelated question model

	CRR			ORR		
	ACV	ACP	AL	ACV	ACP	AL
(p_1, p_2)						
SRSWR						
(0.95, 0.98)	8.3	93.3	0.26	8.1	92.5	0.26
(0.29, 0.95)	8.5	94.6	0.27	8.0	92.6	0.25
(0.77, 0.99)	8.2	92.6	0.26	8.1	92.7	0.26
(0.28, 0.97)	8.4	94.0	0.26	8.1	93.2	0.25
(0.75, 0.98)	8.3	94.6	0.26	8.2	93.8	0.26
(0.57, 0.98)	8.2	94.7	0.26	8.1	94.1	0.26
SRSWOR						
(0.45, 0.74)	8.5	94.4	0.27	7.3	92.2	0.23
(0.38, 0.75)	8.4	95.8	0.27	7.2	92.7	0.23
(0.71, 0.98)	7.1	92.9	0.23	7.0	93.1	0.22
(0.69, 0.97)	7.1	93.5	0.22	6.9	94.3	0.22
(0.17, 0.98)	7.0	91.3	0.22	6.9	91.9	0.22
(0.09,0.96)	7.0	92.3	0.22	6.8	94.0	0.22
(0.19, 0.50)	10.8	97.4	0.34	8.8	94.5	0.28
(0.45, 0.74)	8.5	94.4	0.27	7.3	92.2	0.23
(0.55, 0.81)	8.1	94.9	0.25	7.1	94.4	0.22
(0.89, 0.96)	7.2	92.4	0.23	6.9	93.5	0.22
RHC						
(0.59, 0.99)	11.1	91.9	0.37	11.1	91.7	0.37
(0.28, 0.98)	11.0	92.8	0.37	10.8	91.5	0.37
(0.31, 0.99)	11.2	92.0	0.38	11.1	91.2	0.38
(0.68, 0.97)	11.1	91.9	0.37	10.9	90.5	0.37
(0.04, 0.97)	11.1	91.5	0.37	10.9	89.9	0.37
(0.07, 0.94)	11.5	93.1	0.39	11.1	89.8	0.39
(0.25, 0.97)	11.1	93.7	0.37	10.8	91.4	0.37
(0.36, 0.99)	10.8	92.2	0.36	10.7	91.6	0.36

Acknowledgements

Helpful comments including suggestions to leave out detailed numerical data from three referees leading to this improved draft are gratefully appreciated.

References

Chaudhuri, A., 1987. Randomized response surveys of finite populations: A unified approach with quantitative data. J. Statist. Plann. Inference 15, 157–165.

Chaudhuri, A., 1999. Towards a unified theory of randomized response surveys for dichotomous finite populations. Technical Report ASD/99/36.

Chaudhuri, A., 2001a. Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. J. Statist. Plann. Inference 94, 37–42.

Chaudhuri, A., 2001b. Estimating sensitive proportions from unequal probability sample using randomized responses. Pakistan J. Statist. 17 (3), 259–270.

Chaudhuri, A., 2002. Estimating sensitive proportions from randomized responses in unequal probability sampling. Calcutta Statist. Assoc. Bull. 52 (205–208), 315–322.

Chaudhuri, A., Mukerjee, R., 1985. Optionally randomized response techniques. Calcutta Statist. Assoc. Bull. 34, 225–229.

Chaudhuri, A., Mukerjee, R., 1988. Randomized Response: Theory and Techniques, Marcel Dekker, New York.

Chaudhuri, A., Stenger, H., 1992. Survey Sampling: Theory and Methods, Marcel Dekker, New York.

Chaudhuri, A., Adhikary, A.K., Dihidar, S., 2000. Mean square error estimation in multi-stage sampling. Metrika 52 (2), 115–131.

Greenberg, B.G., Abul-Ela, Abdel-Latif, A., Simmons, W.R., Horvitz, D.G., 1969. The unrelated question randomized response model: theoretical framework. J. Amer. Statist. Assoc. 64, 520–539.

Greenberg, B.G., Kuebler, R.R., Abernathy, J.R., Horvitz, D.G., 1977. Respondent hazards in the unrelated question randomized response model. J. Statist. Plann. Inference 1, 53–60.

Gupta, S., 2001. Validation of optional randomized response technique to circumvent social desirability response bias in sensitive survey questions. International Conference on Statistics, Combinatorics and Related Areas and the Eighth International Conference of Forum for Interdisciplinary Mathematics, December 19–21, 2001, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia.

Horvitz, D.G., Shah, B.V., Simmons, W.R., 1967. The unrelated question randomized response model. Proc. Soc. Sect. Amer. Statist. Assoc. 65–72.

Horvitz, D.G., Greenberg, B.G., Abernathy, J.R., 1976. Randomized response: a data-gathering device for sensitive questions. Internat. Statist. Rev. 44, 181–196.

Kuk, A.Y.C., 1990. Asking sensitive question indirectly. Biometrika 77, 436–438.

Mangat, N.S., 1991. An optional randomized response sampling technique using non-stigmatized attribute. Statistica, Anno LI, vol. 4, pp. 595–602.

Rao, J.N.K., Hartley, H.O., Cochran, W.G., 1962. On a simple procedure of unequal probability sampling without replacement. J. Roy. Statist. Soc. B 24, 482–491.

Saha, A., 2003. On efficacies of Dalenius–Vitale technique with compulsory versus optional randomized responses from complex surveys. Calcutta Statist. Assoc. Bull., 54, 215–216, 223–230.

Warner, S.L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. J. Amer. Stat. Assoc. 60, 63–69.