MODIFICATIONS OF THE RANDOMIZED RESPONSE TECHNIQUE FOR SAMPLING WITHOUT REPLACEMENT

Jong-Ik Kim and John A. Flueck
Temple University

I. INTRODUCTION

Warner's pioneering model (1965) in effect considered the case where both the respondent and question populations were sampled with replacement. However, in real-world situations, one generally desires to sample respondents without replacement and perhaps sample questions without replacement. In this situation, consideration has to be given to modifications of the Warner model and its estimators.

We propose to examine the following four sampling cases:

- Case I, the respondent and question populations are both sampled with replacement;
- Case II, the respondent population is sampled without replacement and the question population is sampled with replacement;
- Case III, the respondent population is sampled with replacement and the question population is sampled without replacement; and finally,
- Case IV, both the respondent and question populations are sampled without replacement.

As is seen, these four cases are simply the four possible combinations of binomial and hypergeometric processes.

We present both the point estimators ($\hat{\pi}$) of π and the variances of $\hat{\pi}$ for all four cases. In addition, some theorems and numerical results are presented for the comparisons of the four cases.

The sampling results of this section also can be extended to other RRT models. In particular, the unrelated question model is modified to handle the above four sampling cases. Lastly, we also show how one can design randomization devices to handle the sampling of questions without replacement.

II. THE FOUR CASES

Let N represent the number of individuals in the finite respondent population, M represent the number of questions in the finite question population and p represent the proportion of the M questions which refer to the sensitive question (Q1). For simplicity, it is assumed that the finite population of N respondents is composed of A who are members of group 1 (G_1) and N-A who are members of group not-1 ($G_{\bar{1}}$) and hence $A/N=\pi$. Correspondingly, the finite population of M questions is composed of B of type Q1 and M-B of type Q2 and hence $B/M=p$. We follow Boruch (1972) and utilize the following notation:

$$x_i = \begin{cases} 1 & \text{if the } i^{th} \text{ respondent belongs to } G_1 \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if the } i^{th} \text{ respondent selects Q1} \\ 0 & \text{otherwise} \end{cases}$$

$$z_i = \begin{cases} 1 & \text{if the } i^{th} \text{ respondent's answer is yes} \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$z_i = x_iy_i + (1-x_i)(1-y_i) = 2x_iy_i-x_i-y_i+1. \quad (2.1)$$

We also define $X = \sum_{i=1}^{n} x_i$, $Y = \sum_{i=1}^{n} y_i$ and $Z = \sum_{i=1}^{n} z_i$, where X is the random variable (r.v.) representing the total number of respondents in the sample having characteristic 1, Y is the r.v. representing the total number of times Q1 is selected by the n respondents, and Z is the r.v. representing the total number of "yeses" reported by the n respondents. Following the standard formula for the variance of a sum of r.v.'s, we have

$$V(Z) = E(\sum_{i=1}^{n} z_i^2)+E\sum_{i\neq j}^{n} z_iz_j)-\{E(\sum_{i=1}^{n} z_i)\}^2. \quad (2.2)$$

This formula can be used for all four cases with only its second term differing from case to case.

2.1.1. Both Respondent and Question Populations Sampled with Replacement (Case I)

This case is Warner's original model (1965) and the theoretical results are well known. They can be summarized as follows:

i). The p.d.f. of the number of "yeses" (Z) is a binomial distribution with parameter λ.

ii). The estimator of π is, $\hat{\pi}=\{\frac{Z}{N} - (1-p)\}/(2p-1)$ $(p\neq\frac{1}{2})$, and is the same for the following three other cases. In short, the change to sampling respondents and/or questions without replacement does not affect the point estimator π.

iii). Warner showed that $E(\hat{\pi})=\pi$, and because the $E(Z)=n\lambda$ irrespective of whether X or Y is binomial or hypergeometric, $E(\hat{\pi})$ is the same for all four cases.

iv). The $V(\hat{\pi})$ is

$$V_1(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \quad (p\neq\tfrac{1}{2}) \quad (2.3)$$

2.1.2. Respondent Population Sampled without Replacement and Question Population Sampled with Replacement (Case II)

Here X is hypergeometrically distributed ($X \stackrel{d}{=} H(n,N,A)$) and Y is still binomially distributed ($Y \stackrel{d}{=} B(n,p)$).

Using equation (2.2), and noting that

$$E(x_i x_j) = \Pr(x_i=1 \text{ and } x_j=1) = \frac{\binom{N\pi}{2}}{\binom{N}{2}} = \frac{\pi(N\pi-1)}{N-1}$$

(the probability of selecting two persons both of whom have the sensitive characteristic out of a total of $N\pi$ such persons), one gets the $V(\hat{\pi})$ for Case II as

$$V_2(\hat{\pi}) = \frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right) + \frac{p(1-p)}{n(2p-1)^2} \quad (p\neq\tfrac{1}{2}). \qquad (2.4)$$

Note, this is (2.3) with a finite population correction (fpc) for sampling respondents without replacement.

2.1.3. Respondent Population Sampled with Replacement and Question Population Sampled without Replacement (Case III)

Now X is binomially distributed and Y is hypergeometrically distributed.

Note, y_i is not independent of y_j, $i\neq j$, and thus the technique of section 2.2.3 is again applied with Mp being the number of sensitive questions in the randomization device. Thus,

$$E(y_i y_j) = \frac{\binom{Mp}{2}}{\binom{M}{2}} = \frac{p(Mp-1)}{M-1}.$$

Using equation (2.2), the variance for this case is

$$V_3(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{4\pi(1-\pi)p(1-p)}{n(2p-1)^2}\left(1 - \frac{M-n}{M-1}\right) + \frac{p(1-p)}{n(2p-1)^2}\left(\frac{M-n}{M-1}\right) \quad (p\neq\tfrac{1}{2}). \qquad (2.5)$$

Note that this variance has the fpc due to the sampling of questions without replacement $\left(\frac{M-n}{M-1}\right)$. If the questions were sampled with replacement, the fpc is one, and this variance reduces to (2.3).

2.1.4. Both Respondent and Question Populations Sampled without Replacement (Case IV)

Both X and Y have hypergeometric distributions.

Now, both x_i and x_j, $\forall$ $i\neq j$, are dependent and similar for y_i and y_j. Using equation (2.2) and the past results for these two cases (section 2.1.3 and 2.1.4),

$$E(x_i x_j) = \frac{\pi(N\pi-1)}{N-1}, \quad E(y_i y_j) = \frac{p(Mp-1)}{M-1}.$$

and thus the variance of $\hat{\pi}$ is

$$V_4(\hat{\pi}) = \frac{\pi(1-\pi)}{n(2p-1)^2}\left[(4p^2-4p)\left\{\frac{N(M-n)-n(M-1)}{(N-1)(M-1)}\right\} + \frac{N-n}{N-1}\right] + \frac{p(1-p)}{n(2p-1)^2}\left(\frac{M-n}{M-1}\right) \quad (p\neq\tfrac{1}{2}). \qquad (2.6)$$

This variance contains fpc's due to sampling both respondents and questions without replacement. It should be noted that as N increases, equation (2.6) approaches equation (2.5) and similarly as M increases, equation (2.6) approaches equation (2.4).

III. COMPARISONS OF THE FOUR CASES

As we have seen, the point estimators of π are the same for all four cases. However, the variance of $\hat{\pi}$ differs from case to case. First, we present some theoretical comparisons of the $V(\hat{\pi})$ among the four cases, and then for selected values of N, M, n, p and π, provide some numerical comparisons to indicate the magnitude of the differences.

Theorem 1. $V_1(\hat{\pi}) > V_i(\hat{\pi})$, $i=2,3,4$, if $p\neq\frac{1}{2}$.

Proof.

It follows from comparing equations (2.3), (2.4), (2.5) and (2.6) and the fact that $\frac{N-n}{N-1}$, $\frac{M-n}{M-1}$ and $\frac{N(M-n)-n(M-1)}{(N-1)(M-1)}$ are all <1.

Theorem 2. Assuming $p\neq\frac{1}{2}$, $V_2(\hat{\pi}) \geq V_4(\hat{\pi})$, provided

$$4\pi(1-\pi) \leq \frac{N-1}{N}.$$

Proof.

Using equations (2.4) and (2.6), the difference between the two variances is

$$V_2(\hat{\pi})-V_4(\hat{\pi}) = \frac{(n-1)p(1-p)}{n(M-1)(2p-1)^2}\left\{-\frac{4\pi(1-\pi)N}{N-1}+1\right\} \quad (p\neq\tfrac{1}{2}).$$

Hence for $p\neq\frac{1}{2}$, $V_2(\hat{\pi}) \geq V_4(\hat{\pi})$ if and only if $4\pi(1-\pi) \leq \frac{N-1}{N}$.

It is interesting to note that when N is large and $\pi\neq.5$, the above condition almost always holds. Hence $V_4(\hat{\pi})$ is rarely greater than $V_2(\hat{\pi})$ for large N.

Theorem 3. Assuming $p\neq\frac{1}{2}$, $V_3(\hat{\pi}) \geq V_4(\hat{\pi})$, provided

$$4p(1-p) \leq \frac{M-1}{M}.$$

Proof.

Using equations (2.5) and (2.6), the difference between the two variances is

$$V_3(\hat{\pi})-V_4(\hat{\pi}) = \frac{\pi(1-\pi)}{n(2p-1)^2}\,\frac{n-1}{N-1}\left\{1 - \frac{4Mp(1-p)}{M-1}\right\} \quad (p\neq\tfrac{1}{2}).$$

By inspection, one can see for $p\neq\frac{1}{2}$, if $1-\frac{4Mp(1-p)}{M-1} \geq 0$, then $V_3(\hat{\pi}) \geq V_4(\hat{\pi})$.

In Table 1 the variances of $\hat{\pi}$ for small values of π (i.e., .1 and .2) are presented. Note that the variances in Cases III and IV are often much smaller (e.g. 63% decrease for certain cases) than Cases I and II. The two former cases are where one samples questions without replacement compared to the two latter cases where sampling of questions is with replacement. Cases I and II are virtually identical (up to at least the third decimal place) and similarly for Cases III and IV. This is because the respondent fpc is close to one (99,900/100,000), and hence in this case the sampling of respondents without replacement (Case II and IV) provides little improvement over sampling with replacement.

As can be seen, as π increases toward $\frac{1}{2}$ the reduction in variance becomes less. Also, as M increases for n held constant, the reduction in variance becomes less because of the increase in the question fpc.

Table 1

Variances for Comparisons of the Four Cases-Warner and Its Modified Cases (n=100)

N	M	CASE 1	CASE 2	CASE 3	CASE 4
		π=.1, p=.7			
100000	100	.01403	.01402	.00563	.00562
100000	150	.01403	.01402	.00844	.00844
100000	200	.01403	.01422	.00985	.00985
		π=.2, p=.7			
100000	100	.01473	.01472	.01000	.00000
100000	150	.01473	.01472	.01159	.01158
100000	200	.01473	.01472	.01237	.01237
		π=.1, p=.6			
100000	100	.06090	.06090	.02250	.02250
100000	150	.06090	.06090	.03539	.03539
100000	200	.06090	.06090	.04180	.04180
		π=.4, p=.8			
100000	100	.00684	.00682	.00667	.00664
100000	150	.00684	.00682	.00673	.00670
100000	200	.00684	.00682	.00676	.00673

IV. MODIFICATIONS OF THE UNRELATED QUESTION MODEL

Greenberg, et al.'s (1969) unrelated question model is one of the important RRT models, and the previous sampling with and without replacement modifications also are readily applied.

In this model the respondent in subsample i, i=1,2, randomly selects one of the following questions:

Q1. I am a member of G_1,

Q2. I am a member of G_2,

and responds either "yes" or "no" (we will assume truthful reporting). Note that characteristic 2 is non-sensitive and unrelated to characteristic 1.

Define

$$x_{ij} = \begin{cases} 1 & \text{if the } j^{th} \text{ respondent in subsample } i, \ i=1,2, \text{ belongs to } G_1 \\ 0 & \text{otherwise} \end{cases}$$

$$v_{ij} = \begin{cases} 1 & \text{if the } j^{th} \text{ respondent in subsample } i, \ i=1,2, \text{ belongs to } G_2 \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} 1 & \text{if the } j^{th} \text{ respondent in subsample } i, \ i=1,2, \text{ selects Q1}, \\ 0 & \text{otherwise} \end{cases}$$

$$z_{ij} = \begin{cases} 1 & \text{if the } j^{th} \text{ respondent in subsample } i, \ i=1,2, \text{ answers "yes"} \\ 0 & \text{otherwise} \end{cases}$$

Then

$$z_{ij} = x_{ij}y_{ij} + v_{ij}(1-y_{ij}), \quad j=1,2,\ldots,n_i; \quad i=1,2.$$

We also define $Z_i = \sum_{j=1}^{n_i} z_{ij}$, $\pi = Pr(x_{ij} \epsilon G_1)$, and $\pi_2 = Pr(x_{ij} \epsilon G_2)$.

4.1. The Estimator of π

Greenberg, et al. presented the point estimator of π for Case I as

$$\hat{\pi} = \frac{(1-p_2)Z_1/n_1+(1-p_1)Z_2/n_2}{p_1-p_2} \quad (p_1 \neq p_2), \qquad (4.1)$$

and he termed this the MLE of π. It also is the method of moments estimator of π for this and the other three cases.

i). The $E(\hat{\pi})$

Defining λ_i = Pr(yes in subsample i), i=1, 2, then $E(z_i) = n_i\lambda_i$, i=1,2 and we can easily show that $\hat{\pi}$ is an unbiased estimator of π for all four cases (Greenberg, et al. gives the proof for Case I).

ii). The $V(\hat{\pi})$

Since the two subsamples are independent, z_{1j} and z_{2j} are independent. Hence, the variance of $\hat{\pi}$ is simply the sum of the individual variances,

$$V(\hat{\pi}) = \frac{(1-p_2)^2V(z_1/n_1)+(1-p_1)^2V(Z_2/n_2)}{(p_1-p_2)^2} \quad (p_1 \neq p_2).$$

4.2. Both Respondent and Question Populations are Sampled with Replacement (Case I)

$$V_1(\hat{\pi}) = \frac{1}{(p_1-p_2)^2}\{(1-p_2)^2\frac{\lambda_1(1-\lambda_1)}{n_1}+(1-p_1)^2\frac{\lambda_2(1-\lambda_2)}{n_2}\}$$
$$(p_1 \neq p_2) \qquad (4.2)$$

which was given by Greenberg, et al. (1969).

4.3. Respondent Population Sampled without Replacement and Question Population Sampled with Replacement (Case II)

For this case, after some algebra we get,

$$V_2(\hat{\pi}) = V_1(\hat{\pi}) - \frac{1}{(p_1-p_2)^2}[(1-p_2)^2\{(1 - \frac{1}{n_1})\ p_1^2\pi(1-\pi)$$
$$+ (1-p_1)^2\pi_2(1-\pi_2)\} + (1-p_1)^2(1 - \frac{1}{n_2})$$
$$\cdot \{p_2^2\pi(1-\pi) + (1-p_2)^2\pi_2(1-\pi_2)\}]/(N-1)$$
$$(p_1 \neq p_2) \qquad (4.3)$$

It should be noted that $V_2(\hat{\pi})$ has an additional term in comparison with $V_1(\hat{\pi})$. This additional term is always negative and contains terms due to the variability of π and π_2.

4.4. Respondent Population Sampled with Replacement and Question Population Sampled without Replacement (Case III)

$$V_3(\hat{\pi}) = V_1(\hat{\pi}) - \frac{1}{(p_1-p_2)^2}\Big[(1-p_2)^2$$

$$\cdot \left(1 - \frac{1}{n_1}\right)\frac{p_1(1-p_1)\ \pi^2}{M_1-1} + (1-p_1)^2$$

$$\left.\cdot \left(1 - \frac{1}{n_2}\right)\frac{p_2(1-p_2)\ \pi^2}{M_2-1}\right] \quad (p_1 \neq p_2). \tag{4.4}$$

$V_3(\hat{\pi})$ also has an additional term compared with $V_1(\hat{\pi})$. Again the additional term is negative and has an expression due to the variability of $\hat{p}$.

4.5. Respondent and Question Populations Sampled without Replacement (Case IV)

This is the most complicated case for computing the variance. However, after some algebra, the following result was obtained:

$$\begin{aligned} V_4(\hat{\pi}) = V_1(\hat{\pi}) - \frac{1}{(p_1-p_2)^2} \Bigg[& (1-p_2)^2\left(1-\frac{1}{n_1}\right) \\ & \cdot \left\{ \frac{N\pi^2 p_1(1-p_1)+M_1 p_1^2 \pi(1-\pi)-\pi p_1(1-\pi p_1)}{(N-1)(M_1-1)} \right. \\ & - \frac{2\pi\pi_2 p_1(1-p_1)}{M_1-1} + \frac{(1-2p_1)\pi_2(1-\pi_2)}{N-1} \\ & \left. + \frac{N\pi_2^2 p_1(1-p_1)+M_1 p_1^2 \pi_2(1-\pi_2)-\pi_2 p_1(1-\pi_2 p_1)}{(N-1)(M_1-1)} \right\} \\ & + (1-p_1)^2\left(1-\frac{1}{n_2}\right)\left\{ \frac{N\pi^2 p_2(1-p_2)+M_2 p_2^2 \pi(1-\pi)-\pi p_2(1-\pi p_2)}{(N-1)(M_2-1)} \right. \\ & - \frac{2\pi\pi_2 p_2(1-p_2)}{M_2-1} + \frac{(1-2p_2)\pi_2(1-\pi_2)}{N-1} \\ & \left. + \frac{N\pi_2^2 p_2(1-p_2)+M_2 p_2^2 \pi_2(1-\pi_2)-\pi_2 p_2(1-\pi_2 p_2)}{(N-1)(M_2-1)} \right\} \Bigg] \end{aligned}$$

$$(p_1 \neq p_2) \tag{4.5}$$

Note, as N increases, $V_4(\hat{\pi})$ approaches $V_3(\hat{\pi})$ and also as M_i, $i=1,2$, increases $V_4(\hat{\pi})$ approaches $V_2(\hat{\pi})$. It is interesting to note that the second term in the above equation has terms due to the variability of $\hat{\pi}$, $\hat{\pi}_2$, $\hat{p}_i$, $\hat{\pi}\hat{p}_i$, and $\hat{\pi}_2\hat{p}_i$, $i=1,2$.

V. COMPARISONS OF $V(\hat{\pi})$ FOR THE FOUR CASES

Theorem 4. $V_1(\hat{\pi}) > V_i(\hat{\pi})$, $i=2,3$, if $p_1 \neq p_2$.

Proof.

It follows from comparing equations 4.2, 4.3 and 4.4 and the fact that the subtrahends in equations 4.3 and 4.4 are negative.

Numerical comparisons were performed under the assumptions that i) the two subsamples are of equal size and ii) the two question populations have the same size (Table 2). Generally, Case IV has the smallest variance among the four, as expected by considering the fpc's. If π_2 is small, the gain in relative efficiency from using the finite population(s) is nominal. However, as π_2 increases, the gain in relative efficiency increases. In particular, when $\pi=.1$, $\pi_2=1.00$, $p_1=.7$, $p_2=.3$, $N=100{,}000$, $M_1=M_2=50$, and $n_1=n_2=50$, the gains of 67% and 67% can be realized for Cases III and IV, respectively.

Table 2

Variances for Comparisons of the Four Cases-Unrelated Question Model and Its Modified Cases ($n_1 = n_2 = 50$)

N	M	CASE 1	CASE 2	CASE 3	CASE 4
$\pi=.1$, $\pi_2=.1$, $p_1=.7$ and $p_2=.3$					
100000	50	.00725	.00725	.00725	.00725
100000	90	.00725	.00725	.00725	.00725
$\pi=.1$, $\pi_2=.5$, $p_1=.7$ and $p_2=.3$					
100000	50	.01775	.01775	.01531	.01531
100000	90	.01775	.01775	.01641	.01641
$\pi=.1$, $\pi_2=1$, $p_1=.7$ and $p_2=.3$					
100000	50	.03087	.03087	.01854	.01854
100000	90	.03087	.03087	.02409	.02408
$\pi=.3$, $\pi_2=.1$, $p_1=.7$ and $p_2=.3$					
100000	50	.01650	.01650	.01589	.01589
100000	90	.01650	.01650	.01616	.01616
$\pi=.3$, $\pi_2=.5$, $p_1=.7$ and $p_2=.3$					
100000	50	.02700	.02700	.02639	.02639
100000	90	.02700	.02700	.02666	.02666
$\pi=.3$, $\pi_2=1$, $p_1=.7$ and $p_2=.3$					
100000	50	.04012	.04012	.03266	.03266
100000	90	.04012	.04012	.03602	.03601

VI. FIELD TRIAL OF A NEW RANDOMIZATION DEVICE

A number of question randomization devices, utilizing sampling without replacement, can be easily designed. One simple approach is to construct a sheet(s) of randomly allocated face down sealed numbers {1,2} of known proportions. The respondent selects one of the numerous sealed numbers, pulls it free of the sheet and looks at the underside. If the number is 1, he answers Q1 and if 2, Q2. He keeps the number.

An "urn version" of this approach uses a container with both red and white balls (red denoting the sensitive question) and a mouth such that whenever the device is placed upside down only one ball comes out of the mouth. The respondent notes its color, answers the appropriate question and keeps the ball. Numerous other sampling-without-replacement devices also are possible.

For a number of reasons, the respondent's perceived protection may be a function of the number of remaining balls or sealed numbers in the sampling-without-replacement randomization device. In this situation, the size of the question population (M) can be larger than the number of respondents (n). However, as M increases the variance increases, and thus a trade-off is needed between increasing M to presumably obtain better respondent cooperation and the resulting increasing variance.

To evaluate the real-world performance of the sampling-without-replacement randomization device (Section II), the original Warner model (1965) and a modification (to allow sampling without replacement) were field tested in the fall of 1977 using a mixture of Philadelphia subway passengers, West Park Hospital employees, and some Temple Law students. The questions were:

Q1. I have received at least one welfare check or food stamps since 1974.

Q2. I have received neither a welfare check nor food stamps since 1974.

Each respondent used both randomization devices with half of the respondents first using the sampling-with-replacement (W.R.) randomization device and the other half first using the sampling-without-replacement (W.O.R.) randomization device.

Table 3 presents the number of "true" and "false" replies for the 54 respondents.

Table 3

The Sample Response for the Warner and Modified Warner Models

	W.R. (Warner)	W.O.R. (Mod. Warner)
True	25	19
False	29	35
Total	54	54

The proportion of people who have received at least one welfare check or food stamps (π) is estimated as follows for the two cases:

$$\text{W.R.,} \qquad \hat{\pi} = \frac{\frac{25}{54} - (1 - \frac{2}{3})}{(2 \cdot \frac{2}{3} - 1)} = \frac{7}{17} = .39$$

$$\text{W.O.R.,} \qquad = \frac{\frac{19}{54} - (1 - \frac{2}{3})}{(2 \cdot \frac{2}{3} - 1)} = \frac{1}{18} = .06$$

It is interesting to note that according to the last census (1970), 8.67% of all the Philadelphia families received some sort of public assistance. The corresponding estimated variances are

$$\text{W.R.,} \qquad v(\hat{\pi}) = .04$$

$$\text{W.O.R.,} \qquad v(\hat{\pi}) = .0094.$$

The data provided by the respondents on the "convenience" and the "perceived protection" of the two randomization methods are summarized (in percentages) in Table 4.

The results show that 69% of the respondents considered the W.O.R. sampling of questions to be an easier way of selecting a question. The overwhelming reason given was that the W.O.R. method directly provided the number of the question they were to answer whereas the W.R. method did not (only the combination of rolling a die and following the directions for selection of question secures the question). The respondents are almost evenly divided in their opinion on the perceived protection provided by the two methods.

Table 4

The Sample Responses on Ease and Perceived Protection for the Warner and Modified Models

	W.R. (Warner)	W.O.R. (Mod. Warner)	No Difference
Easier	20	69	11
Better Protection	31	33	36

VII. CONCLUSIONS

In this paper, modifications of the original Warner model (1965) for sampling without replacement were presented. It was shown that by sampling the respondent and/or question population without replacement, we can reduce the variance of the estimator of the population proportion. In general, Warner's estimator has the largest variance and the estimator obtained from sampling both the respondent and question populations without replacement has the least variance. Numerical comparisons show that sampling questions without replacement can result in large reductions in the estimated variance of π (e.g. 63%).

In addition, since the respondent population typically is finite and generally is sampled without replacement, we recommend the appropriate variance be calculated. In large scale surveys, some modification is also recommended in order to use a sampling-questions-without-replacement randomization device, as can be done without much difficulty.

Greenberg, et al.'s unrelated question model was modified to accomodate the sampling-without-replacement idea. Again, sampling questions without replacement can provide substantial reduction in the variance of π (e.g. 67%).

Finally, a small field trial has suggested that respondents generally prefer the sampling W.O.R. procedure. For more details, see Kim (1978).

REFERENCES

[1] Boruch, R.F., "Relating among Statistical Methods for Assuring Confidentiality of Social Research Data," Social Science Research, 1 (1972), 403-414.

[2] Greenberg, B.G., Aul-Ela, A.A., Simmons, W.R., and Horvitz, D.G., "The Unrelated Question Randomized Response Model: Theoretical Framework," Journal of the American Statistical Association, 64 (1969), 520-539.

[3] Kim, J., "Randomized Response Techniques for Surveying Human Populations," Ph.D. Dissertation, Department of Statistics, Temple University, Philadelphia, Pa. (1978), 1-144.

[4] Warner, S.L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (1965), 63-69.