

STATISTICS AND SURVEYS AS LEGAL EVIDENCE

James A. Bayton, National Analysts, Inc.

The issue as to the disclosure of the names of respondents could be resolved if courts could be persuaded that these respondents do not stand in the conventional legal position of witnesses. Persuasion of the courts to this point would seem to depend upon their acceptance of a particular basic approach to the matter of survey data as legal evidence. The key factor in this approach is the establishment of the fact that the survey data are presented solely as being the result of research activity. The critical point to be established is that since this is the result of a research activity the challenges must be made in terms of research principles. If there is any fundamental matter that is uppermost in a researcher's mind it is that his results must be able to withstand challenge. In fact, one could say that all of the care that goes into the designing of a research project (whether for strictly "academic" or for applied purposes) has the purpose of anticipating challenge to the results which will be produced. The researcher designs his project in terms of research principles; he expects the challenges to be made in terms of research principles. A challenge which takes the form of demanding the names of the respondents comprising the sample is invoking a research principle. What research principle or criterion is involved in such a challenge?

The two essential research principles are validity and reliability. Any specific challenge must be striking at either the validity of the research data or its reliability.

Validity refers to whether the techniques used "tapped" dimensions or parameters that are pertinent to the issue at hand. Obviously, if the issue at hand was some matter related to price, a questionnaire directed only to the aesthetics of brand labels would be of dubious validity. Challenges to validity of survey data can be met by giving the challenger the questions asked and the answers given to the questions (with names of respondents deleted). Inspection of the questions and their answers could lead to quite legitimate claims as to unsatisfactory validity of the survey data.

When, however, the challenge is a demand for the names of the respondents, the attack has to be upon the reliability of the survey data, rather than upon its validity. The heart of this issue is: "You asked a

valid question, but I don't believe the percentage you give as the answer is correct - I believe that something is producing a wrong answer to the problem, in terms of that percentage you show." The search now must turn to locating the "something" that might be causing an error. Once again, we must insist that the challenge to reliability is legitimate.

Reliability of survey data (assuming initial validity) is a function, of course, of two factors - sampling error and bias. Expert investigation of the probability methods used in the sample design and the methods of calculation of the sampling errors would reveal weakness in reliability with respect to the criterion of sampling error. Knowing the actual names of the respondents would make absolutely no contribution to an attack on the reliability of the data from the point of view of sampling error.

The challenge represented in the demand for the names of the respondents must, therefore, be directed toward the possibility of exposing some form of bias in the results. What are the potential sources of bias in sample survey research?

The sample might be biased - a matter of expert investigation of the definition of the universe sampled, method of drawing the sample, characteristics of the sample of respondents actually interviewed (income, age, and/or other relevant characteristics). Names of the respondents are not needed for this challenge.

The questions, per se, might be biased - "Don't you feel that the price of Brand X is much too high?" The question is valid because it refers to the pertinent issue - the price of Brand X; it is obviously biased. Inspection of the questions as written reveal the status of this problem.

The question, per se, is not biased but the interviewer inserts bias by inflection of voice, adding comments, etc. - the interviewers are available to be called as witnesses.

Bias might enter in the analysis (especially coding) of the interview material - this can be inspected in the interview material made available (with names

deleted) and the codes used. If one wishes, he could recode the data and experts could testify as to the merits of any differences in results.

Bias could enter into the interpretation of the data. For example, does 25 percent represent "many people" saying something or does it represent "only a minority." The interpretative phrase used might depend upon the interest of the speaker (or writer) in the issue being studied.

With respect to the names of the respondents, the argument has to be that the respondents were a sample and were of no importance as specific individuals. The particular people who fell into the sample were, at best, fortuitous. Assuming that at some point in the sample design a system of randomization was used, the specific respondents would all have been different, if one had started at a different place in a table of random numbers. The researcher would argue that, with all other factors held constant, the data would be the same (within sampling error) regardless of the shift to a different set of respondents. The respondents in survey research have no more than the status of guinea pigs in biological laboratory research. The respondents are used solely as a sample to permit generalization back to some universe. Their particularity as Henry Smith, Mary Harrison, etc. has no status for the research purpose. (There is, of course, the question of honesty - were the so-called respondents actually interviewed? An independent research organization could make a check on this matter.)

If after expert investigation the challenger still is not convinced that the survey data are reliable, he has available to him the ultimate challenge - replication. Hold all factors constant - definition of the universe, sample design, questionnaire, coding procedure, etc. - and see whether the same results (within sampling error) are obtained.

Even further, the challenger could initially

conduct his own survey, in anticipation of a survey to be presented by the opposition. If this survey produces different results experts could be called to investigate and to testify as to the sources of the difference.

In summary, Arnold King has asked that the Association take a stand on this issue. He poses three possibilities - strict anonymity of respondents; disclosure with the consent of the respondent but restriction upon the use of the information; and disclosure without restriction as to use.

The speaker advocates that the proper rule should be complete anonymity of respondents. It is believed that the respondent's awareness of anonymity makes a positive contribution to the reliability of the data. It is further argued that anyone who questions the results of a survey has open to him the avenues of challenge cited above. Finally, if the rule is to be otherwise, with disclosure demanded, we can anticipate that a valuable source of information in legal matters will be erased.

Without wishing to insult the intelligence of the courts, if survey research data are to be admitted as evidence, the courts will have to become knowledgeable with respect to the rules and principles of research or rely upon expert advice and testimony as to these rules and principles. With respect to the names of the respondents, the heart of this matter is that they do not have value or status as particular individuals. This is especially true when the implication is that only those particular persons who comprised the sample could possibly have had the attitudes revealed by the survey. From the point of view of sampling theory, a start from a different place in a table of random numbers would have produced a different set of particular individuals but the results should be essentially the same, in spite of this change to a new group of particular respondents