

DISCUSSION

D. G. Horvitz, Research Triangle Institute

My evaluation of this paper is, in general, favorable. The authors are to be congratulated for carrying out a type of study which is badly needed by those of us interested in the design of surveys and in the overall quality of survey data. Despite the widely recognized need to be concerned with the magnitude of reporting biases in surveys, significant validation studies have been rather few in number.

The results of this study indicate, at least in part, the magnitude and range of the net bias for questions of a "threatening" nature. The authors prefer to use the term "response distortion" rather than "response bias" and I would appreciate a comment as to why they choose to deviate from usual terminology. Since the estimated "distortions" or "biases" for the random samples (i.e., library card, voter registration, and primary voter) range from .11 to .48 over the four methods of administration, I would be quite concerned about the validity of survey data on these items regardless of the method of administration. The proportions of distorted responses shown in Table 2 for "bankruptcy" and "drunken driving" do not have the same immediate interpretation (of consequences) since the samples for these items were all in the sensitive behavioral category. Thus, if only 2 percent of all drivers have been charged with driving under the influence then an estimate with a .35 reporting bias among "drunk drivers" would lead to an estimate of 1.3 percent in this category among all drivers or a net bias of .7 percent. The relative bias of .7/2 or 35 percent remains large, of course.

While the authors devote a considerable portion of their paper to comparisons of the four methods of administration, I am not certain that very much of their discussion is justified since, as the paper notes, there are few, if any, statistically significant differences. This does not imply that, to quote the paper, "one is free to use whatever procedure is most convenient" end quote. A more appropriate way to proceed, in my opinion, would be to take a total survey design approach in which a portion of the total resources available for data collection is allocated to validating the actual state or behavioral class of a sample of the respondents (provided such validation is possible). Thus, in the case of the item "voting in the most recent primary election," some of the survey budget would be spent on checking official voting records for some random portion of the total sample in order to estimate the net bias in this item. Thus, a two phase or double sampling scheme is envisaged in which a cost function (including the costs of collecting the data directly from a large sample together with the costs of error validation for a subsample) together with the variance function is used to determine the most efficient allocation of resources to the two phases.

This is but a simple example of the total survey design approach in which a balanced

allocation of available resources among the various sources of error is sought. Recently, Judith Lessler, completed a thesis concerned with this very problem. She was able to determine, using total error models for both self-administered and interviewer administered instruments, conditions under which one should (a) allocate their entire resources to a single but faulty measurement method or (b) use a double sampling scheme with some resources allocated to "error-free" (or validation) measurements or (c) allocate all resources to "error-free" measurements. The error models used by Lessler included response variance components as well as sampling variance and net bias in the self-administered case and (in addition) correlated response variance terms in the interviewer administered case.

The optimum allocation (or sample sizes) depends, naturally, on costs as well as on the type and magnitude of the various components of error. Clearly, data on the costs of the various alternative data collection methods as well as the biases and error variances associated with each method are essential to achieving optimum or balanced survey designs. Studies such as just was reported here should emphasize estimating the cost components as well as measurement error components.

The comparison (in this paper) of the randomized response (RR) technique with direct questions apparently is based solely on the estimated bias (response distortion). Since the variance of the RR estimate is larger than for a direct question with the same sample size, it would have been better to compare root mean square errors, rather than just the net biases. Or better yet, root mean square errors for a given data collection investment.

The paper does not describe the RR devices that were used. It clearly should in any published version. The paper also attempts to minimize the potential value of RR procedures which I don't feel is warranted. RR as a device for reducing the bias in responses to threatening questions is in the earliest stages of its development. It has had rather varied success in tests by others, some providing very positive and heartening results; others were complete failures or at best not very good.

It is noted that RR provided an exact answer for the bankruptcy item in this study, but an underreport bias of .35 for "drunk driving." A recent validation test by Folsom with a sample of indicted drunk drivers (DUI's) had rather different results, with 15 percent underreporting with the direct question (self-administered) and 28 percent underreporting with RR. The results here today, taken at face value, seem to contradict Folsom's results. Folsom noted that Locander et al. also used the RR device on several less sensitive questions before the DUI trial which

may have helped to convince respondents that the device was truly random and did in fact protect their privacy.

Folsom concluded that there is a definite need for more extensive research into alternate ways of asking sensitive questions and I agree. This means testing a wide range of questioning methods, including RR. Coin flip devices which have been tried by several researchers offer some promise. One coin flip version, for example, suggested by Dawes instructs the respondent to say "yes" if the coin comes up "heads" and "to answer the sensitive question" if the coin comes up "tails." There is a large variance penalty with this technique--the variance of the estimate of the proportion with the sensitive attribute is four times the variance for a direct question with no response bias. On the other hand, the respondents know that at least half the time (on the average) there will be a "yes" response which offers quite a bit of protection. [Folsom used a device with 35 red, 4 white and 11 blue beads. Thus the probability was only .22 that a "yes" response refers to the selection of a nonsensitive blue bead compared to .50 with the coin flip.]

Locander et al. conclude that "threatening questions have high response biases" and I agree. They also conclude that these response biases "are not greatly affected by the way in which the question is asked, even if privacy is preserved." I cannot agree that this conclusion is warranted from their study. Much more research and testing of alternatives is needed. Ultimately, as I have attempted to say earlier, we need survey designs which will permit the total error to be assessed by the survey data, just as probability samples permit objective assessment of the sampling error. Such survey designs could involve a combination of direct questioning and RR.

We live in a period where concerns about preserving privacy and confidentiality are paramount. As survey researchers we need to be prepared with techniques which will protect respondents to the fullest extent from invasions of privacy. I urge you all to undertake methodological studies--particularly validating studies--of the now very numerous alternative RR procedures for collecting data on personally threatening subjects. These studies are essential since, in my view, without them we may find ourselves in a position of being prevented by law from collecting any data on these subjects in any form or manner whatsoever. This may sound overly pessimistic to many of you. I don't think it is--there is already legislation before the Congress which suggests, if enacted, that the ability to collect important social and behavioral data considered to be of a private nature will be limited in the future.

References Cited

- Dawes, Robyn M., "Guttman Scaling Randomized Responses: A Technique for Evaluating the Underlying Structures of Behaviors to Which People May Not Wish to Admit," Oregon Research Institute, University of Oregon, March 1974 (Mimeo.).
- Folsom, Ralph E., "A Randomized Response Validation Study: Comparison of Direct and Randomized Reporting of DUI Arrests," Final Report, 255U-807, Research Triangle Institute, August 1974.
- Lessler, Judith T., "A Double Sampling Scheme Model for Eliminating Measurement Process Bias and Estimating Measurement Errors in Surveys," Ph.D. Dissertation, University of North Carolina at Chapel Hill, 1974.