SURVEY RESPONSE ERRORS FOR SENSITIVE TOPICS

Kent H. Marquis, M. Susan Marquis, and J. Michael Polich
The Rand Corporation

INTRODUCTION

Social policy research must often use the interview survey method to investigate aspects of life that we prefer to keep private, such as amounts and sources of income, alcoholism, use of drugs, illegal activities, mental illness, and intimate diseases. If respondents do not tell the truth about private matters, and if they have faulty memories, researchers legitimately ask whether interview surveys can be used to study sensitive topics.

Our objective in this paper is to examine how much error there is in responses to sensitive survey questions. Our approach is to estimate response errors from the existing literature on six sensitive subject matter areas: receiving welfare, income, drug use, alcohol use, criminal history, and embarrassing medical conditions.1/ The estimates of average response bias center on zero, suggesting that the typical respondent does not withhold sensitive personal information in these surveys. The unreliability estimates tend to be large, suggesting that response unreliability is the predominant form of response error in sensitive topic survey data.

DEFINITIONS

An individual response error occurs when an answer to a survey question isn't the same as the true value of the characteristic being measured. For example, if a family with a $10,000 income reports its income as $9,000, a response error has occurred.

We define two fundamental kinds of response errors--systematic and random. A systematic error, or bias, is a mistake the respondent is likely to make repeatedly in a variety of circumstances. For example, if the individual would consistently report only 80 percent of his actual alcohol consumption. Random response errors are the inconsistent mistakes the respondents seemingly make at random. So any individual's answer to a survey question contains three elements: true value, systematic errors, and random error.

The set of individual answers obtained from a survey also contains these true values, biases, and random errors. So we will define the two characteristics of response errors in a set of survey answers as:

> Average Response Bias--the average value of the individual systematic response biases across respondents.
>
> Random Response Error Variance--the average value of the variance of the random response errors across respondents.

RESPONSE ERROR ESTIMATORS

We obtained estimates of the two response error parameters by reviewing full-design, criterion validity studies. These are studies that compare questionnaire responses to external criterion information (e.g., official records or biological test results) that is presumtively accurate or at least minimally biased. By full designs, we mean studies that compare all survey responses to criterion values. We do not mention partially designed criterion validity studies--those that question (or analyze) only people with a positive record value; and we ruled out studies that check records only if the respondent admits something sensitive.

To estimate average response bias, we subtract the criterion mean from the survey mean. The two means are estimated on the same sample of respondents. We express the bias result as a percentage of the criterion mean.2/

We use reliability coefficients to give information about estimated random response error variances. Most of our reliability estimates are correlations between survey and criterion values; some are correlations between answers given on two different occasions, and a few are structural equation parameter estimates.

The reliability coefficient reflects the percentage of total measured variance that is due to systematic sources (true score and sytematic response biases). One minus the reliability coefficient reflects the percentage of total measured variance that is due to random response error variance.3/

RESULTS OF THE LITERATURE REVIEW

How good are answers to sensitive questions? We will look at the average response bias and reliability estimates from the literature and observe some unexpected results. In terms of average response bias, the sensitive-question answers look good. In terms of reliability, however, some potential problems are uncovered.

Average Response Bias

For most of the sensitive topics in Fig. 1, the average response bias estimates center around zero or no bias. The estimates are from full-design, criterion validity studies and each asterisk represents an estimate from one such study. Estimates to the left of zero indicate the percentage that the survey responses underestimate the criterion mean; estimates to the right of zero indicate the percentage that the survey responses overestimate the criterion mean.

The group of studies labeled "welfare" asked respondents if they receive particular kinds of government transfer payments, such as Social Security or Aid to Families with Dependent Children (AFDC). Or they ask about the amounts received. Although this is a sensitive topic, only one of the seven bias estimates is importantly negative.

Similarly, the income studies produce average response bias estimates close to zero or slightly positive. In these studies, respondents were asked to report either total income or wage and salary income. Although surveys are known to underestimate income aggregates, this underestimation does not appear to be due to an average response bias. There are other sources of potential estimation error that we do not discuss here (e.g., sampling, definitions, treatment of missing observations). It is possible that these other sources cause survey estimates of income aggregates to be less than

estimates made directly from administrative records.

The drug and alcohol estimates appear more spread out than the income estimates but, again, they don't center on a negative bias as most of us might expect. The drug studies compare self-reports of current illegal drug use with urine test results. Half of these studies yield negative estimates of the average response bias, and the other half yields positive estimates. There are some record-check studies not included in the figure; they yield very large, positive bias estimates and we suspect that the records used as criteria are very incomplete.

Of the six alcohol-use estimates, the right-hand four are from studies asking automobile drivers or alcoholics in treatment if they drank recently. Although both groups have a good reason to deny drinking, the criterion validity studies suggest that they don't cover up. The two negative estimates on the left are from Rand studies comparing reported amounts of recent drinking to breath-test results. Both suggest that self-reports of amounts consumed are underreported by about 25 percent. The breath test is a new technology in social research, so we are reluctant to conclude that these respondents told the truth about whether they drank but underreported amounts; nevertheless, we should be aware that alcoholics may underestimate amounts drunk in the recent past when answering survey questions.

The crime studies compare self-reported arrests and convictions with official records. The bias estimates are more likely to be positive than negative, which suggests, again, that survey respondents will report the socially disapproved features of their past. One study did yield a large, negative bias estimate, but its methods were not very appropriate. Among other problems, its questionnaire did not ask specifically about arrests and convictions.

Our last category includes bias estimates for reports of socially embarrassing health problems, specifically mental illness, diseases of the genitourinary system, and problems in the anal region. These studies yield large, average, response bias estimates that are mostly negative. We suspect that the method of questioning may be at fault: respondents were asked to report their health problems, aided by some probe questions covering a few general disease categories. The criteria, on the other hand, are detailed medical descriptions of definite and suspected primary, secondary, and ruled-out diagnoses, often with several entries per person for each disease. It isn't surprising that this kind of medical detail isn't reported in surveys that ask only general health questions. On the other hand, hemorrhoids is a category that is specifically asked about in the surveys. The two positive bias estimates at the botton of Fig. 1 are for this disease (namely, they are not underreported).

So, it is possible to find a class of embarrassing information that is usually underreported in surveys; but the underreporting is probably caused by poor questionnaire design or the use of inappropriate validating criteria rather than by the sensitivity of the topic.

In sum, taking into consideration all of Fig. 1, average response bias for most of the sensitive topics centers around zero or maybe a small positive value. The estimates do not support the hypothesis that people consistently deny their undesirable attributes in surveys.

Response Reliability

Our main point about response reliability, from Fig. 2, is that it is seldom close to 100 percent for these sensitive topics. The scales in Fig. 2 run from zero on the left (indicating the lowest possible reliability) to 100 on the right (indicating the highest value possible). The closer the estimate is to 100, the more reliable are the survey responses. The estimates are correlations of survey responses with either criterion values or answers to a second (retest) survey on the same topic.

Let's arbitrarily define reliability to be a potential problem if it is lower than 70. Only the income reliability estimates in Fig. 2 are consistently above this value. The rest of the topics yield estimates that are mostly below 70, which suggests to us that special steps are needed to protect users from making seriously distorted estimates of correlations, regressions, and other relationships because of the biasing effects of unreliability on such estimates.

CONCLUSIONS

We reach some unexpected conclusions based on our estimates of survey response errors from the existing literature on sensitive topics. The conclusions suggest that we should redirect the attention we give to survey response errors.

Fifty-two estimates of average response bias were made for six sensitive topics. With some exceptions, the trends indicate either no average response bias or a slightly positive one. The new estimates do not support our initial expectations that respondents will usually deny their socially undesirable attributes. Of the negative bias estimates we found, many can be traced to not asking the respondent a specific question about the desired item of information. But others, such as an underestimation of the amount of alcohol consumed, are less easily explained.

Why do we conclude that average response bias is zero or positive while others believe it is negative? We suspect that others give too much weight to inappropriately designed evaluations, such as aggregate or incomplete comparisons. The comparisons of aggregates from different samples is subject to many sources of error besides a survey response bias. We hypothesize that sampling biases and definitional differences may be more important than is usually assumed. An incomplete comparison of survey and criterion values can be very misleading also. The most common incomplete comparison--evaluating survey responses only if the record indicates the presence of the sensitive characteristic--is almost guaranteed to produce a negative estimate of the average response bias. In its pure form, this comparison cannot detect a positive response bias, and it mistakes up to half of the random response error for a negative response bias. An uncritical acceptance of results from this kind of comparison could underlie the belief that survey responses are negatively biased.

Equally unexpected was the consistent tendency for response reliability to be in the "problem range." The preoccupation of the field with response bias may have distracted us from the real

problem for policy research, namely, the misestimation of the relationships possibilities caused by substantial amounts of random response error in the data.

An implication of these findings is that average response biases need less attention and response reliabilities need a lot more and other sources of error, besides response mistakes, are responsible for survey underestimates of population means. Other sources include frame biases, undercoverage, nonresponse, definition differences and imputation procedures. So our recommendation to survey methodologists is to worry less about response biases in sensitive topic surveys and to focus more on these other kinds of and sources of error.

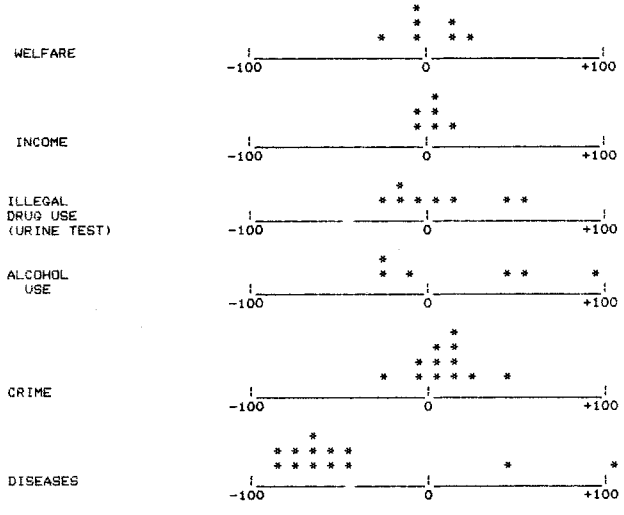
FOOTNOTES

1/ This paper summarizes a review in K. Marquis, et al., Response Errors in Sensitive Topic Surveys: Estimates, Effects, and Correction Options, R-2710/2, The Rand Corporation, Santa Monica, CA, 1981. The complete report provides more detail about each of the studies included in the review.

2/ It would be desirable to estimate the components of the average response bias, such as bias correlated with the true score or with important demographic variables. Since information available in the literature seldom allows us to do this, we estimate the average of these more complicated biases. If the averages are close to zero, it is much less likely that the complex biases exist.

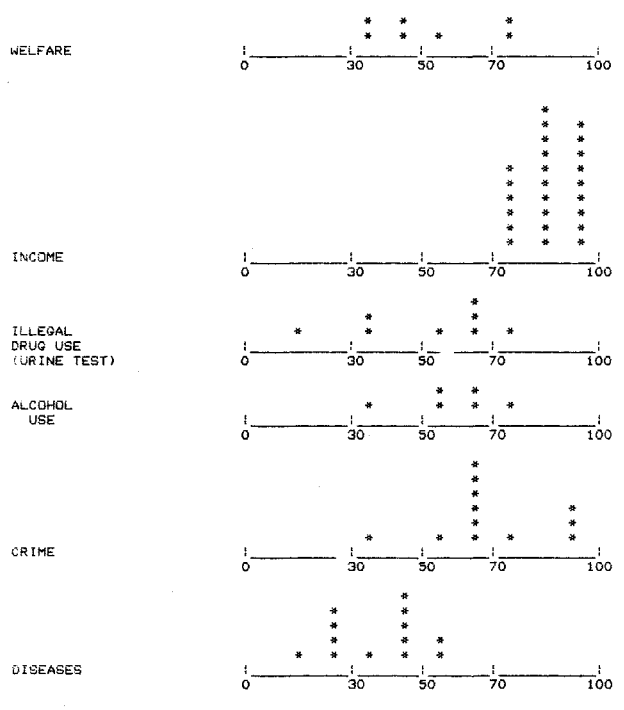
3/ The correlation provides a good reliability estimate if the true score doesn't change between measurements and if the two measures have similar error structures. The correlation overestimates the survey response reliability if the survey data contain more (variable) errors than the other measure.

Fig. 1 - Average Response Bias Estimates



Note: Each * represents an estimate from one study.

Fig. 2 - Response Reliability Estimates



Note: Each * represents an estimate from one study.