RANDOMIZED RESPONSE RATIO ESTIMATES:

Bias and Efficiency

Abdel-Latif A.Abul-Ela and Sultan M. Abdel Hamied

University of Mansoura, EGYPT

1. Introduction

In sample surveys of human populations, refusal to respond or intentional untruthful reporting constitute customary headache to surveyors studying certain social problems which are sensitive in nature. To overcome such difficulties and to ensure greater cooperation on the part of the respondents, Warner [5] developed a new survey technique which he called the Randomized Response Technique. Since then, extensive work has been done towards improving and modifying Warner's model. This paper is a continuation to one of these attempts, refered to as the Randomized Response Ratio Estimate, abbreviated hereafter by RRRE.

The present research is directed towards obtaining a new estimate of the proportion of the sensitive characteristic, say A based on the ratio R. The properties of the estimate R and the RRRE assuming complete truthfull reporting will be studied. Conditions underwhich the precision and accuracy of the RRRE may be increased are presented.

2. Randomized Response Ratio Estimate

The ratio estimate approach aims at using information collected on some auxiliary related characteristic, say B to estimate the proportion of another sensitive characteristic A, with increased precision. This could prove to be true when the auxiliary characteristic B, is less sensitive (or even non-sensitive) than A, the one we are interested in measuring its frequency.

All possible means could be tried to increase the precision of the estimate of R including proper choice of interviewers, field work control and related information from available sources.

Applying a double sampling technique, the proportions $\hat{\Pi}_a$, $\hat{\Pi}_{b1}$ and $R = \Pi_a / \Pi_{b1}$ may be found from the first sample; where $\hat{\Pi}_a$ and $\hat{\Pi}_{b1}$ are the sample estimates of the proportion of A and B, respectively. From the second sample, which may be a subsample from the first one or an independent sample, another estimate of the proportion of B, say Π_{b2} is found. Thus, the sample estimate of the sensitive proportion Π_a may be defined as following.

$$\hat{\Pi}_{a/r} = \hat{R}\,\hat{\Pi}_{b2} \qquad \ldots(1)^*$$

At this point, it should be noted that the Randomized Response Ratio Estimate of A, i.e. $\hat{\Pi}_{a/r}$ has two components; $\hat{R}$ and Π_{b2}. It follows that the properties of each component should be studied before the discussion of the properties of $\Pi_{a/r}$

* $\hat{\Pi}_{a/r}$ refers to the estimate of Π_a using the ratio estimate technique as defined by (1), to differentiate this estimate from other estimates.

2.1. The estimate $\hat{R}$ and its properties

As noted before, R will be obtained from the first sample. The first sample of size n_1 will be subdivided into two separate samples of size n_{11} and n_{12}, one of them to obtain Π_a and the other to obtain $\hat{\Pi}_{b1}$. Using the Warner's dichotomous model [5] twice, and assuming complete truthful reporting, the estimates of Π_a and Π_b take the form :

$$\hat{\Pi}_a = \frac{(p_1-1)+\lambda_1}{2p_1-1}$$

$$\text{for} \quad p_i \neq \tfrac{1}{2}, \; i=1,2 \qquad \ldots(2)$$

$$\hat{\Pi}_{b1} = \frac{(p_2-1)+\lambda_2}{2p_2-1}$$

With variances given by :

$$\mathrm{Var}(\hat{\Pi}_a) = \frac{\Pi_a(1-\Pi_a)}{n_{11}} + \frac{p_1(1-p_1)}{n_{11}(2p_1-1)^2}$$

and, ... (3)

$$\mathrm{Var}(\hat{\Pi}_{b1}) = \frac{\Pi_{b1}(1-\Pi_{b1})}{n_{12}} + \frac{p_2(1-p_2)}{n_{12}(2p_2-1)^2}$$

Where :

P_i : The probability of drawing the question of interest in the first and second part of the first sample, i = 1,2

λ_i : The proportions of "YES" answers in the first and second part of the first sample, respectively.

Thus, R is estimated by :

$$\hat{R} = \frac{\hat{\Pi}_a}{\hat{\Pi}_{b1}} = \frac{(p_1-1)+\lambda_1}{(2p_1-1)} \cdot \frac{(2p_2-1)}{(p_2-1)+\lambda_2} \qquad \ldots(4)$$

It is quite difficult to obtain the exact expression for Var $(\hat{R})$, since that estimate is biased. Instead, an approximate expression for Var $(\hat{R})$ derived by Abul-Ela et al. [1] is the following:

$$\mathrm{Var}(\hat{R}) = E(\hat{R}-R)^2 \qquad \ldots(5)$$

$$= \frac{1}{\hat{\Pi}_{b1}^2}\left[\mathrm{Var}(\hat{\Pi}_a) + \hat{R}^2\mathrm{Var}(\hat{\Pi}_{b1}) - 2\hat{R}\hat{\rho}\sqrt{\mathrm{Var}(\hat{\Pi}_a)\mathrm{Var}(\hat{\Pi}_{b1})}\right]$$

Where :

$\hat{\Pi}_a, \hat{\Pi}_{b1}$ and $\hat{R}$ are the sample estimates of the true population parameters Π_a, Π_b and R respectively, and

$\hat{\rho}$ denotes the estimate of the correlation coefficient between A and B.

Abul-Ela et al. [1] discussed the precision of the estimate of R under certain assumptions and they reached some general conclusions about the conditions under which the efficiency of the ratio estimate could be increased. To complete the discussion started by Abul-Ela to some of the properties of the ratio estimate, an expression for the bias of that estimate will be derived.

It has been noted that the characteristic B could be chosen to be less-sensitive than A, or even non-sensitive. Therefore, Bias $(\hat{R}) = E(\hat{R}-R)$ considering the two possible choices of B will be briefly treated in the following paragraphs.

2.1.1. Case I : Bias $(\hat{R})$ for a less-sensitive Characteristic B

Bias $(\hat{R})$ is defined as following :

$$\text{Bias}(\hat{R}) = E(\hat{R}-R)$$

$$= E\left(\frac{\hat{\Pi}_a}{\hat{\Pi}_{b1}} - R\right)$$

$$= E\left(\frac{\hat{\Pi}_a - R\hat{\Pi}_{b1}}{\hat{\Pi}_{b1}}\right)$$

Finding an exact expression for Bias $(\hat{R})$ from the last equation is rather tedious, since $\hat{\Pi}_a$ and $\hat{\Pi}_{b1}$ varies from sample to sample however, it is possible to derive an approximate expression for Bias $(\hat{R})$ in the following form [2]:

Let $$\frac{1}{\hat{\Pi}_{b1}} = \frac{1}{\Pi_b+(\hat{\Pi}_{b1}-\Pi_b)}$$

$$= \frac{1}{\Pi_b\left[1+\frac{(\hat{\Pi}_{b1}-\Pi_b)}{\Pi_b}\right]}$$

$$= \frac{1}{\Pi_b}\left(1+\frac{\hat{\Pi}_{b1}-\Pi_b}{\Pi_b}\right)^{-1}$$

Expanding $\left(1+\frac{\hat{\Pi}_{b1}-\Pi_b}{\Pi_b}\right)^{-1}$ by Taylor's series, results in :

$$\text{Bias}(\hat{R}) = E\left[(\hat{\Pi}_a-R\hat{\Pi}_{b1})\frac{1}{\hat{\Pi}_{b1}}\right]$$

$$= E\left[(\hat{\Pi}_a-R\hat{\Pi}_{b1})\frac{1}{\Pi_b}\left(1+\frac{\hat{\Pi}_{b1}-\Pi_b}{\Pi_b}\right)^{-1}\right]$$

$$= E\left[(\hat{\Pi}_a-R\hat{\Pi}_{b1})\frac{1}{\Pi_b}\left(1-\frac{\hat{\Pi}_{b1}-\Pi_b}{\Pi_b}+\frac{(\hat{\Pi}_{b1}-\Pi_b)^2}{\Pi_b^2} - \ldots\ldots\right)\right]$$

Limitting ourselves to the first two terms in the expansion, since other terms will be of the order $(1/n)^2$, an approximate formula for Bias $(\hat{R})$ is :

$$\text{Bias}(\hat{R}) = \frac{1}{\Pi_b}E(\hat{\Pi}_a-R\hat{\Pi}_{b1}) - \frac{1}{\Pi_b^2}E(\hat{\Pi}_a-R\hat{\Pi}_{b1})(\hat{\Pi}_{b1}-\Pi_b)$$

But, $\hat{\Pi}_a$ and $\hat{\Pi}_{b1}$ are unbiased estimates , thus

$$\text{Bias}(\hat{R}) = 0 - \frac{1}{\Pi_b^2}\left[E\,\hat{\Pi}_a(\hat{\Pi}_{b1}-\Pi_b) - R\,E\,\hat{\Pi}_{b1}(\hat{\Pi}_{b1}-\Pi_b)\right]$$

$$= -\frac{1}{\Pi_b^2}\left[E(\hat{\Pi}_a-\Pi_a)(\hat{\Pi}_{b1}-\Pi_b) - RE(\hat{\Pi}_{b1}-\Pi_b)^2\right]$$

$$= -\frac{1}{\Pi_b^2}\left[\text{Cov}(\hat{\Pi}_a,\hat{\Pi}_{b1}) - R\,\text{Var}(\hat{\Pi}_{b1})\right]$$

$$= \frac{1}{\Pi_b^2}\left[R\,\text{Var}(\hat{\Pi}_{b1}) - \rho\sqrt{\text{Var}(\hat{\Pi}_a)\text{Var}(\hat{\Pi}_{b1})}\right] \quad \ldots (6)$$

For practical application, the unknown true parameters Π_b and R in the above expression, can be replaced by their sample estimates. Therefore :

$$\text{Bias}(\hat{R}) = \frac{1}{\hat{\Pi}_{b1}^2}\left[\hat{R}\,\text{Var}(\hat{\Pi}_{b1}) - \hat{\rho}\sqrt{\widehat{\text{Var}}(\hat{\Pi}_a)\widehat{\text{Var}}(\hat{\Pi}_{b1})}\right] \quad \ldots (7)$$

From (7), the following conclusions could be reached [3] :

1) Bias $(\hat{R})$ decreases-as expected-as total sample size increases.
2) Bias $(\hat{R})$ decreases as ρ increases from -1 to +1
3) Bias $(\hat{R})$ will be zero if :

$$\hat{R}\,\text{Var}(\hat{\Pi}_{b1}) = \hat{\rho}\sqrt{\widehat{\text{Var}}(\hat{\Pi}_a)\widehat{\text{Var}}(\hat{\Pi}_{b1})}$$

i.e. if : $\hat{\Pi}_a = \hat{\rho}\sqrt{\widehat{\text{Var}}(\hat{\Pi}_a)/\widehat{\text{Var}}(\hat{\Pi}_{b1})}\;\hat{\Pi}_{b1}$

The last remark implies that Bias $(\hat{R}) = 0$ if the regression of $\hat{\Pi}_a$ on $\hat{\Pi}_{b1}$ is a straight line going through the origin, and the quantity $\hat{\rho}\sqrt{\widehat{\text{Var}}(\hat{\Pi}_a)/\widehat{\text{Var}}(\hat{\Pi}_{b1})}$ represents the slope of the regression line.

4) From (5) and (7), $\widehat{\text{Var}}(\hat{R})$ could be written in terms of Bias $(\hat{R})$ as following :

$$\text{Var}(\hat{R}) = \frac{1}{\hat{\Pi}_{b1}^2}\left[\text{Var}(\hat{\Pi}_a) - \hat{R}\hat{\rho}\sqrt{\widehat{\text{Var}}(\hat{\Pi}_a)\widehat{\text{Var}}(\hat{\Pi}_{b1})}\right] + \hat{R}\,\text{Bias}(\hat{R}) \quad \ldots\ldots (8)$$

Assuming, for simplicity, that $n_{11}=n_{12}=1000$, say, $p_1=p_2=p$ and under complete truthful reporting, Bias $(\hat{R})$ is numerically investigated in table (1). Results given in that table show that :

a) For the same values of Π_a , Bias $(\hat{R})$ tends to decrease rapidly with increasing values of ρ, p and Π_b
b) For positive values of ρ , the ratio estimate tends to under-estimate Π_a especially with ρ coming closer to "One".
c) Bias $(\hat{R})$ diminishes rapidly, the more ρ approaches zero from either side. This may be explained by the very nature of the ratio estimate and the dependence of the unbiased estimates of Π_a and Π_b on the biased estimate of R.

Based on these results, the following recommendations may help increasing the precision of $\hat{R}$:

1. Choose the two characteristics A and B highly positive correlated. Low positive or negative correlated A and B, however, are not excluded. In this case, respondents could feel more confident to report about the sensitive trait A.
2. Choose P as high (or as low) as possible, however, it should not be too high (or as low) to arouse suspicions of the respondent.
3. Choose the second characteristic B to be more frequent than the usually unknown characteristic A. In any case, the last condition is expected to be satisfied, since, the less-sensitive or non-sensitive trait is anticipated to be more frequent than the sensitive one.

TABLE(1) Bias $(\hat{R})$ for some values of Π_a, Π_b P and ρ (B is less-sensitive)

Π_a	P	Π_b	ρ = -1	-.5	0	.1	.2	.5	1
.1	.6	.2	.2301	.1535	.0770	.0617	.0465	.0005	-.0761
		.6	.0200	.0114	.0029	.0012	-.0053	-.0057	-.0142
		.8	.0107	.0095	.0012	.0003	-.0007	-.0036	-.0084
	.9	.2	.0103	.0070	.0037	.0030	.0025	.0004	-.0028
		.6	.0009	.0006	.0002	.0001	.0000	-.0002	-.0006
		.8	.0005	.0003	.0001	.0000	.0000	-.0001	-.0004
.2	.6	.2	.3080	.2310	.1540	.1385	.1232	.0770	.0000
		.6	.0229	.0144	.0058	.0041	.0023	-.0028	-.0114
		.8	.0120	.0072	.0024	.0014	.0005	-.0024	-.0072
	.9	.2	.0150	.0113	.0075	.0067	.0060	.0037	.0000
		.6	.0012	.0003	.0004	.0003	.0002	-.0001	-.0005
		.8	.0006	.0003	.0001	.0001	.0000	-.0001	-.0004

2.1.2. Case II : Bias $(\hat{R})$ for a non-sensitive Characteristic B

Assuming B is a non-sensitive characteristic but related to A, therefor Π_b will be estimated from direct interviewing. Thus, the estimates $\hat{\Pi}_a$ and $\hat{\Pi}_{b1}$ will be obtained using the first sample as a whole.

Bias $(\hat{R})$ will take the same form as given by (7), with the sample estimate of Var $(\hat{\Pi}_{b1})$ given by :

$$\text{Var}\ (\hat{\Pi}_{b1}) = \hat{\Pi}_{b1}\ (1-\hat{\Pi}_{b1})/n_1 \qquad \ldots\ (9)$$

Considering this assumption, Bias $(\hat{R})$ will be numerically investigated under truthful reporting. Table (2) presents a summary of this investigation.

As expected, Bias $(\hat{R})$ decreases rapidly so long as B is considered non-sensitive. All other conclusions and recommendations about Bias $(\hat{R})$ mentioned in the previous Case I still hold.

2.2. The estimate $\hat{\Pi}_{b2}$ and its properties

As noted before $\hat{\Pi}_{b2}$ estimates the proportion of characteristic B from the second sample of size n_2. Assuming complete truthful reporting, then the randomized response estimate [5] of Π_{b2} is :

$$\hat{\Pi}_{b2} = \frac{(p_3-1)+\lambda_3}{(2p_3-1)}\ ,\ p_3 \neq \tfrac{1}{2} \qquad \ldots .\ (10)$$

This estimate is unbiased with sample estimate of the variance [5] :

$$\text{Var}\ (\hat{\Pi}_{b2}) = \frac{\hat{\Pi}_{b2}(1-\hat{\Pi}_{b2})}{n_2} + \frac{p_3(1-p_3)}{n_2(2p_3-1)^2} \qquad \ldots\ (11)$$

where : $\Pi_{b1} = \Pi_{b2} = \Pi_b$, and

p_3 : The probability of drawing the question referring the respondent to characteristic B in the second sample.

λ_3 : The proportion of "YES" answers reported by respondents when questioned about B in the second sample.

TABLE (2) Bias $(\hat{R})$ for some values of Π_a, Π_b, and ρ (B is non-sensitive)

Π_a	P	Π_b	ρ = -1	-.5	0	.1	.2	.5	1
.1	.6	.2	.0257	.0138	.0020	-.0004	-.0027	-.0098	-.0217
		.6	.0034	.0017	.0011	-.0001	-.0005	-.0015	-.0032
		.8	.0015	.0008	.0000	-.0001	-.0003	-.0007	-.0014
	.9	.2	.0068	.0044	.0020	-.0015	.0010	-.0003	-.0027
		.6	.0007	.0004	.0001	.0001	.0000	-.0002	-.0005
		.8	.0003	.0002	.0000	.0000	.0000	-.0001	-.0003
.2	.6	.2	.0277	.0158	.0040	.0016	-.0007	-.0078	-.0197
		.6	.0036	.0019	.0002	-.0001	-.0004	-.0015	-.0032
		.8	.0015	.0008	.0001	-.0001	-.0002	-.0007	-.0014
	.9	.2	.0095	.0067	.0040	.0035	.0030	.0013	-.0015
		.6	.0009	.0006	.0002	.0001	.0000	-.0002	-.0005
		.8	.0004	.0002	.0001	.0000	.0000	-.0001	-.0003

3. Properties of the Randomized Response Ratio Estimate $(\hat{\Pi}_{a/r})$

In this section, the properties of the RRRE of A. i.e. $\hat{\Pi}_{a/r}$ as given by (1) will be discussed in brief, considering the two different cases about the choice of the second characteristic B :

Case I : B is less-sensitive than A.

Case II : B is non-sensitive

3.1. Case I : B is Less-Sensitive than A

Assuming that A is a sensitive characteristic and B is a related but less-sensitive trait, two independent simple random samples of size n_1 and n_2 are drawn with replacement. As noted before, the estimate $\hat{R}$ is computed from the sub-sample of size n_{11} from the first sample and $\hat{\Pi}_{b2}$ computed from the second sample. The first sample is expected to be larger than the second, because the first sample will be used to obtain the estimates $\hat{\Pi}_a$ and $\hat{\Pi}_{b1}$, in addition, A is expected to be less frequent than B.

Thus, the RRRE of Π_a can be estimated in terms of $\hat{R}$ and $\hat{\Pi}_{b2}$ from the following relation.

$$\hat{\Pi}_{a/r} = \hat{R}.\ \hat{\Pi}_{b2} = (\hat{\Pi}_a\ /\hat{\Pi}_{b1}\ ).\ \hat{\Pi}_{b2}$$

Assuming complete truthful reporting in both samples, the unbiased estimates $\hat{\Pi}_a, \hat{\Pi}_{b1}$ and $\hat{\Pi}_{b2}$ are as given before.

3.1.1 Var $(\hat{\Pi}_{a/r})$:

The estimate $\hat{\Pi}_{a/r}$ is dependent on the biased estimate $\hat{R}$. Evedently so, $\hat{\Pi}_{a/r}$ is expected to be biased, and an exact expression for its variance is quite difficult. Therefore, an approximate expression for var $(\hat{\Pi}_{a/r})$ will be derived as following [2,3,4] :

$$\hat{\Pi}_{a/r} - \Pi_a = \hat{R}\,\hat{\Pi}_{b2} - \Pi_a$$
$$= (\hat{\Pi}_a/\hat{\Pi}_{b1})\,\hat{\Pi}_{b2} - \Pi_a$$
$$= \left(\frac{\hat{\Pi}_a}{\hat{\Pi}_{b1}}\Pi_b - \Pi_a\right) + \frac{\hat{\Pi}_a}{\hat{\Pi}_{b1}}(\hat{\Pi}_{b2} - \Pi_b)$$
$$= \frac{\Pi_b}{\hat{\Pi}_{b1}}(\hat{\Pi}_a - R\hat{\Pi}_{b1}) + \frac{\hat{\Pi}_a}{\hat{\Pi}_{b1}}(\hat{\Pi}_{b2} - \Pi_b)$$

In order to obtain an approximate expression for the Var $(\hat{\Pi}_{a/r})$, it is possible to replace the factor $(\Pi_b/\hat{\Pi}_{b1})$ by unity in the first term of the above equation. Also, it is possible to replace the factor $(\hat{\Pi}_a/\hat{\Pi}_{b1})$ in the second component in the same equation by the population ratio $R = \Pi_a/\Pi_b$ [2], considering that $\hat{\Pi}_a$, $\hat{\Pi}_{b1}$ and $\hat{\Pi}_{b2}$ are unbiased estimates.

Thus, the last equation can be written in the form :

$$\hat{\Pi}_{a/r} - \Pi_a = (\hat{\Pi}_a - R\,\hat{\Pi}_{b1}) + R\,(\hat{\Pi}_{b2} - \Pi_b)$$

It is clear that the first term is due to the first sample, and the other term comes from the second sample. If the two samples are drawn independently, then the approximate formula for Var$(\hat{\Pi}_{a/r})$ will be :

$$\text{Var}\ (\hat{\Pi}_{a/r}) = \text{Var}(\hat{\Pi}_a - R\,\hat{\Pi}_{b1}) + R^2\ \text{Var}\ (\hat{\Pi}_{b2} - \Pi_b)$$
$$= [\text{Var}\ (\hat{\Pi}_a) + R^2\text{Var}(\hat{\Pi}_{b1}) - 2R\rho\sqrt{\text{var}(\hat{\Pi}_a)\text{Var}(\hat{\Pi}_{b1})}] + [R^2\ \text{var}(\hat{\Pi}_{b2})] \quad \ldots\ (12)$$

Where Var $(\hat{\Pi}_a)$, Var $(\hat{\Pi}_{b1})$ and Var $(\hat{\Pi}_{b2})$ are as given before, and the sample estimates can be used instead of the unknown true parameters.

From (5) and (12) , Var $(\hat{\Pi}_{a/r})$ may be written:

$$\text{Var}(\hat{\Pi}_{a/r}) = \hat{\Pi}^2_{b1}\ \text{Var}\ (\hat{R}) + \hat{R}^2\ \text{Var}\ (\hat{\Pi}_{b2}) \quad \ldots\ (13)$$

3.1.2. Bias $(\hat{\Pi}_{a/r})$:

To derive an approximate formula for Bias $(\hat{\Pi}_{a/r})$ where $\hat{\Pi}_{a/r}$ is given by (1), $\hat{\Pi}_{a/r}$ may be written as following [3] :

$$\hat{\Pi}_{a/r} = [(\hat{R}-R)+R]\ [(\hat{\Pi}_{b2} - \Pi_b) + \Pi_b]$$
$$= (\hat{R}-R)(\hat{\Pi}_{b2} - \Pi_b) + \Pi_b\ (\hat{R}-R) + R(\hat{\Pi}_{b2} - \Pi_b) + R\ \Pi_b$$
$$E\ (\hat{\Pi}_{a/r}) = E\ (\hat{R}-R)(\hat{\Pi}_{b2} - \Pi_b) + \Pi_b\ E\ (\hat{R}-R) + R\ E(\hat{\Pi}_{b2} - \Pi_b) + R\ \Pi_b$$

Since $\hat{\Pi}_{b2}$ is an unbiased estimate and $E(\hat{R}-R) = \text{Bias}(\hat{R})$ as given before, then :

$$E\ (\hat{\Pi}_{a/r}) = \text{Cov}\ (\hat{R}, \hat{\Pi}_{b2}) + \Pi_b\ \text{Bias}\ (\hat{R}) + R\ \Pi_b$$
$$E\ (\hat{\Pi}_{a/r}) - R\ \Pi_b = \text{Cov}\ (\hat{R}, \hat{\Pi}_{b2}) + \Pi_b\ \text{Bias}\ (\hat{R})$$
$$\therefore\ \text{Bias}\ (\hat{\Pi}_{a/r}) = \rho_{\hat{R},\hat{\Pi}_{b2}}\sqrt{\text{Var}(\hat{R})\ \text{var}(\hat{\Pi}_{b2})} + \Pi_b\text{Bias}\ (\hat{R}) \quad \ldots(14)$$

3.1.3. Precision of the RRRE, $\hat{\Pi}_{a/r}$:

In this section, the precision of the RRRE$(\hat{\Pi}_{a/r})$ will be numerically investigated for different values of parameters of interest. For simplicity, it is assumed that :

(i) $p_1 = p_2 = p_3 = p$

(ii) $n_1 = 2000$ and will be divided into two samples of equal size $(n_{11} = n_{12} = n_1/2 = 1000)$ to obtain the estimates $\hat{\Pi}_a$ and $\hat{\Pi}_{b1}$

(iii) $n_2 = 1000$ to obtain the estimate $\hat{\Pi}_{b2}$

From equations (13) and (14) and for these parameters values, tables (3) and (4) will present the results of this investigation.

From table (3), the following trends could be easily observed :

(a) For the same values of Π_a, Var $(\hat{\Pi}_{a/r})$ decreases for increasing values of P, Π_b and ρ . The lowest values are always reached the more ρ is coming closer to "one".

(b) High (or low) values of p helps decreasing the value of Var $(\hat{\Pi}_{a/r})$. This result is due to the contribution of the randomizing device to the variance, which decreases with increasing p.

In additon, table (4) shows that except for moderate relationship between A and B, Bias $(\hat{\Pi}_{a/r})$ is almost negligible and comes closer to "zero" with ρ approaching $\pm$ 1.

These results are generally in agreement with Abul-Ela et al . [1] results. Also our previous conclusions concerning Bias $(\hat{R})$, except for the situation when ρ approaches zero from either side, are close to those above mentioned about Var $(\hat{\Pi}_{a/r})$

TABLE (3) Var$(\hat{\Pi}_{a/r})$ for Π_b from 0.2 to 0.8, for some values of Π_a , p and ρ. (B less-sensitive)

Π_a	P	Π_b	ρ = -1	-.5	0	.3	.5	1
.1	.6	.2	.01529	.01223	.00917	.00733	.00611	.00305
		.4	.00995	.00841	.00687	.00595	.00533	.00379
		.6	.00849	.00746	.00644	.00583	.00542	.00439
		.8	.00781	.00704	.00628	.00582	.00552	.00475
	.9	.2	.00064	.00051	.00038	.00031	.00025	.00012
		.4	.00042	.00035	.00028	.00024	.00021	.00014
		.6	.00034	.00030	.00025	.00022	.00020	.00016
		.8	.00030	.00027	.00024	.00022	.00021	.00018
.2	.6	.2	.03080	.02464	.01848	.01479	.01232	.00616
		.4	.01547	.01237	.00928	.00742	.00619	.00209
		.6	.01168	.00962	.00755	.00631	.00545	.00342
		.8	.01001	.00847	.00693	.00601	.00539	.00385
	.9	.2	.00150	.00120	.00090	.00072	.00060	.00030
		.4	.00083	.00066	.00049	.00039	.00032	.00015
		.6	.00061	.00049	.00038	.00031	.00027	.00015
		.8	.00049	.00041	.00034	.00029	.00027	.00019

TABLE (4) Bias ($\hat{\Pi}_{a/r}$) for Π_a = 0.1,0.2, for some values of Π_b, p and ρ (B less-sensitive)

Π_a	ρ	Π_b	p=.6	p=.9	Π_a	ρ	Π_b	p=.6	p=.9
.1	- 1	.2	.00000	.00002	.2	- 1	.2	.00001	.00001
		.4	.00000	.00003			.4	.00000	.00002
		.6	.00000	.00004			.6	.00000	.00002
		.8	.00000	.00005			.8	.00000	.00002
	-.5	.2	.01042	.00060		-.5	.2	.01953	.00096
		.4	.00276	.00019			.4	.00527	.00031
		.6	.00125	.00008			.6	.00241	.00015
		.8	.00070	.00005			.8	.00136	.00012
	.5	.2	.01335	.00066		.5	.2	.03079	.00097
		.4	.00314	.00019			.4	.00676	.00040
		.6	.00136	.00008			.6	.00286	.00017
		.8	.00075	.00004			.8	.00155	.00010
	1	.2	.00001	-.00001		1	.2	.00000	.00000
		.4	.00000	-.00005			.4	.00000	-.00021
		.6	.00000	-.00007			.6	.00000	-.00016
		.8	.00000	-.00014			.8	.00000	-.00014

3.2. Case II : B is a Non-sensitive Characteristic

For a non-sensitive characteristic B, the first sample will be used totally to estimate Π_a , using the randomized response tehnique, meanwhile the other estimate $\hat{\Pi}_{b1}$ will be estimated from the same sample by direct interviewing, and the estimate $\hat{R}$ is computed. The second estimate $\hat{\Pi}_{b2}$ will be found from the second sample by questioning respondents about their affiliation to group B directly, Therefore : $\hat{\Pi}_{a/r}$ will be :

$$\hat{\Pi}_{a/r} = \hat{R} \; \hat{\Pi}_{b2} = \left(\frac{\hat{\Pi}_a}{\hat{\Pi}_{b1}}\right) \hat{\Pi}_{b2}$$

where :

(i) $\hat{\Pi}_{b1} = (\acute{n}_1/n_1)$ and $\hat{\Pi}_{b2} = (\acute{n}_2/n_2)$

(ii) $\acute{n}_1$ and $\acute{n}_2$: the number of "YES" answers to the non-sensitive question in the first and second sample, respectively.

The approximate expressions for the variance and the bias of $\hat{\Pi}_{a/r}$ will be as given by (13) and (14) taking into account that Var ($\hat{\Pi}_{b1}$) and Var ($\hat{\Pi}_{b2}$) will be the regular binomial variance , i.e. :

$$\text{Var } (\hat{\Pi}_{bi}) = (\hat{\Pi}_{bi}(1- \hat{\Pi}_{bi})/n_i \text{ , for } i = 1,2$$

Tables (5) and (6) shows in brief, the sizeable gains in efficiency of the RRRE when the second characteristic B is non-sensitive, assuming that n_1=2000, n_2=1000 and all respondents tell the truth.Such results are quite expected, since the first sample is used, as a whole, to find $\hat{\Pi}_a$ and $\hat{R}$. In addition, the contribution of the randomizing device to Var ($\hat{\Pi}_{bi}$) denoted by $p_i\ (1-p_i)/[n_i(2p_i-1)^2]$ is eliminated because of direct interviewing.

TABLE (5) Var ($\hat{\Pi}_{a/r}$) for Π_b from 0.2 to 0.8, for some values of Π_a , p and ρ (B non-sensitive)

Π_a	P	Π_b	ρ = -1	-.5	0	.3	.5	1
.1	.6	.2	.00359	.00335	.00310	.00300	.00285	.00261
		.4	.00336	.00321	.00306	.00297	.00291	.00276
		.6	.00325	.00315	.00305	.00299	.00295	.00285
		.8	.00316	.00310	.00304	.00300	.00298	.00292
	.9	.2	.00026	.00022	.00017	.00014	.00012	.00007
		.4	.00019	.00016	.00013	.00011	.00010	.00008
		.6	.00016	.00014	.00012	.00011	.00010	.00008
		.8	.00013	.00012	.00011	.00010	.00010	.00009
.2	.6	.2	.00431	.00381	.00332	.00303	.00283	.00233
		.4	.00378	.00347	.00317	.00299	.00287	.00256
		.6	.00353	.00332	.00312	.00300	.00292	.00271
		.8	.00334	.00322	.00310	.00303	.00298	.00286
	.9	.2	.00061	.00050	.00039	.00032	.00028	.00017
		.4	.00037	.00031	.00024	.00020	.00017	.00011
		.6	.00028	.00032	.00019	.00016	.00015	.00010
		.8	.00022	.00019	.00017	.00016	.00015	.00012

TABLE (6) * Relative Efficiency of RRRE with B non-sensitive to RRRE with B less-sensitive for some given values of Π_a , Π_b , P and ρ

Π_a	P	Π_b	ρ = -1	-.5	0	.5	1
.1	.9	.2	246	232	223	208	150
		.4	221	219	215	210	200
		.6	212	214	208	200	200
		.8	231	225	218	210	200
.2	.9	.2	246	240	231	214	176
		.4	224	213	204	188	136
		.6	218	213	200	180	150
		.8	223	216	200	180	158

* Relative Efficiency defined by

$$= \frac{\text{Var}(\hat{\Pi}_{a/r})(\text{L.S.})}{\text{Var}(\hat{\Pi}_{a/r})(\text{N.S.})} \times 100$$

The similarity of the observed trends of the bias and variance values of $\hat{R}$ and $\hat{\Pi}_{a/r}$ indicated in this paper and as presented in tables (1,2 , ...,6) suggests the following recommendations: The accuracy of the Randomized Response Ratio Estimate ($\Pi_{a/r}$) is increased, when :

(1) B is more frequent than A, and
(2) B is less-sensitive and is better to be non-sensitive, and
(3) A and B are highly positive correlated.
Considering the magnitude of the variance and the bias presented by table(7), it is seen that Var ($\Pi_{a/r}$) is dominating the MSE of the estimate. Therefore, the early noted recommendation to choose A and B such that ρ is close to 0 to minimize the bias is of no significance.
(4) Choose P as high (or as low) as possible, without being too revealing.

TABLE (7) Contributions of $[\text{Bias}(\hat{\Pi}_{a/r})]^2$ to MSE $(\hat{\Pi}_{a/r})$* for different values of parameters, (B less-sensitive)

Π_a	P	Π_b	ρ = -1	-.5	.5	1
.1	.6	.2	.0000	.8794	2.8339	.0000
		.4	.0000	.0903	.1835	.0000
		.6	.0000	.0201	.0332	.0000
		.8	.0000	.0057	.0090	.0000
	.9	.2	.0000	.0588	.1600	.0000
		.4	.0000	.0000	.0000	.0000
		.6	.0000	.0000	.0000	.0000
		.8	.0000	.0000	.0000	.0000
.2	.6	.2	.0000	1.5243	7.1450	.0000
		.4	.0000	.2234	.7313	.0000
		.6	.0000	.0603	.1484	.0000
		.8	.0000	.0212	.0445	.0000
	.9	.2	.0000	.0750	.1500	.0000
		.4	.0000	.0000	.0313	.0000
		.6	.0000	.0000	.0000	.0000
		.8	.0000	.0000	.0000	.0000

$$* \quad \frac{[\text{Bias}(\Pi_{a/r})]^2}{\text{MSE}} \times 100$$

4. Conclusion

Randomized Response Technique has been used to estimate the proportion Π_a of individuals with stigmatizing characteristic A. This has been possible by getting information on such a characteristic by indirect interviewing method on a probability basis. Feeling self protected, the chances of cooperation from respondent's side increase, and better estimates are expected.

An additional source of self-protection arises from estimating Π_a based on information collected on some auxiliary related less-sensitive or even non-sensitive characteristic B.

Thus, the RRRE of Π_a , i.e. $(\hat{\Pi}_{a/r})$ is derived and its properties are studied. Numerical investigations of Bias $(\hat{\Pi}_{a/r})$ and Var $(\hat{\Pi}_{a/r})$ helped in reaching some rules about the optimal values of the estimating parameters Π_b, P and ρ which increase the accuracy of the estimate $\hat{\Pi}_{a/r}$. In our presentation, repeated sampling of the same population with two independent samples has been used to estimate Π_a. The question is open as to whether it is more practical to get the information on A and B, from the same sample. It should be noted in this connection that chances of bias resulting from correlated answers and less cooperation from interviewers may be increased.

REFERENCES

[1] Abul-Ela, Abdel-Latif A. & Dakrouri, Hala M. Randomized Response Models : A Ratio Estimator Proceeding of the Social Statistics Section, American Statistical Association, 1980,PP.205-8

[2] Cochran, W.G. Sampling Techniques, 3rd ed., John Wiley and Sons, Inc., New York, 1977.

[3] Jessen, R.J., Statistical Survey Techniques, John Wiley and Sons, Inc. New York, 1978.

[4] Raj, D. Sampling Theory, Mc-Graw Hill, Inc., 1979.

[5] Warner, S.L., Randomized Response : A Survey Technique for Eliminating Evasive Answer Bias. JASA, 60, 1965, PP. 63-69