A RANDOMIZED RESPONSE TECHNIQUE FOR INVESTIGATING SEVERAL SENSITIVE ATTRIBUTES

Ajit C. Tamhane, Northwestern University

1. INTRODUCTION AND SUMMARY

Most of the literature on randomized response (RR) techniques has been concerned with the study of a single sensitive attribute. However, very often, social researchers are interested in studying several sensitive attributes together. Therefore it is necessary to develop privacy preserving techniques which would allow statistical inference to be made concerning marginal as well as joint distributions of the attributes. Only recently attention of the survey statisticians has been focussed on this particular problem.

In his dissertation, Barksdale (1971) proposed and analyzed some RR techniques for investigating two sensitive dichotomous attributes. In particular, he considered a repeated (for each attribute) application of the Warner's original technique (see also Clickner and Iglewicz 1976 and Drane 1976), a repeated application of the Simmons' unrelated question technique (Greenberg et al. 1969) and a third technique which is as follows: The two statements concerning the two sensitive attributes are phrased so that a "Yes" response to one of the two statements would be nonstigmatizing. (E.g., the two statements might be "I have never smoked marijuana" and "I am an alcoholic.") The interviewer presents both the statements to the respondent on two occasions. On each occasion, the respondent picks one of the two statements at random, unknown to the interviewer, but according to some known probability (different for each occasion) and responds to it. This procedure maintains the privacy of the respondent and yet allows the researcher to compute the estimates of the marginal and bivariate probabilities of the attributes from the observed frequencies of "Yes-Yes," "Yes-No," "No-Yes," and "No-No" responses.

In a survey dealing with $t \geqq 2$ sensitive attributes, a repeated application of any RR technique for a single attribute, such as the Warner's technique, involves t trials per respondent. If t is large then this procedure becomes tedious, costly and leads to degradation in cooperation on the part of respondents. Also the estimating equations involve all the joint probabilities which the researcher is not often interested in. On the other hand, the technique described in the previous paragraph can be easily extended to $t > 2$ case with the number of trials per respondent restricted to $r \leqq t$ if the researcher's interest lies in only up to r-variate joint probabilities. Quite often, $r = 2$ will suffice for the purposes of the research.

In Section 2 of the present paper we extend the above technique (henceforth referred to as the multiple RR trials technique or the M-technique) to the case of $t > 2$ sensitive dichotomous attributes. But we restrict to only $r = 2$ trials per respondent to keep the algebra simple and also since $r = 2$ appears to be the most useful case from a practical viewpoint. The estimates derived by Barksdale do not satisfy the natural restrictions on the marginal and bivariate probabilities; also no procedure for testing independence between the attributes is provided in his work. We provide a correct statistical analysis of the extended technique and also give a test of pairwise independence for any set of pairs of attributes.

In Section 3 we carry out a numerical comparison of the multiple RR trials technique with some competing techniques in terms of the trace of the variance-covariance matrix of the estimator vector for the marginal and joint probabilities of the attributes. To make this a just comparison, it is necessary to keep fixed some measure of privacy afforded to the respondent. In Section 3.2 such a measure is defined which extends to $t > 1$, the corresponding notion for $t = 1$ due to Leysieffer and Warner (1976). No clear winner is indicated by the numerical comparisons which are made for the $t = 2$ case. But if the proportions in the population possessing the sensitive attributes are small (which is often the case) and the respondent jeopardy levels are moderate, i.e., not too high or low (which is also often the case) then the multiple RR trials technique appears to dominate. This technique has one drawback however, which is that it fails to attain certain low levels of respondent jeopardy. Still, in view of the practical advantages pointed out earlier, the multiple RR trials technique definitely merits a consideration in any survey dealing with several sensitive attributes.

2. MULTIPLE RR TRIALS TECHNIQUE

2.1 Notation and Description of Technique: Consider $t \geqq 2$ dichotomous attributes $A_1, A_2, \ldots, A_t$; we shall assume that all the attributes are sensitive but obviously that need not be so. Let $\theta_{i_1 \ldots i_u}$ denote the unknown proportion of individuals in the target population which possesses the attributes $A_{i_1 \ldots i_u}$ $(1 \leqq i_1 < \cdots < i_u \leqq t$, $1 \leqq u \leqq t)$. The researcher's interest lies in making statistical inference (estimation and hypothesis testing) concerning the θ's.

For employing the multiple RR trials technique, the statements must be phrased so that a "Yes" response to some statements would be nonstigmatizing whereas a "No" response to the others would be so. Without loss of generality, we shall assume that the first $s < t$ statements are phrased "I possess the attribute A_i" $(1 \leqq i \leqq s)$, a "No" response to each one of which would be nonstigmatizing; the remaining $t - s$ statements are phrased "I do not possess the attribute A_i" $(s + 1 \leqq i \leqq t)$, a "Yes" response to each one of which would be so. An appropriate choice of s would be $\cong t/2$. Let $\pi_{i_1 \ldots i_u}$ be defined in the same manner as $\theta_{i_1 \ldots i_u}$ but with respect to the modified attributes B_i which are either original A_i $(1 \leqq i \leqq s)$ or the complements of the A_i $(s + 1 \leqq i \leqq t)$. It is clear that the θ's can be obtained from the π's and vice versa and therefore we shall consider the equivalent problem of estimation of the π's.

As remarked in the previous section we shall assume that the researcher is interested only in the marginal and bivariate probabilities, i.e., $\pi_i\,(1 \leq i \leq t)$ and $\pi_{ij}\,(1 \leq i < j \leq t)$, respectively. Thus there are $t(t+1)/2$ unknown parameters to be estimated and only 2 trials may be performed per respondent. We now describe the technique.

A total sample of n individuals (which may be assumed to be a simple random sample drawn with replacement) is divided into $b \geq 1$ subsamples; the value of b will be specified in the following section. Let $n_1, n_2, \ldots, n_b$ be the subsample sizes with $\sum_{h=1}^{b} n_h = n$.

Each individual is presented all the t statements and asked to respond to one statement picked at random according to some randomizing device, but not reveal his choice of the statement to the interviewer. This procedure is repeated with another randomizing device and both the responses are recorded. Let $P_{hi}^{(\ell)}$ denote the (known) probability that an individual drawn from the hth subsample picks, on the ℓth trial, the ith statement $(1 \leq i \leq t)$; obviously we have $\sum_{i=1}^{t} P_{hi}^{(\ell)} = 1$ for $1 \leq h \leq b$ and $\ell = 1, 2$.

2.2 Estimation of the π's: Suppose that the responses are coded so that a score of $2^{\ell-1}$ is assigned to a "Yes" response on the ℓth trial and a score of 0 is assigned to a "No" response. Then the total score, say ν, completely identifies the individual's response. E.g., $\nu = 3$ corresponds to a "Yes-Yes" response, $\nu = 2$ corresponds to a "No-Yes) response etc. Let $\lambda_{h\nu}$ denote the probability of obtaining a score of ν for an individual drawn from the hth subsample. Then we have the following equations.

$$
\begin{aligned}
\lambda_{h1} &= \sum_{i=1}^{t} P_{hi}^{(1)}(1 - P_{hi}^{(2)})\pi_i - \sum_{i=1}^{t}\sum_{j=i+1}^{t}(P_{hi}^{(1)}P_{hj}^{(2)} + P_{hj}^{(1)}P_{hi}^{(2)})\pi_{ij} \\
\lambda_{h2} &= \sum_{i=1}^{t} P_{hi}^{(2)}(1 - P_{hi}^{(1)})\pi_i - \sum_{i=1}^{t}\sum_{j=i+1}^{t}(P_{hi}^{(1)}P_{hj}^{(2)} + P_{hj}^{(1)}P_{hi}^{(2)})\pi_{ij} \\
\lambda_{h3} &= \sum_{i=1}^{t} P_{hi}^{(1)}P_{hi}^{(2)}\pi_i + \sum_{i=1}^{t}\sum_{j=i+1}^{t}(P_{hi}^{(1)}P_{hj}^{(2)} + P_{hj}^{(1)}P_{hi}^{(2)})\pi_{ij} \\
\lambda_{h0} &= 1 - \lambda_{h1} - \lambda_{h2} - \lambda_{h3},
\end{aligned} \tag{2.1}
$$

for $1 \leq h \leq b$. In the vector notation, if $\underset{\sim}{\lambda} = (\lambda_{11}, \lambda_{12}, \lambda_{13}, \ldots, \lambda_{b1}, \lambda_{b2}, \lambda_{b3})'$ and $\underset{\sim}{\pi} = (\pi_1, \ldots, \pi_t, \pi_{12}, \pi_{13}, \ldots, \pi_{t-1,t})'$ then (2.1) can be expressed compactly as

$$\underset{\sim}{\lambda} = \underset{\sim}{R}\,\underset{\sim}{\pi}, \tag{2.2}$$

where the elements of the matrix $\underset{\sim}{R}$ are given by the following equations: For $1 \leq h \leq b$ and $1 \leq i \leq t$ we have,

$$
\begin{aligned}
\underset{\sim}{R}_{3h-2,i} &= P_{hi}^{(1)}(1 - P_{hi}^{(2)}), \\
\underset{\sim}{R}_{3h-1,i} &= P_{hi}^{(2)}(1 - P_{hi}^{(1)}), \quad \underset{\sim}{R}_{3h,i} = P_{hi}^{(1)}P_{hi}^{(2)},
\end{aligned} \tag{2.3}
$$

and for $1 \leq i < j \leq t$ if $k = it - i(i+1)/2 + j$ then we have

$$
\begin{aligned}
\underset{\sim}{R}_{3h-2,k} &= -(P_{hi}^{(1)}P_{hj}^{(2)} + P_{hj}^{(1)}P_{hi}^{(2)}) \\
&= \underset{\sim}{R}_{3h-1,k} = -\underset{\sim}{R}_{3h,k}.
\end{aligned} \tag{2.4}
$$

To find b, the total number of subsamples, necessary to estimate the t marginal probabilities $\{\pi_i\}$ and $\binom{t}{2}$ bivariate probabilities $\{\pi_{ij}\}$, consider an extreme case (and a most favorable one from the statistician's viewpoint) where the P-values can be chosen either equal to zero or one (which corresponds to the "direct response" case). By choosing $P_{hi}^{(1)} = 1$ and $P_{hj}^{(2)} = 1$ for different pairs (i, j) for different subsamples h, it is easy to see that all the parameters can be estimated using $\binom{t}{2}$ subsamples and no less number of subsamples would do. An extension of this argument shows that, even for general P-values, to estimate all the parameters, at least $\binom{t}{2}$ subsamples are required. In other words, by suitably choosing the P's, the matrix $\underset{\sim}{R}$ defined in (2.3) and (2.4) can be made to have a full column rank only if $b \geq \binom{t}{2}$. Let us then assume that $b \geq \binom{t}{2}$ and that $\underset{\sim}{R}$ is a full column rank matrix.

We propose to obtain the maximum likelihood estimator (MLE) of $\underset{\sim}{\pi}$ from the observed data $\{n_{h\nu}\}$ where $n_{h\nu}$ = the number of individuals from the hth subsample having a score of $\nu(0 \leq \nu \leq 3)$; $\sum_{\nu=0}^{3} n_{h\nu} = n_h\,(1 \leq h \leq b)$. The usual method of first obtaining the unrestricted MLE (UMLE) of $\underset{\sim}{\lambda}$ (i.e., the UMLE of $\lambda_{h\nu} = n_{h\nu}/n_h$ for $0 \leq \nu \leq 3$, $1 \leq h \leq b$) and then obtaining the UMLE of $\underset{\sim}{\pi}$ by "solving" (2.2) is not applicable for two reasons in the present context:

1. Matrix $\underset{\sim}{R}$ can be chosen to be a square full rank matrix only for $t = 2$. For $t > 2$, in general, there is no unique solution in $\underset{\sim}{\pi}$ to (2.2).
2. Even in the case where the UMLE of $\underset{\sim}{\pi}$ can be obtained by the above method, the resulting estimator may not satisfy the natural restrictions on the π's namely that

$$
\begin{aligned}
&0 \leq \pi_i \leq 1 \;\; \forall i \;\text{ and,} \\
&\max(0, \pi_i + \pi_j - 1) \leq \pi_{ij} \leq \min(\pi_i, \pi_j) \;\; \forall (i, j).
\end{aligned} \tag{2.5}
$$

From a theoretical viewpoint, the UMLE of $\underset{\sim}{\pi}$ may even be inadmissible as shown in the case of the Warner's technique for a single attribute by Singh (1976).

Therefore we must find the restricted MLE (RMLE) of $\underset{\sim}{\pi}$, say $\hat{\underset{\sim}{\pi}}$. We propose to obtain $\hat{\underset{\sim}{\pi}}$ directly by maximizing the likelihood function

$$L \;\alpha\; \prod_{h=1}^{b}\prod_{\nu=0}^{3}(\lambda_{h\nu})^{n_{h\nu}} \tag{2.6}$$

subject to (2.5). In (2.6) the $\lambda_{h\nu}$ are given in terms of π by (2.1). Denote the restricted maximum of L by L^*. The constraint set (2.5) is linear in the π's and the objective function $\log_e L$ can be easily checked to be concave in the π's. The resulting nonlinear programming (NLP) problem is thus well structured and can be solved quite economically on a computer using one of the commonly available algorithms.

2.3 Properties of $\hat{\underset{\sim}{\pi}}$: The RMLE $\hat{\underset{\sim}{\pi}}$ is biased in small samples but is asymptotically (as $n_h \to \infty \ \forall h$) unbiased. The asymptotic variance-covariance matrix of $\hat{\underset{\sim}{\pi}}$ (which is also the exact variance-covariance matrix of the UMLE of $\underset{\sim}{\pi}$) is given by the inverse of the information matrix $\underset{\sim}{\mathcal{J}}$; we give below an expression for the elements of the upper left $t \times t$ principal submatrix of $\mathcal{J}$: For $1 \leq i, j \leq t$ we have

$$\underset{\sim}{\mathcal{J}}_{ij} = -E\{\frac{\partial^2 \log L}{\partial\pi_i \partial\pi_j}\} = \sum_{h=1}^{b} n_h \sum_{\nu=0}^{3} \frac{1}{\lambda_{h\nu}} (\frac{\partial\lambda_{h\nu}}{\partial\pi_i})(\frac{\partial\lambda_{h\nu}}{\partial\pi_j}).$$

The remaining elements of $\underset{\sim}{\mathcal{J}}$, which would involve $\partial\lambda_{h\nu}/\partial\pi_{ij}$ terms, can be obtained in an analogous manner. The various derivatives can be evaluated easily using (2.1).

For $t = 2$, the expressions for the asymptotic variances and covariances can be written down explicitly and they may be found in Barksdale (1971). Large sample hypothesis testing concerning the π's can be carried out using the expressions for the variances and covariances with $\underset{\sim}{\lambda}$ replaced by its RMLE $\hat{\underset{\sim}{\lambda}} = \underset{\sim}{R}\,\hat{\underset{\sim}{\pi}}$.

2.4 Test of Independence: First we note that testing pairwise independence between the original attributes, say A_i and A_j, is equivalent to testing pairwise independence between the corresponding modified attributes. In fact, if ρ_{ij} denotes the correlation between A_i and A_j and η_{ij} denotes the correlation between the corresponding modified attributes then $|\rho_{ij}| = |\eta_{ij}|$ for $1 \leq i < j \leq t$. Therefore we shall consider the problem of testing independence between pairs of modified attributes.

Suppose that it is desired to test the hypothesis $H_{\mathcal{J}}: \pi_{ij} = \pi_i \pi_j$ for all pairs (i, j) in a certain set $\mathcal{J}$. We can use the generalized likelihood ratio method to test this hypothesis as follows: Compute the maximum of the likelihood function L in (2.6) subject to the following constraints on the π's

$$0 \leq \pi_i \leq 1 \ \forall i,$$

$$\max(0, \pi_i + \pi_j - 1) \leq \pi_{ij} \leq \min(\pi_i, \pi_j) \ \forall (i, j) \notin \mathcal{J} \tag{2.7}$$

$$\pi_{ij} = \pi_i \pi_j \qquad \forall (i, j) \in \mathcal{J}.$$

Denote the corresponding maximum of L by $L^*_{\mathcal{J}}$. Then under $H_{\mathcal{J}}$ asymptotically $-2 \log_e (L^*_{\mathcal{J}}/L^*)$ has a chi-square distribution with f degrees of freedom (d.f.), where f is the number of pairs in set $\mathcal{J}$.

2.5 Choice of the P's: For fixed h and ℓ, the $P^{(\ell)}_{hi}$ should be chosen as different from 1/t as possible. In fact, for large t, the length of the questionnaires can be cut down by choosing $P^{(\ell)}_{hi} = 0$ for different sets of statements for different subsamples. Assuming that the researcher is equally interested in all the attributes, it seems that, the P's should be chosen symmetrically as far as possible. For $t = 2$, such a symmetric choice is provided by $P^{(1)}_{11} + P^{(2)}_{11} = 1$; subject to this restriction, $P^{(1)}_{11}$ and $P^{(2)}_{11}$ may be chosen as far away from 1/2 as the researcher dares. Obviously the actual choice will depend on the average educational and social sophistication of the population. A pilot survey should be carried out to test different randomizing devices (different P's) as well as the questionnaire itself.

3. COMPARISON WITH SOME COMPETING TECHNIQUES

3.1 Brief Description of the Competing Techniques: We shall consider two techniques in competition with the M-technique developed above: a repeated application of the Warner's technique (W-technique) and a repeated application of the Simmons' unrelated question technique (S-technique).

In the W-technique t trials are performed per respondent. On the ith trial the interviewer presents the respondent with a pair of statements: "I possess the attribute A_i" and "I do not possess the attribute A_i." The respondent picks one of the two statements at random according to known probabilities P_i and $1 - P_i$ ($P_i \neq 1/2$) respectively, and without revealing his choice to the interviewer, responds to it. This procedure is repeated for $i = 1, 2, \ldots, t$. Suppose that the responses are coded so that a score of 2^{i-1} is assigned to a "Yes" response on the ith trial and a score of 0 is assigned to a "No" response ($1 \leq i \leq t$) and let ν denote the total score. Then ν ($0 \leq \nu \leq 2^t - 1$) completely identifies the individual's response. The π's can then be estimated from the observed frequencies $\{n_\nu\}$ where n_ν = the number of individuals in the sample having a score of ν; $\sum_{\nu=0}^{2^t-1} n_\nu = n$.

In the S-technique also t trials are performed per respondent. On the ith trial the interviewer presents the respondent with a pair of statements "I possess the attribute A_i" and "I possess the attribute Y_i" where Y_i is some unrelated and innocuous attribute. The respondent picks one of the two statements at random according to known probabilities P_i and $1 - P_i$ respectively, and without revealing his choice to the interviewer, responds to it. This procedure is repeated for $i = 1, 2, \ldots, t$. Again using the same scoring system as in the previous paragraph, the π's can be estimated from the observed frequencies $\{n_\nu\}$ if the fraction in the population possessing the attribute Y_i, say β_i, is known for $1 \leq i \leq t$.

3.2 A Measure of Respondent Jeopardy: Recently Leysieffer and Warner (1976) have developed a measure of the jeopardy of respondent's privacy in the case of a single sensitive attribute. Here we shall extend their approach to the case of $t \geqq 2$ sensitive attributes: Consider the 2^t mutually exclusive and collectively exhaustive groups into which the population is divided depending on the possession or nonpossession of different attributes and denote these groups by $A_1A_2\cdots A_t$, $A_1^cA_2\cdots A_t$, $\cdots$, $A_1^cA_2^c\cdots A_t^c$ where the notation is obvious. Consider, say, the group $A_1A_2\cdots A_t$. By using the Bayes' theorem in the same manner as Leysieffer and Warner (1976) it can be shown that a measure of information resulting from response ν in favor of $A_1A_2\cdots A_t$ against $(A_1A_2\cdots A_t)^c$ is given by

$$g(\nu;A_1A_2\cdots A_t) = P(\nu|A_1A_2\cdots A_t)/P(\nu|(A_1A_2\cdots A_t)^c). \quad (3.1)$$

Thus the response ν can be regarded as jeopardizing with respect to the group $A_1A_2\cdots A_t$ (and not jeopardizing with respect $(A_1A_2\cdots A_t)^c$) if $g(\nu;A_1A_2\cdots A_t) > 1$ and not jeopardizing with respect to either $A_1A_2\cdots A_t$ or $(A_1A_2\cdots A_t)^c$ if $g(\nu;A_1A_2\cdots A_t) = 1$. Now to get a measure of the worst jeopardy of the privacy of an individual in group $A_1A_2\ldots A_t$ we define the jeopardy function for that group as

$$g(A_1A_2\cdots A_t) = \max_{\nu} g(\nu;A_1A_2\cdots A_t). \quad (3.2)$$

The jeopardy functions for other groups can be defined in an identical manner.

The parameters of each RR technique should be chosen so that the jeopardy function values for different groups do not exceed some prespecified upper bounds. We note here that these jeopardy function values will depend in general on the unknown θ's (in contrast to the case of $t = 1$). Therefore some apriori guesses at the values of the θ's will be necessary to compute their values.

3.3 Jeopardy Functions for Competing Techniques: Using the definitions (3.1) and (3.2), we shall derive the expressions for the jeopardy functions associated with the W-, S- and the M-techniques for $t = 2$. Here we shall consider only the following special case of practical interest. (The general case with $t \geqq 2$ is quite straightforward but algebraically messy and is hence omitted.) For the W-technique we take $P_1 = P_2 = P_W$ (say) where $P_W > 1/2$ without loss of generality. For the S-technique we take $P_1 = P_2 = P_S$ (say) and $\beta_1 = \beta_2 = \beta$ (say). For the M-technique we take $P_{11}^{(1)} = 1 - P_{11}^{(2)} = P_M$ (say) where $P_M > 1/2$ without loss of generality.

Define additional notation as follows: $Q_W = 1 - P_W$, $Q_S = 1 - P_S$, $Q_M = 1 - P_M$, $\gamma = 1 - \beta$ and $\theta^*_{12} = 1 - \theta_1 - \theta_2 + \theta_{12}$. Then the expressions for the jeopardy functions (using W, S and M to index the jeopardy functions for the W-, S- and the M-techniques respectively) are as follows. (The details of their derivations are given in an unabridged version of this paper available with the author.)

(i) W-technique:

$$g_W(A_1A_2)=P_W^2(1-\theta_{12})/\{P_WQ_W(1-\theta_{12}-\theta^*_{12})+Q_W^2\theta^*_{12}\}$$

$$g_W(A_1^cA_2)=P_W^2(1-\theta_2+\theta_{12})/\{P_WQ_W(\theta_{12}+\theta^*_{12})+Q_W^2(\theta_1-\theta_{12})\}$$

$$g_W(A_1A_2^c)=P_W^2(1-\theta_1+\theta_{12})/\{P_WQ_W(\theta_{12}+\theta^*_{12})+Q_W^2(\theta_2-\theta_{12})\}$$

$$g_W(A_1^cA_2^c)=P_W^2(1-\theta^*_{12})/\{P_WQ_W(1-\theta_{12}+\theta^*_{12})+Q_W^2\theta_{12}\}.$$

(ii) S-technique:

$$g_S(A_1A_2)=(P_S+Q_S\beta)^2(1-\theta_{12})/\{Q_S\beta(P_S+Q_S\beta)(1-\theta_{12}-\theta^*_{12})+Q_S^2\beta^2\theta^*_{12}\}$$

$$g_S(A_1^cA_2)=(P_S+Q_S\beta)(P_S+Q_S\gamma)(1-\theta_2+\theta_{12})/\{Q_S\gamma(P_S+Q_S\beta)\theta_{12}+Q_S^2\beta\gamma(\theta_1-\theta_{12})+Q_S\beta(P_S+Q_S\gamma)\theta^*_{12}\}$$

$$g_S(A_1A_2^c)=(P_S+Q_S\beta)(P_S+Q_S\gamma)(1-\theta_1+\theta_{12})/\{Q_S\gamma(P_S+Q_S\beta)\theta_{12}+Q_S^2\beta\gamma(\theta_2-\theta_{12})+Q_S\beta(P_S+Q_S\gamma)\theta^*_{12}\}$$

$$g_S(A_1^cA_2^c)=(P_S+Q_S\gamma)^2(1-\theta^*_{12})/\{Q_S\gamma(P_S+Q_S\gamma)(1-\theta_{12}-\theta^*_{12})+Q_S^2\gamma^2\theta^*_{12}\}.$$

(iii) M-technique:

$$g_M(A_1A_2)=P_M^2(1-\theta_{12})/Q_M^2\theta^*_{12}$$

$$g_M(A_1^cA_2)=(1-\theta_2+\theta_{12})/P_MQ_M(\theta_{12}+\theta^*_{12})$$

$$g_M(A_1A_2^c)=(1-\theta_1+\theta_{12})/P_MQ_M(\theta_{12}+\theta^*_{12})$$

$$g_M(A_1^cA_2^c)=P_M^2(1-\theta^*_{12})/Q_M^2\theta_{12}.$$

3.4 Equating the Jeopardy Functions for the Competing Techniques: Our approach here will be to first equate the jeopardy functions for the four different groups for the competing techniques and obtain their equivalent parameter values, i.e., their P-values and the β-value for the S-technique. (Clearly the parameter values yielded by the four sets of equations will not in general be consistent. Therefore some criterion such as guaranteeing the lowest jeopardy level will be necessary in order to arrive at a unique parameter value for each technique.) The next step in our approach will be to compute for each technique a measure of its performance based on these parameter values. We have taken the measure of performance to be the trace of the variance-covariance matrix of the estimator vector. We note here that because of the special symmetric case that we are considering for each technique, no optimization in the sense of Leysieffer and Warner (1976) is possible.

First we equate the $g_W(\cdot)$'s with the respective $g_S(\cdot)$'s and we obtain that $P_S = 2P_W - 1$ and $\beta = 1/2$. Next we equate the $g_W(\cdot)$'s with the respective $g_M(\cdot)$'s and solve the resulting quadratic equations for P_M in terms of $g_W(\cdot)$'s. We give below the condition that must be satisfied by $g_W(\cdot)$ in each case for the solution to be feasible (i.e., $1/2 \leqq P_M \leqq 1$) and the corresponding expression for P_M. For notational convenience we have defined the following quantities: $k_1 = g_W(A_1A_2)/(1 - \theta_{12})$, $k_2 = g_W(A_1^cA_2)/(1 - \theta_2 + \theta_{12})$, $k_3 = g_W(A_1A_2^c)/(1 - \theta_1 + \theta_{12})$ and $k_4 = g_W(A_1^cA_2^c)/(1 - \theta_{12}^*)$. We have

$$g_W(A_1A_2)=g_M(A_1A_2)\Rightarrow P_M=(k_1\theta_{12}^*-\sqrt{k_1\theta_{12}^*})/(k_1\theta_{12}^*-1)$$
$$\text{if } g_W(A_1A_2)\geqq(1-\theta_{12})/\theta_{12}^*. \tag{3.3a}$$

$$g_W(A_1^cA_2)=g_M(A_1^cA_2)\Rightarrow P_M=[1+\{1-4/k_2(\theta_{12}+\theta_{12}^*)\}^{1/2}]/2$$
$$\text{if } g_W(A_1^cA_2)\geqq4(1-\theta_2+\theta_{12})/(\theta_{12}+\theta_{12}^*). \tag{3.3b}$$

$$g_W(A_1A_2^c)=g_M(A_1A_2^c)\Rightarrow P_M=[1+\{1-4/k_3(\theta_{12}+\theta_{12}^*)\}^{1/2}]/2$$
$$\text{if } g_W(A_1A_2^c)\geqq4(1-\theta_1+\theta_{12})/(\theta_{12}+\theta_{12}^*). \tag{3.3c}$$

$$g_W(A_1^cA_2^c)=g_M(A_1^cA_2^c)\Rightarrow P_M=(k_4\theta_{12}-\sqrt{k_4\theta_{12}})/(k_4\theta_{12}-1)$$
$$\text{if } g_W(A_1^cA_2^c)\geqq(1-\theta_{12}^*)/\theta_{12}. \tag{3.3d}$$

It is only fair to point out that one drawback with the M-technique might be that it cannot match the W- and S-techniques at low levels of jeopardy. Also if unknown to the statistician, either $\theta_{12} = 0$ or $\theta_{12}^* = 0$ or both then at least one of the conditions on $g_W(\cdot)$ in (3.3) is certainly violated and there is no hope for matching the M-technique with the others in terms of the jeopardy values. In practice it is likely that θ_{12} (the proportion in the population possessing both the sensitive attributes A_1 and A_2) will be small whereas θ_{12}^* will be large. Hence it is likely that only the condition on $g_W(A_1^cA_2^c)$ in (3.3d) will be violated and it will not be possible to guarantee that $g_W(A_1^cA_2^c) = g_M(A_1^cA_2^c)$. However, this would be of no consequence since usually the upper limit on $g(A_1^cA_2^c)$ will be very large (even infinity) since $A_1^cA_2^c$ is a completely innocuous group.

Now the P_M-values given by (3.3a) - (3.3d) will in general be unequal. We follow the convention of guarding the individuals in the most sensitive group A_1A_2, i.e., controlling $g(A_1A_2)$ for each technique. Therefore we take the P_M-value given by (3.3a). Thus if the condition on $g_W(A_1A_2)$ in (3.3a) is satisfied then the corresponding P_M-value would be feasible and all the three techniques would be matched in terms of their jeopardy values for the A_1A_2 group.

3.5 Numerical Results: Define the trace inefficiency of a RR technique as the ratio of the trace of the (asymptotic) variance-covariance matrix of its estimates θ_1, θ_2, and θ_{12} to the corresponding quantity for the direct response technique when both the techniques use the same sample size n. This latter quantity is given by

$$\{\theta_1(1 - \theta_1) + \theta_2(1 - \theta_2) + \theta_{12}(1 - \theta_{12})\}/n.$$

The expressions for the traces of the variance-covariance matrices of the UMLE's of $\underset{\sim}{\theta}$ (which can be regarded as the asymptotic variance-covariance matrices of the RMLE's of $\underset{\sim}{\theta}$) for the W- and the S-technique are given respectively, in Clickner and Iglewicz (1976) and Barksdale (1971).

It can be checked that for the choice $P_S = 2P_W - 1$ and $\beta = 1/2$, the expressions for the variance-covariance matrices for the W- and the S-techniques are identical and therefore the two techniques are equivalent; this extends the corresponding result for t = 1 by Leysieffer and Warner (1976). Hence we only consider the comparison between the W- and M-techniques.

The values of P_W were obtained from Table 3 of Clickner and Iglewicz (1976) where they have computed them so that the W-technique attains selected levels of trace inefficiency (namely 1.25, 2.5, 5.0 and 10.0) for selected values of θ. The corresponding values of P_M were computed from (3.3a) which guarantees that $g_W(A_1A_2) = g_M(A_1A_2)$ but does not in general guarantee the equality of the jeopardy levels for the other groups. Using these P_W and P_M values the trace inefficiencies for the two techniques were computed. The results of these computations are displayed in Table I. The values of P_M reported are rounded off in the third decimal place.

An inspection of the results reveals that if θ_1 and θ_2 are small (which would usually be the case for sensitive attributes) and P_W is in the range 0.7 - 0.8 (which are the values most frequently used in practice) then the M-technique indeed dominates the W-technique. However, for large values of θ_1 and θ_2 leading to small values of θ_{12}^* the situation is reversed and the M-technique has either very large values for trace inefficiency or in a few cases the M-technique is even nonexistent.

An explanation of this phenomenon is as follows: First, consider the variation with respect to θ_{12}^*. It is easy to check that for fixed P_W and θ_{12}, the P_M-value (as given by (3.3a)) decreases with θ_{12}^* which leads to high values of the trace inefficiency and in some instances even the nonexistence of the M-technique. Next consider the variation with respect to P_W. We note that, in general (i.e., except for the case $\theta_{12} + \theta_{12}^* = 1$), $P_M < 1$ even when $P_W = 1$ and therefore by a continuity argument we would expect the W-technique to dominate the M-technique for P_W-values in the neighborhood of 1. For fixed $\underset{\sim}{\theta}$, as P_W decreases, P_M decreases too. But for the intermediate values of P_W, it is possible for the M-technique to dominate the W-technique. As P_W decreases even further, P_M approaches 1/2 and therefore leads to very high values of the trace inefficiency for the M-technique.

No clear indication of the dependence of the trace inefficiency on ρ_{12} is evident in this table. It is known, however, that for the W- and the S-techniques, the variances of $\hat{\theta}_1$ and $\hat{\theta}_2$ are not affected by the correlation; in fact, the corresponding formulae are the same as though these attributes were studied independently.

ACKNOWLEDGEMENT

The author wishes to thank Professor Michael Rubinovitch for a helpful discussion and Professor Robert Boruch for providing support for this research through Grant No. NIE-C-74-0115 from the National Institutes of Education.

REFERENCES

Barksdale, W. B. (1971), "New Randomized Response Techniques for Control of Nonsampling Errors in Surveys," unpublished Ph.D. thesis, Department of Biostatistics, University of North Carolina, Chapel Hill.

Clickner, R. P. and Iglewicz, B. (1976), "Warner's Randomized Response Technique: The Two Sensitive Questions Case," Report 5, Department of Statistics, Temple University; also presented at the Annual Meeting of the American Statistical Association at Boston.

Drane, W. (1976), "On the Theory of Randomized Responses to Two Sensitive Questions," Communications in Statistics-Theory and Methods, A5(6), 565-74.

Greenberg, B. G. Abul-Ela, A. A., Simmons, W. R., and Horvitz, D. G. (1969), "The Unrelated Question Randomized Response Model: Theoretical Framework," Journal of the American Statistical Association, 64, 520-29.

Leysieffer, F. W. and Warner, S. L. (1976), "Respondent Jeopardy and Optimal Designs in Randomized Response Models," Journal of the American Statistical Association, 71, 649-56.

Singh, J. (1976), "A Note on the Randomized Response Technique," paper presented at the Annual Meeting of the American Statistical Association at Boston.

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

TABLE I

TRACE INEFFICIENCIES

θ_1	θ_2	θ_{12}	ρ_{12}	P_W	P_M	W-technique	M-technique
.05	.05	.0125	.2105	.988	.967	1.2449	1.6038
				.939	.911	2.5050	2.8568
				.878	.851	4.9937	4.7318
				.813	.789	9.9919	7.8812
.10	.05	.0250	.3059	.982	.953	1.2549	1.5824
				.918	.881	2.5022	2.7575
				.847	.812	5.0209	4.5591
				.781	.750	10.0041	7.5315
.20	.15	.0750	.3151	.963	.901	1.2785	1.5903
				.869	.798	2.4988	2.7253
				.789	.724	4.9772	4.5618
				.727	.669	9.9388	7.7277
.25	.05	.0375	.2649	.973	.912	1.2522	1.6233
				.888	.811	2.5019	2.8843
				.810	.738	4.9880	4.8655
				.745	.680	10.0223	8.3834
.25	.25	.0625	.0000	.962	.857	1.2511	1.6269
				.858	.729	2.5005	3.1724
				.777	.657	4.9902	6.1222
				.716	.606	10.0204	12.5720
.25	.25	.2500	1.0000	.953	.953	1.2478	1.2251
				.837	.837	2.4951	2.0434
				.756	.756	5.0077	3.2044
				.699	.699	10.0397	4.9160
.40	.05	.0250	.0468	.971	.872	1.2525	1.6838
				.882	.750	2.5027	3.2818
				.803	.675	4.9789	6.2459
				.738	.621	10.0390	12.7284
.55	.25	.1250	-.0580	.958	.784	1.2509	1.8043
				.848	.633	2.4987	5.2679
				.766	.561	4.9943	21.4589
				.706	.516	10.0363	294.7612
.75	.05	.0250	-.1325	.978	.784	1.2563	1.8443
				.904	.623	2.5031	6.6806
				.828	.538	4.9948	62.3038
				.760*	----	10.0363	------
.75	.70	.5250	.0000	.959	.676	1.2522	4.1165
				.850	.504	2.4924	7059.5656
				.766*	----	4.9977	------
				.705*	----	9.9647	------

Note: The M-technique does not exist for the starred P_W-values and the corresponding $\underset{\sim}{\theta}$ vectors.