

Cognitive Aspects of Health Surveys for Public Information and Policy

STEPHEN E. FIENBERG, ELIZABETH F.
LOFTUS, and JUDITH M. TANUR

Carnegie-Mellon University;

University of Washington;

State University of New York, Stony Brook

SOURCES OF DATA DIFFER GREATLY IN THEIR quality and in their usefulness in helping to project the effects that will result from policy decisions. Often the policy maker requires information on cause-and-effect relationships, and other times it is only information on background levels and conditions that is critical to informed public policy decisions. Much of the survey data collected by United States government agencies originated in requests by government officials for policy purposes, and almost all such survey data have the potential of providing information that can be used to challenge basic government policy. These statements are especially true in the area of health.

The purposes of this article are: (1) to reiterate what we believe to be the proper uses of sample survey data for health policy purposes, (2) to caution against drawing causal conclusions on the basis of survey data, and (3) to note the relevance, to those making use of health survey data, of an understanding of cognitive processes that underlie respondents' responses to survey questions. In Fienberg, Loftus, and Tanur (1985a) we described some recent efforts to reexamine health survey methodology from the perspective of cognitive psychology. In the present article, through a pair of examples, we attempt to demonstrate

the relevance of the cognitive perspective to the interpretations of survey data made in the public health policy arena. Then, in a separate section we consider survey questions focusing on the subjective assessment of risk.

Our position on the value of health survey data in informing public policy should not be misinterpreted. We believe that such data are important for policy purposes and are often of high quality. Moreover, we are also aware that health policy makers and administrators must often make decisions on the basis of whatever information or data are available. Nevertheless, our position is a cautionary one. People carrying out statistical analyses, often of an elaborate nature, have a tendency to find the answers they set out in search of and they are quick to ignore problematic aspects of and sources of uncertainty in the data they analyze. Our cautionary position leads us, in the final section of the article, to comment on the roles of public and private surveys in informing public policy.

Surveys and Causal Interpretations

Surveys are not appropriate for showing cause-and-effect relations, although sometimes policy makers need to use the observations as if they were the equivalent of an experiment. For example, one could take a survey regarding respondents' medical care utilization and how it relates to health insurance. The survey might show a negative correlation between the size of deductibles and the utilization of medical resources. This information could then help in formulating legislative cost-containment strategies that could influence the structure of private health insurance benefits. This is exactly what Wilensky, Farley, and Taylor (1984) do using data from the National Medical Care Expenditure Survey. But it is only after the policy changes they recommend are actually implemented that we will be able to see if their analyses based on survey data lead to the predicted changes. Even then, causal attribution will be difficult in a nonstationary environment. Thus, we say that sample surveys are useful for *informing* public policy and generating new policy ideas rather than for *making* public policy. If it is crucial that we know the answer to questions about how health care utilization reacts to different payment schemes

before implementing such schemes system-wide, then it is almost certain that we need to conduct a randomized controlled experiment, such as was done in the Health Insurance Study conducted by the Rand Corporation (for details, see Newhouse 1974; Newhouse et al. 1981; Fienberg, Singer, and Tanur 1985).

There are now many compelling examples of the feasibility of randomized controlled experimentation in areas of public health and medicine, beginning with the famous Salk poliomyelitis vaccine experiment of the 1950s (Meier 1972), in which children were randomly selected to receive an injection of either the vaccine or a placebo, up through the far more complex health insurance experiment mentioned above.

Why do we need to do experiments? As Hoaglin et al. (1982, 28–29) note:

In causal work policy analysts often use statistical devices called regression methods for analyzing observational studies, and this leads to difficulties. . . . Consider height versus weight for U.S. male adults. We can certainly fit a regression line of height on weight . . . to predict height. In a sense, the regression line serves the two purposes of predicting and summarizing the relation between the two variables in the population *as it stands*. . . We know from many sad experiences that in adults changes in weight produce practically no change in height. Thus, the regression for predicting height from weight was misleading for predicting what would happen when weight changes.

In commenting upon the use of randomized clinical trials in surgery and anesthesia, Gilbert, McPeck, and Mosteller (1977) note that although randomized clinical trials are not the only strong form of evidence about therapies in humans, weakly controlled investigations may not give the same results as better controlled ones. For example, in a reexamination of 53 different investigations involving the surgical operation of portacaval shunt, they noted an inverse association between the degree of enthusiasm of the investigator for the operation and the degree of experimental control. These and other such results, they claim, lend strong support for Hugo Muench's "law" (Bearman, Loewenson, and Gullen 1974), which says essentially that "nothing improves the performance of an innovation as much as the lack of controls."

Often health policy must be formulated and administered on the best available information, and in situations where the clinical trial evidence is weak at best. A good illustration of this point is found in the controversy surrounding conclusions and recommendations of the National Institutes of Health (NIH) consensus panel on blood cholesterol and heart disease (see Kolata 1985). The panel recommended that "lowering cholesterol can reduce the incidence of coronary artery disease and save lives," on the basis of strong conclusions to this effect in many observational studies but weak and equivocal findings in several crucial experiments. Several statisticians voiced the following view of the NIH policy position: The clinical trial evidence simply does not support the strong statement the panel made but, on the basis of all the information available, the dietary recommendations for adults to reduce cholesterol intake is sound public policy.

To recap our position, we quote again for Hoaglin et al. (1982, 45):

The well-conducted controlled trial is the most definitive method of investigating causal relationships both in the laboratory and in the field. It offers the best evidence for comparing the relative effectiveness of a limited number of programs, or program variants. . . . In some cases time and funding constraints may make a comparative trial infeasible. Where feasible and practical, however, controlled trials in a policy area will tend in the long run to increase our knowledge and understanding of the effectiveness of various interventions, and hence the effectiveness of policies made on the basis of this knowledge.

Some Examples of Health Policy Issues Recently Addressed Using Survey Data

In Fienberg, Loftus, and Tanur (1985a) we focused on cognitive aspects of content and methodology of the National Health Interview Survey (NHIS). But there are many other government-sponsored and private surveys that focus primarily on health policy issues, as well as a host of other surveys that include health-related questions. Here we examine two health surveys, one public and one private, and we consider, among other things, cognitive aspects of some of the questions at the center of controversial policy debates.

Does Calcium Intake Reduce Hypertension?

Reduced consumption of calcium and potassium is the primary nutritional marker of hypertension. . . . Diets low in sodium are associated with higher blood pressures, while high sodium diets are associated with the lowest blood pressures. (McCarron et al. 1984).

These surprising research conclusions have provoked strong critical reactions (e.g., see Kolata 1984; Feinleib, Lenfant, and Miller 1984). What remained hidden in the public discussion of the controversy were the design of the survey and the particular questions, the responses to which were used as input to the analysis.

The data analyzed by McCarron et al. came from NCHS's National Health and Nutrition Survey (NHANES), in particular from NHANES I (the first version of the survey conducted from 1971 through 1974). The survey used multistage probability samples of the United States population designed and executed by the Bureau of the Census for the National Center for Health Statistics (NCHS). The Bureau of the Census carried out household interviews, and NCHS then performed direct physical examinations, clinical and laboratory tests, and related measurement procedures in examination centers, using specially trained medical interviewers and examiners. Everything that had to be read or interpreted (e.g., blood chemistries, EKGs) was sent to a centralized location. The response rate for the household interview was almost 99 percent; medical history questionnaires were completed for 88 percent of the designated sample respondents; but only 74 percent of the designated sample respondents provided information on the nutritional component and 70 percent in the actual health examination. NHANES requires self-reporting for everyone aged 12 and older. Because considerable information is available for respondents who did not provide the nutritional component or did not take the health examination there is the potential for indepth analysis of nonresponse bias, some analyses of which have been performed under NCHS.

With respect to nutrition, NHANES included four types of data:

- Information concerning dietary intake—taken from 24-hour recall interviews and food frequency questionnaires, both administered by an interviewer who is a trained dietitian;

- Hematological and biochemical tests—a sizable battery of such tests, with processing at the mobile examination centers where necessary, but for the most part at a central nutrition laboratory established at the Centers for Disease Control;
- Body measurements—an especially important battery in connection with infants, children, and youths, where growth may be affected by nutritional deficiencies;
- Various signs of high risk of nutritional deficiency—based on clinical examinations.

The health (as distinguished from nutrition) component of the NHANES program includes detailed examinations, tests, and questionnaires (developed to obtain a measure of prevalence levels of specific diseases and conditions). These vary with the particular program and have included such conditions as chronic rheumatoid arthritis and hypertensive heart disease. Important normative health-related measurements, such as height, weight, and blood pressure, are also collected. Recorded blood pressure values are based on the average of three readings taken as part of the physical examination (no recall is involved).

Findings from NHANES have been presented primarily through publication of individual reports in the National Center for Health Statistics *Vital and Health Statistics*, Series 11, but data tapes are also available to the public, and it was these tapes that were used by McCarron et al. in their analyses.

McCarron et al. began with just under 21,000 respondents for whom data were available on blood pressure and 24-hour nutrient consumption. Exclusion of persons under 18 years of age reduced the sample to just under 14,000 and elimination of pregnant women, individuals on a low-salt diet, and those who responded positively to the high blood pressure question (“During the past 6 months, have you used any medicine, drugs, or pills for . . . high blood pressure?”) reduced the sample to about 10,400 (a number that Feinleib, Lenfant, and Miller 1984, were unable to replicate). For this specially selected subsample, McCarron et al. chose the upper 10th percentile of systolic blood pressure for age-sex-race subgroups to classify individuals as hypertensive, and then compared hypertensives and “normotensives” using direct standardization (see, e.g., Bishop, Fienberg, and Holland 1975) within age-race-sex groupings. (They also did some analyses using a 160 mmHg. systolic blood pressure cutoff). But in the most

crucial analyses the standardization for age was not attempted, and it was these comparisons that were used to show that “calcium was the nutrient for which reduced intake was most consistent in hypertensive individuals,” and “hypertensives tended to consume less sodium than normotensives,” regardless of the definition of hypertension and regardless of how the background variables were controlled. Despite adding a disclaimer that these relations cannot be accepted as proof of causation, the authors have done little to temper causal conclusions from being broadcast by the press (e.g., see a report on an interview with one of the authors—“Study says calcium can cut risk of high blood pressure” [*Peninsula Times-Tribune*, November 6, 1984]).

Feinleib, Lenfant, and Miller (1984) are critical of both conceptual and statistical problems in the McCarron et al. analysis, including the lack of direct standardization in the crucial analyses. In some alternative analyses they illustrate the lack of robustness of the original conclusions to the form and specification of the choice of underlying statistical model. It *does seem to matter* how you control for background variables using regression.

Only a few critics have expressed reservations regarding the appropriateness of the use of the survey data in the analyses. Questions were asked by trained dieticians about nutrient intake during the 24-hour period of the day preceding the interview, based on recall of amounts of different foods consumed, with cross-references to the answers to other questions on the usual frequency of intake of certain foods—not on actual measurement of nutrient intake, although estimation of portion size was aided by the use of three-dimensional models. Despite this care, in light of what is known from cognitive psychology about problems of recall, it seems to us a very difficult task to list everything we ate yesterday in such a way that nutrients could be estimated accurately. Further, no one to our knowledge has expressed any concern over what may be substantial problems of recall for questions (on medications and conditions) used to define the “population of interest.” There is also a questionable assumption that the food eaten in the previous 24 hours is typical of the individual’s intake over days, weeks, and even years. In fact, NHANES includes a physical measurement of serum calcium that could have been used in place of the recall item.

For us, this example illustrates the inappropriate use of fairly high-quality survey data in an attempt to reach causal conclusions, which if they should turn out to be correct, would be of far-reaching importance.

We would need to look to a randomized controlled experiment that varied diets for treatment groups for firm results. There was such an experiment—known as the Multiple Risk Factor Intervention Trial (MRFIT)—but its results were not compelling. The growing conviction, among both laymen and health care providers, that diet affects health, caused the control group in MRFIT to change its diet over the course of the study and thus to appear more similar to the treatment group than was intended in advance of the experiment. Thus, even high-quality experiments do not necessarily work out the way investigators expect.

Cost Containment for Health Care

“Our recent survey shows that doctors are ‘out of step with the rest of the nation’ in their views on rising health care costs” was the claim of a vice-president of the Equitable Life Assurance Society (*Pittsburgh Post-Gazette*, June 8, 1984). In the current debate over how to restrain the cost of doctors’ fees under Medicare and other federally sponsored medical service programs, government officials are proposing new forms of payment to doctors such as a flat, all-inclusive payment for the doctors’ services associated with each type of illness, rather than separate payments for each individual service. One of the few sources of information available on public attitudes toward such cost containment is the Equitable Healthcare Survey (EHS), designed and administered by Louis Harris and Associates (a private polling and market research organization) and sponsored by the Equitable Life Assurance Society, a company prominent in designing, promoting, and marketing new forms of health care coverage (for details, see the Equitable Healthcare Survey 1983, 1984).

The EHS was not one but a set of six surveys of different participants in the United States health care system. The respondents consisted of a national sample of 1,500 adults, a national sample of 100 physician leaders, a sample of 100 hospital administrators, a sample of 50 senior health insurance executives, a sample of 250 corporate benefit officers, and a sample of 26 union leaders responsible for health care benefits. Those six concurrent surveys were followed by a seventh survey with a sample of 500 practicing physicians.

In the initial set of surveys, interviews lasted approximately 30 minutes on average and were conducted by telephone using essentially the same questionnaire. The claimed response rates were 57 percent

for the public sample, 78 percent for the physician leaders, and 83 percent for the other groups combined. Unfortunately, the reports do not explicitly say how these rates were calculated—various methods of calculating response rates yield very different rates. In addition, the achieved sample sizes of 1,500, 100, 100, 50, 250, and 26 suggest the use of forms of quota and other nonrandom-sampling techniques. By comparison with government-sponsored sample surveys, the reported response rates are poor, but they are at a higher level than that of many private polling organizations. The followup survey of 500 practicing physicians was drawn from a list of 2,500 names obtained from the American Medical Association and involved some form of nonrandom selection within clusters. No response rate was reported.

The questionnaires for the EHS contained a variety of questions including items on (1) general attitudes toward the United States health care system, (2) perceptions of the reasons for health care cost escalation, and (3) attitudes toward various cost-containment policies and programs in terms of their effectiveness and acceptability. Among the questions asked of the national sample of adults in the first category were:

Was there any time in the last year . . . that you felt you or a member of your family living with you needed medical help but did not get it for some reason?

and

How many visits, excluding overnight hospital stays, did you and all of the members of your family living with you make to a doctor's office, outpatient clinic, hospital emergency room, or any other place for medical care during the past twelve months?

Both of these questions involve aspects of recall (with the latter closely resembling questions on the NHIS discussed in Fienberg, Loftus, and Tanur 1985a). The first of these questions involves, in addition, at least two kinds of judgment by the respondent—judgment about need for medical help and judgment that (sufficient) care was not received (perhaps because the patient did not feel better after the help was received). There could also be an earlier judgment about affordability of care involved. The problem with the use of such questions in the context of the EHS is that there are no detailed questions on health conditions and illnesses to assist respondents in providing reasonable

responses (c.f. the discussion in Fienberg, Loftus, and Tanur 1985b on how the reporting of conditions relating to trouble with feet increased with probes on specific problems).

The EHS also included a general question on self-perception of the health of the respondents. The responses to the question on medical care visits for those respondents to EHS who claimed excellent/good health status as compared with those who claimed fair/poor status, are remarkably similar (NHIS results show a difference on this point). More substantial differences occurred in response to the question, "Have you or anyone in your family living with you been a patient overnight in a hospital during the past twelve months?," with 30 percent of those in the excellent/good group reporting "yes," as compared with 48 percent in the fair/poor group.

The physicians and other professionals were asked the question:

It is clear that some people incur more health care costs than others. What percentage of all health care costs do you think is incurred by the 10% of the population who are the biggest consumers and who incur the highest costs?

This question is interesting from a cognitive perspective. Not only does a sensible answer to it require considerable judgment, but it also requires an elaborate calculation. Many respondents will fail to comprehend what they are being asked. Few will recognize the concentration curve (known technically as the "Lorenz curve") nature of the question. That only 9 percent of the physician leaders and 17 percent of the practicing physicians responded that they were "not sure" is surprising. The 38 percent "not sure" response of union leaders suggests that they were being more candid about their confusion. Some might say that this EHS question was an attempt to measure opinions about fictions, while others would say that it was simply incomprehensible.

Finally, we turn to the key cost-containment question, which is unfortunately embedded in a long list of questions on vague proposals, each asking:

Do you think it will be very effective, somewhat effective, not very effective, or not at all effective in *controlling costs* or *discouraging use* of nonessential services?—A system in which the fees paid to doctors and hospitals for treating all patients with particular types of diagnoses are fixed [Emphasis in the original].

This is a double-barreled question in which respondents are asked to rate the effectiveness of something *either* for cost control *or* for discouraging use of nonessential services as if the two things are the same. It is difficult to imagine what an answer to this question means other than some global feeling about the relation between fees and services.

Of nonphysicians, between 73 percent and 84 percent said very or somewhat effective, as compared with about 50 percent for physicians. Moreover, in response to another question, 71 percent of hospital administrators and 76 percent of the public sample responded that such a system would be very or somewhat acceptable, whereas 37 percent of the practicing physicians and 32 percent of the physician leaders did so. These responses—as advertised—do indicate that the views of doctors are out of step with those of the rest of the nation, but do they provide any real basis for informed policy changes? Does the answer to a throwaway question (for which the different groups probably have different contextual interpretations) in a 30 minute interview indicate how doctors and patients really feel about the proposal? Even if accurate feelings are being reported, there are still the problematic links among knowledge, attitudes, and behavior that social scientists have long worried about. Do these answers indicate anything about how the doctors and patients *will act* under the proposed new system—in their choice of diagnoses and the utilization of care and services? Even *good* survey data can do little to answer the key behavioral questions regarding response to new cost-containment practices.

Even with a randomized controlled experiment to examine such behavioral questions, the data for evaluating the outcomes are likely, in many instances, to involve some form of self-report. Thus, for both surveys and experiments, the investigator has to pay considerable attention to the form and context of questions in order to elicit data with as little error or bias as possible.

The Risky Business of Survey Assessments of Risk

Americans greatly overestimate the health dangers of birth control pills despite scientific evidence . . . according to a new Gallup Poll released Tuesday by the American College of Obstetricians and Gynecologists. In the survey, 76 percent of the women and 62 percent of the men said they believed there were “substantial risks”

in taking the pill, the poll found. And 31 percent of the women and 27 percent of the men named cancer as the most serious consequence of pill use, which the group said also was incorrect (*Los Angeles Times*, March 6, 1985).

Public opinion polls often ask respondents to estimate the risk associated with diseases, technical innovations, medicines, and nuclear power plants. The results of such questions are often used as part of press releases designed to influence public opinion and even public policy. The resulting newspaper articles contain provocative statements, such as those about the risks associated with the use of birth control pills quoted above. What should the intelligent reader make of such survey results?

Our purpose in this section is not to present a critical evaluation of the Gallup Poll questions on the risks associated with birth control pills, even though this might be easy to do. Rather our intent is to point to the importance of actual knowledge and context in the production of a judgment or estimate of risk. (Again, while in daily life there is surely a relation between the judgment of risk and behavior regarding that risk, the link is not a simple one.) Clearly the survey about birth control pills was not measuring individuals' risk assessments, but was attempting to reflect some aspects of attitudes toward birth control pills that might well encompass religious attitudes and beliefs.

If we are interested in eliciting estimates of risk or even perceptions of risk, then some recent findings in the cognitive psychology literature are of special interest. Survey researchers have long known that seemingly small changes in the wording of questions can lead to dramatic differences in the responses elicited (e.g., see the discussion of this topic in Turner and Martin 1985, and in the paper by Bradburn and Danis in Jabine et al. 1984). One striking example of this comes from the work of Tversky and Kahneman (1981) who showed that people make different decisions depending upon how the choices put to them are framed. To see this most clearly, consider two of the problems Tversky and Kahneman have used in their research:

Imagine that the United States is preparing for the outbreak of an unusual Asian disease which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed with these consequences:

- If program A is adopted, 200 people will be saved,
- If program B is adopted, there is a 1/3 probability that 600

people will be saved, and a $2/3$ probability that no people will be saved.

Which program would you choose?

Most people play it safe, that is, they are “risk averse.” The certainty of saving 200 lives is more attractive than a risky prospect that has a 1 in 3 chance of saving 600 lives.

A second group of people read the same scenario, but with a different set of alternatives:

—If program C is adopted, 400 people will die.

—If program D is adopted, there is a $1/3$ probability that nobody will die, and a $2/3$ probability that 600 people will die.

Objectively, programs C and D are identical respectively to programs A and B, described in terms of number of lives saved (programs A and B) or number of lives lost (programs C and D). But faced with the second set of alternatives, people tend to reverse their decisions and choose the second program; the certain death of 400 people becomes less acceptable than the 2 in 3 chance that 600 will die. According to Tversky and Kahneman, the preferences in this example illustrate a common pattern: When gains are involved people tend to avoid risks; faced with losses, they are more willing to take risks.

The implications of this work are straightforward. Survey researchers must ask themselves whether the particular framing of the question has produced the choices that are made, and whether a different framing would lead to a different response. Note that the framing we refer to here includes not only the wording of the particular question but also the entire context in which that question is embedded. Such framing differences can often explain why two surveys that appear to be asking a very similar question of similar respondents produce vastly different results. The differences in results can be especially dramatic when the assessment of risk is involved.

Conclusions

Over the past 20 years we have seen substantial growth in the use of survey data to measure historical trends (e.g., in illness and disability) and to answer questions relating to public health policy. Much of

this growth has been in the public sector, in large part under the sponsorship of the National Center for Health Statistics (NCHS), but there has also been an increasing use of surveys dealing with health issues in the private sector. Bridging the gap between the two are government-sponsored health surveys, such as the National Medical Care Utilization and Expenditure Survey (NMCUES), that are conducted by organizations such as the National Opinion Research Center and the Research Triangle Institute under contract with NCHS and the Health Care Financing Administration. This latter class of surveys more closely resembles government-conducted surveys in methodology employed, response rates, and attention to problems of nonresponse and questionnaire design.

Since 1981 the federal government budget for health-related surveys has been cut substantially and several important surveys have been altered, delayed, or even cancelled (e.g., see Baseline Data Corp. 1984). For example, the National Ambulatory Medical Care Survey (NAMCS), which was originally designed to be annual, has become triennial and, as noted in Fienberg, Loftus, and Tanur (1985b), samples only physicians in office-based practice despite the fact that much ambulatory care is being delivered to hospital out-patients and in free-standing ambulatory centers. Similarly, although NMCUES was originally planned to be biennial or triennial, we can now expect at least an eight-year gap, with the next survey scheduled for 1988. There will also be at least a ten-year gap between national samples of the National Health and Nutrition Examination Survey (NHANES).

At the same time that these delays compromise the timeliness of the data derived from these carefully designed and implemented health-related surveys, government policy makers are looking to them for information to guide critical policy choices in the health area. We have argued that survey data cannot provide all of the answers to their questions since surveys measure things as they stand, and thus are not appropriate for showing cause-and-effect relations. Nonetheless, sample surveys are often useful for *informing* public policy in a variety of ways. The current governmental stance toward the funding of survey research and data collection in the health area undercuts the government's ability to make informed policy choices. Continuation of such a stance will force the government to rely upon survey data emanating from the private sector—funded by the very organizations that are likely to be affected by policy changes based on the collected data.

There are both advantages and disadvantages associated with a reliance on government-sponsored surveys for answers to policy-related questions. Among the disadvantages listed by Cartwright (1983) are:

1. Such surveys are unlikely to ask questions that would challenge basic government policy.
2. The production of data may be seen as the end product, with little or no resources devoted to analysis.

To these we would add that

3. Government surveys tend to be conservative in their orientation, and are often slow to change.

Yet changes in content and structure of surveys need to be made with great care (there is the usual tension between improvement in data quality and comparability of data over time), and the availability of quality data on topics of interest should serve to stimulate academic researchers and others to carry out those analyses left undone by government statisticians.

One of the major advantages of public surveys is that they are subjected to broad public and professional scrutiny and continuing critique. There is great danger in basing public policy on data collected (whether by surveys or other means) in an environment where there is little opportunity for public and professional examination, critique, and restructuring of both survey methods and questionnaire content and format. For government-sponsored surveys (such as NHIS, NHANES, and NMCUES) the survey data as well as substantial information on nonresponse and other checks on validity are publicly available for reanalysis and examination. This public milieu encourages outside scrutiny of all phases of survey methodology from questionnaire construction through to analysis and reporting. The milieu is absent in settings that focus on quick and dirty results. Many private survey organizations are capable, if approached, of carrying out a survey that would meet the "public criteria" to which we refer (and some actually do so on a regular basis). Unfortunately, there have been few *privately commissioned* surveys that come anywhere near to meeting the "public criteria," in part because of the great expense involved in quality survey work. It is, therefore, not surprising to find that it is the government-sponsored surveys like the NHIS that are the focus of

innovative research efforts that will ultimately improve the entire survey enterprise, both public and private.

References

- Baseline Data Corporation. 1984. *The Federal Statistical System, 1980 to 1985: A Report Prepared for the Congressional Research Service, House of Representatives Committee on Government Operations*. Washington.
- Bearman, J.E., R.B. Loewenson, and W.H. Gullen. 1974. Muench's Postulates, Laws, and Corollaries, or Biometricians' Views of Clinical Studies. *Biometrics Note No. 4 (April)*. Bethesda, Md.: Office of Biometry and Epidemiology, National Eye Institute, National Institutes of Health.
- Bishop, Y.M., S.E. Fienberg, and P.W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.
- Cartwright, A. 1983. *Health Surveys in Practice and in Potential: A Critical Review of Their Scope and Methods*. London: King Edward's Hospital Fund for London.
- Equitable Healthcare Survey. 1983. *Options for Controlling Costs*. New York: Equitable Life Assurance Society.
- . 1984. *II. Physicians' Attitudes toward Cost Containment*. New York: Equitable Life Assurance Society.
- Feinleib, M., C. Lenfant, and S.A. Miller. 1984. Hypertension and Calcium. *Science* 226:384–86.
- Fienberg, S.E., E.F. Loftus, and J.M. Tanur. 1985a. Cognitive Aspects of Health Survey Methodology: An Overview. *Milbank Memorial Fund Quarterly/Health and Society* 63(3):547–64.
- . 1985b. Recalling Pain and Other Symptoms. *Milbank Memorial Fund Quarterly/Health and Society* 63(3):582–97.
- Fienberg, S.E., B. Singer, and J.M. Tanur. 1985. Large Scale Social Experimentation in the U.S.A. In *A Celebration of Statistics, The ISI Centenary Volume*, ed. S.E. Fienberg and A.C. Atkinson. New York: Springer-Verlag. (In press.)
- Gilbert, J.P., B. McPeck, and F. Mosteller. 1977. Statistics and Ethics in Surgery and Anesthesia. *Science* 198:684–89.
- Hoaglin, D.C., R.J. Light, B. McPeck, F. Mosteller, and M.A. Stoto. 1982. *Data for Decisions*. Cambridge, Mass.: Abt Books.
- Jabine, T., M. Straf, J.M. Tanur, and R. Tourangeau. 1984. *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Washington: National Academy Press.
- Kolata, G. 1984. Does a Lack of Calcium Cause Hypertension? *Science* 226:705–6.

- . 1985. Heart Panels' Conclusions Questioned. *Science* 227:40–41.
- McCarron, D.A., C.D. Morris, H.J. Henry, and J.L. Stanton. 1984. Blood Pressure and Nutrient Intake in the United States. *Science* 224:1392–98.
- Meier, P. The Biggest Public Health Experiment Ever The 1954 Field Trial of the Salk Poliomyelitis Vaccine. In *Statistics: A Guide to the Unknown*, ed. J.M. Tanur et al., 2–13. San Francisco: Holden-Day.
- Newhouse, J.P. 1974. A Design for a Health Insurance Experiment. *Inquiry* 11:5–27.
- Newhouse, J.P., W.G. Manning, C.N. Morris, L.L. Orr, N. Duan, E.B. Keeler, A. Leibowitz, K.H. Marquis, M.S. Marquis, C.E. Phelps, and R.H. Brook. 1981. Some Interim Results from a Controlled Trial of Cost-sharing in Health Insurance. *New England Journal of Medicine* 305:1501–7.
- Turner, C.F., and E. Martin, eds. 1985. *Surveying Subjective Phenomena*. Vol. 1. New York: Russell Sage Foundation.
- Tversky, A., and D. Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211:453–58.
- Wilensky, G., P.J. Farley, and A.K. Taylor. 1984. Variation in Health Insurance Coverage: Benefits vs. Premiums. *Milbank Memorial Fund Quarterly/Health and Society* 62(1):53–81.

Acknowledgments: The preparation of this paper was supported in part by several National Science Foundation grants: No. SES-8119219 to Carnegie-Mellon University, No. SES-8119138 to the Research Foundation of the State University of New York, No. BNS-8023404 to the University of Washington, and by BNS-8011494 while the first author was a Guggenheim Fellow at the Center for Advanced Study in the Behavioral Sciences. We are grateful to Norman Bradburn, John Bunker, Ann Cartwright, Mary Grace Kovar, Judith Lessler, Dorothy Rice, Monroe Sirken, and several referees for comments and suggestions that helped to shape the final version of this paper.

Address Correspondence to: Professor Stephen E. Fienberg, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213.