Yosef Hochberg, Jane C. Stutts, and Donald W. Reinfurt. (1976). The Randomized Response Technique: A Review and Application. Chapel Hill, NC: University of North Carolina Highway Safety Research Center.

highway
Safety
Research
Center

university
of north
carolina

chapel hill, n.c.

THE RANDOMIZED RESPONSE TECHNIQUE:
A REVIEW AND APPLICATION

Yosef Hochberg

Jane C. Stutts

Donald W. Reinfurt

May 1976

Prepared for the Department of Transportation,
National Highway Traffic Safety Administration
under Contract No. DOT-HS-4-00897 (SY085).

Report No. DOT-HS-4-00897-4

DOT-HS-4-00897
May 1976

THE RANDOMIZED RESPONSE TECHNIQUE:

A REVIEW AND APPLICATION

Yosef Hochberg

Jane C. Stutts

Donald W. Reinfurt

MAY 1976

FINAL REPORT
July 1, 1974 - May 31, 1976

Prepared for

DEPARTMENT OF TRANSPORTATION
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION
Washington, D.C. 20590

Technical Report Documentation Pag

1. Report No. DOT-HS-4-00897-4	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The Randomized Response Technique: A Review and Application		5. Report Date May, 1976
		6. Performing Organization Code
7. Author(s) Yosef Hochberg, Jane C. Stutts, Donald W. Reinfurt		8. Performing Organization Report No. UNC/HSRC-76/5/2
9. Performing Organization Name and Address The Highway Safety Research Center University of North Carolina Chapel Hill, NC 27514		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DOT-HS-00897
12. Sponsoring Agency Name and Address Department of Transportation National Highway Traffic Safety Administration 400 Seventh Street, S.W. Washington, DC 20590		13. Type of Report and Period Covered Final Report July 1, 1974 - May 31, 1976
		14. Sponsoring Agency Code

15. Supplementary Notes

The opinions, findings and conclusions expressed in this publication are those of the authors and not necessarily those of the National Highway Traffic Safety Administrati

16. Abstract

The purpose of this report is to describe a relatively new interviewing technique and examine its potential use in the field of highway safety research. Thi technique, which intends to encourage truthful responses in personal interviews on sensitive issues, is known as the Randomized Response Technique (RRT). After a brief introduction in Chapter I, Chapter II outlines the theoretical development of the technique, from its inception by Warner in 1965. Chapter III presents the results of some of the more pertinent field research, conducted on such topics as illegitimacy, abortion and drug usage, and then Chapter IV describes HSRC's efforts to extend application of the technique in the area of highway safety.

From its experience with the RRT, HSRC concludes that much additional field research is needed before the technique can be of any great value to highway safety researchers.

Also included in the report are two appendices which extend the theoretical development of the technique. Appendix A delineates two variations on a two stage randomized response scheme (TSRRS), while Appendix B outlines procedures for using a RRT on a subsample of a larger sample of possibly misclassified responses, to obtain a more reliable estimate of the parameter of interest.

17. Key Words Randomized Response Technique Randomizing Device		18. Distribution Statement Available to the Public	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages iii, 92	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

ACKNOWLEDGMENTS

The authors very gratefully acknowledge the efforts of Ms. Lucy Smith and Ms. Jane Youngblood, who were responsible for carrying out the three pilot randomized response studies described in Chapter IV. Though the studies did not prove successful, this certainly was not due to any lack of dedication on their part.

Also, appreciation is expressed to Dr. Bernard G. Greenberg, Dean of the UNC School of Public Health, for his helpful guidance in the initial phase of the project, to Mr. Ralph Folsom for his inputs during the preparation of this report, and to Dr. James R. Abernathy, for his review of the draft of the final report.

Finally, special thanks are due the typists, Ms. Barbara Crockett and Ms. Peggy James, whose skills are especially evident in the Appendices to the report.

I. INTRODUCTION

When conducting sample surveys of human populations, one must contend with several sources of non-sampling bias. These include failure to locate or visit some units in the sample, failure to recontact the "not-at-homes", and inability of respondents to provide the information requested (Cochran, 1965).

Two sources of non-sampling bias that may be of particular importance when the survey involves questions about sensitive or highly personal matters are refusal to respond (refusal bias) and intentionally giving misleading responses (evasive answer or response bias) (Warner, 1965). Respondents may quite naturally be reluctant to respond truthfully when asked questions concerning, for example, some socially deviant or illegal behavior. In sample surveys of this sort, it is extremely important that the interviewer obtain the cooperation of the respondent so that the data obtained will be accurate and complete.

Traditionally, cooperation has been sought by attempting to gain the respondent's confidence and assure him of the anonymity of his response. This approach, however, has generally not proven effective. A far more promising approach to encouraging cooperation and truthful replies in sample surveys of sensitive matters was introduced by Warner (1965). This approach is known as the "randomized response technique (RRT)".

Basically, the randomized response technique attempts to encourage cooperation and truthful replies (and thus decrease refusal and response bias) by allowing the respondent to select a question on a probability basis from two or more questions, without revealing to the interviewer which question he has chosen. Since all of the questions have the same range of responses (e.g., "yes" or "no"), the interviewer has no way of knowing which question is being answered. Thus, the respondent can answer truthfully without fear of "revealing" himself to the interviewer.

In fact, since the "randomized response" approach calls for direct answers to randomly selected questions, a more appropriate name for the technique might be the "randomized question technique (RQT)". However, for the purposes of this report, the name originally suggested by Warner and subsequently adopted in later reports on the technique will be used.

With the information provided by the respondent, along with knowledge of the probability distribution used in the design, an estimate of the proportion of the population with the sensitive characteristic can be computed. If one assumes that the respondents all answered truthfully since their privacy had been assured, then this estimate should be unbiased with respect to response errors.

In the following sections of the report, the development of the randomized response technique over the past decade is outlined. Some examples of research utilizing the technique are then presented. Finally, HSRC's attempts to apply randomized response techniques in the field of highway safety are described.

II. Development of the Randomized Response Technique

The Warner Design

In his initial formulation of the randomized response procedure, Warner (1965) assumed the population sample to be divided into two mutually exclusive and complementary classes -- those with the sensitive attribute (Group A) and those without (Group B). The person being interviewed would use a randomizing device to select one of two statements of the form:

I belong to Group A

I do not belong to Group A

He would then answer, "yes" or "no," whether the statement selected was characteristic of him.

As an example, Group A might consist of individuals who have knowingly cheated on income tax returns, while Group B might be those who have never knowingly cheated on their returns. The randomizing device might be a spinner that points to "A" with probability p, and "B" with probability (1-p). If the respondent spins the spinner and it points to "A," he must answer, "yes" or "no," whether he has ever knowingly cheated on his income tax returns. If the spinner points to "B," he answers, "yes" or "no," whether he has never cheated on his returns.

If λ is the probability of a "yes" answer, π_A the proportion of the total population belonging to the stigmatized group, p the probability of responding to the sensitive statement, and (1-p) the probability of responding to the non-sensitive statement, then

$$\lambda = \Pr\{\text{"yes" answer}\}$$

$$= \Pr\{\text{spinner points to "a" and a "yes" answer given}\}$$
$$+ \Pr\{\text{spinner points to "B" and a "yes" answer given}\}$$

$$= \Pr(\text{"A" selected}) \Pr(\text{"yes"} | A)$$
$$+ \Pr(\text{"B" selected}) \Pr(\text{"yes"} | B)$$

$$= p\,\pi_A + (1-p)(1-\pi_A)$$

The estimated value of π_A is then given by

$$\hat{\pi}_A = \frac{\hat{\lambda} - (1-p)}{2p-1}, \quad p \neq \tfrac{1}{2}$$

where $\hat{\lambda}$ is the proportion of "yes" responses actually observed. Finally, an estimate of the variance of π_A is given by

$$V_A = \hat{\text{Var}}(\hat{\pi}_A) = \frac{\hat{\pi}_A(1-\hat{\pi}_A)}{n} + \frac{p(1-p)}{n(2p-1)^2},$$

where n is the number of people participating in the survey. Note that this estimate is actually the sum of the variance associated with direct questioning (assuming completely truthful replies) and the variance due to the randomized response technique itself. Also, note that this variance will be large when p is close to 0.5 and/or when n is small.

In considering the randomized response technique, the choice of p is of primary importance, since it largely determines the extent to which those interviewed are likely to cooperate and respond truthfully. The value of p also controls the sample size, n, required for a given level of precision. Clearly, low values of p do not offer as much information per respondent, and thus require a larger sample size for a given level of precision. At the same time, however, high values of p (those close to 1) are not as effective in encouraging respondent cooperation. The respondent has the feeling that "the deck is loaded against him." Thus, the rule for selection of p seems to be a compromise - namely,to select it low enough to assure cooperation, but high enough so that an extraordinarily large sample is not necessary.

To examine the effectiveness of the randomized response technique vs. direct questioning, Warner first assumed that individuals interviewed with the randomized response approach would respond truthfully, but that there would be less than truthful reporting if a direct approach were adopted. He then compared the mean square errors (the variance plus the square of the bias) of each approach, for various values of p and for various probabilities of truthful reporting if questioned directly. He concluded that except for the cases where the bias using direct questioning is negligible (i.e., near zero) there appears to be sizable potential gains through the randomized response. He left open, however, the task of determining which types of randomized response techniques might prove most effective.

Since Warner first introduced the randomized response procedure in 1965, there have been many attempts to extend, modify and improve the technique. Generally, more has been done to develop the technique "on paper" than to apply it in the field. As will be shown later, however, field application has not been neglected. In the following sections of this paper, some of the important developments in the technique will first be reviewed, followed by a description of some field applications.

The Trichotomous Randomized Response Model

Abul-Ela, Greenberg, and Horvitz (1967) extended the original Warner technique to the case where the population is divided into three mutually exclusive groups according to some characteristic, with either one or two of these groups being stigmatizing. Let π_1 π_2 , and π_3 represent the true proportions of the population in groups A, B, and C, respectively, where

$$\sum_{j=1}^{3} \pi_j=1.$$

There must be three unique probabilities of selecting the statement characterizing each of the three groups. In addition, since two independent, non-overlapping samples of size n_1 and n_2 are required to estimate values of the π_j , two different sets of randomizing probabilities are required. Thus, if p_{ij}= the probability of selecting the statement denoting the j^{th} group (j = 1, 2, 3) for the i^{th} sample (i = 1, 2) where

$$\sum_{j=1}^{3} p_{ij}=1$$

then the p_{ij} must be selected such that

$$(p_{11} - p_{13})(p_{22} - p_{23}) \neq (p_{12} - p_{13})(p_{21} - p_{23}).$$

As in Warner (1965), mean square errors of the trichotomous randomized response estimates were compared with mean square errors of the regular trinomial estimates (i.e., those obtained using standard direct questioning approaches) to obtain a measure of efficiency for this technique. Contrary to Warner, however, Abul-Ela _et al._ did not assume that there would be completely truthful reporting with the randomized response design. Besides this "ideal" situation, they also considered the case where there is less than completely truthful reporting by both groups, but where the randomized response respondents are at least as truthful as their direct question counterparts.

The mean square error efficiency (MSEE) was defined as the ratio of the mean square error of the regular (direct questioning) technique to the mean square error of the randomized technique, and calculated for various combinations of p's along with different probabilities of completely truthful reporting. This was done for the case where

$$\underline{\pi} = (\pi_1, \pi_2, \pi_3) = (.80, .05, .15).$$

Results showed that the trichotomous randomized response technique has potential advantages over the direct question approach, even when respondents are not reporting completely truthfully in either mode of questioning. The results also indicated that the relative efficiency of the new technique improves as the p_{ij}'s progressively differ from one-third.

The results of a field trial using the trichotomous randomized response technique are given in Abul-Ela _et al._ and will be discussed in Chapter III of this report.

The Unrelated Question Design

It seems reasonable to assume that a respondent would be more likely to answer truthfully if asked to reply to one of three questions (as in the trichotomous design described above) than to one of two questions (as in Warner's original design). Provided there is only one stigmatizing question, the additional question would intuitively seem to lessen the chance of self-incrimination. Simmons (cited in Greenberg *et al*., 1974) suggested a somewhat different approach to encouraging respondent cooperation and thus improving upon the original Warner design. He noted that the respondent might be even more truthful if provided the opportunity of responding to either of two questions, one of which was completely innocuous and unrelated to the stigmatizing question. For example, the following two questions might be included in a randomized response survey on the prevalence of shoplifting:

1) I have intentionally taken goods from a store without paying for them.

2) I was born in North Carolina.

The theoretical framework for this unrelated question randomized response technique was developed by Greenberg, Abul-Ela, Simmons, and Horvitz (1969). Expressions for estimates of π_A, the proportion of the population with the sensitive characteristic A, were derived for the following two cases:

1) π_Y, the proportion of the population with the non-sensitive characteristic Y, is unknown.

2) π_Y is known.

In the first case, two independent samples of size n_1 and n_2, respectively, are required to estimate π_A. In addition, the probability for choosing the sensitive question must be different for the two samples. If p_i is the probability of selecting the sensitive question in sample i, i = 1, 2, and λ_i is the probability of a "yes" response in sample i, then

$$\lambda_i = p_i\pi_A + (1-p_i)\pi_Y \qquad i=1,2 \tag{2.1}$$

From (2.1), one can derive the estimate of π_A, based on the two samples, namely

$$\hat{\pi}_A = \frac{\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1)}{p_1-p_2}$$

where $\hat{\lambda}_i$ is the observed proportion of "yes" responses in sample i, i = 1, 2. An estimate of the variance of $\hat{\pi}_A$ is then given by

$$V_A = \hat{Var}(\hat{\pi}_A) = \frac{1}{(p_1-p_2)^2}\left[\frac{\hat{\lambda}_1(1-\hat{\lambda}_1)(1-p_2)^2}{n_1} + \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)(1-p_1)^2}{n_2}\right]$$

where n_i is the size of sample i, i = 1, 2.

To minimize the variance and improve the efficiency of their two sample unrelated question design, the authors made recommendations for selecting values for certain parameters. Generally, they suggested that (1) the non-sensitive group, Y, be such that $|\pi_Y-0.5|$ is a maximum and π_Y falls on the same side of 0.5 as π_A; 2) p_1 be as close to 1 or 0 as possible without jeopardizing cooperation of the respondent; and 3) p_2 be chosen so that $p_2 \quad 1 - p_1$. The authors also suggested that n_1 and n_2 be selected such that the larger portion of the total sample size would be allocated to the case where $p > 0.5$ (i.e., where the respondents are more likely to be asked the sensitive question than the non-sensitive one). Furthermore, n_1 and n_2 should be selected to satisfy the following relationship:

$$\frac{n_1}{n_2} = \sqrt{\frac{\hat{\lambda}_1(1-\hat{\lambda}_1)(1-p_2)^2}{\hat{\lambda}_2(1-\hat{\lambda}_2)(1-p_1)^2}}$$

where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are "educated guesses" for λ_1 and λ_2.

Adopting these optimal parameter values for a sample case where $p_1 = 1 - p_2 = .20$, $\pi_A = .20$, $\pi_Y = .10$, and $n = 1000$ (with n_1 and n_2 optimally allocated), the authors used the ratio of the mean square errors to determine the effectiveness of the two sample unrelated question technique vs. the original Warner design. The results revealed that, whenever the assumed proportion of truthful responses was at least as great for the unrelated question technique as for the Warner design, the former was considerably more effective.

A simpler case of the unrelated question design arises when the proportion (π_Y) of the population with the non-sensitive characteristic is known in advance (i.e., case (2)). For this case, Greenberg et al. (1969) showed that only one sample is required to estimate the proportion of the population with the sensitive attribute, and $\hat{\pi}_A$ and $\hat{Var}(\hat{\pi}_A)$ are obtained as follows:

$$\hat{\pi}_A = \frac{\hat{\lambda} - (1-p_1)\pi_Y}{p_1}$$

$$V_A = \hat{Var}(\hat{\pi}_A) = \frac{\hat{\lambda}(1-\hat{\lambda})}{np_1^2} \qquad \text{(where } \pi_Y \text{ is given).}$$

Furthermore, the authors showed that this unrelated question design with π_Y known is even more efficient than the same design with π_Y not known, since observations are not wasted on that sample where the probability of selecting the sensitive question is lowest. This increase in efficiency remains even when there is some minor error in the "known" value of π_Y .

Finally, Greenberg et al. (1969) indicated how the unrelated question method might be extended to account for more than two population groups. They noted, however, that the optimal selection of parameter values in these cases would be much more complicated than in the binomial case.

Moors (1971) proved the validity of the various recommendations made by Greenberg et al. for all parameter selections, with the exception of the choice of p_2 when π_Y is not known. Instead of making p_1 and p_2 sum to one, Moors showed that choosing $p_2 = 0$ resulted in a much higher precision for the original two sample unrelated question model, and had several practical advantages as well. When $p_2 = 0$, the unrelated question randomized response device is employed in one of the independent samples, while the second sample is used solely to estimate π_Y. In this case,

$$\hat{\pi}_A = \frac{\hat{\lambda}_1 - (1-p_1)\hat{\lambda}_2}{p_1}$$

since $\hat{\pi}_Y = \hat{\lambda}_2$.

Moors showed that, for values of π_Y between 0.1 and 0.5, the increase in efficiency from setting $p_2 = 0$ was about 80 percent for $p_1 = 0.7$, 25 percent for $p_1 = 0.8$, and 5 percent for $p_1 = 0.9$. He also showed that, even for the worst possible choice of π_Y (namely, $\pi_Y = 0.5$) his "optimized unrelated question model" proved superior to Warner's technique.

Clearly, Moors' case where an independent sample is used to estimate an unknown π_Y is very similar to the case described by Greenberg et al. (1969) where π_Y is known in advance. In the former, an "estimated" value of π_Y (i.e., $\hat{\lambda}_2$) is used to estimate π_A , while in the latter, a "known" value of π_Y is used. Other than this, there is no difference in the calculation of $\hat{\pi}_A$.

Morton (cited in Greenberg, Horvitz, Abernathy, 1974) demonstrated that, at least theoretically, two samples are never necessary since knowledge of π_Y can always be incorporated into the randomizing device. In this approach, the respondent randomly selects one question from a field of three, rather than two, questions. One of these three questions deals with the sensitive characteristic, another requires a "yes" response, and a third requires a "no" response. Thus, this approach can be termed the "built-in unrelated question RRT."

As an example, the randomizing device might be a plastic box containing red, white and blue beads. The questions corresponding to these three colors of beads might be as follows:

red: I have had an induced abortion.

white: The color of this bead is white.

blue: The color of this bead is white.

The respondent uses the randomizing device to confidentially select one of the three types of beads, and then answers only the question corresponding to that color bead.

If p_1, p_2 and p_3 represent the known probabilities of selecting a red, white or blue bead, respectively, where $p_1 + p_2 + p_3 = 1$, then, using the same notation defined previously,

$$\pi_Y = \frac{p_2}{p_2+p_3} \qquad (2.2)$$

and since $\lambda = p_1\pi_A + p_2$,

$$\hat{\pi}_A = \frac{\hat{\lambda}-p_2}{p_1} \qquad (2.3)$$

Substituting in (2.3) the value of p_2 found in (2.2) and replacing $(p_2 + p_3)$ by $(1 - p_1)$, one obtains

$$\hat{\pi} = \frac{\hat{\lambda} - (1-p_1)\pi_Y}{p_1}$$

which is the same estimate found in the one sample unrelated question design where π_Y is known in advance. The variances likewise are identical.

Folsom, Greenberg, Horvitz and Abernathy (1973) introduced an interesting variation on the general unrelated question randomized response design when π_Y is not known in advance. They termed this model the "two alternate question RRT" since it incorporates one sensitive question (A) and two non-sensitive, unrelated alternate questions (Y_1 and Y_2) to yield two unbiased estimates of π_A. One sample of respondents uses a randomizing device to select either the sensitive question (A) or the first non-sensitive question (Y_1). This sample also responds directly to the second non-sensitive question (Y_2). Similarly, a second sample responds directly to the first non-sensitive question (Y_1), but also uses a randomizing device to determine whether to answer the sensitive question (A) or the second non-sensitive question (Y_2). Table 2.1 illustrates this design.

Table 2.1 Randomized response technique from Folsom et al. (1973).

Survey mode	Sample 1	Sample 2
Randomized Questioning	Question A Question Y_1	Question A Question Y_2
Direct Questioning	Question Y_2	Question Y_1

Each sample yields an independent estimate of π_Y. The authors showed that the optimum estimator, a weighted average of these two estimates, has minimum variance when the two sample sizes are equal and when the alternate questions are positively associated with the sensitive question. Furthermore, this estimator was found to be never _less_ efficient than Moors' optimized version of the two sample (one) unrelated question design, but never _more_ efficient than the single sample unrelated question model with π_Y known.

In summary, of the various unrelated question randomized response designs presented, Morton's "built-in unrelated question RRT" and Greenberg's "unrelated question RRT with π_Y known" would appear to be the most effective. The next most effective unrelated question technique would be Folsom's "two alternate question RRT," followed by Moors' "optimized unrelated question RRT." The original unrelated question RRT introduced by Greenberg appears to be the least effective, but is still an improvement over the Warner design. Note, however, that since these relationships are based solely on a comparison of mean square errors, they do not necessarily reflect the relative effectiveness of these techniques for obtaining information in the field This point will be examined in a later chapter.

Two Stage Randomized Response Schemes (TSRRS)

As discussed in previous sections, Abul-Ela _et al_. (1967) and Greenberg _et al_. (1969) demonstrated how the randomized response technique could be extended to the case where the population is divided into three or more exclusive groups. Abul-Ela _et al_. extended Warner's original randomized response design to the trichotomous case, while Greenberg _et al_. showed how the unrelated question technique could allow for three or more groups. In addition, Warner (1971) suggested a competing model to that of Abul-Ela _et al_., based on the general linear model.

Recently, it was noted that the efficiency of all three of these approaches is lowered because there is no differential application of randomized response procedures to the stigmatized and non-stigmatized groups (see Appendix A). Clearly, the most efficient randomized response schemes would be those that maximize protection for individuals in the stigmatizing groups while at the same time minimizing protection for individuals not in the stigmatizing groups. Thus, the problem becomes one of extracting a portion of the total sample not belonging to the stigmatizing group.

In order to accomplish this, a two-stage randomized response scheme (TSRRS) was proposed, in which the second stage is conditioned on the individual's response in the first stage. A detailed discussion of this approach is included as Appendix A. Two variations of TSRR's are presented. In the first scheme, information from individuals' randomized responses to a first stage of questioning is used to elicit further information in the second stage using direct questioning. In the second variation of the TSRRS, the individual's direct response to the first stage questioning serves as a basis for eliminating a portion of the sample from a second stage, where a conventional randomized response approach is utilized.

The Contamination Model

The various unrelated question designs described earlier represent one line of approach to modifying and improving the original Warner randomized response technique. An alternative approach was suggested by Boruch (1972). Boruch proposed that error be introduced into the classificatory data by presenting the respondent a single sensitive question and instructing him to either lie or tell the truth depending on the outcome of a randomization device. If Φ_p and Φ_n represent known probabilities of obtaining false positive (saying they belong to the sensitive group when in fact they do not) and false negative replies, respectively, then the probabillty of a "yes" response with Boruch's contamination design is given by

$$\lambda = \pi_A(1-\Phi_n) + (1-\pi_A)\Phi_p.$$

Thus, π_A is estimated by

$$\hat{\pi}_A = \frac{\hat{\lambda}-\Phi_p}{1-\Phi_p-\Phi_n}$$

where, as earlier, $\hat{\lambda}$ is the proportion of "yes" responses actually observed. The estimated variance of π_A is given by

$$\hat{Var}(\hat{\pi}_A) = \frac{\hat{\lambda}(1-\hat{\lambda})}{n(1-\Phi_p-\Phi_n)^2}$$

In comparing the effectiveness of the contamination method with the original Warner design and the Greenberg unrelated question approach when π_Y is known in advance, Boruch first determined conditions under which the approaches would yield equivalent estimates. Thus, for example, there is no conceptual difference between results based on the Warner model and those based on the misclassification model when the contamination parameters Φ_p and Φ_n are both equal to the probability of selecting the non-sensitive statement in the Warner design.

When the various conditions of equivalence are not present, Boruch concluded that the contamination method is the most efficient in certain restricted ranges, namely, when the population proportion with the sensitive characteristic (π_A) is either very high or very low, and when the probability of selecting the sensitive question is not close to 0.5. Otherwise, he concluded that the unrelated question approach with π_Y known was the most efficient.

Boruch proceeded to argue, however, that the contamination method may prove more effective than <u>any</u> standard randomized response approach when put to the test in the field. This would be because the subject being interviewed would perhaps accept the process of "lying" easier than he would the more complicated processes behind the other randomized response approaches. Therefore, the subject should be more inclined to cooperate and respond truthfully when questioned by the contamination approach. Also, Boruch suggested that, since the contamination design is based on the more familiar misclassification and error measurement models as opposed to the less developed randomized response models, it might be more readily accepted and simpler to use for the researcher.

Finally, the author suggested that the greatest use of the contamination method may rest, not in obtaining information on sensitive matters through personal interviews, but in overcoming problems of disclosure associated with the use of large data banks. This is because the method readily lends itself to the development of simple computer programs that can generate contaminated data prior to release to the researcher.

Multiple Trials Design

Clearly, using more than one trial per respondent will reduce the variance and increase the effectiveness of a RRT. A unique approach to incorporating the concept of multiple trials into the randomized design was suggested by Chow and is described in Greenberg, Horvitz and Abernathy (1974). This approach is based on a unique randomizing device that, in a single trial, has an effect which is equivalent to that of several trials with the Warner technique. Basically, this device consists of a spherical bottle or flask with a long, narrow neck, containing beads of at least two different colors. Only a given number of beads (say k) can "line up" in the neck of the bottle at one time, and thus the total number of beads of each color must be greater than k.

As an example, the bottle might contain nine red and six white beads (with the red beads corresponding to the sensitive category), and the neck might have exactly five positions, so that k = 5. During the interview , the respondent shakes the bottle and then turns it so that five beads enter the neck. Without letting the interviewer see the beads and without mentioning any color, he then reports the number of beads in the neck corresponding to the group to which he belongs. Thus, if he belongs to the sensitive group and a total of three red and two white beads are visible in the neck of the bottle, his correct response is "three".

If p is the proportion of red beads and z_i is the number of beads reported by the i-th respondent (z_i = 0, 1, 2..., k), then the mean number of beads reported ($\bar{z}$) will have expected value

$$E(\bar{z}) = \pi_A p + (1-\pi_A)(1-p)$$

and π_A will be estimated by

$$\hat{\pi}_A = \frac{\bar{z}-(1-p)}{(2p-1)} \qquad \text{where } \bar{z} = \frac{1}{kn} \sum_{i=1}^{n} z_i$$

with

$$Var(\hat{\pi}_A) = \frac{\hat{\pi}_A(1-\hat{\pi}_A)}{n} + \frac{K-k}{K-1}\left[\frac{p(1-p)}{kn(2p-1)^2}\right]$$

where K is the total number of beads and k the number of beads that can move into the neck of the bottle at one time.

Note that, for the single trial case where k = 1, this variance is identical to the variance of the Warner technique. For larger values of k, however, the variance associated with the bottle device is proportionately smaller.

A modification of this approach, presented in Liu, Chow and Mosley (1975), allows for a population which is divided into t mutually exclusive groups. This approach requires that there be t different color beads and that the number of beads of each color be different. In addition, all of the beads must be able to fall into the neck of the bottle. That is, if there are a total of m beads, then there must be m locations along the neck. A diagram of this device, termed "The Hopkins' Randomizing Device - Model III", is shown in Figure 2.1.

As before, the respondent shakes the bottle and allows the beads to move into the neck. This time, however, he responds by giving the location of the first bead from the bottom of the neck corresponding to the group to which he belongs. For example, if there are six red (R), three white (W),and two blue (B) beads, with red being the "sensitive" color, and the beads fall into the neck in the order WBRRRWRRBWR, then the correct response for a subject possessing the sensitive trait is "three".

Liu et al. argue that such a randomizing device is particularly applicable for obtaining discrete quantitative data, since all that is required is a different color ball corresponding to each group or number category. In fact, since this second "multiple trials" approach readily allows for t > 2 groups of any sort, it might most appropriately be viewed as an alternative to other multichotomous designs.

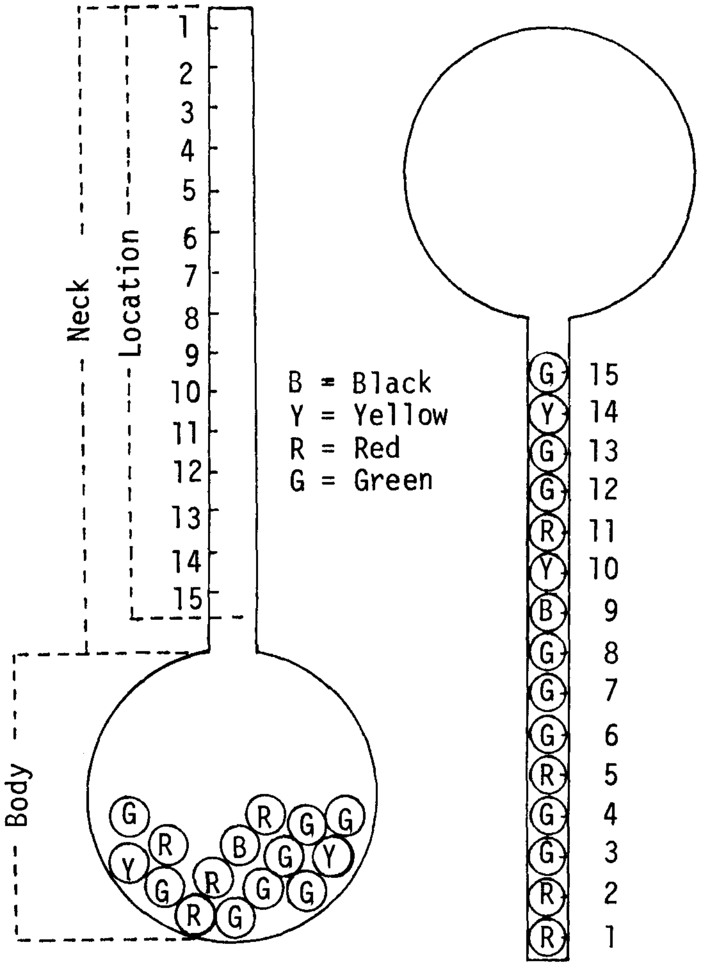


Figure 2.1. "The Hopkins' Randomizing Device - Model III" (<u>JASA</u>, <u>70</u>, p. 330).

Comparison of Randomized Response Techniques

Greenberg et al. (1974) compare the effectiveness of six different randomized response designs in estimating the proportion of a population with a sensitive characteristic. The designs compared are the "Warner design", the "single unrelated question design" with π_Y known, the "two unrelated questions design" (Folsom), the "contamination design", and two variations of the "multiple trials designs".

A "design effect" is computed for each of the six designs, which is the ratio of the variance of the design to the mean square error of the estimated proportion obtained by directly questioning the same number of respondents. These design effects are computed for various combinations of π_A (.05, .10, 20, .40), p (.90, .80, .70, .50) and sample size n (100, 400, 800). It was assumed that certain proportions (.50 , .70, .80, .90, 1.00) of the respondents with the sensitive characteristic would not answer truthfully if questioned directly, but that all respondents with the sensitive characteristic would answer truthfully when questioned with a randomizing technique. Also, it was assumed that respondents not belonging to the sensitive group would always answer truthfully.

Results of the comparisons indicate that the initial "Warner design" is far less efficient than the other randomized response designs, but that no design is superior for all combinations of π_A, p, and n. The "single unrelated question design" with π_Y known was slightly more efficient than the "two unrelated questions design" with π_Y unknown. And, while the "multiple trials designs" were more efficient than the "unrelated question designs" for higher values of π_A, they were less efficient for lower values of π_A. The "contamination design" was generally less efficient than the "unrelated question" or "multiple trials designs", but still much more efficient than the "Warner design".

As might be expected, the effectiveness of randomized response techniques relative to direct questioning increases as the proportion of respondents who would not answer truthfully when questioned directly increases. The effectiveness of these techniques also increase as the sample size increases, and as the probability of selecting the sensitive question increases. Finally, the effectiveness of randomized response techniques relative to direct questioning were shown to increase as π_A, the proportion of the population with the sensitive characteristic, increases from 0.05 to 0.40. Presumably, for levels of π_A greater than 0.40, a randomized technique would not be necessary since there would be little, if any, stigma attached to belonging to the "sensitive" group.

Greenberg *et al*. conclude that, whenever there is a chance that the respondent will not answer a direct question truthfully, use of a randomized response technique should be considered. However, they suggest that the particular randomized response technique employed (outside of the Warner technique) is not as critical as values selected for the probability of selecting the sensitive question (p) and the sample size (n).

Perhaps even more important than whether or not a randomized response design appears effective "on paper" is whether or not it "works" in the field. The purpose of any randomized response technique is to encourage the respondent being interviewed on a sensitive subject matter to cooperate and respond truthfully. Thus, the only way to directly ascertain the utility of a design that looks attractive on paper is to test it in the field on a question that might be validated in some other way. Obviously, a randomized response technique which is perceived by the respondent as an attempt to trick or deceive is not going to be as effective in the field as another randomizing technique that is more easily understood and readily accepted.

When comparing the relative field effectiveness of different randomized response techniques, one must also consider the varying effects of subject matter, respondent populations, methods of presentation, etc. A particular randomized response technique may work very well in one setting, but be completely ineffective in another. Thus, the overall relative effectiveness standings of techniques might be expected to vary with alterations in the field environment.

Chapter III describes some of the field research that has been conducted using the randomized response technique, and points out some of the limitations of this survey approach.

III. FIELD APPLICATIONS OF THE RANDOMIZED RESPONSE TECHNIQUE

Introduction

While there has been considerable theoretical advancement in randomized response methodology, until recently there were relatively few reports documenting use of the technique in actual data collection. Such field investigations are essential for evaluating the true effectiveness of the technique in obtaining reliable data on sensitive topics.

In the past, field investigations of randomized response techniques have taken several forms. Much of the earlier research was aimed at directly testing the effectiveness of a particular RRT in obtaining truthful responses for a given topic and from a given population. These studies compared the results obtained with the randomized response methods to corresponding results from an independent source. Other more recent randomized response studies have been conducted to secure information previously unobtainable. While these studies do not constitute a direct test of the effectiveness of the techniques, they usually give some indication of the practicality of employing various randomized response techniques in certain areas of investigation.

Still other studies have been conducted primarily to obtain public reaction and to test the general feasibility of new randomized response techniques, randomizing devices, etc. And finally, there have been a few attempts to compare the field effectiveness of various randomized response and direct question approaches.

Following is a description of some of the more significant research directed at testing the field effectiveness of various randomized response techniques.

Applications in Areas of Illegitimacy, Abortions and Extramarital Sex

Abul-Ela, Greenberg and Horvitz (1967).

One of the earliest applications of the randomized response technique is briefly reported by Abul-Ela *et al.* (1967) and involves the initial trial of the trichotomous randomized response model. The objective of the study was to test the effectiveness of the technique in estimating the proportion of unwed mothers in North Carolina who had had a reported live birth between October, 1964 and October, 1965. The three groups used for the trichotomous design were 1) females married at time they became pregnant, 2) females who married during pregnancy, and 3) females still unwed at time of delivery. Over 3,000 households were contacted by some 31 interviewers. The randomizing device employed was a deck of cards.

Estimates of illegitimacy were computed for the total sample and by race, residence (urban - rural), and educational level. These results were compared with known information obtained from birth certificates. While the authors do not present any figures on this comparison, they do make some conclusions regarding future use of the randomized response technique. These include a warning against using cards as a randomizing device except by skilled interviewers and intelligent respondents, along with a recommendation that, regardless of the randomizing device used, interviewers understand the technique and appreciate its purpose. The authors also note that bias was quite evident in their survey due to non-randomized response and inaccurate reporting by respondents. Thus, this early test of a randomized response technique was certainly not altogether successful.

Greenberg, Abul-Ela, Simmons and Horvitz (1969).

A field test of the unrelated question randomized response design was reported by Greenberg *et al.* (1969) in conjunction with the development of the technique. The test was conducted during the fall of 1965, before recommendations had been developed for optimal choices of π_Y (proportion having the non-sensitive characteristic) and sample size allocation. This study was also directed at estimating the frequency of illegitimate births. The authors investigated a sample of 148 households for which information on marital status of mother was available. The sensitive statement in this case was, "There was a baby born in this household after January 1, 1965, to an unmarried woman who was living here." The non-sensitive, unrelated statement was, "I was born in North Carolina." Respondents answered, "true" or "false", to whichever statement was selected using an unreported randomizing device.

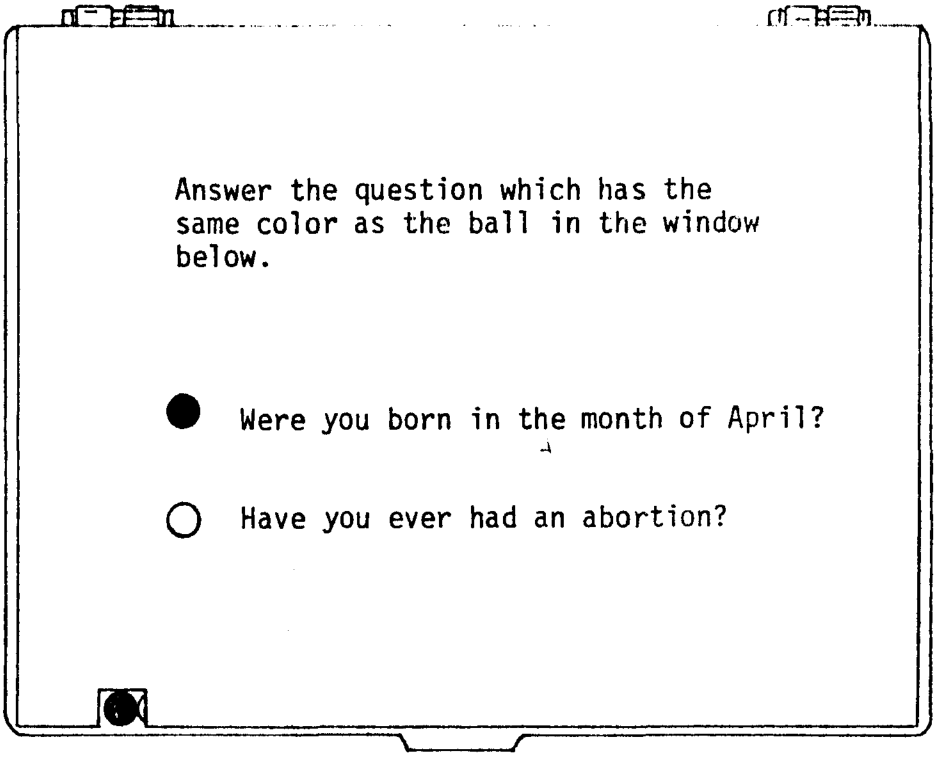
In contrast to the previously reviewed study on illegitimacy, results of this study showed amazing accuracy in estimating the frequency of illegitimacy in the households contacted. For example, the proportion of illegitimacy among white respondents sampled was 7.7 percent according to information on birth certificates, and 7.4 percent as estimated by the randomized response technique. The slight underestimate of illegitimacy rates in these and other studies is not surprising. Provided that the stigmatizing question is indeed perceived as such by the respondent, then it would be expected that the RRT would never overestimate the proportion of the population with the stigmatizing attribute. That is, those responding to the stigmatizing question would never say that they belonged to the stigmatizing group when in fact they did not. At the same time, however, they might still be reluctant to respond correctly when indeed they did belong to this stigmatizing group.

Abernathy, Greenberg
and Horvitz (1970).

In this study, the authors used the single sample, unrelated question randomized response technique (π_Y known) to estimate the proportion of women 18-24 years of age having an abortion during the previous year. They also used a two sample, unrelated question design to estimate the proportion of women 18 years of age and over who had ever had an abortion. The estimates were based on two non-overlapping samples involving approximately 2900 eligible women residing in central, urban North Carolina.

The particular randomized response device employed was one developed by Greenberg in conjunction with earlier research on the randomized response technique. Briefly, this device consists of a sealed transparent plastic box containing different color beads. If the randomized response design is a dichotomous model (e.g., the single, unrelated question design), there are two different color beads; if it is a trichomous model (e.g., Abul-Ela et al., 1967), there are three different color beads, etc. The sensitive and non-sensitive questions are printed on a piece of paper covering the lid of the box. Each question is coded by one of the bead colors. The respondent shakes the box, and allows one of the beads to move up a built-in ramp and appear in a "window" clearly outlined in one corner of the box. He then answers only the question corresponding to the color of the bead in the window. Thus, the number of beads of each color determines the probability of selecting a given question. A diagram of this "box-and-bead" device is shown in Figure 3.1.

For Abernathy et al's study of induced abortion rates, the sensitive statement for one sample was, "I was pregnant at some time during the past 12 months and had an abortion which ended the pregnancy," while the sensitive statement for the other sample was, "At some time during my life I had an abortion which ended the pregnancy." For both samples, the non-sensitive statement (for which there was an independent estimate) was, "I was born in the month of April."



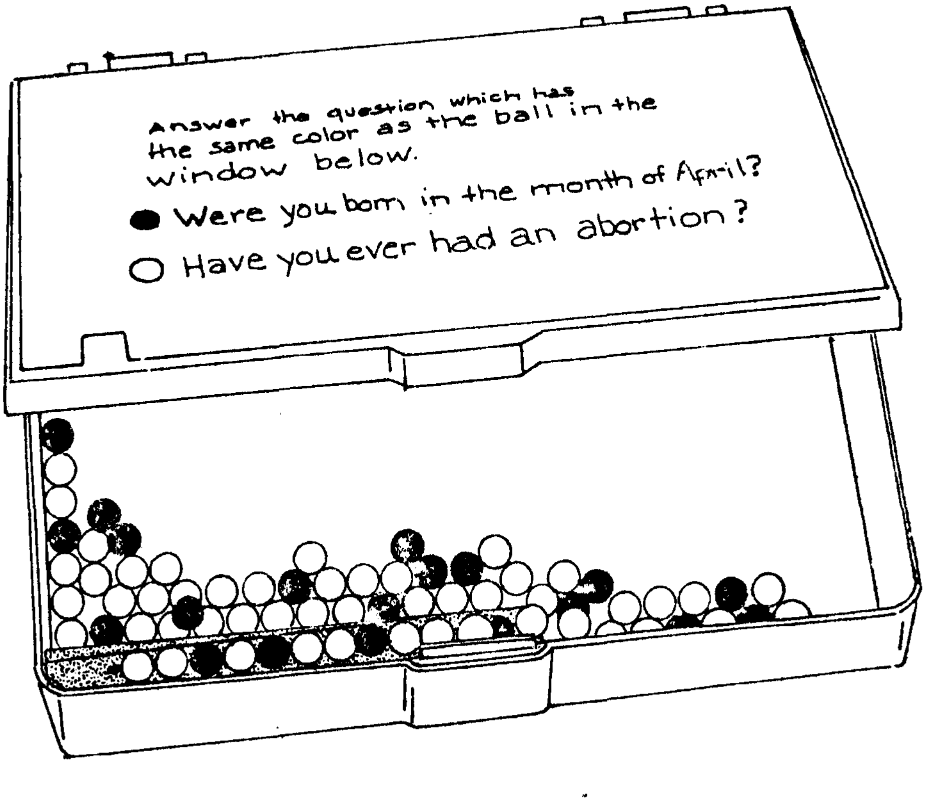


Figure 3.1. The "box-and-bead" randomizing device.

The randomizing device employed in each sample contained 35 red and 15 blue beads, where the red beads were associated with the sensitive statement.

Prior to the randomized response portion of the interview, steps were taken to establish rapport between respondent and interviewer. The randomized response technique and rationale behind its use was explained, along with the necessity of field testing. The interviewer also provided an opportunity for the respondent to familiarize herself with the randomizing device.

Results of the investigation were used to compute estimates of the proportion of women having an abortion either during the past year or during their lifetime. These estimates were presented for the total population, as well as by race, marital status, age category and educational level. While there were no direct data with which to compare these results (and hence the effectiveness of the survey technique), the estimates were generally in line with available rates reported in other studies on abortion.

In addition, an indication of the acceptance of this survey technique was obtained from questioning the respondents themselves. Shortly before the randomized response portion of the interview, respondents were asked whether or not they felt one of their friends would answer truthfully if asked by an interviewer if she had ever had an abortion. At the conclusion of the interview, respondents were questioned whether they believed other people would think there was a trick to the box or that the interviewer could know which question was being answered. Responses to these questions generally showed a remarkable degree of faith in the technique, and clearly indicated the utility of employing such a technique when seeking information on a topic as sensitive as abortion. In response to the preliminary question on whether a friend would truthfully answer a direct question on abortion, 67 percent reported "no," 17 percent reported "yes," and 16 percent were undecided. Regarding the post-randomized response questions, 60 percent of the respondents felt that the randomized approach would not reveal their personal situation, 20 percent felt it would, and another 20 percent were undecided. The authors concluded that these results "clearly indicate that the randomized response procedure is a worthwhile tool in the hands of the survey designer."

Finally, it should be noted that this particular study on abortion was actually part of a large-scale 1968 field survey, which yielded data on several additional topics, including oral contraceptive use and emotional problems. These findings are reported in Greenberg, Abernathy, and Horvitz (1970).

Krótki and Fox (1974).

A comprehensive fertility study which replicated many aspects of the Greenberg et al (1970) study, was conducted by Krótki and Fox (1974) at the University of Alberta. The 1,045 women surveyed were either questioned in a standard interview situation, questioned using a randomized response procedure, or given an anonymous mail-back questionnaire. The questions covered such topics as abortion, illegitimacy, pre-marital sexual intercourse and use of contraceptives. In addition, opinions concerning use of the RRT were ellicited from those interviewed with that approach, using basically the same set of questions as in the North Carolina field survey.

Results revealed that there was "a greater tendency for women to report 'sensitive' events if the questions were randomized than if they are part of a self-administered questionnaire, and a greater tendency in both of these cases than in the standard interview situation." The response rate for the RRT was also substantially higher than for the mail-back questionnaire. This finding might be expected, however, since those interviewed represented a "captured audience."

In response to the opinion questions on the RRT, the Alberta residents were more doubtful than the North Carolinians about the legitimacy of the technique (only 58 percent were sure that their privacy had been guaranteed), and generally more willing to answer the sensitive questions directly (e.g., some 68 percent thought their friends would respond to a direct question on abortion). In fact, the authors noted that one perceptive interviewer had reported that "only a few of her interviewees did <u>not</u> indicate the question they were answering in the RRT 'game'". These findings led the authors to reduce their estimation of the importance of the randomized response method.

<u>Liu, Chow and Mosley (1975).</u>

As discussed in Chapter II, the Hopkin's Randomizing Device is readily applicable to questions involving grouped quantitative data. To test the field use of the device, Liu <u>et al</u>. (1975) recently conducted a (limited) field survey directed at estimating the number of partners, other than their legally married spouse, with whom a group of 34 male and female graduate students had had sexual relations in their lifetime. There were four different color balls in the flask, corresponding to the following four mutually exclusive

1) Those who had never had such an experience.
2) Those who had had such an experience with one partner.
3) Those who had had such an experience with two partners.
4) Those who had had such an experience with three or more partners.

Results were presented, but there was no discussion of their appropriateness or of subjects' reactions to being interviewed with the new technique. In truth, with its small sample base, the only conclusions that could possibly be derived from the study would be feelings of the appropriateness of the technique, based on the reactions of those interviewed. It seems reasonable that the "numerical" results of any randomized response study based on such a small sample size would be of little value since the proportion indeed answering the sensitive question might easily differ considerably from the expected proportion based on the setup of the randomizing device.

Greenberg, Kuebler,
Abernathy and Horvitz (1971).

Greenberg et al. were the first to apply the randomized response technique to obtain quantitative, non-categorical data. Using two samples of respondents and a modification of the unrelated question randomized response technique, they estimated the mean number of abortions in a population of women in urban North Carolina along with the mean income of heads of households.

The questions pertaining to the abortion issue were:

1) How many abortions have you had during your lifetime?

2) If a woman has to work full-time to make a living, how many children do you think she should have?

And the questions on income were:

1) About how much money in dollars did the head of this household earn last year?

2) About how much money in dollars do you think the average head of household your size earns in a year?

Note that, in each of these sets of questions, responses to the non-sensitive and sensitive questions are in the same unit of measure and have roughly equivalent ranges, thereby encouraging respondent cooperation.

Over 900 women were interviewed regarding the abortion issue while an independent sample of 1600 women were asked to respond to the income questions. The randomizing device employed with both samples was the plastic box and bead device described in connection with the Abernathy et al. (1970) study.

Estimates of mean number of abortions and mean income were computed for the sample as a whole and for whites and non-whites separately. The details of the computation procedure are thoroughly discussed in the report (see JASA, 66, p. 244-245). Basically, however, the authors noted that for each topic, the problem was one of statistically separating the two "pure" distributions (corresponding to responses to either the sensitive or non-sensitive questions) from the overall distribution of responses obtained. Once the sensitive and non-sensitive distributions are separated, then it is a "simple" process to calculate the mean and variance of each.

The results of the study appeared reasonable, and reflected expected white - non-white trends. Also, the authors noted that only 1 percent of those contacted for the abortion issue refused to respond, while only 3 percent refused to cooperate with the income interview. These relatively low refusal rates indicate that the randomized response technique might be an effective approach in securing quantitative (as well as qualitative) data in situations requiring subject questioning on such "sensitive" issues as abortions and income.

Applications in Drinking-Driving Area

Gerstel, Moore, Folsom and King (1970).

From 1970-1973, a series of four personal interview surveys on drinking - driving attitudes was conducted as part of a drinking - driving countermeasure program in Mecklenberg County, North Carolina. The surveys were directed at obtaining baseline and follow-up information on variations in drinking - driving characteristics of a representative sample of some 1500 residents. The results of the first of these surveys are documented in Gerstel, Moore, Folsom and King (1970).

For this survey, both direct questioning and randomized response approaches were used to obtain true - false responses to the following two statements:

1) During the past year, I have received one or more citations or tickets for serious traffic violations such as speeding, reckless or drunken driving.

2) During the past year, I have driven an automobile within an hour after having as many as four alcoholic drinks.

The particular randomized response technique employed was the "built-in unrelated question technique" suggested by Morton as cited in Greenberg et al. (197). The randomizing device used was the plastic box-and-bead device which, in this study, contained 35 red beads, 6 white beads, and 9 blue beads. The red beads corresponded to the sensitive statement, while the white beads corresponded to the (true) statement, "This bead is white," and the blue beads corresponded to the (false) statement, "This bead is white".

A comparison of the direct question and randomized response reporting rates revealed that respondents were more willing to admit that the two sensitive statements were characteristic of them if questioned with a randomized response technique. This was especially so for males, where percentage increase in reporting jumped from 11.8 to 20.5 percent, and from 20.4 to 35.0 percent for the serious violation and driving-after-drinking questions, respectively. Thus, the randomized response technique appears to have had a positive effect on truthful reporting for this study.

Folsom, Greenberg, Horvitz, and Abernathy (1973).

Folsom, et al. (1973) report on a 1971 field application of the two alternate questions randomized response model, also conducted in conjunction with the drinking - driving surveys in Mecklenberg County, North Carolina. Using as the sampling frame the population of drivers 16 years of age and older who were not alcohol abstainers, the authors used the two alternate question randomized response technique to estimate the percentage of drivers responsible for an automobile accident during the previous year.

Each person interviewed was asked to toss a coin, and, if the coin landed "heads up", to respond to the sensitive statement or, if it landed "tails up", to respond to one of two non-sensitive statements. The sensitive statement was, "I had an automobile accident during the past year in which I was at fault", while the non-sensitive statements were:

1) I was born in the month of April.
2) I lived in North Carolina but not in Mecklenberg County in 1966.

The response rate for the over 400 individuals contacted was virtually 100 percent. Results were presented, but no discussion of the appropriateness of the estimates or problems with the technique was included. This is perhaps not unexpected since the study was intended primarily to be an application rather than a test of the technique.

Folsom (1974).

Encouraged by the apparently successful application of the randomized response technique in the Mecklenberg County drinking and driving attitude surveys, Folsom (1974) designed a randomized response validation study. This study was aimed at comparing direct questioning and randomized response reporting rates for driving under the influence of alcohol.

The randomizing device employed was a plastic box containing red, white and blue beads. The red beads corresponded to the sensitive question, "Have you been arrested for driving under the influence of alcohol during the past year?". The white and blue beads both corresponded to the non-sensitive question, "Is the bead in the window blue?".

Knowledge of π_Y was incorporated into the randomizing device, and thus the particular randomized response design being tested was the "built-in unrelated question RRT."

There were two different randomizing devices, corresponding to two different proportions of red, white and blue beads. In the "blue" randomizing device, there were 35 red, 4 white and 11 blue beads, while in the "yellow" randomizing device, the number of blue and white beads were reversed so that there were 35 red, 11 white and 4 blue beads.

The target population for the study were persons residing in the surrounding three county area who had been arrested for driving under the influence (DUI) during the previous eight months. In addition, some 90 individuals who had not been arrested for DUI were included to give a total sample size of 900. Each sample member was assigned to one of three groups, according to whether he would be questioned directly (16.7 percent), questioned using the "blue" randomizing device (33.3 percent) or questioned using the "yellow" randomizing device (50.0 percent).

For each respondent questioned using the randomized response technique, a "trail" run preceded the actual test run, and both responses were recorded. Based on the proportion of beads in each of the randomizing devices, the probability that two "yes" responses indicated at least one affirmative answer to the sensitive question was 0.95 for the "blue" device, and 0.99 for the "yellow" device.

Results from 283 individuals who eventually responded to the DUI question failed to support use of the randomized response approach over direct questioning for the particular population/topic of this study. Using the driving record for validation, Folsom found that 84.0% of those directly questioned correctly reported that they had been arrested for DUI over the past year, but (assuming that the randomizing device functioned properly) only 71.6% of those questioned with one of the randomized response methods reported their true situation.

After discussing several factors that might have contributed to such an outcome, the author concluded that efforts to optimize the precision of the randomizing device (e.g., by recording response on trial runs, using a high proportion of beads corresponding to the sensitive question, etc.), along with limiting its use to a single very sensitive area may have seriously jeopardized the credibility of the method.

Application in Drug Usage Area

In contrast, a recent study by Goodstadt and Gruson (1975) does appear to support use of a randomized response approach over direct questioning. This study was directed at estimating usage of various drugs by high school students, using both traditional direct questioning and the two sample unrelated question randomized response approach (π_Y not known).

Questionnaires presenting a series of drug-related questions were randomly distributed to over 800 high school students in Ontario, Canada. The stigmatizing question asked the number of times over the past months the student had used each of six different drugs. The corresponding non-sensitive question asked the number of times during the same period the student had (1) watched TV, (2) visited a library, (3) visited a museum, or (4) attended a play, (5) a rock concert, or (6) a classical concert. Approximately half of the students were asked to respond directly to the drug usage question. The remaining students were divided into two groups and asked to use the last digit of their telephone number as a randomizing device for determining whether they would answer the sensitive or non-sensitive question. For one group, the sensitive question was answered if this number was a 0, 1 or 2, and the non-sensitive question answered if it was between 3 and 9. For the second group, the instructions were reversed, so that students whose telephone number ended in a digit between 3 and 9 answered the sensitive question, while those whose number ended in a 0, 1 or 2 answered the non-sensitive question.

The results from 840 completed questionnaires supported the authors' hypothesis that estimates of reported drug use derived from the randomized response procedure would be "significnatly higher" than those obtained by the traditional direct questioning approach. Not only did more students report using drugs, but they also reported more frequent drug use when questioned with the randomized response technique. Moreover, there was a much lower refusal rate among those questioned with the randomized response approach.

Results of the survey generally did not support a second hypothesis, namely, that the randomized response procedure would prove even more effective when asking about the more socially sensitive drugs (e.g., hallucinogens as opposed to alcohol). The authors concluded that self-reported drug use may be significantly underestimated when standard inquiry procedures are utilized.

Summary

In the field applications reviewed, the randomized response technique was used to obtain information on illegitimacy and abortion rates, use of the contraceptive pill, emotional problems, income levels, pre- and extramarital sexual behavior, frequency of serious traffic violations (including driving after drinking), accident liability, DUI (driving under the influence of alcohol) arrests, and drug usage. In addition, randomized response techniques have been used to investigate such topics as organized crime, voting behavior, and unreported deaths.

Results of these studies seem to indicate that randomized response techniques can be effective tools for obtaining information on such sensitive topics. However, this effectiveness is clearly dependent on a number of factors. These include the particular topic being investigated, the population involved and the randomized response technique employed. Also important are the choice of relevant parameters (including values of p, sample size allocations, etc.), the selection of the randomizing device (e.g., beads, cards or coin toss), and the choice of the interviewing procedures (including any explanation of the RRT, justification for its use, etc.).

Generally, when compared with direct questioning, randomized response procedures seem most effective when used to investigate the more sensitive topics. Also, it seems that the simplicity of the randomizing device and the readiness of its acceptance by the respondent are key determinants of the effectiveness of the technique.

One of the greatest disadvantages of the RRT, at least at this stage of its development, is its dependence on "face-to-face" interviewing situations. This dependence decreases the cost effectiveness of the technique when compared with more standard interviewing procedures which can operate via telephone, mail, etc.

Clearly, if a direct questioning procedure can be used with sufficient accuracy, then it should be used -- not a RRT. The RRT was designed for use in situations where direct questioning could not be used, i.e., in sensitive areas where it is impractical to ask the question directly. In this respect, it is not really in competition with direct questioning.

Along this line, it is conceivable that as society becomes more and more "open" regarding certain issues, there would be less need for interviewing techniques such as the RRT. Indeed, this was suggested in the Alberta Families Study (Krótki and Fox, 1974). Thus, a number of years ago a randomized response approach might have been the only practical way of obtaining information on, say, illegitimate births. Now, however, the direct questioning method seems quite sufficient for many such issues.

In way of summary, research on the RRT has demonstrated that the technique can be successfully applied. However, success is by no means guaranteed, as evidenced by several of the studies reviewed. If the RRT is to be applied extensively in future survey situations, then additional field research is needed to refine the technique. Presently, there is a great deal of "uncertainty" associated with the RRT, which limits its usefulness.

The concluding chapter of this report describes HSRC's attempt to further extend the application of the RRT in the area of highway safety, by using the technique primarily to obtain information on seat belt usage.

IV. HSRC'S APPLICATIONS OF THE RANDOMIZED RESPONSE TECHNIQUE

HSRC employed the randomized response technique in conjunction with an investigation of safety belt usage and effectiveness. In order to determine the true effectiveness of safety belts in reducing injuries from accidents, one must have reliable information on both seat belt usage and level of injury. In the past, measures of belt usage and injury have most frequently been obtained from the police accident report files. This data source is not entirely satisfactory, however, since it has been shown that independent investigations of belt usage and of injury level for a given sample of accident victims can yield widely discrepant results. Clearly, failure to account for these misclassification errors when using police report information will produce erroneous measures of seat belt effectiveness.

An alternative to using police report information for measures of belt usage is to contact each individual in the accident sample and ask him if he was wearing his seat belt at the time of the accident. If the sample is quite large, however, this approach is too time consuming and costly. Also, there is the matter of assuring that those contacted are responding truthfully to the inquiry.

If belt usage by accident victims is indeed a relatively sensitive issue (perhaps because of insurance implications), it would seem reasonable to utilize a randomized response approach at least on a trial basis. This can be done on a relatively small but representative sample of the accident population. If successful, the results can then be used to determine an improved estimate of belt usage for the entire population. This process of estimating a Bernoulli parameter from a sample of misclassified responses (the accident report information) and a subsample of randomized responses is described in detail in Appendix B.

Following are descriptions of three independent efforts by HSRC to utilize the randomized response technique to examine errors in seat belt usage reporting as found on the accident report form, and to demonstrate the usefulness of the technique as applied to a highway safety problem. The initial plan was to test the technique on a small sample and then, if this proved successful, to conduct a randomized response study of greater magnitude to allow for adjusting seat belt usage data for the entire accident population under study. Accordingly, the initial subsamples selected were relatively small and not necessarily representative. Unfortunately, as will be seen, the results of the various pilot studies by no means warranted further extension of these efforts!

Randomized Response Pilot Study #1

As stated earlier, the three randomized response pilot studies were directed at comparing police-reported belt usage with belt usage rates obtained using a RRT for selected samples of drivers who had recently been involved in accidents. The sample for the initial pilot study consisted of some 108 drivers in the Chapel Hill and Raleigh-Durham areas of North Carolina. The particular randomized response technique employed was the unrelated question technique (π_y not known). The randomizing device was primarily the box-and-bead device used in the RTI studies, although a deck of cards was used as an alternative for a number of respondents.

This initial pilot study was afflicted by several practical problems. One of these was locating the materials needed to assemble the plastic box-and-bead randomizing devices. Several such devices were obtained from RTI for use in the pilot study. Anticipating a larger, follow-on study, efforts were made to secure the necessary ingredients (clear plastic boxes, anti-static beads in two colors, wooden ramps). This proved more than a minor task. Perhaps, fortunately, the results of the pilot surveys using the RRT in the area of belt usage among accident victims obviated the need of obtaining a sizable number of these box-and-bead devices.

The major problem afflicting this first pilot study was the extreme difficulty experienced in contacting the 108 sample individuals for an interview. This might have been anticipated, as these were people involved in accidents! The names and addresses of these individuals, along with information on seat belt usage, was obtained from recent accident report forms supplied by the local police departments. The plan was to contact these individuals via telephone to arrange a personal ("face to face") interview. The following phone conversation served as a guide:

> "Hello (name of person contacted). I am (interviewer's name) with the University of North Carolina, Highway Safety Research Center. In cooperation with the Chapel Hill Police Department, we have been given the names of people who were recently involved in accidents in this area. As part of a pilot survey, we are experimenting with a visual questionnaire technique for the National Highway Traffic Safety Administration. We are interested in obtaining accurate information about accidents. It would be very helpful if I could see you for about 10 or 15 minutes sometime in the next day or two to answer these questions. Can you suggest a good time for us to meet? ....
> Good. May I have directions to find you? ....
> Let me give you my name again along with my telephone number in case you have a change of plans. Thank you."

If an individual's name was not listed in the directory, as is often the case especially in university towns, the following alternatives were utilized:

1) Call information operator.
2) For women, check the directory for the same street address to obtain husband's or father's telephone number.
3) Call registered owner of vehicle, even if not the driver.

Once a telephone number was obtained, several attempts to reach the individual were made at different times of the day. If a party was successfully contacted, an interview was arranged for the earliest convenient time.

Initially, only those individuals in the immediate Chapel Hill area were to be contacted. However, this area is characterized by a disproportionately large number of students who proved to be especially difficult to reach and, once contacted, unreliable in keeping interview appointments. As a result, the pilot study was extended to include individuals residing in the Raleigh-Durham area as well. Unfortunately, results here were not much better!

Out of the total of 108 potential respondents, 71 or 65.7% could not be reached. Another 15 (13.9%) were contacted, but for various reasons never interviewed. Only 22 individuals, or roughly 20% of the original sample, were successfully contacted and interviewed, and 7 of these could only be interviewed over the phone. These results are presented with some further breakdowns in Table 1.

The "sensitive" question for those interviewed was, "Were you wearing your seat belt during your recent accident?" When asked over the phone, this question was always prefaced by the statement, "You don't have to answer this particular question if you don't want to", and no randomizing technique was employed.

Table 4.1. HSRC Pilot Study #1

Respondents:		
Not reached	71	(65.7%)
No listing or number not in service	35	
Incorrect or unpublished number	6	
Busy, not at home	30	
Reached, but not interviewed	15	(13.9%)
Unable to schedule interview (including refusals)	11	
Interview scheduled but subject did not show	4	
Interviewed	22	(20.4%)
Using randomizing device	15	
Directly questioned over phone	7	
Total	108	

For the 15 individuals who were actually interviewed using the RRT, 5 chose to use the box-and-bead device, while 10 selected the card device. For both devices, the probability of selecting the belt usage question was the same. For the box-and-bead device, there were 30 red and 10 blue beads (with red corresponding to the belt usage question), and for the card device, the belt usage question appeared on 3 out of every 4 cards. Thus the probability of selecting the critical question with either device was 0.75.

Since this first pilot study was intended only as a test of the method, no formal analysis of the results was attempted. Indeed, none could have been made, since the single small sample did not allow for an estimate of the population proportion with the non-sensitive characteristic.

Of some value, however, were the reactions and comments elicited by the randomized response technique. The prevailing attitude of those interviewed seemed to be, "If you want to know whether or not I was wearing my seat belt, why don't you just ask me?" Respondents generally did not seem to understand why the randomized response technique was being used, and some even thought it "ridiculous" that the interviewer would go to such lengths to allow them to "shake the plastic box" or "pick a card".

Thus, the general indication from this initial pilot study was that seat belt usage was not a sensitive enough issue to warrant use of the randomized response technique. Before abandoming the project, however, HSRC decided to try a considerably different and less expensive approach to using a RRT to obtain estimates of seat belt usage. These efforts are recorded in the following section.

Randomized Response Pilot Study #2

The second randomized response pilot study was designed to overcome some of the difficulties encountered in the initial investigation. Perhaps the greatest of these difficulties was making the initial contact with the person to be interviewed. This was because only the person's name and address was included on the accident report form, and correct telephone numbers were very difficult to obtain. Accordingly, it was decided to avoid the situation entirely by devising a randomized response technique that could be carried out through the mail. Furthermore, since the issue of seat belt usage did not seem to be sufficiently stigmatizing to warrant use of a randomized response technique, two additional questions pertaining to speeding and drinking were included.

In order to allow for a trichotomous response to these issues, a two-stage randomized response scheme (TSRRS) was needed (see Appendix A). A penny was selected as the randomizing device, since pennies are readily available, cheap, and easily distributed through the mail.

A copy of the entire questionnaire for this second randomized response pilot study is included as Appendix C. These questionnaires were mailed to a sample of 204 drivers in the Triangle area (Chapel Hill, Raleigh, Durham) who had been involved in accidents during the previous month. The names and addresses of these individuals were obtained from public records maintained by the Department of Motor Vehicles. Only those individuals for whom information on seat belt usage was available were included in the sample.

Response to the efforts of this second pilot study was likewise very discouraging. Out of the 204 questionnaires mailed, only 37, or

18.1% were returned. Furthermore, only 9 of these, or less than 25% of those returned, were evidently completed correctly. Eighteen, or approximately half of the individuals who responded to the survey, gave direct responses to the questions (i.e., they did not follow directions). Indications of such direct responses included underscoring the question being answered, writing out the response instead of marking the appropriate space, drawing lines to the question answered, and various written comments. In addition, 2 of the 37 questionnaires were returned blank, and another 8 included a mixture of "correct", direct, and blank responses.

Perhaps more enlightening were the comments offered by the respondents. Ten respondents gave their own account or description of their accident. Seven indicated that they were confused by the questionnaire, or thought it rather "stupid" or "ridiculous". Another three respondents expressed hostility towards the questionnaire, with one even returning the the penny! And finally, seven individuals questioned the purpose of the questionnaire and/or use of the randomized response technique.

Generally, it seemed particularly difficult for respondents to accept that HSRC truly was interested in the overall distribution of the responses to the three questions (from which the desired estimates could be derived) and not in specific replies to the questions or actual accounts of what happened. Clearly, the randomized response technique is not easily explainable by mail. Use of the two stage technique undoubtedly contributed to the misunderstandings in this particular study.

With these lessons learned, HSRC planned and executed one final pilot study of the randomized response technique.

Randomized Response Pilot Study #3

For this final pilot study, every effort was made to make the randomized response questionnaire as simple and clear as possible. Only a single question on seat belt usage was included, and a simple one-stage design was utilized. In addition, the cover letter specifically requested that the respondent follow the directions given, even if he did not mind answering directly. See Appendix D for a sample questionnaire.

In an attempt to make the issue of seat belt usage more sensitive, and hence more applicable to use of a randomized response technique, only individuals driving 1974 model cars at the time of the accident were included in the sample population. It was thought that these individuals might be more reluctant to admit not wearing their belt, since this would indicate that the interlock system had been disconnected. A statewide sample of 63 such drivers was obtained from the North Carolina Department of Motor Vehicles.

Since HSRC was interested in determining if respondents were following the randomized response directions, as well as if individual reports of belt usage differed from police reports, a method was devised to allow a check on this without requiring respondents' names on the questionnaires. Instead of using names, each individual was uniquely coded according to the particular size and color paper of his questionnaire. Thus, for each form returned, it was known whether the month or birth or seat belt question should have been answered, and whether police had reported "belt" or "no belt" for that individual.

Unfortunately, results of the pilot study were only slightly more encouraging then results of the previous studies. Of the 63 questionnaires mailed, only 16, or 25.4%, were returned. Provided directions were correctly followed, 12 of these were responses to the belt usage question, and 4 responses to the non-sensitive month-of-birth question. Regarding the belt usage question, the driver report (i.e., response on questionnaire) agreed in all cases with the police report (4 belt wearers - 8 non-wearers).

Regarding the 5 responses to the month-of-birth question, one individual responded incorrectly, indicating that he probably was giving a direct response to the seat belt question. While this was the only case of clear misunderstanding of the randomized response technique, 4 other individuals did indicate the question to which they were responding, thus showing some misunderstanding of the instructions.

In conclusion, it appears from this and the two pilot studies previously described that the randomized response technique cannot be effectively used to determine safety belt usage of individuals involved in accidents. One primary reason is that the issue of belt usage does not appear to be sufficiently sensitive to warrant use of such a sophisticated technique. Also, there is the added difficulty of locating the designated individual or, if conducting the study via mail, adequately but concisely explaining the technique. Thus, while the randomized response technique may be effectively applied in other areas of highway safety, it would not appear to be effective in the area of seat belt investigations.

V. DISCUSSION AND RECOMMENDATIONS

The original intent of HSRC's investigation of the randomized response technique was to apply the technique to an appropriate subsample of accident-involved individuals, in order to investigate classification and response errors and biases in estimating belt usage rates and injury reduction capabilities. To this end, three pilot studies were conducted, and some additional theory developed (see Appendices A and B). However, as has been noted, none of the pilot studies warranted a full-scale extension.

A lack of sensitivity in the subject matter (i.e., seat belt usage) was obviously a key reason for this failure. But even if the subject matter had been appropriate, the successful application of the technique would still have been severely restricted by the tremendous difficulty in contacting the subjects and obtaining their responses. This same problem was cited by Folsom (1974) in his report on DUI arrests, and indeed is characteristic of human surveys where a re-designated population must be contacted.

Primarily for this reason, HSRC feels that further development of the RRT might best be directed at extending it beyond the "face-to-face" encounter situation, so that it can be adequately explained and applied over the telephone, via mail, etc. While there are certain research situations that may require face-to-face interviews, the RRT would have much greater applicability if it could be effectively applied on the larger scale made possible by mail questionnaires, telephone surveys, and the like. Not only would such an extension enable more subjects to be contacted, but it would do so at reduced costs and time.

The final application described in Chapter II, a 1975 study of student drug usage by Goodstadt and Gruson, was a step in this direction. In this apparently successful field study, a RRT was collectively applied to a group of subjects, each provided with "a detailed example of the completed procedure with instructions." The randomizing device (last digit of the subject's telephone number) was appropriately selected for self administration. Although there was someone present to administer the questionnaire, each subject could complete the form on his own, provided he could read and follow directions.

Obviously, an advantage in the above study is that the subjects were part of a "captured audience." Particularly in the field of highway safety research, this will seldom be the case. However, it is conceivable that a similar questionnaire could be successfully administered over the phone or via mail.

A problem that then arises, though, is that of motivating the subjects to respond. Certainly, this was a very real problem for the two HSRC mail questionnaire studies. Thus, there are at least three additional parameters which must be considered in extending the RRT for use through the public media . These include:

1) Producing an adequate verbal or written explanation of the technique, including directions that are relatively easy to follow.

2) Selecting an appropriate randomizing device (i.e., one that is inexpensive, readily available and easy to apply, as well as not loaded against the respondent).

3) Providing sufficient motivation for subjects to respond (especially when relying on the mail).

In summary, HSRC recommends that further research on RRT be directed at its practical application, with emphasis on extending it beyond the usual face-to-face interview situation. While the technique may have the most potential in certain areas of social and psychological research, HSRC does not feel that the technique is too promising in the field of highway safety research. The issues in this field are generally not of a highly sensitive and personal nature, and when human surveys are called for, a direct questioning technique would appear to prove more effective.

REFERENCES

Abernathy, J.R., Greenberg, B.G. and Horvitz, D.G. (1970). Estimates of induced abortion in urban North Carolina. Demography, 7, 19-29.

Abul-Ela, A.A., Greenberg, B.G., Horvitz, D.G. (1967). A multi-proportions randomized response model. Journal of the American Statistical Association, 62, 990-1008.

Boruch, R.F. (1972). Relations among statistical methods for assuring confidentiality of social research data. Social Science Research, Vol. 1, 403-414.

Bross, I. (1954). Misclassification in 2 x 2 tables. Biometrics, 10, 478-486.

Cochran, W.G. (1965). Sampling Techniques. New York: John Wiley and Sons, Inc.

Diamond, E. and Lilienfeld, A. (1962). Effects of errors in classification and diagnosis in various types of epidemiological studies. American Journal of Public Health, 10, 2106-2110.

Folsom, R.E. (1974). A randomized response validation study: Comparison of direct and randomized reporting of DUI arrests. Research Triangle Institute Final Report (#25U-807), Research Triangle Park, NC.

Folsom, R.E., Greenberg, B.G., Horvitz, D.G. and Abernathy, J.R. (1973). The two alternate questions randomized response model. Journal of the American Statistical Association, 68, 525-530.

Gerstel, E.K., Moore, P., Folsom, R.E. and King, D.A. (1970). Mecklenberg County Drinking-Driving Attitude Survey, 1970. (DOT-FH-11-7538) Report prepared by Statistics Research Division, Research Triangle Park, NC.

Goodstadt, M.S. and Gruson, V. (1975). The randomized response technique: A test on drug use. Journal of the American Statistical Association, 70, 814-818.

Greenberg, B.G., Abernathy, J.R. and Horvitz, D.G. (1970) A new survey technique and its application in the field of public health. Milbank Memorial Fund Quarterly, 48, 39-55.

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, 520-539.

Greenberg, B.G., Horvitz, D.G. and Abernathy, J.R. (1974). A comparison of randomized response designs. Reliability and Biometry, Siam, Philadelphia, 787-815.

Greenberg, B.G., Kuebler, R.R. Jr., Abernathy, J.R. and Horvitz, D.G. (1971). Application of the randomized response technique in obtaining quantitative data. Journal of the American Statistical Association, 66, 243-250.

Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969). Analysis of categorical data by linear models. Biometrics, 25, 489-504.

Koch, G.G., Imrey, P.B. and Reinfurt, D.W. (1972). Linear model analysis of categorical data with incomplete response vectors. Biometrics, 28, 663-692.

Krótki, K.J. and Fox, B. (1974). The randomized response technique, the interview, and the self-administered questionnaire: An empirical comparison of fertility reports. American Statistical Association Proceedings of the Social Statistics Section, 367-371.

Liu, P.T., Chow,L.P. and Mosley, W.H. (1975). Use of the randomized response technique with a new randomizing device. Journal of the American Statistical Association, 70, 329-333.

Moors, J.J.A. (1971). Optimization of the unrelated question randomized response model. Journal of the American Statistical Association, 66, 627-629.

Tenebein, A. (1970). A double sampling scheme for estimating from binomial data with misclassifications. Journal of the American Statistical Association, 65, 1350-1361.

Tenebein, A. (1971). A double sampling scheme for estimating from binomial data with misclassifications of sample size determination. Biometrics, 27, 935-944.

Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60, 63-69.

Warner, S.L. (1971). The linear randomized response model. Journal of the American Statistical Association, 66, 884-888.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 57, 348-368.

APPENDIX A

TWO STAGE RANDOMIZED RESPONSE SCHEMES FOR ESTIMATING A MULTINOMIAL*

Introduction

Warner's (1965) original randomized response technique for dichotomous data was extended by Abul-Ela et al. (1967) to the case of a multinomial with $t \geq 3$ groups, r of which are stigmatizing. A competing approach for estimating a multinomial via the randomized response technique is discussed by Warner (1971). Yet a third approach for the multinomial is considered by Greenberg et al. (1969) and is based on using an alternate independent question.

The purpose of this paper is to outline some alternative schemes for estimating the t group proportions when $r < t - 1$, using only one sample. Their realizations for any sampled individual constitute two-stage schemes. The second stage is conditional on the random individual's response in the first stage.

First, the three existing models are briefly summarized to provide a reference for development of the new procedures and resulting estimators. These new schemes are then identified as special cases of Warner's (1971) general linear randomized response model, thereby yeilding alternative estimators based on modified generalized least squares method due to Zellner (1962).

* Published in Communications in Statistics, 4 (11), 1021-1032 (1975) by Yosef Hochberg.

Summary of Three Existing Models

(1) The Approach in Abdel-Latif A. Abul-Ela et al. (1967).

Let π_i be the population proportion of the i-th group, $i = 1,\ldots,t$, $\sum_{i=1}^{t} \pi_i = 1$. Put $\underset{\sim}{\pi}$: $(t-1) \times 1 = (\pi_1,\ldots,\pi_{t-1})'$. This approach uses t-1 samples as follows. A matrix (p_{ij}): $(t-1) \times t = (\underset{\sim}{p}_1,\ldots,\underset{\sim}{p}_t)$ has to be determined such that $\sum_{j=1}^{t} p_{ij} = 1$ for $i = 1,\ldots,t-1$ and $\underset{\sim}{P}$: $(t-1) \times (t-1) = (\underset{\sim}{p}_1-\underset{\sim}{p}_t,\ldots,\underset{\sim}{p}_{t-1}-\underset{\sim}{p}_t)$ is non-singular. The p_{ij}'s determine the randomized scheme here, where p_{ij} is the chance for an individual in the i-th sample to randomly select the question: "Are you in group j?", $i = 1,\ldots,t-1$; $j = 1,\ldots,t$. Let n_i denote the i-th sample size, n_{i1} denote the number of 'yes' responses in the i-th sample, and put $\underset{\sim}{d} = \left(\frac{n_{11}}{n_1},\ldots,\frac{n_{t-1,1}}{n_{t-1}}\right)'$. It is easily verified that the MLE's of the π_i's are given by

$$\hat{\underset{\sim}{\pi}} = \underset{\sim}{P}^{-1}\underset{\sim}{c}, \quad \hat{\pi}_t = 1 - \sum_{i=1}^{t-1} \hat{\pi}_i$$

where $\underset{\sim}{c} = \underset{\sim}{d} - \underset{\sim}{p}_t$.

(2) The Approach in Warner (1971).

This is a special case of Warner's (1971) general 'linear randomized response model.' In the general setup á simple random sample of size m is drawn and for the i-th individual in the sample the realization $\underset{\sim}{x}_i = t \times 1$ of a random vector $\underset{\sim}{x}$ cannot be observed. Rather, one observes $\underset{\sim}{y}_i$: $q \times 1 = \underset{\sim}{T}_i\underset{\sim}{x}_i$, $i = 1,\ldots,m$, where the $\underset{\sim}{T}_i$: $q \times p$'s are random independent matrices and independent of the $\underset{\sim}{x}_i$'s. Assuming that the expectations of the $\underset{\sim}{T}_i$'s are known, the problem is that of estimating the expectation vector of $\underset{\sim}{x}$. Warner (1971) discusses generalized least squares estimates for the mean vector of $\underset{\sim}{x}$. For the problem of estimating a multinomial Warner proposes the following scheme. Let $\underset{\sim}{x}_i = (x_{i1},\ldots,x_{it})'$, where $x_{ij} = 1$ or 0 according as the i-th individual belongs to the j-th group or not. Here q = t, where $\underset{\sim}{T}_i$ for all i=1,...,m, is the random t x t matrix whose possible configurations are formed by permuting the columns of the t x t identity matrix. Then the random transform amounts to directing the individual's response according to which group he is in depending on the random (unknown to the interviewer) realizations of the $\underset{\sim}{T}_i$'s.

(3) <u>The Approach in Greenberg et al. (1969)</u>.

This approach is very similar to the first. It is obtained from the first by giving a zero chance for an individual in any of the samples to be faced with the question: "Are you in group t?", and instead replaces this question by an alternate question, "Do you possess characteristic Y?", where Y is unrelated to the characteristic by which $\pi_1,\ldots,\pi_t$ were formed. This is a modification of the dichotomous unrelated question randomized response model. If π_Y (proportion in population of individuals with characteristic Y) is known, only t - 1 samples are necessary to estimate the π_i's; otherwise, t samples are required.

Two observations follow:

1. These three procedures are all quite involved. First, in all of them inversion of matrices must take place. Also, the task of choosing the p_{ij}'s and n_i's in first and third methods and the chance probabilities attached to the various realizations of the $\underset{\sim}{T}_i$'s in the second method is difficult since no guiding theory exists.

2. In all these schemes there is a loss in efficiency resulting from their low sensitivity to the relation between r and t. Clearly, good randomized response schemes give protection to individuals in stigmatizing groups while minimizing protection (i.e., removing uncertainty) for individuals in non-stigmatizing groups. The ability to design such efficient schemes depends very much on the relations between r and t. This is clarified in the next section where two new schemes are outlined, with a double stage interview of an individual where the second stage depends on the random individual's response in the first stage. Thus, these procedures will be referred to as: Two Stage Randomized Response Schemes or -- TSRRS.

Two Stage Randomized Response Schemes (TSRRS)

The case $r = 1$ is treated first, and without loss in generality the first group is assumed to be the stigmatizing one.

Scheme 1. Take a simple random sample of n individuals. Use a randomization device which gives a chance p for an individual to be faced with the question: "Are you in group 1?" and chance $1 - p$ to be commanded to say the word "yes." This is the first stage. All individuals who answered "no" in this stage, say, n_o of them, are directly asked in the second stage: "In what group are you?" since, clearly, they don't belong to the stigmatizing group. All other $n - n_o$ individuals who answered "yes" are protected. For these individuals there is no second stage. Let n_i denote the number of individuals among the n_o who responded "no" in the first stage, who identified themselves as belonging to group $i = 2,3,\ldots,t$. Let λ denote the probability of a "no" response in the first stage.

$$\lambda = p(1 - \pi_1)$$

The MLE estimate of λ is $\frac{n_o}{n}$ and thus the MLE of π_1 is

$$\hat{\pi}_1 = 1 - \frac{n_o}{np}$$

Now, conditional on n_o $(n_2, n_3, \ldots, n_t)'$ where $\sum_{i=2}^{t} n_i = n_o$ is a multinomial with cell probabilities

$$\frac{\pi_2}{1-\pi_1}, \frac{\pi_3}{1-\pi_1}, \ldots, \frac{\pi_t}{1-\pi_1} .$$

Thus the estimates of the population proportions of the non-stigmatizing groups are given by

$$\hat{\pi}_i = \frac{n_i}{n_o}(1-\pi_1) = \frac{n_i}{pn}, \quad i = 2,3,\ldots,t.$$

These estimates are unbiased

$$E[E(\hat{\pi}_i | n_o)] = E\left[\frac{\pi_i n_o}{(1-\pi_1)pn}\right] = \pi_i, \quad i = 2,3,\ldots,t$$

$$E(\hat{\pi}_1) = 1 - \frac{n(1-\pi_1)p}{np} = \pi_1.$$

The variances are given by

$$\text{Var}(\hat{\pi}_1) = \frac{(1 - \pi_1)(1 - p + p\pi_1)}{np}$$

$$\begin{aligned} \text{Var}(\hat{\pi}_i) &= E[\text{Var}(\hat{\pi}_i | n_o)] + \text{Var}[E(\hat{\pi}_i | n_o)] \\ &= \frac{\pi_i(1 - \pi_1 - \pi_i)}{(1 - \pi_1)pn} + \frac{\pi_i^2(1 - p + p\pi_1)}{(1 - \pi_1)pn} \\ &= \frac{\pi_i(1 - \pi_i p)}{pn}, \quad i = 2,3,\ldots,t. \end{aligned}$$

Scheme 2. This is the scheme used in the first mail questionnaire survey reported in Chapter IV. In the first stage, a direct question is presented to all individuals in the sample: "Are you in group 3?, 4?,...,t? or in either 1+2? Let n_1 be the number of those who fell in group i, $i = 3,4,\ldots,t$, and $n_{1,2}$ be the number of those in either the first or second group. All $n_{1,2}$ individuals then undergo a second stage in which a randomized scheme is used in order to estimate $\pi_j/(\pi_1 + \pi_2)$, $j = 1,2$. The second stage is then a conditional randomized response scheme for dichotomous data and one may use any of the available techniques for this stage. The scheme used here is where one chooses p_1,p_2,p_3 for the chances that an individual undertaking the second stage will have to answer the question "Are you in group 1?", will have to say "yes", or will have to say "no", respectively. Clearly, $p_1+p_2+p_3=1$. Here

$$\hat{\pi}_i = \frac{n_i}{n}, \quad i = 3,4,\ldots,t$$

which are the best possible estimates of these unknown quantities. Let m_1 denote the number of individuals who say "yes" in the second stage.

Then, estimate $\frac{\pi_1}{\pi_1 + \pi_2}$ from the relation

$$p_1 \frac{\pi_1}{\pi_1 + \pi_2} + p_2 \hat{=} \frac{m_1}{n_{1,2}} .$$

Thus $\hat{\pi}_1 = \frac{n_{1,2}}{n} \left(\frac{m_1}{n_{1,2}} - p_2\right)/p_1 = \frac{m_1 - n_{1,2}p_2}{np_1}$; $\hat{\pi}_2 = \frac{n_{1,2}}{n} - \hat{\pi}_1$.

The $\hat{\pi}_i$'s for $i = 3,4,\ldots,t$ are clearly unbiased and have minimum variance. The $\hat{\pi}_i$, $i = 1,2$, are unbiased.

$$E[E(\hat{\pi}_1|n_{1,2})] = E\left[\frac{(\frac{\pi_1}{\pi_1 + \pi_2} p_1 + p_2)n_{1,2} - n_{1,2}p_2}{np_1}\right] = \pi_1 ,$$

hence $E(\hat{\pi}_2) = \pi_2$.

$$Var(\hat{\pi}_1) = E[Var(\hat{\pi}_1|n_{1,2})] + Var[E(\hat{\pi}_1|n_{1,2})]$$

$$= \frac{\pi_1 + \pi_2}{n\ p_1^{\ 2}} [p_1 \frac{\pi_1}{\pi_1 + \pi_2} + p_2][1 - p_1 \frac{\pi_1}{\pi_1 + \pi_2} - p_2]$$

$$+ \frac{\pi_1^{\ 2}}{(\pi_1 + \pi_2)n} (1 - \pi_1 - \pi_2)$$

These TSRRS's are now generalized to $r > 1$, $r < t - 1$.

Scheme 1. Without loss of generality, assume that the r stigmatizing groups are the first r. Let p_i be the probability provided by some randomizing mechanism that an individual will be questioned: "Are you in group i?" and $1 - p_i$ be the chance that he will have to say: "yes", $i = 1,\ldots,r$. Let N_o denote the number of individuals in the sample who responded with a "no" to all r questions. These N_o individuals are clearly not in any stigmatizing group and thus a second stage in which they are asked to reveal their group is appropriate. All other $N - N_o$ individuals are protected.

Let λ_i denote the probability of a "yes" response on the i-th question, $i = 1,\ldots,r$, and let o_i be the number of "yes" responses for the i-th question

$$\lambda_i = \pi_i p_i + 1 - p_i \ , \ i = 1,\ldots,r.$$

A possible estimate of λ_i is $\frac{o_i}{N}$, $i = 1,\ldots,r$, yielding

$$\hat{\pi}_i = \frac{n_{yi} + n(p_i - 1)}{n\, p_i} \quad , \; i = 1,\ldots,r.$$

[Note, $\frac{o_i}{N}$ as an estimate of λ_i is unbiased and asymptotically best, based on the asymptotic normality of the o_i's. For finite samples the common distribution of the o_i's is not simple.]

Let N_i denote the number of individuals among the N_o who reported that they belong to the i-th group in the second stage, $i = r+1,\ldots,t$. Since, conditional on N_o, $(N_{r+1},\ldots,N_t)$ is a multinomial with cell proportions $\pi^*_{r+1},\ldots,\pi^*_t$, where

$$\pi^*_\ell = \frac{\pi_\ell}{\sum_{v=r+1}^{t} \pi_v} \;, \; \ell = r+1,\ldots,t$$

then

$$\hat{\pi}_i = \frac{N_i}{N_o}\left(1 - \sum_{j=1}^{r} \hat{\pi}_j\right) \;, \; i = r+1,\ldots,t.$$

These estimates are unbiased. The individual variances can be computed in a manner similar to the above. However, care must be taken of the covariances among the $\hat{\pi}_j$'s.

Scheme 2. Form the groups $(\pi_1 + \pi_2 + \pi_3 + \ldots + \pi_{r+1})$, $\pi_{r+2},\ldots,\pi_t$. In the first stage ask the individuals directly to which of these t - r groups they belong. (It is understood that π_{r+1} is not too low to bias an individual's response.)

Let $N_{1,r+1}$ be the number of individuals who identified themselves in the combined first r + 1 groups. Denote by N_i the number of those who identified themselves in the i-th group, $i = r+2,\dots,t$.

Let p_{i1}, p_{i2}, p_{i3} be the chances that an individual among the $N_{i,r+1}$ will have to: answer the question "Are you in group i?"; say "yes"; say "no", $i = 1,\dots,r$.

Let θ_i be the probability of a "yes" response to the i-th 'question' in second stage, and let u_i be the number of such a response to that 'question.' Possible estimators of the θ_i's are

$$\hat{\theta}_i = \frac{u_i}{N_{1,r+1}}, \quad i = 1,\dots,r.$$

From equating $\theta_i = \left[\frac{\pi_i}{\sum_{j=1}^{r+1} \pi_j}\right] p_{i1} + p_{i2}$, $i = 1,\dots,r$, one obtains

$$\hat{\pi}_i = \frac{N_{1,r+1}}{N} \frac{\frac{u_i}{N_{1,r+1}} - p_{i2}}{p_{i1}} = \frac{u_i - N_{1,r+1}p_{i2}}{N\, p_{i1}}, \quad i = 1,\dots,r.$$

$$\hat{\pi}_{r+1} = \frac{N_{1,r+1}}{N} - \sum_{i=1}^{r} \hat{\pi}_i .$$

The Relation to Warner's (1971) General Linear Model

On identifying the TSRRS's described above as special cases of Warner's (1971) general linear randomized response model one may obtain alternative estimators based on Zellner's (1962) modified generalized least squares estimates. Following are three examples.

Example 1: Suppose $t = 3$, $r = 1$ and the first group is the stigmatizing one. Here, the second scheme is demonstrated, using Warner's (1965) original approach for randomizing in the second stage.

$$\underset{\sim}{x}_i = \begin{cases} (1,0,0)' & \pi_1 \\ (0,1,0)' & \pi_2 \\ (0,0,1)' & \pi_3 \end{cases}, \quad i = 1,\ldots,N$$

Choose $\underset{\sim}{T}_i$ as a 2x3 random matrix with distribution

$$\underset{\sim}{T}_i = \begin{cases} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} & p \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & 1 - p \end{cases}, \quad i = 1,\ldots,N$$

$$\underset{\sim}{y}_i = \begin{cases} (0,0)' & \pi_1(1-p) + p\pi_2 \\ (1,0)' & \pi_1 p + \pi_2(1-p) \\ (0,1)' & \pi_3 \end{cases}, \quad i = 1,\ldots,N$$

where

p is the randomization chance in the second stage; (0,1) denotes identification of an individual with the third group in the first stage; (1,0) is the response 'yes,' from an individual in the second stage; while (0,0) is a 'no' response from individuals who undertook a second stage.

Example 2: Here, the second scheme is demonstrated when $r = 1$, $t = 3$ as above, but when an alternate question on a characteristic Y is used in the second stage to randomize with probability p ((1-p)) an answer to the question: "Are you in group 1?" ("do you have characteristic Y?"). Using same $\underset{\sim}{y}_i$'s with same interpretation as in Example 1 yields the following:

$$\underset{\sim}{x}_i = \begin{cases} (1\ 0\ 0\ 0\ 0\ 0)' & \pi_Y\pi_1 \\ (0\ 1\ 0\ 0\ 0\ 0)' & \pi_Y\pi_2 \\ (0\ 0\ 1\ 0\ 0\ 0)' & \pi_Y\pi_3 \\ (0\ 0\ 0\ 1\ 0\ 0)' & (1-\pi_Y)\pi_1 \\ (0\ 0\ 0\ 0\ 1\ 0)' & (1-\pi_Y)\pi_2 \\ (0\ 0\ 0\ 0\ 0\ 1)' & (1-\pi_Y)\pi_3 \end{cases}, \quad i = 1,\ldots,N$$

$$\underset{\sim}{T}_i = \begin{cases} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} & p \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} & 1 - p \end{cases}, \quad i = 1,\dots,N$$

$$\underset{\sim}{y}_i = \begin{cases} (0,0)' & \pi_2 p + (1-p)(1-\pi_Y)(\pi_1+\pi_2) \\ (1,0)' & \pi_1 p + (1-p)\pi_Y(\pi_1+\pi_2) \\ (0,1)' & \pi_3 \end{cases}, \quad i = 1,\dots,N.$$

Example 3: This final example demonstrates the first scheme where $t = 4$ and $r = 2$ (the first two). On letting '1' refer to a 'yes' answer, '0' to a 'no' answer and the position in a vector to the individual's group

$$\underset{\sim}{x}_i = \begin{cases} (1\ 0\ 0\ 0) & \pi_1 \\ (0\ 1\ 0\ 0) & \pi_2 \\ (0\ 0\ 1\ 0) & \pi_3 \\ (0\ 0\ 0\ 1) & \pi_4 \end{cases}, \quad i = 1,\dots,N$$

$$\underset{\sim}{y}_i = \begin{cases} (1\ 1\ 0\ 0) & (1-p_1)(1-p_2) + p_1(1-p_2)\pi_1 + (1-p_1)p_2\pi_2 \\ (1\ 0\ 0\ 0) & (1-p_1)p_2(1-\pi_2) + p_1p_2\pi_1 \\ (0\ 1\ 0\ 0) & (1-p_2)p_1(1-\pi_1) + p_1p_2\pi_2 \\ (0\ 0\ 1\ 0) & p_1p_2\pi_3 \\ (0\ 0\ 0\ 1) & p_1p_2\pi_4 \end{cases}, \quad i = 1,\dots,N$$

$$\underset{\sim}{T}_i = \begin{cases} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & p_1p_2 \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & p_1(1-p_2) \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & p_2(1-p_1) \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & (1-p_1)(1-p_2) \end{cases}, \quad i = 1,\dots,N.$$

In all TSRRS's the distributions of the $\underset{\sim}{T}_i$'s are the same for all $i = 1,\ldots,N$. Hence, as discussed in Warner (1971) p. 885, the application of Zellner's (1962) method is easy.

Additional Comments

Another approach which appears to be a compromise between the two TSRRS's given above for $r > 1$ is as follows: First, use a randomizing device for choosing between the question, "Are you in any stigmatizing group?" or arbitrarily answering "yes." Those who say "yes" are then asked several more randomized questions, as in the first stage of Scheme 1 where $r > 1$.

Also, as noted above, when $1 < r < t - 1$, the estimates are justified by large sample theory. A very convenient tool for the analysis of data in such designs (and in more complicated sampling schemes which combine responses to direct and randomized questions) is the least squares approach as discussed for general categorical data in Grizzle et al. (1969). This has been noted in Folsom et al. (1973), p. 530.

APPENDIX B

ESTIMATING A BERNOULLI PARAMETER FROM A SAMPLE OF MISCLASSIFIED RESPONSES AND A SUBSAMPLE OF RANDOMIZED RESPONSES

Introduction

It appears that, in the various publications on the use of the Randomized Response Technique (RRT), it has always been assumed that the experimenter has available to him only the sample of randomized responses from which to draw inferences. However, in many applications, the randomized response technique is used when an original, usually relatively large, sample is available. If this original sample is based on misclassified responses due to some stigma in the issues under study, then a subsample of individuals can be taken from the original sample (with the individual possibly misclassified responses available) for application of the randomized response technique.

This appendix discusses some efficient methods for estimating the Bernoulli parameter of a stigmatizing response, based on the simultaneous classification of the sub-sampled individuals according to their misclassified and randomized responses, along with the original total sample of responses, some of which are misclassified.

The Double Sampling Technique

Tenenbein (1970) proposed a double sampling technique for the problem of estimation using categorical data which is subject to misclassification errors. His approach is based on the model in Bross (1954) for misclassification errors. The following experimental situation is assumed. There are two classification devices available. One device is expensive to apply and gives correct results. The other device is relatively inexpensive but fallible. Such experimental situations were considered by many other writers, e.g., Diamond and Lilienfeld (1962), where the true expensive classification device is a physician's examination whereas the fallible classifier is a questionnaire completed by the patient. The double sampling scheme involves the following steps:

a. Obtain the fallible classifications on a "large" number of units (say, N).

b. Obtain, in addition, the true classifications on a subsample of n out of the N units.

c. Combine (a) and (b) efficiently for estimating the Bernoulli parameter under study.

Note that in many problems, stage (a) might simply amount to accessing an existing file (e.g., statewide accident file).

In many experimental situations, there is no exact device for measuring the true response. Often only the individual knows the true response. If the response has a stigmatizing nature, estimates based on a direct questionnaire are biased due to errors of misclassification. A reasonable approach in such cases would be to use the randomized response technique on a subsample of individuals in phase (b).

Clearly, the two-stage approach will be more efficient than using only the randomized response subsample when the sample of phase (a) is readily available. In other cases, the double sampling plan might still be considered a good strategy depending on certain parameters. This is discussed in the following section.

A Double Sampling Plan with Misclassified and Randomized Responses

Introduction and notation.

The randomized response procedure to be used in this discussion is the one that uses an unrelated question of a known proportion. This technique can always be used by artificially forming responses with known probabilities (see Greenberg et al, 1974).

The data in the subsample can be summarized as in Table B.1.

Table B.1. Frequencies in the subsample.

		Misclassified Responses		
		No	Yes	Total
Randomized Response	No	n_{00}	n_{01}	$n_{0\cdot}$
	Yes	n_{10}	n_{11}	$n_{1\cdot}$
	Total	$n_{\cdot 0}$	$n_{\cdot 1}$	n

A "Yes" and "No" for the misclassified responses implies belonging or not to the stigmatizing group, respectively. The "Yes" and "No" for the randomized response are the literal responses when the individual is asked, "Do you belong to the stigmatizing group?" with probability p_1 and, with probability 1-p, he is asked an unrelated question which has probability w of a "Yes" response and 1-w of a "No" response. Both $0<p<1$ and $0<w<1$ are assumed known.

Corresponding to Table B.1. is the following table of population proportions:

Table B.2. Population proportions

		Misclassified Responses		
		No	Yes	Total
Randomized Responses	No	β_{00}	β_{01}	$1-\lambda$
	Yes	β_{10}	β_{11}	λ
	Total	$1-\phi$	ϕ	1

On letting $\alpha_{0,1}$ denote the probability that the fallible classifier gives "No" when the truth is "Yes", letting $\alpha_{1,0}$ denote the reverse error, and letting π denote the true proportion of individuals in the stigmatizing group, the following relationships obtain:

$$\lambda = p\pi + (1-p)w$$

$$\beta_{00} = (1-\pi)(1-\alpha_{1,0})[p+(1-p)(1-w)] + \pi\alpha_{0,1}(1-p)(1-w)$$

$$\beta_{01} = \pi(1-\alpha_{0,1})(1-p)(1-w) + (1-\pi)\alpha_{1,0}[p + (1-p)(1-w)]$$

$$\beta_{10} = (1-\pi)(1-\alpha_{1,0})(1-p)w + \pi\alpha_{0,1}[p + (1-p)w]$$

$$\beta_{11} = \pi(1-\alpha_{0,1})[p + (1-p)w] + (1-\pi)\alpha_{1,0}(1-p)w.$$

The N-n remaining individuals for which information is only available from the fallible classifier can be represented as follows:

	No	Yes	Total
Frequency	Y	X	N-n
Expected Proportion	$1-\phi$	ϕ	1

Maximum likelihood estimates of $\alpha_{0,1}$, $\alpha_{1,0}$, and π.

First, let

$\xi = \beta_{11}/\phi$ = Conditional probability of getting a "Yes" for the randomized response when the fallible response is "Yes".

$\Psi = \beta_{10}/(1-\phi)$ = Conditional probability of getting a "Yes" for the randomized response when the fallible response is "No".

The common probability distribution function of X, Y, n_{00}, n_{10}, n_{01} and n_{11} is given by

$$(\text{constant})\phi^{X+n_{\cdot 1}}(1-\phi)^{Y+n_{\cdot 0}}\xi^{n_{11}}(1-\xi)^{n_{01}}\Psi^{n_{10}}(1-\Psi)^{n_{00}} .$$

The maximum likelihood estimators (MLE) of ϕ, ξ, Ψ are given by $\hat{\phi}$, $\hat{\xi}$, $\hat{\Psi}$, respectively, where

$$\hat{\phi} = \frac{X+n_{\cdot 1}}{N} ; \qquad \hat{\xi} = \frac{n_{11}}{n_{\cdot 1}} ; \qquad \hat{\Psi} = \frac{n_{10}}{n_{\cdot 0}} \tag{B.1}$$

Now, since $(\phi, \xi, \Psi) \overset{1:1}{\longleftrightarrow} (\lambda, \xi, \Psi)$, the MLE $(\hat{\lambda})$ of λ is obtained from

$$\lambda = \phi\xi + (1-\phi)\Psi$$

by substituting values of $\hat{\phi}$, $\hat{\xi}$, and $\hat{\Psi}$ from (B.1) to give

$$\hat{\lambda} = \frac{X+n_{\cdot 1}}{N}\frac{n_{11}}{n_{\cdot 1}} + \frac{\lambda+n_{\cdot 0}}{N}\frac{n_{10}}{n_{\cdot 0}}$$

Then, the MLE ($\hat{\pi}$) of π can easily be obtained by solving for π in

$$\lambda = p\pi + (1-p)w$$

and substituting $\hat{\lambda}$ for λ as follows:

$$\hat{\pi} = \frac{\hat{\lambda} - (1-p)w}{p} \qquad \text{(B.2)}$$

To obtain the MLE's of $\alpha_{0,1}$ and $\alpha_{1.0}$ consider the following:

$$\xi = \frac{\beta_{11}}{\phi} = \frac{\pi(1-\alpha_{0,1})[p+(1-p)w]+(1-\pi)\alpha_{1,0}(1-p)w}{\phi}$$

$$\Psi = \frac{\beta_{10}}{1-\phi} = \frac{(1-\pi)(1-\alpha_{1,0})(1-p)w+\pi\alpha_{0,1}[p+(1-p)w]}{1-\phi}$$

On letting $\underline{\theta} = (\xi,\Psi)'$,

$$\underline{b} = \begin{bmatrix} \dfrac{\pi(p+(1-p)w]}{\phi} \\ \\ \dfrac{(1-\pi)(1-p)w}{1-\phi} \end{bmatrix}, \quad \underline{A} = \begin{bmatrix} \dfrac{-\pi[p+(1-p)w]}{\phi} & \dfrac{(1-\pi)(1-p)(1-w)}{\phi} \\ \\ \dfrac{\pi[p+(1-p)w]}{1-\phi} & \dfrac{-(1-\pi)(1-p)w}{1-\phi} \end{bmatrix}$$

and $\underline{\alpha}' = (\alpha_{0,1}\ \alpha_{1,0})$, it follows that

$$\underline{\alpha} = \underline{A}^{-1}(\underline{\theta}-\underline{b}) \qquad \text{(B.3)}$$

Substituting $\hat{\phi}$, $\hat{\xi}$, $\hat{\Psi}$, and $\hat{\pi}$ for ϕ, ξ, Ψ, and π in (B.3) yields the MLE ($\hat{\alpha}$) of α. (It is easily verified that $\underline{A}$ is non-singular provided $w \neq \frac{1}{2}$.)

The asymptotic variance of $\hat{\pi}$.

The asymptotic covariance matrix of $(\hat{\phi}, \hat{\xi}, \hat{\Psi})$ is the inverse of the corresponding information matrix. It can be verified that $\hat{\phi}$, $\hat{\xi}$, $\hat{\Psi}$ are asymptotically independent and

$$\mathrm{Var}(\hat{\phi}) = \frac{\phi(1-\phi)}{N}$$

$$\mathrm{Var}(\hat{\xi}) = \frac{\xi(1-\xi)}{n\phi}$$

$$\mathrm{Var}(\hat{\Psi}) = \frac{\Psi(1-\Psi)}{n(1-\phi)}$$

Since $\hat{\lambda} = \hat{\phi}\hat{\xi} + (1-\hat{\Phi})\hat{\Psi}$, one can obtain the asymptotic variance by $\hat{\lambda}$ by linearization, which gives

$$V(\hat{\lambda}) = (\xi-\Psi)^2 \frac{\phi(1-\phi)}{N} + \frac{\phi\xi(1-\xi}{n} + \frac{(1-\phi)\Psi(1-\Psi)}{n} \tag{B.4}$$

From this expression and the relation between $\hat{\lambda}$ and $\hat{\pi}$ (i.e., $\hat{\pi} = \frac{\hat{\lambda}-(1-p)w}{p}$), the asymptotic variance of $\hat{\pi}$ is obtained.

Next, the variance of $\hat{\lambda}$ can be expressed in terms of the $\alpha_{0,1}$ and $\alpha_{1,0}$. First, let

$\theta_{0,1}$ = Conditional probability of getting a "No" response on the fallible classifier when the randomized response is "Yes."

$\theta_{1,0}$ = Conditional probability of getting a "Yes" response on the fallible classifier when the randomized response is "No."

It then follows, as in Tenenbein (1970), that

$$V(\hat{\lambda}) = \frac{\lambda(1-\lambda)}{n}\left[1-\frac{\lambda(1-\lambda)}{\phi(1-\phi)}(1-\theta_{0,1}-\theta_{1,0})^2\right] + \frac{[\lambda(1-\lambda)]^2}{N\phi(1-\phi)}(1-\theta_{0,1}-\theta_{1,0})^2 \quad (B.5)$$

Using any two of the four identities involving the B_{ij}'s, $i,j = 0,1$, straightforward computations give

$$\theta_{0,1} + \theta_{1,0} = \frac{p\pi(1-\pi)(\alpha_{0,1}+\alpha_{1,0})+(\lambda-\pi p)(1-\lambda)+p\pi(\pi-\lambda)}{\lambda(1-\lambda)} \quad (B.6)$$

Also,

$$1-\theta_{0,1}-\theta_{1,0} = p\,\frac{\pi(1-\pi)}{\lambda(1-\lambda)}(1-\alpha_{0,1}-\alpha_{1,0})$$

which can now be substituted in (B.5) to give

$$V(\hat{\lambda}) = \frac{\lambda(1-\lambda)}{n}\left[1 - p^2\,\frac{\pi(1-\pi)}{\phi(1-\phi)}\,\frac{\pi(1-\pi)}{\lambda(1-\lambda)}(1-\alpha_{0,1}-\alpha_{1,0})^2\right]$$

$$+ p^2\,\frac{\pi(1-\pi)}{N}\,\frac{\pi(1-\pi)}{\phi(1-\phi)}(1-\alpha_{0,1}-\alpha_{1,0})^2 \quad (B.7)$$

It is interesting to note the following:

(i) If $p = 1$ (i.e., $\lambda = \pi$), this case reduces to that of Tenenbein (1970).

(ii) If $0<p<1$ and no error is involved in using the fallible classifier, one does not get $V(\hat{\lambda}) = p^2\pi(1-\pi)/N$, because in such a case $\hat{\lambda}$ is not the MLE of λ. In this case ξ and Ψ are fixed constants

$$\xi = p + (1-p)w$$

$$\Psi = (1-p)w.$$

Clearly the MLE for π in this case is that of ϕ with variance $\pi(1-\pi)/N$.

The least squares approach.

The estimators considered thus far are MLE's, similar to those in Tenenbein (1970). Consider now the least squares estimators (LSE) based on Grizzle et al. (1969). These are obtained along the lines of Koch et al. (1972). The strategy here is to obtain LSE's of the β_{ij}'s from which the estimators of λ, $\alpha_{0,1}$ and $\alpha_{1,0}$ are obtained.

Let

$$\underline{n}' = (n_{00}, n_{01}, n_{10}, n_{11}),$$

$$\underline{N}' = (Y, X)$$

$$\underline{p} = (\underline{n}/n)$$

$$\underline{p}_N = (\underline{N}/(N-n))$$

$$\underline{p}_G' = (\underline{p}', \underline{p}_N').$$

Then

$$E(\underline{p}_G') = (\pi', \pi_.')$$

where $\pi' = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$ and $\pi_.' = (1-\phi, \phi)$.

The covariance matrix of $\underline{p}_G$ is given by

$$\underline{V}(\underline{p}_G) = \begin{bmatrix} (D_{\underline{\pi}} - \underline{\pi}\underline{\pi}')/n & \underline{0}:4\times 2 \\ \underline{0}:2\times 4 & (D_{\underline{\pi}.} - \underline{\pi}.\underline{\pi}.')/(N-n) \end{bmatrix},$$

where $D_{\underline{a}}$ is diagonal with $\underline{a}$ as the elements on the diagonal. Let

$$\underline{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & 0 & 0 \\ & \text{(Symmetric)} & & & 0 & 0 \\ & & & & & 1 \end{bmatrix}$$

and write $\underline{F} = \underline{A}\underline{p}_G$. Next let the model be $E(\underline{F}) = X\beta$, where

$$\underline{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

and $\underline{\beta} = (\beta_{01}, \beta_{10}, \beta_{11})$. The LSE of $\underline{\beta}$ is given by

$$\underline{b} = \hat{\underline{\beta}} = (\underline{X}'\underline{V}_F^{-1}\underline{X})^{-1}\underline{X}'\underline{V}_F^{-1}\underline{F}$$

where $\underline{V}_F = \underline{A}'\underline{V}(\underline{p}_G)\underline{A}$. The resulting estimated covariance of $\underline{b}$ is

$$\underline{V}_b = (\underline{X}'\underline{V}_F^{-1}\underline{X})^{-1} .$$

Having obtained the estimates of the β_{ij}'s ($\hat{\beta}_{01}=b_1$, $\hat{\beta}_{10}=b_2$, $\hat{\beta}_{11}=b_3$ and $\hat{\beta}_{00}=1-\beta_{01}-\beta_{10}-\beta_{11}=1-b_1-b_2-b_3$), one obtains $\hat{\lambda}=\hat{\beta}_{10}+\hat{\beta}_{11}$, $\hat{\phi}=\hat{\beta}_{01}+\hat{\beta}_{11}$. Finally, $\hat{\pi}$, $\hat{\alpha}_{0,1}$ and $\hat{\alpha}_{1,0}$ are derived as previously with the MLE's.

<u>The coefficient of reliability and efficiency of the two-stage procedure</u>.

The discussion here is based on the MLE in order to parallel the development given in Tenenbein (1970). Let k_R denote the squared correlation coefficient between the randomized and the fallible responses (where 1 and 0 are attached to "Yes" and "No", respectively).

One obtains

$$k_R = \frac{\lambda(1-\lambda)}{\phi(1-\phi)} \, (1-\theta_{1,0}-\theta_{0,1})^2 .$$

On letting k_T denote the squared correlation between the true and the fallible responses,

$$k_T = \frac{\pi(1-\pi)}{\phi(1-\phi} (1-\alpha_{1,0}-\alpha_{0,1})^2 .$$

As in Tenenbein (1970), using (B.5) and the relationship between $\hat{\lambda}$ and $\hat{\pi}$, one obtains

$$V(\hat{\pi}) = \frac{\lambda(1-\lambda)}{p^2} \{ \frac{1}{n} [1-k_R] + \frac{1}{N} k_R \} . \qquad (B.8)$$

If only the randomized responses are utilized, (B.8) reduces to

$$V(\hat{\hat{\pi}}) = \frac{\lambda(1-\lambda)}{np^2}$$

To study the efficiency of the two stage procedure versus the use of only the randomized responses for equal cost, one must first obtain the best allocation of observations, i.e., that allocation which achieves minimum $V(\hat{\pi})$ for a given cost. This is equivalent to minimizing $V(\hat{\lambda})$ for a given cost which follows along the line of Tenenbein (1970). Thus, on letting c_m and c_r be the costs per unit sampling of a misclassified response and of a randomized response, respectively, $R = c_r/c_m$, $R = n/N$, and

$$f_0 = \text{Min} \left\{ \left(\frac{1-k_R}{k_R R} \right)^{\frac{1}{2}} , 1 \right\} ,$$

the best allocation is

$$n = n_0 \left[\frac{Rf_0}{Rf_0+1} \right] \quad ; \quad N = (n_0-n)R$$

where $n_0 = C_0/c_r$ and C_0 is the total available cost. The comparison of the two methods can thus be conducted along the lines of Tennenbein (1970) and thus will not be pursued here. Similarly, for a three stage-type sampling as in Tenenbein (1971), where a pilot sample is taken to estimate k_R which then determines the best allocation.

Note that since n<N, for any $k_R>0$ the double sampling plan is more efficient when cost is not considered (as is approximately the case when the total sample of misclassified responses are already available on file). It is interesting to compare the efficiency of $V(\hat{\pi})$ with that of $V(\hat{\hat{\pi}})$ for that important case. Since

$$e = \text{efficiency} = \frac{V(\hat{\pi})}{V(\hat{\hat{\pi}})} = 1-k_R(1-\frac{n}{N})$$

the efficiency (e) depends on k_R and n/N. However, k_R is clearly a function of p, w, $\alpha_{0,1}$, $\alpha_{1,0}$ and π.

In any particular problem, the possibility of biased estimates resulting from randomized responses should also be considered. Since there are so many parameters, the various tables are not considered here. However, in any specific problem the experimenter should make a decision, based on an appropriate pilot study, regarding the course of action to be taken, namely

(i) Use only the sampling of misclassified responses.

(ii) Use only the randomized responses.

(iii) Use the double sampling scheme.

Such a decision should be based on one's best guesses of the following: relative costs, π, the errors $\alpha_{0,1}$, $\alpha_{1,0}$, the bias in the randomized response estimators, and the parameters p and w.

APPENDIX C

Cover Letter and Questionnaire for
Randomized Response Pilot Study #2

THE UNIVERSITY OF NORTH CAROLINA
HIGHWAY SAFETY RESEARCH CENTER
CHAPEL HILL 27514

TO: North Carolina drivers involved in accidents in December 1974

FROM: The University of North Carolina Highway Safety Research Center

DATE: January 9, 1975

SUBJECT: Highway safety

The Highway Safety Research Center, in cooperation with the U.S. Department of Transportation, is studying certain aspects of highway crashes. Obviously, the accident-involved driver is the best source of such information. To locate drivers involved in recent accidents, we randomly sampled December accident reports from the State files which contain records of all highway accidents occurring in North Carolina. We would very much appreciate your cooperation in completing this brief experimental questionnaire.

The questions involve speed, seat belt usage, and possible alcohol involvement. As these questions are rather sensitive to most persons, you will be asked to use a coin-flipping experiment which will guarantee the confidentiality of your response (as you will see later).

Attach Penny "a penny for your thoughts"

We are completely dependent on the help and cooperation of people like yourself if we are to find out more about accidents and thus help drivers avoid accidents. If for any reason you do not wish to participate in this survey, you of course need not do so. However, if you do respond, do not put your name on the questionnaire or on the enclosed envelope. In any event, the penny is yours with our gratitude. If you have any questions, please call HSRC collect at (919) 933-2202 and refer to the "**Randomized Response Survey.**"

QUESTIONNAIRE

Please read each question completely and then follow the instructions below.

I. What was your speed just before your accident occurred?

(A) More than 5 mph UNDER the speed limit (for example, you were traveling below 30 mph in a 35 mph zone)?

(B) WITHIN 5 mph of the speed limit (for example, you were traveling 38 mph in a 35 mph speed zone)?

(C) More than 5 mph OVER the speed limit (for example, you were traveling over 40 mph in a 35 mph speed zone)?

Did you choose answer (B)? ____Yes ____No

If "Yes", continue to Question II. If "No", flip the enclosed penny 2 times and, in the space provided below, if 2 HEADS appear, answer Question (A) above; otherwise if 1 HEAD, 1 Tail or 2 TAILS appear, answer Question (C) above.

____Yes ____No

II. Had you been drinking before your accident?

(A) I had no alcoholic beverages before my accident.

(B) I only had one drink in the two hour period before my accident.

(C) I had more than one drink in the two hour period before my accident.

Did you choose answer (B)? ____Yes ____No

If "Yes", continue to Question III. If "No", flip the enclosed penny 2 times and, in the space provided below, if 2 HEADS appear, answer Question (A) above; otherwise if 1 HEAD, 1 TAIL or 2 TAILS appear, answer Question (C) above.

____Yes ____No

III. Was your car equipped with lap and shoulder belts? ____Yes ____No

If no, lap belts only? ____Yes ____No

If your car did not have any seat belts, you are finished. Otherwise, please answer the following question.

What safety belts were you wearing at the time of the accident?

(A) Lap and shoulder belts.

(B) Lap belt only.

(C) No belt.

Did you choose answer (B)? ____Yes ____No

If "Yes", you are finished. If "No", flip the enclosed penny 2 times and, in the space provided below, if 2 HEADS appear, answer Question (A) above; otherwise if 1 HEAD, 1 TAIL or 2 TAILS appear, answer Question (C) above.

____Yes ____No

Do you have any questions or comments? ______________________________

____

____

Thank you very much for your cooperation.

APPENDIX D

Cover Letter and Questionnaire for
Randomized Response Pilot Study #3

THE UNIVERSITY OF NORTH CAROLINA
HIGHWAY SAFETY RESEARCH CENTER
CHAPEL HILL, NORTH CAROLINA 27514

April 18, 1975

Dear North Carolina driver:

The University of North Carolina Highway Safety Research Center is studying seat belt usage among North Carolina drivers who have been recently involved in accidents. The questionnaire may look somewhat strange or unusual. The design, which is called the "randomized response technique," has been used before in statistical studies where it was thought the questions might be too personal to ask directly. We feel some people may find our seat belt questions somewhat touchy to answer directly so we are using the randomized response technique here.

Even if you would not mind answering this questionnaire directly, please follow the directions given so that we might be able to judge the effectiveness of this method as applied to questions about highway safety.

There is no need to place your name on the questionnaire or the envelope, for we are not interested in individual responses but only in the overall group response.

If you have any questions on this survey or about the Highway Safety Research Center, please call (919) 933-2202 collect and refer to the "Randomized Response Survey."

Thank you for your cooperation.

SEAT BELT USAGE QUESTIONNAIRE

Read all three parts before attempting to answer in the box at right.

IF you were born in either March or April, write "yes here

IF you were born in either November or December, write "no" here

OTHERWISE, answer the following question:

Were you wearing your seat belts when you had your recent accident?

We appreciate your cooperation. Please return this sheet in the enclosed stamped self-addressed envelope. It will help us greatly both in studying the effectiveness of seat belts and determining the usefulness of questionnaires of this type.

Comments: ______________________________________

____

____

____