Reprint from:

SOCIOLOGICAL METHODS AND RESEARCH

"A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda"

by U.N. Umesh and Robert A. Peterson

Books
Journals
Newsletters
University Papers
Annual Series

2455 Teller Road
Newbury Park
California 91320

Phone (805) 499-0721
Fax (805) 499-0871
Cable SagePub
Telex (510) 1000799

Sage Publications, Inc.

Warner (1965) introduced the Randomized Response Method (RRM) 20 years ago. This method has been advocated as a useful tool in eliciting sensitive information. Much of the early research focused on various parameter estimation methods. This article provides a comprehensive review of applications of the randomized response method with emphasis on recent publications, identifies issues now being studied, and suggests future research directions. As such, the article addresses method validation, respondent jeopardy, and new applications. In so doing, the method is considered in the broad context of obtaining responses to sensitive questions. Judging by the number of articles published, randomized response method continues to be of interest in numerous and diverse disciplines.

A Critical Evaluation of the Randomized Response Method

Applications, Validation, and Research Agenda

U. N. UMESH
Washington State University
ROBERT A. PETERSON
University of Texas at Austin

A quarter of a century has passed since Warner (1965) introduced the randomized response method (RRM) as a means of obtaining truthful answers to questions about sensitive attributes. The original randomized response method required that respondents use a randomizing device to choose one of two statements of the form:

— I have sensitive attribute A (selected with probability p).
— I do not have sensitive attribute A (selected with probability $1 - p$).

AUTHOR'S NOTE: *The authors thank J. Scott Long for providing suggestions on how best to configure the manuscript for this journal.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 20, No. 1, August 1991 104-138

Without revealing which statement was selected, respondents answer "yes" or "no" according to the statement selected and their actual status with respect to attribute A.

The variance of the estimated proportion of respondents with attribute A consists of two components. The first component is equivalent to the variance associated with binomial estimation encountered in traditional surveys; the second is due to the use of the randomizing device. Consequently, the estimated variance is larger than that derived from direct questioning about attribute A. In spite of this, the reduction in response bias from increased truthfulness was expected to produce a lower mean squared error of estimate.

The Warner design suffers from several deficiencies: The estimates are statistically inefficient, estimates of proportions are not always in the 0-1 range, and respondents continue to be somewhat apprehensive of being stigmatized. Over time, researchers attempted to rectify these defects. Research has evolved from identifying estimators with low(er) variances to focusing on method validation, respondent jeopardy, new randomizing devices, multivariate calculations and Bayesian estimation. One improved design asks about an attribute that is unrelated to the sensitive one (Bourke and Moran 1988; Greenberg, Abul-Ela, Simmons, and Horvitz 1969; Moors 1971). In this design, respondents use a randomizing device to choose one of two questions of the form:

– Do you have sensitive attribute A?
– Do you have nonsensitive attribute B?

Respondents then appropriately answer the question selected.

In almost all studies that measure sensitive questions with RRM, the basic assumption that the method increased both the propensity to respond, and the propensity to respond more accurately than direct questioning, has not been challenged. Rather, the RRM estimates were *expected* to be more valid than those derived from direct questioning. Researchers assume that when underreporting is motivated by a concern on the part of the respondent about the embarrassment or adverse consequences should their identities be known, RRM will help in mitigating such concerns. Although the large number of published statistical extensions are intellectually interesting, expectations about the validity of RRM, and hence its practical utility, may have been overblown. As we show in this article, the field application and validation of RRM provide mixed support for its usefulness.

The rest of the article is divided into sections based on issues rather than on individual methods. The first section deals with statistical developments of the method. The next section discusses applications of RRM. This is followed by a discussion of the validity of the method in different areas. The final section is devoted to future research avenues for the method.

STATISTICAL DEVELOPMENTS

For an extensive review of the statistical aspects of RRM, the reader is referred to Chaudhuri and Mukherjee (1988) and Fox and Tracy (1986). This section briefly summarizes the statistical extensions of the method.

CONTAMINATION MODEL

An alternative to the basic randomized response design has been proffered to increase its statistical efficiency. This alternative design, known as the contamination model, asks respondents to modify their responses in a prearranged fashion. Several variations of the contamination model have been developed. Among the variations, the additive constants model is the most widely used (Himmelfarb and Edgell 1980; Pollock and Bek 1976). In an application of this model (Himmelfarb and Edgell 1980), respondents were handed a deck of numbered cards and asked an open-ended question about household income. Participants were asked to provide the sum of income and the number on the card to the interviewer. Subsequently, an unbiased estimate of the population mean was obtained by subtracting the population mean of the distribution of the cards from the mean of the respondents' answers. In this way, response anonymity is protected and the respondents' modified responses can be transformed (changed back) by the researchers, at the aggregate level, to obtain the "truth."

In the additive constants model, a random value Y is added to the "true answer" X to produce response Z. The distribution of Y is known, hence

$$\mu_z = \mu_x + \mu_y \tag{1}$$

so that

$$\overline{X} = \overline{Z} - \mu_y \tag{2}$$

where $\overline{X}$ and $\overline{Z}$ are in the observed mean responses, μ_x, μ_y and μ_z are the respective population means of X, Y and Z. Since $\overline{X}$ is obtained by subtracting a constant μ_y from the random variable $\overline{Z}$,

$$\text{var}\,(\overline{X}) = \text{var}\,(\overline{Z}) - 0 = \sigma_z^2/n \tag{3}$$

The multiplicative contamination model (Pollock and Bek 1976) is very similar to the additive model except that the sensitive attribute value X is multiplied by a random value Y to produce response Z. The contamination model has been extended to the estimation of the distribution of a sensitive trait (in addition to its mean and variance) by Poole (1974). Poole and Clayton (1982) provided a multivariate extension of the Poole (1974) design. Recently, Scheers and Dayton (1988) used a covariate that is nonsensitive to improve the accuracy of the RRM. The correlation between the sensitive and nonsensitive question responses improves the overall estimate of the sensitive attribute. Although their method is not strictly a contamination model, the analysis in the two methods follows similar principles.

In general, the contamination model is a promising design. The standard errors of estimation are lower than the commonly used unrelated question design. In addition to improved efficiency, contamination designs possess another potential advantage that is often overlooked. By using a randomly determined number to add to or multiply with their response, respondents may be more likely to believe they can conceal the true nature of the sensitive attribute than by randomly selecting a question to answer.

MULTIPLE SENSITIVE ATTRIBUTES

In practice, it is common for a study to simultaneously address many topics, several of which may be of a sensitive nature. Clickner and Iglewicz (1980) extended the Warner (1965) model using two sensitive questions. Respondents are asked to respond to the first sensitive question in a manner similar to the Warner (1965) method:

— I have sensitive attribute A.
— I do not have sensitive attribute A.

Respondents are then presented with a second randomizing device and asked to answer a question relating to a second sensitive attribute. The outcome probabilities of the two randomizing devices need not be the same. Bourke (1981) has extended the method to the case of sensitive questions with multiple levels and Tamhane (1981) has suggested a procedure for more than two sensitive questions. Tamhane's (1981) procedure requires each respondent to provide answers to only two questions selected at random from the entire set of sensitive questions. Consequently, the strain on the respondent is reduced. The answers of respondents to different subsets of questions are pooled to estimate the overall mean responses. The trade-off between the increased quantity of information from use of the other procedures and the expected increase in quality of information from the Tamhane procedure needs to be evaluated.

TWO-STAGE DESIGNS

Two-stage designs are used to estimate not only the proportion of a population possessing a sensitive attribute, but also the extent to which it is possessed. The first stage reveals the proportion of respondents possessing the sensitive attribute, whereas the second-stage estimates the extent to which the respondents have the sensitive attribute.

Reinmuth and Geurts (1975) proposed a randomized response design to estimate the mean frequency of shoplifting among shoplifters in a shopping center. They devised a two-stage procedure involving two samples. The first stage used the unrelated question method to estimate the proportion of the population consisting of shoplifters. The second stage used the Greenberg, Kuebler, Abernathy, and Horvitz (1971) quantitative design to estimate the mean frequency of shoplifting across the population.

Lamb and Stem (1978) successfully validated this two-stage procedure using college students. The proportion of respondents who failed a course and the average number of courses failed by a student were estimated. A ratio estimator was used to obtain the mean number of courses failed by those who had failed at least one course. The two-stage design is a specialized design applicable whenever conditional means have to be estimated.

MEASURES OF ASSOCIATION

Although early research concentrated on finding efficient estimators of means, variances, and proportions, some recent work has focused on estimating measures of association through RRMs. Fox and Tracy (1984) derived estimates of the Pearson product moment correlation coefficient by assuming that randomized response observations can be treated as individual-level scores that are contaminated by random measurement error. Kraemer (1980) derived a bivariate correlation between an attribute with polytomous responses and an attribute with quantitative responses using data derived from the Warner design. Edgell, Himmelfarb, and Cira (1986) extended these results to correlations between two sensitive questions using the unrelated question design or the additive constants design. An interesting extension would be the comparison of the relative accuracy of estimating correlation coefficients with randomized response method versus direct questioning.

HYPOTHESIS TESTING

Hypothesis testing is another application of the randomized response method. For instance, it could be hypothesized that the proportion of a population convicted of drunk driving is .3. Levy (1976) suggested using the large sample properties of estimated proportions obtained from randomized response methods. Hypothesis testing is accomplished using the commonly applied normal approximation

$$Z = \frac{\pi_j - \pi}{[\mathrm{Var}(\pi_j)]^{1/2}} \tag{4}$$

where π_j is the estimator of the population proportion and Z is asymptotically normal. Similarly, two sample tests can also be performed. Hypotheses regarding multiple populations can be tested using the weighted chi-squared approach (Levy 1977).

Hypothesis testing using randomized response methods is generally feasible and no more difficult than using direct questioning surveys. However, the sample sizes required are generally larger in randomized response designs than in direct questioning because of the added error from the randomizing process.

BAYESIAN ESTIMATION

Several researchers have suggested using prior information about the sensitive attribute in a Bayesian framework to increase the accuracy of the estimates. Often, information from prior studies employing direct questioning is available before the RRM is to be applied. This information might have produced estimates that are generally downward biased, but at least provides approximate parameter values. The prior distribution is combined with the current data to obtain a Bayesian posterior distribution.

Winkler and Franklin (1979) illustrated a Bayesian estimation procedure for the RRM using the basic Warner design. They viewed Bayesian estimation as important because the very process of randomization (or contamination with error) reduces the effective amount of sample information. Gunel (1985) used the Winkler and Franklin (1979) procedure to compare the relative accuracy of the randomized response method and direct questioning. In general, as the probability (P) of being asked the sensitive question increased, the randomized response estimator became more accurate relative to the direct questioning estimator.

Winkler and Franklin's (1979) procedure has been extended and illustrated by Spurrier and Padgett (1980) and Pitz (1980). Spurrier and Padgett (1980) demonstrated that a Bayesian procedure was superior to the straightforward RRM when the sample size is small or when a relatively narrow range can be specified for the prior distribution of the parameter.

O'Hagan (1987) has suggested the use of Bayes linear estimates for the randomized response design. The Bayes linear estimator solves many of the problems associated with previous Bayesian approaches: The estimator is distribution free, neither prior distributions nor likelihoods need be specified fully and only first- and second-order moments need to be specified.

One consideration for users of the Bayesian procedure is that the prior distribution assumes a more significant role in randomized response designs than in traditional questioning because the process of randomization reduces the effective sample size. Therefore, users of Bayesian estimation in randomized response designs should consider this fact when setting priors and, in general, should use more diffuse priors.

APPLICATIONS

Although the RRM has been in use for over 20 years, there have been few published substantive applications. In many studies where the method was used, the primary purpose was to test a variant or illustrate a statistical derivation. As a result, applications are generally small-scale and appear to be primarily demonstrative. Numerous applications stress the comparison of RRM results with those using direct questioning. These studies have focused on the differences of the two approaches rather than on applying the RRM to solve a problem.

The lack of substantive applications of the RRM is an indication of the relative novelty of the approach, particularly among researchers in the social sciences. This view is corroborated by the continued publication of research that compares the RRM with direct questioning, particularly in journals outside the discipline of statistics. Another explanation for the paucity of substantive applications is that the randomized response method is not fully understood by all researchers in the social sciences. Thus journal reviewers and editors may question the appropriateness of the method in any particular application.

Too rigorous a definition of what constitutes a substantive application would preclude most randomized response studies from being discussed here. Therefore, in an attempt to present a broad mix of randomized response studies, applications are discussed even if they only represent one aspect of a study. The studies discussed are not meant to be exhaustive, only representative of a wide variety of applications. The studies are classified under subject areas. Some of them have been described in earlier sections, but the discussion in this section is limited to applications issues.

Table 1 provides a comparison of the incidence of sensitive behavior estimated by the RRM and direct questioning. Whenever possible, proportions and quantitative responses have been compared separately. Studies that did not have enough data for calculating the magnitude and significance of the differences have been excluded. Most studies that compared the two methods used separate samples for each method. Contrary to popular belief, the RRM does not consistently yield higher estimates of sensitive behavior. The lack of significant differences is not an indication that either method is equally valid. Such a conclusion can be derived only by comparing the reported incidence of sensitive behavior from each method with true population values.

CRIMINAL BEHAVIOR

Tracy and Fox (1981) applied the RRM to estimate the mean number of arrest records of individuals with at least one arrest. The randomizing device consisted of a lucite container with a sealed neck holding 25 red and 25 white balls. Each white ball had a number printed on it, and these numbers ranged from 0 through 8 with a prespecified distribution. The survey was administered to 410 respondents, all of whom had been previously arrested. The randomized response method estimate of .75 arrests is only half the value of the true number of arrests of 1.45 obtained from court records. Those who

TABLE 1: Comparison of Estimates Obtained from Randomized Response Method and Direct Questioning

Researchers	*Sensitive Attributes*	*One or Two Sample Test*	*Direct Questioning Estimates*	*RRM Estimates*	*Magnitude or Significance of Differences Between Direct Questioning and RRM*
Barth and Sandler (1976)	alcohol consumption	two	0.63	0.85	Randomized response estimate was significantly higher ($p < .01$).
Brewer (1981)	drug use; a total of 15 comparisons between two methods	two	0.0 to 0.585	–0.36 to 0.494	Randomized response method consistently produced lower estimates; of which 11 were significantly lower ($p < .01$).
Buchman and Tracy (1982)	honesty in accounting practices based on seven questions	two	0.0 to 0.046	0.0 to 0.132	Randomized response method consistently indicated higher incidence rate but only one of the differences was significant ($p < .03$).
Himmelfarb and Lickteig (1982)	56 questions on sensitive social topics	two	0.027 to 0.99	–0.005 to 0.984	For a majority (60%) of the questions, the randomized response estimates were higher, although most differences were insignificant.
Lamb and Stem (1978)	failing college courses—proportion of respondents and number of courses failed	two	0.288 and 1.602	0.363 and 1.460	Differences were not significant.
Madigan, Abernathy, Herrin, and Tan (1976)	concealment of death in household	one	0.028 and 0.040 in urban and rural areas	0.074 and 0.081	Higher estimates existed for RRM in both urban and rural areas ($p < .05$).
Scheers and Dayton (1987)	academic cheating—5 measures	two	0.026 to 0.294	0.154 to 0.482	The five direct questioning estimates were 30% to 83% lower than RRM ($p < .01$).

(continued)

TABLE 1 Continued

Researchers	*Sensitive Attributes*	*One or Two Sample Test*	*Direct Questioning Estimates*	*RRM Estimates*	*Magnitude or Significance of Differences Between Direct Questioning and RRM*
Shimizu and Bonham (1978)	legal abortions	two	0.006 to 0.008	0.026	Nationally, the number of legal abortions is 3-4 times higher using RRM ($p < .05$).
Stem and Steinhorst (1984)	5 measures of honesty —binomial and quantitative measures	two	0.024 to 0.266 for binomial and 0.050 to 0.727 for quantitative	0.007 to 0.262 for binomial and 0.173 to 1.286 for quantitative	None of the differences were significant.
Volicer and Volicer (1982)	alcohol use and non-compliance among hypertensives	two	0.2 and 0.15 proportions, 3.9 and 4.3 for quantitative responses	0.26 and 0.23 for proportions, 8.7 and 4.2 for quantitative responses	No significant differences in proportion; mean number of drinks was significantly more using RRM ($p < .05$).
Volicer et al. (1983)	alcoholism	two	0.219	0.34	Aggregate randomized response significantly greater than direct questioning; Differences were significant for older respondents, with low seniority and those holding lower positions.
Zdep and Rhodes (1976)	incidence of beating of children	two	0.035	0.15	The randomized response estimates were significantly higher ($p < .01$).
Zdep, Rhodes, Schwarz, and Kilkenny (1979)	marijuana use for various demographic categories	two	0.0 to 0.53 range for various demographic groups	0.03 to 0.48	Slightly higher incidence reported with randomized response; differences widen for older age groups.

NOTE: This table excludes studies discussed in the applications section that did not provide enough information to calculate significance.

had been arrested only once were very reluctant to report the arrest, whereas those with multiple arrests gave accurate responses when asked about the number of times they were arrested. Overall, estimates using the RRM were less biased than estimates using direct questioning. The authors concluded that the reduction in bias more than compensated for the increased variance of the randomized response estimator as compared to direct questioning.

Geurts (1980) used the unrelated question design to estimate the incidence among students of shoplifting, cheating on exams, and the expressed need for non-Caucasian professors. Two coins were used as the randomizing device. The questions were administered to 67 students, mostly of Japanese ethnicity. The randomized response estimator produced a surprisingly high rate of cheating on exams (90%). One student in seven admitted to shoplifting at least once in the past year. The study demonstrated the relative ease with which a small-scale survey using the RRM can be designed and administered.

Beldt, Daniel, and Garcha (1982) surveyed middle managers and student-employees about employee theft. The researchers used a randomizing device that consisted of a list of colors which was presented to the respondents from which they chose a color at random. Contrary to what was expected, the randomized response estimators indicated lower levels of stealing than did direct questioning. This finding might reflect the authors' choice of randomizing device.

ALCOHOL/DRUG USAGE

The stigma associated with alcohol and drug use might influence some individuals unwilling to not report their consumption and others to underreport their usage-frequency and quantity used. Barth and Sandler (1976) used the unrelated question design to estimate the proportion of high school seniors who had consumed the equivalent of at least 50 glasses of alcoholic beverages in the previous year. A significantly higher proportion of students (.85) reported consuming alcohol using randomized response method as opposed to direct questioning (.63). Similarly, Volicer and Volicer (1982) used the unrelated question design to estimate the percentage of adults who consumed alcohol every day. The percentage of respondents who drank every day was estimated to be 26%, a figure not statistically different from direct questioning estimates. However, a Phase II study on the quantity of alcohol consumed resulted in estimates that were substantially larger using RRM as compared with direct questioning (8.7 vs. 3.9 drinks per week). The differences were pronounced for women, for respondents over 52 years of age, and for those with over 12 years of education.

Volicer, Cahill, Neuburger, and Arntz (1983) used the unrelated question design to estimate the incidence of alcoholism among women employees. The estimated incidence of alcoholism was significantly higher using the RRM (34.3%) as compared with direct questioning (21.9%). The differences in estimates widened for older employees, employees in lower grades, and those with low seniority.

Duffy and Waterton (1984) used the randomized response method to estimate the distribution function of alcohol consumption. Most respondents had great difficulty standardizing beer, wine, and spirits into a common unit of measurement, and liquor consumed at home was more difficult to standardize than that consumed at a commercial establishment. The authors concluded that the additive constants randomized response model (e.g., Himmelfarb and Edgell 1980; Pollock and Bek 1976) was unsuitable for measuring the quantity of alcohol consumed.

Zdep, Rhodes, Schwarz, and Kilkenny (1979) used the unrelated question design to estimate marijuana usage for a sample of 2,084 adults. Randomized response estimates of usage were slightly higher than direct questioning estimates. Interestingly, for adults between the ages of 18 and 25, direct questioning produced a higher estimate of marijuana use than did the RRM. The authors attributed this seemingly unusual result to marijuana usage being fashionable among this age group, with negative connotations for nonusage of marijuana. For all other age groups, randomized response estimates of marijuana usage were substantially greater than those from direct questioning.

Brewer (1981) obtained the incidence of marijuana use from 365 respondents in the age group 15 to 39. The author was surprised to find that the randomized response estimates of incidence of marijuana use were lower than direct questioning estimates. He questioned the view that the RRM is useful in eliciting accurate responses to sensitive questions. An alternative explanation might be that the preponderance of young respondents in the sample skewed the results similar to what was observed by Zdep et al. (1979).

Akers, Massey, Clarke, and Lauer (1983) used the randomized response method to estimate the incidence of cigarette smoking among students. Their study had an interesting application of the bogus-pipeline procedure, wherein respondents are told that the truthfulness of their responses can be physiologically checked by the researchers. The estimates from direct questioning, bogus pipeline procedure, and RRM were approximately the same. The authors concluded that respondents might not have considered smoking to be a sensitive behavior. Therefore, their responses did not vary with the method of interviewing used.

EDUCATION

Lamb and Stem (1978) used a two-stage model to estimate the mean number of students who failed a course. The unrelated question design was administered to 121 students for the qualitative question of failing a course and to 128 students for the quantitative question of number of courses failed. The RRM was very successful in eliciting responses; only 4% of the students surveyed refused to respond. A ratio estimator approach underestimated the true number of courses failed by 25%. Failing a course was not a sensitive issue, while the number of courses failed was.

Stem and Steinhorst (1984) administered the unrelated question design to 350 students using mail questionnaires. Five types of particularly dishonest cheating were studied. The RRM produced higher estimates of cheating for some of the types and lower estimates for other types as compared with direct questioning. However, these differences were not significant. Students who were administered the RRM rated exam cheating to be significantly less serious an offense than did those in the direct questioning group. Apparently, students viewed exam cheating as a minor offense but felt socially pressured to rate it as a serious offense when directly questioned. The estimated proportion of cheaters in this study was much smaller than in the Geurts (1980) study, but this discrepancy could be partly explained by the different measures of cheating used in the two studies.

Scheers and Dayton (1987) used the unrelated question design to estimate the incidence of cheating behaviors. They observed a consistent inverse relationship between level of cheating and grade point average; the group with the lowest grade point average had the highest incidence of each of the five types of cheating. The estimated incidence of cheating using direct questioning was consistently smaller than from using the RRM. However, the differences were not statistically significant.

ACCOUNTING

Buchman and Tracy (1982) evaluated the honesty in accounting practices with a mail survey using a sample of 170 respondents. The RRM consistently produced larger estimates (but not significant) of improper accounting practices than did direct questioning (only one of seven measures was slightly lower for the RRM). However, the response rate was lower for the RRM (35%) when compared with direct questioning (50%). The authors recommended two improvements in the administration of the RRM. First, the instructions for the randomizing device should be made easy to follow. Second, when administering the directed "yes" response method, the value of answering "yes" (rather than not responding) must be made clear to respondents. Otherwise, respondents provide an answer only when the outcome of the randomizing device is the sensitive question because they see no value in answering "yes" or "no" when the outcome directs them to do so.

DEMOGRAPHY

Madigan, Abernathy, Herrin, and Tan (1976) used the unrelated question design to estimate the incidence of deaths in a province. As the reporting of deaths was viewed negatively in that society, the RRM was used in an attempt to elicit honest responses. The sample consisted of 4,000 households. The response rate was 99%. The RRM was successful in getting household members to admit to a death rate more than double that obtained from direct questioning.

Shimizu and Bonham (1978) estimated the number of women having abortions in a 12-month period using a multistage probability sampling of 4,926 women in the age group 15 to 44 years. The response rate was 98.5%. The number of abortions estimated from the RRM was far in excess of estimates prepared by national health agencies or other sources that used direct questioning as a method of obtaining responses. A major contribution of this study is in pointing out the flaws in routine statistics of sensitive attributes maintained by government health agencies.

PERSONAL AND SOCIAL BEHAVIOR

Zdep and Rhodes (1976) used a national probability sample of 2,000 adults to estimate the incidence of intentionally beating children. Although 15% of the respondents admitted beating their children when a RRM was used, the corresponding estimate with a confidential self-administered questionnaire was only 3%-4%. Further, the RRM produced a higher response rate (98%) than the self-administered questionnaire (88%).

Himmelfarb and Lickteig (1982) measured the incidence of 56 sensitive attributes of a personal nature. Similarly, Soeken and Macready (1985) used a subset of 10 of these attributes in their study of 480 undergraduate students. The sensitive personal attributes ranged from lying on the current questionnaire to kissing on the first date. Using a sample size of 198, Himmelfarb and Lickteig (1982) concluded that the RRM yielded estimates of sensitive behavior that were higher than direct questioning estimates. Soeken and Macready (1985) found that as the probability of being asked the sensitive question increased, respondent jeopardy actually decreased, that is, respondents reported higher incidences of sensitive behavior. This result belies the conventional belief that jeopardy increases as the probability of having to answer the sensitive question increases when using the unrelated question design.

OWNERSHIP OF PRODUCTS

Stem and Steinhorst (1984) conducted a telephone survey of 554 households about telephone ownership. The randomizing device was a telephone directory and 8.7% of the respondents reported having a nonleased telephone. This estimate was fairly close to independent telephone company estimates of 10%-12%. All other randomized response estimates relating to further questions on telephone ownership were poor and some were even negative. Stem and Steinhorst (1984) attributed the "breakdown" of the RRM to the questions not being sufficiently sensitive. Further, a significant nonresponse problem occurred due to the randomizing method (telephone book) used.

MEDICINE

Soeken and Macready (1986) used the unrelated question design to estimate the sensitive behavior (toward patients) of 270 graduate nursing students and faculty. Several variations of the basic RRM design were used to assess the incidence of three sensitive attributes. In general, their results indicated that incidence estimates of sensitive attributes were larger using the RRM than using direct questioning. Conversely, the incidence estimates of nonsensitive attributes were larger using direct questioning than using the RRM.

Overall, the RRM has been applied in a variety of areas. No major problems in administering the survey were reported in any of the studies. As observed in many of the studies, the degree of sensitivity of the attributes determined the success of the randomized response approach. Estimates of nonsensitive or low sensitive behavior using randomized response were sometimes inaccurate. However, randomized response estimates were not necessarily higher than estimates from direct questioning even when sensitive attributes were studied.

VALIDITY OF THE RANDOMIZED RESPONSE METHOD

As mentioned previously, the preponderance of research on RRMs has been concerned with deriving model estimators or establishing the statistical efficiency of alternative models. An underresearched topic is the validity of RRMs. Regardless of their statistical elegance, if the methods do not provide valid estimates, they are of no practical value. This section reviews evidence on the validity of RRMs, where validity refers to the extent to which they produce estimates of sensitive attributes that correspond to true values in the underlying population. Unfortunately, the number of randomized response validation studies is limited. The principal difficulty in validating the RRM is the unavailability of population parameters or true individual values for the attribute(s) of interest.

Several issues need to be addressed regarding the validity of the RRM:

(a) What is the method's achieved level of validity?
(b) What are the differences between using individual and population figures for validation?
(c) How many reported validation studies are true validation studies?
(d) Are officially published data acceptable sources of validation criteria?
(e) How does the design of a randomized response study cause jeopardy or affect the risk of self-incrimination?

These issues are discussed in the context of the validation studies reported in this section.

VALIDITY OF RANDOMIZED RESPONSE ESTIMATES

Seven studies that address the validity of the RRM are summarized in Table 2. Contrary to common beliefs (and claims), the validity of the RRM does not appear to be very good. The most accurate proportion estimate (by Lamb and Stern 1978) underestimated the corresponding population parameter by 5%, although a second sensitive attribute was underestimated by 12% to 25%. Other studies produced even poorer results. The estimation error was 35% in the Locander, Sudman, and Bradburn (1976) study and 28% for the Kulka, Weeks, and Folsom (1981) study. Tracy and Fox (1981) produced an estimate that was one-quarter of the true number, whereas Horvitz, Shah, and Simmons (1967) produced an estimate that was 20 times larger than the true value.

(text continues on p. 125)

TABLE 2: Summary of Validation Studies

Authors	*Type of Design*	*Randomizing Device Used*	*Questions Asked (All)*	*Value of p/Sample Sizes*	*Results of Validation*	*Comments of Validation*	*Actual Values*	
							Estimated	*Actual*
Horvitz, Shah, and Simmons (1967)	unrelated question design—estimating proportions with two trials	(a) deck of 50 cards with questions printed on them; (b) sealed plastic box with colored beads corresponding to the two questions	In the past 12 months there was a baby born in this household to an unmarried woman who was living here at that time.	$p_1 = .7$ $p_2 = .3$ $n_1 = 63$ $n_2 = 85$	Overestimated by 20 times. Pilot study done initially was quite accurate.	Inaccuracy attributed to mechanical faults such as the deck of cards being mixed.	.14	.008
Locander, Sudman, and Bradburn (1976)	unrelated question design—estimating proportions	plastic box with 50 beads, 35 red and 15 green	a) ownership of library card b) voter registration c) voting behavior d) bankruptcy filing e) drunk driving	$p = .7$ $n = 50$ or 100	Randomized response produced lower estimates from subjects who were known to have the sensitive characteristics in some cases but for the threatening questions on drunk driving and bankruptcy, randomized response was better.	The error rate was quite high for all methods.	error rate = 11% to 35%	

Lamb and Stem (1978)	unrelated question design—estimating proportions and quantities	bag with red and white marbles	a) ever failed college-level courses b) how many college-level courses failed	$p_1 = .75, .25$ $p_2 = .75, .25$ $n_1 = 121$ $n_2 = 128$	The randomized response technique was a relatively accurate technique. The proportion of those who failed was more accurately estimated than number of courses failed.	The accuracy might be due to the use of students. The number of courses failed appears to be a more sensitive topic than the proportion of failures, and, thus, less accurately estimated.	.363 .563 1.460	$\pi = .347$ $\mu_1 = .641$ $\mu_1/\pi = 1.97$
Tracy and Fox (1981)	unrelated question/ quantitative data design	lucite container with 25 red and 25 white balls, white balls have numbers from 1 to 8 printed on them	number of adult arrests by Philadelphia police involving formal booking	$p = .5$ $n = 410$	Underestimated arrest records by about 50%. Was worse than direct questionnaire for those with only 1 arrest.	The poor backgrounds of subjects with arrest records might have led to substantial underreporting.	.77	1.45
Kulka, Weeks, and Folsom (1981)	directed question design estimating proportions	50 beads in a box; three colors, red, white and blue	arrested for DWI in past year	$p = .7$ $n = 283$	Randomized response results were significantly worse than direct questioning.	Poor responses in randomized response attributed to the lack of careful administering of the survey. Respondents might have felt apprehensive about providing an answer because the jeopardy arising from the design parameters was quite high.	.844	1.00

(continued)

TABLE 2 Continued

Authors	Type of Design	Randomizing Device Used	Questions Asked (All)	Value of p/Sample Sizes	Results of Validation	Comments of Validation	Actual Values	
							Estimated	Actual
Aakers, Massey, Clarke, and Lauer (1983)	unrelated question design—estimating proportions	deck of cards numbered 1 through 100	I have smoked cigarettes at least once or twice.	$p = .7$ $n = 577$	Those who smoked recently had elevated levels of salivary thiocynate in their system.	Given the inexact relationship between smoking and thiocynate levels it is difficult to get exact numerical estimates of smoking.	—	—
Edgell, Himmelfarb, and Duchan (1982)	directed response randomized response design	KIM-1 microprocessors	seven sensitive questions	— $n = 32$	As the degree of sensitivity increased, respondents didn't answer the questions as directed—even when they were required to answer yes (or no) they answered differently.	Method does not say if the subjects were honest in answering the sensitive questions but only that they disregarded the directive to answer yes (or no).	—	—

Two final studies listed in the table are examples of novel validation procedures that test an aspect of the RRM. In one study (Akers et al. 1983), elevated levels of thiocyanate in saliva were used as evidence of recent cigarette smoking. It is important to note the study does not indicate all responses to randomized response questions are honest; the only conclusion is that the randomized response study results correlated with respondents' smoking habits. Edgell, Himmelfarb, and Duchan (1982) used a computer to validate the RRM. Respondents were asked to answer yes or no when certain numbers appeared on a computer screen and answer the sensitive question when other numbers appeared on the screen. Respondents systematically flouted the yes (no) requirement as frequently as 25% of the time when the alternate question (which they did not have to answer on the trial) was very sensitive. These results are particularly relevant for evaluating studies that assume respondents are completely honest when responding to nonsensitive questions.

In contrast, a meta-analysis of studies on sensitive topics such as income, drug use, and health showed direct questioning to yield surprisingly accurate results (Marquis, Marquis, and Polich, 1986). Although the estimates in each study fluctuated quite a bit from actual population values, there was no evidence of systematic underreporting (except for health questions). Caution must be used when interpreting these results. The studies dealt with sensitive topics, but there is no way to determine if the questions were as sensitive as those used in the randomized response validation studies.

INDIVIDUAL VERSUS POPULATION VALIDATION CRITERIA

For purposes of validation, comparisons can be made between reported responses and true values at either the individual level or the aggregate level. It is often easier to make comparisons at the aggregate level because of availability and accuracy of aggregate data. For instance, Blair, Sudman, Bradburn, and Stocking (1977) compared the reported alcohol consumption in a national probability sample with national consumption statistics derived from tax receipts and concluded that there was substantial underreporting of alcohol consumption.

The divergence of sample estimates from population values can be attributed to two distinct errors: response bias and response variance (Peterson and Kerin 1981). Response bias is a systematic error that is not reduced by increasing the sample size. In contrast, response variance may be due to fluctuations within the specific study conditions. In a typical survey, response variance may be due to differences that exist among interviewers. Response bias is more difficult to investigate because it requires validating data obtained from external sources for each survey respondent (as done in the validation studies). Response variance is typically investigated by comparing responses of various subgroups of survey participants in an investigation. Response variance is illuminated by the observed difference between randomized response estimates and direct questioning estimates in studies listed in the Applications section.

The validity of randomized response estimates can be determined more accurately when they are compared with true values at the individual level than at the aggregate level. An estimate derived using the RRM would likely differ from aggregate population data even if there were no self-reported errors in the estimates. These differences could be due to sampling error, nonsampling error, or both. By comparing randomized response estimates with true aggregate values, no unequivocal conclusions are possible about the validity of the estimates. However, when randomized response estimates are compared with true values at the individual level, definite conclusions are possible. No matter how biased the sampling plan, individual-level true values always equal respondent-reported values provided there are no response biases. Any difference between the individual-level true values and respondent-reported values can be attributed solely to misreporting without being concerned with the degree of sampling and nonsampling errors. Thus validity is better determined when evaluated at the individual level.

Another problem with using aggregate measures is that aggregation produces measures of central tendency such that negative response biases can offset or compensate for positive response biases (as observed by Goldstein 1971). Even if aggregate reported sample means are equal to true population means, respondents who overreport their behavior might balance underreporting of other respondents. In a study of income reporting (Peterson and Kerin 1980), half of the respondents misreported their income, however, the number underreporting and the number overreporting were about equal. Overall, using individual-level comparisons of sample values and true values is better than making aggregate-level comparisons because there are fewer sources of error in the individual-level comparisons.

STUDIES THAT TRULY VALIDATE RANDOMIZED RESPONSE ESTIMATES

We have included only seven studies as true examples of validation studies. Of these, two are considered to validate only an aspect of the RRM rather than the entire procedure. Studies that compared the RRM with other forms of questioning concluded that the higher estimates obtained by randomized response validated the method. In our opinion, these are not validation studies. A true validation study must compare the randomized response estimate and the actual value.

A study by Buchman and Tracy (1982) is typical of the multitude of quasi-validation studies of the RRM (although the study is not positioned as such). In their study, the RRM was compared with a standard direct questioning procedure. The randomized response estimates turned out to be larger than the estimates from direct questioning. Specifically, the reported incidence of improper auditing by accountants was much lower when using direct questioning than when using a RRM. Thus, the researchers concluded, as most other researchers do, the RRM was superior to direct questioning. Such a comparison, though, does not constitute a validation of the RRM.

A higher reported incidence of a sensitive behavior is not necessarily more accurate than a lower reported incidence. Without any knowledge of true values (population or individual) to validate the results, the randomized response estimates could just as likely be overestimates as underestimates. RRMs could lead to parameter overestimates if the randomizing device makes people report behaviors in which they do not indulge (at least not to the extent they indicate), if it causes respondents to incorrectly answer the nonsensitive question, or if it confuses them to such an extent that they provide incorrect answers.

USE OF PUBLISHED DATA FOR VALIDATION

Published data, even in prominent national publications, are often based on survey results. The use of such data (often official) for validation purposes is questionable. For instance, Zdep et al. (1979) validated their randomized response results with the sixth annual report from the secretary of HEW to the U.S. Congress on marijuana and health. However, report results were based on a self-reported survey (with its inherent misreporting) and there is no guarantee that they reflect actual population behavior. Population figures used for validation are appropriate criteria only if they are obtained through nonsurvey procedures – for example birth records in hospitals, grades from registrars' offices, and so forth.

Survey results can be used for validation if the responses are from a third party who has knowledge of the sensitive attribute of the population of interest. For example, a survey of doctors about the sensitive medical condition of their patients might be used to validate the answers obtained from a survey of patients using the RRM. Even so, several limitations apply. Paired data between two populations, such as patients and doctors, are typically not available. Third-party answers are also prone to misreporting. Mechanical errors increase when matching third-party responses with sample responses. Finally, serious ethical issues are raised when third-party information is obtained about individuals without informing them that such data have been obtained.

JEOPARDY

Jeopardy is the perceived threat or the consequence of disclosure from answering sensitive questions. Answering questions about sensitive attributes, no matter how much the responses are disguised, is probable cause for embarrassment. Respondents are likely to feel uncomfortable when answering questions about either a sensitive or a nonsensitive attribute because the interviewer might suspect the response corresponds to possession of the sensitive attribute. Lanke (1976) defined the lack of protection when answering "yes" as

= P[S | yes]

or the probability of possessing the sensitive attribute given that the respondent said "yes," and

the lack of protection when answering "no" as

= P[S | no]

or the probability of possessing the sensitive attribute given that the respondent said "no."

Bourke (1984) has indicated that jeopardy is reduced if a design is "symmetric." In other words, answering "yes" or "no" should not provide any information about the respondents' possession of a sensitive attribute. The commonly applied unrelated question design is not symmetric. If the two questions respectively deal with drunk driving conviction and date of birth being in April, the answer of "no" is always safer than one of "yes" because it is never associated with having the sensitive attribute.

There is obviously a trade-off between respondent jeopardy and statistical efficiency. In the unrelated question design, respondent jeopardy can be minimized if the probability (p) of being asked the sensitive question is made very small (for example, 0.01). However, this does not provide useful information because the standard deviation of the randomized response estimate increases substantially as the probability of being asked decreases. With such low values of p, the vast majority of respondents do not answer the sensitive attribute question and consequently, not much can be inferred about the sensitive attribute. Similarly, researchers must remember that in spite of apparent statistical efficiency, too high a level of respondent jeopardy will likely lead to more untruthful answers to questions about sensitive attributes. Fox and Tracy (1986) go into considerable detail on how to effectively and efficiently protect survey respondents and minimize survey bias.

In summary, researchers have modeled jeopardy by the statistical probabilities of (the interviewer or researcher) being able to infer the sensitive attribute of respondents. Although Bayesian rules have been derived whereby jeopardy is related to the probability of possessing the sensitive attribute, there are two major problems with them. First, respondents are generally poor at estimating probabilities (cf. Tversky and Kahneman 1974), particularly those of a complex nature. In RRM applications, respondents have been known to not check on the actual values of the probabilities (Duffy and Waterton 1988). Thus respondents might view certain response options as being more risky even though statistically they are not. Second, jeopardy has to be measured as it occurs in the minds of respondents. Some respondents may feel no jeopardy in answering a question whereas the same question might cause great embarrassment for others (e.g., Peterson and Ridgway 1986). For some respondents, even answering "no" to a sensitive question might cause discomfort.

CONCLUSIONS AND RESEARCH AGENDA

The RRM is increasingly being used in survey research, although it is nowhere as popular as direct questioning. Its applications encompass a broad range of social science areas, and research on it no longer focuses on only finding better estimation techniques.

RRM was designed for eliciting responses to sensitive questions, hence it has found numerous applications in areas such as criminal justice and substance abuse. Although some studies provided evidence of a higher incidence of the sensitive attribute when the RRM was used, direct questioning was just as successful in eliciting sensitive responses in other studies. RRM has also been used in areas such as education, demography, and medicine. Again, there was no consistent evidence that RRM produced higher estimates of sensitive attributes.

The results from validation studies add to the findings from application studies. In many cases, estimates using RRM were found to vary considerably from actual values. Therefore, if a higher incidence of a sensitive attribute is obtained using RRM, it is improper to assume that this estimate is necessarily accurate. Both RRM and direct questioning might yield inaccurate estimates. Further research is necessary to make RRM an accurate procedure for estimating the incidence of sensitive attributes.

For any survey method to have widespread use in the future, it must: (a) be easy for the respondents to comprehend, (b) be easy to administer, (c) provide valid (accurate) and reliable estimates of parameters of interest, (d) be easy to integrate into computer-based interviewing, and (e) be cost efficient in terms of design, administration time, and analysis.

The RRM suffers seriously in the validation area and is somewhat worse than direct questioning with respect to simplicity and low cost. Lack of validation is the hardest problem to overcome. So far, researchers have not been concerned with reliability of the RRM. Research is needed to estimate both the validity and the reliability of the RRM.

It is appropriate and perhaps inevitable that a review of the literature in the area of obtaining responses to sensitive questions be followed by a call for more research. The research reviewed in the preceding pages provides many examples of areas that have been investigated minimally or not at all. Three broad areas that need investigation are briefly described below.

ASSUMPTIONS OF THE RANDOMIZED RESPONSE METHOD

The RRM was developed with the assumption respondents will be honest because their answers to sensitive questions cannot be linked to them. However, the few validation studies conducted so far suggest there is some degree of misreporting for sensitive questions as well as for the nonsensitive questions in the RRM. Misreporting of answers, be it from dishonesty, confusing instructions, or mechanical errors, produces incorrect estimates regardless of whether direct questioning or the RRM has been used. The assumption of honesty is the cornerstone of the RRM. Research must be directed toward determining if this assumption is true. Specifically, research must be conducted to answer the following questions:

1. Under what circumstances do respondents provide dishonest answers to sensitive questions? Is this dishonesty systematically related to the questions asked, demographic characteristics, type of randomizing device used or the specific design?

2. How believable (to the respondent) is the concept that the randomizing device provides anonymity? Is believability related to factors that influence honesty?
3. Do designs that reduce respondent jeopardy (Kuk 1990) relieve respondent anxiety? The concept of jeopardy researched so far relies heavily on the assumption that respondents are able to mentally calculate probabilities associated with each design and, further, use Bayes's theorem to calculate the degree of self-incrimination from each outcome. Given the evidence presented by psychological researchers (cf. Nisbett and Ross 1980; Tversky and Kahneman 1974) that even well-educated respondents have difficulty in computing even the simplest of probabilities, how realistic is this assumption? What is the relationship between statistical jeopardy and psychological jeopardy, and what design factors mediate this relationship? Do randomizing devices that provide the same level of statistical jeopardy provide the same level of psychological jeopardy? Studies that provide guidelines for minimizing psychological jeopardy are in order.

INTEGRATION WITH ALTERNATIVE TECHNIQUES

Randomized response is but one method of obtaining data on sensitive attributes. Other methods include observation, projective techniques, modification of question wording and sentence structure, and foot-in-the-door techniques. At issue is not how best to use the RRM, but how best to obtain accurate responses to questions regarding sensitive attributes. Several questions stand out in the broader context of obtaining data on sensitive attributes. These include the following:

1. With respect to response rates and accuracy, how successful are alternative methods at obtaining responses? Empirical studies are needed to validate each of the methods under different design conditions. The results should provide guidelines for research situations in which the RRM is best suited.
2. Is it possible to create scales that rate the sensitivity of attributes and/or questions? Respondents have been known to rate questions on certain topics as more threatening than others (Peterson and Ridgway 1986). Factors studied should include nature of topic area, individual characteristics, structure and design of questions, and question/method interaction.
3. Does assurance of anonymity produce a psychological feeling of security? Observational methods relying on physiological responses

could be applied to measure perceptions of anonymity and levels of discomfort. Do RRMs produce more accurate responses than verbal assurances of anonymity?

VALIDATION OF RANDOMIZED RESPONSE RESULTS

There have been too few studies empirically validating the RRM. Some researchers, content with observing that the RRM was designed to provide valid results, assume that it does so. Others appear to assume that the RRM is valid because estimates of sensitive behavior using the method are generally higher than when using direct questioning procedures. Validation studies are necessary, because the few conducted so far do not show the RRM to have a high degree of validity. The following aspects of validity need to be investigated:

1. Better methods of validation should be devised. Due to the sensitive nature of the topic area, external criteria for comparisons are hard to find. More creativity is needed in selecting criteria and being able to match responses to these criteria. Records of sensitive behavior are generally confidential and difficult to obtain. Physiological tests using skin resistance, eye movement, and brain waves to validate responses might alleviate the problem of obtaining confidential records from private or public sources. In this regard, an important beginning has been made by Stem and Lamb (1990). Several such studies are called for. Efforts must be made to obtain independent, wholly unfalsifiable measures of behavior and these should be compared with RRM results.
2. Questions challenging the basic assumptions of the randomized response method need to be answered. Are there systematic reasons why some people refuse to acknowledge a sensitive attribute? Are refusals related to such things as the outcome of the randomizing device, the design of the study, or demographic characteristics? Why do respondents misreport nonsensitive attributes when asked in the context of a RRM?
3. The issue of randomizing devices has received, at best, perfunctory discussion in the literature. Randomizing devices are expected to provide anonymity as specified and the exact probabilities as designed. This may not often be the case. If the actual probabilities produced by a randomizing device are different from the designed probabilities, estimates of the sensitive attributes will be incorrect. Horvitz, Shah, and Simmons (1967) alluded to improper functioning

of their randomizing device for their poor estimates. Performance of the RRM by mail is in part contingent on the ease of use of the randomizing device.

The RRM can easily be integrated into computer-based interviewing. To maintain respondents' belief in their anonymity, the randomizing device should be external to the computer. If the random process is generated by the computer, the respondents' answers can be checked for accuracy by comparing the responses to what was generated by the computer. The method is compatible with the recent practice of conducting interviews without interviewers, using personal computers in shopping malls.

More research is necessary before the success of the RRM in any study design can be anticipated prior to its use. Given the uncertainties surrounding the appropriateness of the RRM in any particular study, many researchers may hesitate to replace traditional direct questioning with the more complex RRM. It does take a certain amount of time for respondents to become familiarized with the notion of a randomizing process and to be convinced that the method truly provides anonymity. Because of this, the duration of the interview must increase. Fewer interviews can be completed per time period using the RRM than using direct questioning. Researchers must then consider this question: Using a smaller sample, does the randomized response estimate have a lower mean squared error of estimate than does traditional direct questioning? The adoption of a new research methodology is similar to the adoption of a new technology: Unless the new technology is substantially better than the old technology, inertia will make people unwilling to give up the old technology.

REFERENCES

Akers, R. L., J. Massey, W. Clarke, and R. M. Lauer. 1983. "Are Self-Reports of Adolescent Deviance Valid? Biochemical Measures, Randomized Response, and the Bogus Pipeline in Smoking Behavior," *Social Forces* 62:234-51.

Barth, J. T. and H. M. Sandler. 1976. "Evaluation of the Randomized Response Technique in a Drinking Survey." *Journal of Studies on Alcohol* 37:690-93.

Beldt, S. F., W. W. Daniel, and B. S. Garcha. 1982. "The Takahasi-Sakasegawa Randomized Response Technique: A Field Test." *Sociological Methods & Research* 11:101-11.

Blair, E., S. Sudman, N. M. Bradburn, and C. Stocking. 1977. "How to Ask Questions About Drinking and Sex: Response Effects in Measuring Consumer Behavior." *Journal of Marketing Research* 14:316-21.

Bourke, P. D. 1981. "On the Analysis of Some Multivariate Randomized Response Designs for Categorical Data." *Journal of Statistical Planning and Inference* 5:165-70.

———. 1984. "Estimation of Proportions Using Symmetric Randomized Response Designs." *Psychological Bulletin* 96:166-72.

Bourke, P. D. and M. A. Moran. 1988. "Estimating Proportions from Randomized Response Data Using the EM Algorithm." *Journal of the American Statistical Association* 83:964-68.

Brewer, K.R.W. 1981. "Estimating Marihuana Usage Using Randomized Response—Some Paradoxical Findings." *Australian Journal of Statistics* 23:139-48.

Buchman, T. A. and J. A. Tracy. 1982. "Obtaining Response to Sensitive Questions: Conventional Questionnaire versus Randomized Response Technique." *Journal of Accounting Research* 20:263-71.

Chaudhuri, A. and R. Mukherjee. 1988. *Randomized Response: Theory and Techniques*. Statistics: Textbooks and Monographs, Vol. 85. New York: Marcel Dekker, Inc.

Clickner, R. P. and B. Iglewicz. 1980. "Warner's Randomized Response Technique: The Two Sensitive Questions Case." *South African Statistical Journal* 14:77-86.

Duffy, J. C. and J. J. Waterton. 1984. "Randomized Response Models for Estimating the Distribution Function of a Quantitative Character." *International Statistical Review* 52: 165-71.

———. 1988. "Randomized Response vs. Direct Questioning: Estimating the Prevalence of Alcohol Related Problems in a Field Survey." *Australian Journal of Statistics* 30:1-14.

Edgell, S. E., S. Himmelfarb, and D. J. Cira. 1986. "Statistical Efficiency of Using Two Quantitative Randomized Response Techniques to Estimate Correlation." *Psychological Bulletin* 100:251-56.

Edgell, S. E., S. Himmelfarb, and K. L. Duchan. 1982. "Validity of Forced Responses in a Randomized Response Model." *Sociological Methods & Research* 11:89-100.

Fox, J. A. and P. E. Tracy. 1984. "Measuring Associations with Randomized Response." *Social Science Research* 13:188-97.

———. 1986. *Randomized Response: A Method for Sensitive Surveys*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-058. Beverly Hills, CA: Sage.

Geurts, M. D. 1980. "Using a Randomized Response Research Design to Eliminate Non-Response and Response Biases in Business Research." *Journal of the Academy of Marketing Science* 8:83-91.

Goldstein, I. L. 1971. "The Application Blank: How Honest Are the Responses?" *Journal of Applied Psychology* 55:491-92.

Greenberg, B. G., A. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz. 1969. "The Unrelated Question Randomized Response Model: Theoretical Framework." *Journal of the American Statistical Association* 64:520-39.

Greenberg, B. G., R. R. Kuebler, Jr., J. R. Abernathy, and D. G. Horvitz. 1971. "Application of the Randomized Response Technique in Obtaining Quantitative Data." *Journal of the American Statistical Association* 66:243-50.

Gunel, E. 1985. "A Bayesian Comparison of Randomized and Voluntary Response Sampling Models." *Communications in Statistics, Theory and Methods* 14:2411-35.

Himmelfarb, S., and S. E. Edgell. 1980. "Additive Constants Model: A Randomized Response Technique for Eliminating Evasiveness to Quantitative Response Questions." *Psychological Bulletin* 87:525-30.

Himmelfarb, S. and C. Lickteig. 1982. "Social Desirability and the Randomized Response Technique." *Journal of Personality and Social Psychology* 43:710-17.

Horvitz, D. G., B. V. Shah, and W. R. Simmons. 1967. "The Unrelated Question Randomized Response Model." Proceedings of the Social Statistics Section, American Statistical Association, 65-72.

Kraemer, H. C. 1980. "Estimation and Testing of Bivariate Association Using Data Generated by the Randomized Response Technique." *Psychological Bulletin* 87:304-8.

Kuk, A.Y.C. 1990. "Asking Sensitive Questions Indirectly." *Biometrika* 77(2):436-38.

Kulka, R. A., M. F. Weeks and R. E. Folsom, Jr. 1981. "A Comparison of the Randomized Response Approach and Direct Question Approach to Asking Sensitive Survey Questions." Working Paper, Research Triangle Institute, NC.

Lamb, C. and D. E. Stem, Jr. 1978. "An Empirical Validation of the Randomized Response Technique." *Journal of Marketing Research* 15:616-21.

Lanke, J. 1976. "On the Degree of Protection in Randomized Interviews." *International Statistical Review* 44:197-203.

Levy, K. J. 1976. "The Randomized Response Technique and Large Sample Pairwise Comparisons Among the Parameters of *k* Independent Binomial Populations." *The British Journal of Mathematical and Statistical Psychology* 29:257-62.

———. 1977. "The Randomized Response Technique and Large Sample Comparisons Among the Parameters of *k* Independent Binomial Populations." *Psychological Bulletin* 84:244-46.

Locander, W., S. Sudman, and N. Bradburn. 1976. "An Investigation of Interview Method, Threat, and Response Distortion." *Journal of the American Statistical Association* 71: 269-75.

Madigan, F. C., J. R. Abernathy, A. N. Herrin, and C. Tan. 1976. "Purposive Concealment of Death in Household Surveys in Misamis Oriental Province." *Population Studies* 30:295-303.

Marquis, K. H., M. S. Marquis, and J. M. Polich. 1986. "Response Bias and Reliability in Sensitive Topic Surveys." *Journal of the American Statistical Association* 81:381-89.

Moors, J.J.A. 1971. "Optimization of the Unrelated Question Randomized Response Model." *Journal of the American Statistical Association* 66:627-29.

Nisbett, R. and L. Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgement*. Englewood Cliffs, NJ: Prentice-Hall.

O'Hagan, A. 1987. "Bayes Linear Estimators for Randomized Response Models." *Journal of the American Statistical Association* 82:580-85.

Peterson, R. A., and R. Kerin. 1980. "Household Income Data Reports in Mail Surveys." *Journal of Business Research* 8:301-13.

———. 1981. "The Quality of Self-Report Data: Review and Synthesis." Pp. 5-20 in *Review of Marketing*, edited by B. Enis and K. Roering. Chicago: American Marketing Association.

Peterson, R. A. and N. Ridgway. 1986. "A Note on the Perception of Threatening Questions." Proceedings of American Statistical Association, 433-37.

Pitz, G. F. 1980. "Bayesian Analysis of Random Response Models." *Psychological Bulletin* 87:209-12.

Pollock, K. H., and Y. Bek. 1976. "A Comparison of Three Randomized Response Models for Quantitative Data." *Journal of the American Statistical Association* 71:884-86.

Poole, W. K. 1974. "Estimation of the Distribution Function of a Continuous Type Random Variable Through Randomized Response." *Journal of the American Statistical Association* 69:1002-5.

Poole, W. K. and C. A. Clayton. 1982. "Generalizations of a Contamination Model for Continuous Type Random Variables." *Communications in Statistics—Theory and Methods* 11:1733-42.

Reinmuth, J. E. and M. D. Geurts. 1975. "The Collection of Sensitive Information Using a Two-Stage, Randomized Response Model." *Journal of Marketing Research* 12:402-7.

Scheers, N. J. and C. M. Dayton. 1987. "Improved Estimation of Academic Cheating Behavior Using the Randomized Response Technique." *Research in Higher Education* 26:61-69.

———. 1988. "Covariate Randomized Response Models." *Journal of the American Statistical Association* 83: 969-74.

Shimizu, I. M. and G. S. Bonham. 1978. "Randomized Response Technique in a National Survey." *Journal of the American Statistical Association* 73:35-39.

Soeken, K. L., and G. B. Macready. 1985. "Randomized Response Parameters as Factors in Frequency Estimates." *Educational and Psychological Measurement* 45:89-100.

———. 1986. "Application of Setwise Randomized Response Procedures for Surveying Multiple Sensitive Attributes." *Psychological Bulletin* 99:289-95.

Spurrier, J. D. and W. J. Padgett. 1980. "The Application of Bayesian Techniques in Randomized Response." Pp. 533-44 in *Sociological Methodology*, edited by Karl F. Schuessler. San Francisco: Jossey-Bass.

Stem, D. and C. W. Lamb, Jr. 1990. "The Relative Efficiency of Randomized Response Models in Respondent Anxiety Reduction," Working Paper, College of Business and Economics, Washington State University.

Stem, D. and R. K. Steinhorst. 1984. "Telephone Interview and Mail Questionnaire Applications of the Randomized Response Model." *Journal of the American Statistical Association* 79:555-64.

Tamhane, A. C. 1981. "Randomized Response Techniques for Multiple Sensitive Attributes." *Journal of the American Statistical Association* 76:916-23.

Tracy, P. E. and J. A. Fox. 1981. "The Validity of Randomized Response for Sensitive Measurements." *American Sociological Review* 46:187-200.

Tversky, A. and D. Kahneman. 1974. "Judgement Under Uncertainty: Heuristics and Biases." *Science* 185:1124-31.

Volicer, B. J., M. H. Cahill, E. Neuburger, and G. Arntz. 1983. "Randomized Response Estimates of Problem Use of Alcohol Among Employed Females." *Alcoholism: Clinical and Experimental Research* 7:321-26.

Volicer, B. J. and L. Volicer. 1982. "Randomized Response Technique for Estimating Alcohol Use and Noncompliance in Hypertensives." *Journal of Studies on Alcohol* 43:739-50.

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

Winkler, R. L. and L. A. Franklin. 1979. "Warner's Randomized Response Model: A Bayesian Approach." *Journal of the American Statistical Association* 74:207-14.

Zdep, S. M. and I. N. Rhodes. 1976. "Making the Randomized Response Technique Work." *Public Opinion Quarterly* 40:531-37.

Zdep, S. M., I. N. Rhodes, R. M. Schwarz, and M. J. Kilkenny. 1979. "The Validity of the Randomized Response Technique." *Public Opinion Quarterly* 43:544-49.

U. N. Umesh is an associate professor of marketing at Washington State University. He received his Ph.D. in business administration from the University of Washington, his MBA from the University of Rochester and a B-Tech in mechanical engineering from the Indian Institute of Technology, Madras. His publications have appeared in journals such as Psychometrika, Journal of Marketing, Decision Sciences *and* Educational and Psychological Measurement.

Robert A. Peterson holds the John T. Stuart III Centennial Chair in Business Administration and the Charles C. Hurwitz Fellowship at the IC^2 Institute of the University of Texas at Austin. Peterson is a former chairman of the Department of Marketing Administration at the University of Texas at Austin and has been the editor of the Journal of Marketing Research. *He received his doctorate from the University of Minnesota. He has authored or coauthored more than 120 articles and books. His publications have appeared in more than three dozen journals, including* Management Science, *the* Journal of Business, *the* Journal of Marketing Research, *and the* Journal of Marketing. *His books and monographs include* Marketing Research, Strategic Marketing Problems *(with Roger A. Kerin),* The Role of Affect in Consumer Behavior *(with Wayne Hoyer and William Wilson), and* Models for Innovation Diffusion *(with Vijay Mahajan).*