# Cognitive Aspects of Health Survey Methodology: An Overview

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

Carnegie-Mellon University: University of Washington; State University of New York. Stony Brook

**R** ESPONDENTS TO THE NATIONAL HEALTH Interview Survey (NIHS) are asked:

1. During the past 12 months, about how many days did illness or injury keep you in bed for more than half of the day?

To understand the issues that concern us in this and the following three articles, we encourage readers to attempt to answer that question for themselves and to take note of their thought processes as they do so.

How did you construct your answer? Did you start at this time last year and mentally scan a calendar and "check off" sick days? Did you do a similar calendar check working backwards in time starting from yesterday? Did you visualize your bedroom and try to estimate how many days you spent there sick in the last 12 months? Did you visualize a hospital room? Did you focus on particular illnesses and injuries and count up days separately for each of them? How did you recall those illnesses and injuries? Did you include episodes of illness for which you did not consult a doctor? How long is "half of the day" for you?

Milbank Memorial Fund Quarterly/Health and Society, Vol. 63, No. 3, 1985 © 1985 Milbank Memorial Fund

Assessing how respondents understand and answer such questions is crucial if the data derived from surveys is to be useful in judging the health status of the nation and in helping guide health policy. And these data are indeed used. At the beginning of the twentieth century, it was customary to measure health status primarily by basic vital statistics on mortality and standardized mortality ratios. These and other related indices are still widely used by epidemiologists and others. Yet health professionals recognized that counting the dead was simply not enough. Today, information about the nature of disease, illness, and health status (as well as the organization and utilization of services and facilities) comes in a variety of forms and from a variety of sources. In the United States, the National Health Survey Act of 1956 provided a legislative mandate for the development of a continuing survey on the amount, distribution, and effects of illness and disability in the United States, and the utilization of related health care services. In part because of the success of the survey program initiated by the act, the sample survey is currently the most widespread method of information collection in the health area.

The past 25 years have seen the development of a wide variety of sample surveys dealing with health and medical care. A number of these surveys are sponsored by the federal government (for example, under the direction of the National Center for Health Statistics), while others are funded and implemented by the private sector. In addition, aspects of health-related questions occur in other surveys not directly related to health. For example, respondents in the National Crime Survey are asked about the health consequences of criminal victimizations they have experienced. Because these questions require answers from respondents, because respondents are people, and because people use comprehension, recall, and judgment to answer questions, there is a common set of problems involving these cognitive tasks that besets virtually all of these surveys (as well as surveys dealing with matters other than health). These problems are important and require attention, for the quality of data derived from a survey depends on the quality of the answers we get from the respondents.

This article and the three articles which follow it describe various aspects of the interface between cognitive psychology, statistics, and survey design as it has emerged during recent years. The present article presents an overview of the topic from the perspective of health information. The second article, by Lessler and Sirken (1985), focuses on an attempt to implement suggestions emanating from the cognitive sciences/survey interface in the context of an ongoing health survey. The third article, by Fienberg, Loftus, and Tanur (1985a) focuses on the specific problem of recall of pain and other symptoms, which is of interest both in the context of health surveys and in its own right. The final article, also by Fienberg, Loftus, and Tanur (1985b), turns to the issue of the use of survey data in health areas to address issues of public policy. In addition to discussing the general problem of drawing inferences from cross-sectional survey data, that article also examines the cognitive aspects of the critical survey questions that lie at the heart of two specific health policy debates.

The sample survey is a form of observational study that can be used (1) to answer specific questions about the population related to measurements taken at the time (e.g., the average height of American males); (2) to provide information about interrelations among variables in the population as it stands (e.g., salt intake and hypertension), and to give hints about causal relations; (3) to provide a baseline for comparisons with future measurements; and (4) to measure changes through a sequence of surveys (e.g., changes in average salt intake and the incidence of hypertension). Because surveys are not usually able to gather information on a continuous basis, questions often involve recall of events and experiences from the past. Faulty recall is a major source of nonsampling error in surveys (see the appendix for a further discussion of sampling and nonsampling errors). The focus in this article and those that follow is primarily on cross-sectional population-based surveys as opposed to longitudinal or record-based surveys.

We use the term "survey" in two different, but related senses. The first is in the context of the sample survey as a method of gathering information to be generalized to a population. The focus of that use is on the methodology of sampling appropriately to allow for generalization. The other sense is the use of an interview to collect data about individual behavior or attitudes. Surveys, in this second sense, are used in a variety of contexts including randomized field experiments. The cognitive aspects examined here deal mainly with interview content and structure rather than the sample survey nature of the data collection. Indeed, the problems considered afflict not only survey interviews but also all other forms of data collection that depend on self-reports (or reports for related individuals). In fact, many of the problems of health surveys are quite similar to problems physicians face when they interview patients in order to obtain medical histories or reports about symptoms. Inferences drawn from sample survey data depend as much on the quality of the survey interview design as they do on the probability methods used to select the sample members.

The sample survey as a method of data collection based on probability techniques has its origin in the 1930s. For a detailed discussion of the slow introduction of ideas on probabilities into the domain of surveys, and of the notion of "representative" sampling, see Kruskal and Mosteller (1980). But once introduced, these ideas had such a dramatic impact that the sample survey has been characterized in a recent National Research Council report as "the single most important information-gathering invention of the social sciences" (Adams, Smelser, and Treiman 1982).

There is a tradition of research in the area of survey interview design that has drawn, sometimes indirectly, on research findings in other areas of the behavioral and social sciences. But little of this tradition has involved the direct collaboration of cognitive scientists and survey researchers. At a time when the sample survey as a method of data collection is viewed as a fundamental scientific instrument, we believe that it is important to reexamine in a more formal manner what the cognitive sciences have to offer for survey interview structure and content.

#### Some Examples of Health-related Survey Questions

We begin by considering some further examples of questions on health topics that evoke interest in the cognitive aspects of survey design. Several problems are posed by these examples:

- Would you say [that your] health in general is excellent, very good, good, fair, or poor? National Health Interview Survey (NHIS)
- 3. Have your feet been painful or caused you any discomfort during the past four weeks?

Survey by Institute of Community Studies, London

4. List all foods you ate yesterday (in the 24-hour period from midnight to midnight) and the sizes of the portions. (This

question is asked by a trained dietitian who uses a collection of three-dimensional food portion models to help the respondent estimate the amounts of various foods consumed.)

National Health and Nutrition Examination Survey (NHANES) 5. During the past 6 months, have you used any medicine, drugs, or pills for . . . high blood pressure? (NHANES)

6. 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?

Equitable Healthcare Survey (EHS)

 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 percent of the population who are the biggest consumers and who incur the highest costs? (EHS)

The first question, asked on the National Health Interview Survey (NHIS), is used as part of a battery of questions designed to measure overall population levels of disability. What cognitive strategies do respondents use to recall days of illness or injury? How accurate are people's estimates? Do people make some kind of judgment calculation rather than attempt to recall at all? Do other questions on specific illnesses and injuries asked earlier in the interview influence the calculation and, if so, how? We turn to a discussion of these and other issues raised in connection with NHIS in a separate section below.

Other questions on this list call more obviously for judgment on the part of the respondent. For example, questions 5 and 6 have components of judgment as well as recall, while question 2 forces the respondent to make an overall judgment, based on difficult processes of comparison. Question 7 raises serious issues about comprehension, requires substantial judgment, and implicitly asks for a complicated calculation or at least a judgment about the result of that calculation. We consider this question in greater detail in a companion article (Fienberg, Loftus, and Tanur 1985b) focusing on the roles of public and private surveys in informing public policy.

The common threads that tie all seven questions together are their focus on health and the fact that, either implicitly or explicitly, responses to them depend upon cognitive processes that have been the subject of research of a sort quite different from that typically carried out in a survey setting.

### Some History of the Movement to Examine Cognitive Aspects of Survey Methodology

One would think that because surveys depend upon respondents reporting on their actions, events that happen to them, their opinions, and their attitudes, and because such reports depend upon such cognitive processes as recall, comprehension, and judgment, that there would be a long history of interaction between those who do research with surveys and those who do research on relevant cognitive processes. But such is not the case. Surely survey researchers have often been trained in various branches of psychology-though more often social psychology than cognitive psychology----and certainly people specializing in psychology, as well as other disciplines, have carried out surveys. But perhaps the earliest formally organized interaction between survey researchers and cognitive scientists was a two-day workshop in September 1980 organized by Albert Biderman at the Bureau of Social Science Research (BSSR) with support from the Bureau of the Census and the Bureau of Justice Statistics. This workshop concentrated on issues arising in the redesign of the National Crime Survey. There was strong agreement among those participating that knowledge of cognitive processes would be useful in designing survey questionnaires, and further, that cognitive researchers could benefit by thinking of ways to use surveys as a setting for experiments to test their theories.

As a logical outgrowth of the BSSR workshop, the Committee on National Statistics of the National Research Council, with support from the National Science Foundation, organized the Advanced Research Seminar on Cognitive Aspects of Survey Methodology that was held in St. Michaels, Maryland, in the spring of 1983. The seminar brought together a small group of cognitive scientists, survey researchers, statisticians, and personnel from government agencies carrying out and analyzing surveys. The purpose of the seminar was to explore ways in which the insights of the cognitive sciences could be of use in designing and understanding survey research, and—as the other direction of a two-way street—it explored ways in which large-scale surveys could become vehicles for research in the cognitive sciences. While there were dire predictions that the participants at the seminar would have trouble talking to each other across disciplinary boundaries, the participants found a dialogue easy to establish and came away from the seminar bursting with new research ideas to carry out themselves or to put on the agenda for the new cross-disciplinary field. For details of the seminar, see Jabine et al. (1984).

In part, the seminar focused on a particular survey, the National Health Interview Survey (NHIS) carried out by the Bureau of the Census for the National Center for Health Statistics. Many suggestions arose at the seminar for the applications of the cognitive sciences toward its possible improvement. We describe these below. A research project is currently being carried out at the National Center for Health Statistics that will design part of such a survey using traditional survey design methods in tandem with the facilities of the cognitive psychology laboratory. This project is described by Lessler and Sirken (1985).

#### Current Research on Cognitive Aspects of Surveys

There are several other projects underway that have as participants members of the seminar and that were inspired by the seminar activities.

#### Cognitive Processes in Attitude Surveys

With the assumption that knowledge, beliefs, and attitudes about issues often form organized cognitive structures, called schemata, and that for enduring issues that appear regularly in opinion surveys there may be several competing schemata available in a culture, one project sets out to explore which schema would guide respondents' answers to attitude questions and how the choice of a schema is affected by variations in question wording and context. For example, there are probably at least two schemata that could guide a respondent's answer to a question seeking attitudes toward Medicaid or Medicaid recipients. One schema could be characterized as "welfare cheats and chiselers," while the other would involve concepts of "medical indigents," the "working poor," and the "truly needy." Many respondents probably hold both these schemata to some degree or other; if prior questioning activates mainly the "cheats and chiselers" schema, then a respondent is likely to report a negative attitude toward Medicaid. Activation of the second schema would likely lead to a report of a positive attitude.

Another set of studies will examine the cognitive processes used to make estimates of frequencies and quantities, and explore the processes used by respondents when revising initial estimates. These studies are related to the studies of attitude surveys by the similarity of the introspective processes that respondents must use to construct answers.

A third set of studies will explore "affective carry-over effects" in which feelings associated with the response to one attitude question may influence responses to later questions. For example, asking about disabilities early on could put some respondents in a bad mood and affect their judgments of well-being. Also to be explored are situations where respondents entertain multiple possible explanations for their own and others' behavior. Lastly, these investigations will explore the distinction between symbolic and instrumental attitudes, which differ in that in the former, beliefs are more rigidly attached to attitudes and emotions may play a larger role. Thus, the two types of attitudes may have different relations to demographic and other variables which may be important to distinguish in surveys.

# A National Inventory of Cognitive Abilities

Just as it is useful to have national norms on blood pressure or weight, so it would be useful to have national norms on certain cognitive abilities. A project to do this would develop a set of cognitive tasks to be administered to a national sample from the United States population. Analyses of the data might then permit establishment of agespecific norms for memory. The data would serve not only to describe the population but could be used to establish bench marks for the diagnosis of such maladies as Alzheimer's disease.

#### Recall of Autobiographical Events

Still other related research has already been carried out. For example, Fathi, Schooler, and Loftus (1984) learned about how people retrieve personal information to answer survey-type questions by using "protocols." Protocol research requires people to think out loud as they answer questions; the verbalizations produced are called protocols, and can be transcribed and analyzed. In this work, protocols were gathered while subjects answered the question "In the last 12 months, how many times have you gone to a doctor, or a dentist, or a hospital, or utilized any health care specialist or facility?" This general question was followed by more specific questions inquiring about certain medical specialists. From the gathered protocols, Fathi, Schooler, and Loftus discovered that subjects often corrected their original responses, and remembered additional information after first indicating they had retrieved all relevant incidents. For example, one subject recalled that she had been to the dentist and to the doctor when asked the general question. Later, when asked specifically about dentists, she added another visit to the dentist to have her teeth cleaned. This type of protocol research can lead to practical suggestions for maximizing the amount of information retrieved by suggesting the kind of retrieval cues that are needed to produce additional memories.

More recently, Loftus and Fathi (1985) examined the order in which people recall autobiographical events that happen repeatedly (e.g., doctor visits, academic exams). They found that when retrieving information about academic exams, people's memories were better if they retrieved beginning with the most recent incident. This method of backward search may be superior than, say, a forward search because in the backward case the first few items searched for are easier to retrieve, and thus provide a better starting point for retrieval of the entire chain. Interestingly, when retrieving other kinds of information such as health care visits, subjects seemed to find it easier to recall in the forward direction (Fathi, Schooler, and Loftus 1984). This apparent discrepancy raises questions about whether it is possible to generalize about retrieval-order effects. There may be certain classes of retrieval tasks in which the natural order of retrieval is backward. In retrieving academic exam information, for example, since exams are fairly independent events, people might well be expected to begin by retrieving the most recent and available instance. With health care visits, on the other hand, there is more likely to have been some causal relationship between the various visits (e.g., you broke your ankle, so you went to the orthopedic specialists, who told you to go to the radiologist and get some X-rays taken, etc.). It is easy to see how findings of this type might profitably be applied to the gathering of information on surveys.

# Self vs. Proxy Reporting

One final example concerns the storage and retrieval of information about oneself versus another person. In the NHIS, respondents are sometimes asked to retrieve information not only about themselves but also about other family members. An individual might be asked to recall the doctor visits of a spouse or child, for example. Psychological findings have shown that information about the self is apparently more affect-laden, more familiar, more complex, and stored differently than information about others. The implication is that the best retrieval techniques for maximizing the recovery of personal memories might not be best when the goal is to maximize the recovery of information about another person. This is not to say that there is no similarity in how self versus other information is stored and retrieved (in fact, one recent finding is that people's estimates for the duration of events in their own past were not appreciably different from their estimates for the duration of past events in their roommates' lives; see the description of this project by Lee Ross in Jabine et al. 1984). We are merely suggesting that one must be cognizant of possible differences in how self versus other information is best gathered. Empirical research from cognitive psychology can illuminate this issue and suggest ways to optimize both types of information gathering.

#### Some Cognitive Dimensions of the National Health Interview Survey

The National Health Interview Survey (NHIS), sponsored by the National Center for Health Statistics (NCHS) and conducted by the Bureau of the Census, is a primary source for national estimates of acute illnesses and injuries, disability days, limitation of activity due to chronic conditions, and health care utilization. As such, the survey is often used to inform public policy in the health area. The NHIS uses a multistage probability sample of the United States population. NHIS data are collected through personal household interviews, and all adult members of the household 17 years of age and over who are at home at the time of the interview are invited to participate and respond for themselves. For individuals not at home at the time of the interview and for children, an adult family member provides the information. Interviews average 45 minutes per household. The 1981 survey completed interviews with 95 percent of approximately 42,000 eligible households. (See National Center for Health Statistics 1985 for a detailed description of the survey design and sampling procedures, their evolution over the past decade, and copies of questionnaire materials.)

The original 1957 Health Interview Survey questionnaire was the product of considerable work, and over the years since the inception of the survey, NCHS has initiated new research efforts on related topics and has sponsored a large number of methodological and evaluation studies, especially in the area of the reporting of chronic conditions. Jabine (1985) provides a wide-ranging review of these efforts. Further methodological progress for NHIS, as well as for other surveys can, we believe, be achieved by merging the perspectives of cognitive psychology and survey research. Drawing heavily on Jabine et al. (1984), we discuss issues specific to NHIS according to content area of the questionnaire—utilization of medical services, health conditions, and restrictions on activity. We then speculate about respondents' overall rating of their health.

One aspect of the NHIS questionnaire requires articulation at the outset. The current content, procedures, and specific questions used in the NHIS grew out of evolving conceptual definitions for health and illness. For example, if no action to seek diagnosis or care is taken in response to a symptom, that symptom is not considered to be linked to an illness episode. Some of the discussion below is aimed at a reconsideration of this conceptual framework while other parts deal with alternative ways of asking questions within the existing framework.

### Utilization

Two interrelated issues arise in considering reports of utilization. What is the process by which people decide to seek help? What influences underreporting of utilization? When someone decides to seek help depends, among other factors, on the nature of the condition or problem, the person's view of the medical system and how he or she relates to it, and the mechanism for paying for medical care. People may "schedule" their illnesses, seeking help when it is convenient or when it is about to become extremely inconvenient. The relation between a person and the medical system may be mediated by a household "gatekeeper," someone who usually calls the doctor and makes the appointments for the family.

These issues are interesting in their own right and they also have implications for survey methodology, especially how we might attempt to deal with underreporting. For example, questions that focus on the decision-making process might reduce underreporting of medical utilization. Such questions would also aid in identifying those people in the household who are most knowledgeable about utilization—the decision makers, the gatekeepers, the people who pay the bills or fill out the insurance forms. The reference period used for recall is also relevant to the issue of underreporting. For hospitalizations, NHIS has used a thirteen-month reference period in some years, a twelvemonth reference period in others, and in a recent pretest, a six-month period. It might be worthwhile to examine the estimated distributions of discharges by month under the different reference periods and to compare them to estimates based on hospital records.

#### Conditions

The NHIS questionnaire contains a series of items concerning medical conditions; for some of these, cognitive psychology suggests possible improvements. One problem with these questions is the terminology itself. Respondents may know they have a medical problem (for example, a bad back) without recognizing the medical term used on NHIS for it. Self-report data on medical conditions might be more accurate if the items asked for symptoms rather than conditions (see Fienberg, Loftus, and Tanur 1985a). A general item might be added to NHIS to ask for symptoms that bother the respondents but for which they do not know the cause. Allowing respondents to describe each problem in their own terms before the interview proceeds to more structured items might also reduce underreporting for injuries and their sequelae or for acute illnesses. The interviewer could then assist the respondent in answering the structured items in the light of the shared knowledge in unstructured form.

Even for the same individual, there may be several scripts or schemata for different types of health events (chronic conditions versus injuries), and different question orders or wordings may be needed to prompt the fullest recall of different types of conditions. If standard condition lists continue to be used, it might be easier to put them on individual cards and to group them according to conditions that tend to occur together. Respondents may find it easier to sort cards than to listen to lengthy lists, enabling them to deal with more items; grouping the conditions may facilitate retrieval.

There has been considerable research on the topic of reporting of sensitive topics and NCHS has been in the vanguard of this work. More embarrassing conditions (e.g., herpes) and less serious chronic

conditions (e.g., sinus trouble) may be especially prone to underreporting. One way to estimate the amount of underreporting is to compare prevalence rates based on NHIS data with those of other surveyssuch as the National Health and Nutrition Examination Survey (NHANES), which includes medical examinations, and the National Medical Care Utilization and Expenditure Survey (NMCUES), which incorporates checks of physician records-or with expert rankings of condition by prevalence. It would, of course, facilitate the comparisons if a common set of conditions were used, with the NHIS condition items being included in the NHANES interview so that the relation between self-reports of conditions and medical diagnoses could be established. We note that, on an aggregate level, prevalence estimates of hypertension based on self-reports in NHIS differ from those based on examinations in NHANES. Individual level analyses linking the two surveys would be necessary, however, to estimate the relation between individual self-reports and individual conditions. Of course, the usual standard used in discussing over- or underreporting is the corresponding medical record, but investigators doing validation studies should bear in mind that doctors' records, like respondents' reports, are prone to error.

The NHIS includes few items that assess mental health and does not include any of the standard scales that measure depression, stressful life events, or physical symptoms associated with stress (e.g., somaticization scales). Because of this omission, the NHIS data cannot be used to monitor trends in the mental health of the nation or to assess the relationship between physical conditions and psychological states. This has led to ad hoc arrangements for the collection of such mental health data and to possibly erroneous generalizations to achieve national estimates that are then used informally for policy purposes (see New York Times 1984).

### Restricted Activity

Items concerned with restrictions in activity brought on by illness or injury—such as the question readers were encouraged to answer at the start of this article—also have a subjective side that engenders a considerable concern about underreporting. Perhaps random probes could be used to find out how people interpret the term "restricted activity," (e.g., to learn whether respondents include mental illness when they think about "illness or injury"). For the questions on the loss of days from work (e.g., "During the past 12 months, about how many days did illness or injury keep you in bed for more than half of the day?"), underreporting might be reduced if respondents were asked first to report all days lost from work for any reason and then to say why each day was lost. Another approach might begin the restricted-activity section with questions about normal activities during the reference period. For each activity that they normally engage in, respondents would be asked whether it was curtailed or extended during the reference period and the reason for the change. (Some activities, such as reading or watching television, may increase during periods of illness.) It might also be useful to broaden the scope of the restricted-activity questions by asking respondents whether they had carried out their major activities during the reference period with less than their customary efficiency and, if so, why the change occurred.

#### The Subjective Side of Health

Notably omitted from the NHIS are items on emotional stress and mental illness. If the NHIS is viewed in part as a survey of attitudes, then the most serious omission is in the area of conceptions of health and illness. NHIS supplements could provide answers to many questions about the subjective side of health: What conditions do people include under the headings of health, illness, and injury? What health-related conditions are regarded as nonevents? How do the schemata or scripts, the cognitive frameworks or formats that psychologists use to describe patterns of recall, for one kind of health event (e.g., an injury) differ from those for other kinds (e.g., an acute illness or chronic condition)? How do different subcultures differ in their conceptions of health and illness and how do their taxonomies for illness differ? How do emotional states affect physical health?

One NHIS item asks respondents to rate their overall health:

2. Would you say [that your] health in general is excellent, very good, good, fair, or poor?

No other single question provokes so much discussion among analysts. How do respondents make this judgment? Part of the answer probably involves a comparison process; respondents may compare their current health with their health at other times, or they may compare themselves with other people of the same age. But perhaps people who are chronically ill recall most easily, for purposes of comparison, others who are chronically ill. Judgments of overall health are no doubt influenced by objective conditions but the influence may be limited (e.g., respondents who have successfully adjusted to long-term conditions may discount them in evaluating their health) and perceptions of objective conditions may be as much influenced by the overall judgment as the reverse. Research on underreporting of conditions demonstrates the impact of the overall evaluation on the reporting of conditions: underreporting is greater for respondents who see themselves as healthy. Global judgments in other domains typically integrate information from several dimensions, but little is known about the subjective dimensions of health.

It would not be surprising to find that the self-perceived health status item was affected by question context. The correlation between the condition items and ratings of overall health might be increased if the condition items came first in the interview. Even if there were a correlation under both question orders, a "positivity" bias might be expected, with respondents seeing themselves as healthier than their answers to the condition items would warrant. On the other hand, listing all of their conditions before rating their health might lead respondents to lower estimates of their overall health.

Actually, the respondent for the household is asked to rate the overall health status of all family members, except for the other adults present. Thus, the discussion above about how respondents make their judgments must be reinterpreted when it comes to judgments about the overall health of others in the family. (Recall the discussion above of self-reporting versus reporting for others.) We might expect to see a high correlation between the health status of the respondent (e.g., the mother) and the health status reported by that respondent for children. There is also an issue regarding the possible effect of answering for several people on the level of ill health that may be reported or perceived.

#### New Directions

Where have all of these suggestions for the NHIS led? The involvement of cognitive psychologists in a fresh examination of the NHIS has raised fundamental questions about the ways in which we gather selfreport data, especially in the health area. This involvement and these questions have stimulated new research efforts at the National Center for Health Statistics and the Bureau of the Census, one of which is described by Lessler and Sirken (1985).

The research work in connection with the NHIS is only one example of a new interdisciplinary focus on survey research involving projects at several universities and survey centers. In the long run, these efforts are sure to strengthen the quality of survey data in the health area and contribute to more informed contributions to public policy forums.

#### Appendix: Sampling and Nonsampling Errors in Surveys

Surveys pose questions to respondents—questions of varying degrees of difficulty for respondents to answer. But in all cases, the answers from the respondents are used to make estimates that are correct *for the population* of which the respondents represent a sample. One is not interested in estimates that are correct only for the people surveyed, nor in answers that are incorrect even for the people surveyed and hence, of course, incorrect for the population.

Two broad sources of error can arise in surveys: sampling error and nonsampling error. Sampling error arises from the very act of sampling. Even if all the procedures carried out to measure the quantity of interest on the members of this sample produced perfect accuracy, the luck of the draw might have given us a different sample. A different sample would, of course, include different people and would, therefore, be likely to yield slightly different estimates. Conceptually, the measurement of this variation over samples is the measurement of sampling error. Operationally, we estimate the variability associated with sampling errors over all possible samples by using the variation in a particular set of sample data and the size of the sample n. The usual measure of sampling variability, the standard error, decreases with  $1/\sqrt{n}$ , when a simple random sample has been taken. But most large-scale surveys use much more complicated sampling designsusually some form of multistage area probability sampling-rendering the usual formulas for standard errors inappropriate. Estimation of standard errors in these cases is usually based on either design effects or some version of half-sample procedures. Design effects are essentially deflation factors that reduce the sample size achieved by the complicated

sampling procedure in order to approximate the size of an equivalent simple random sample. Half-sample methods break the total sample repeatedly in random halves and carry out estimation procedures separately on each half. The variability among the resulting estimates provides a useful estimate of the standard error of the estimate. NCHS (1985) describes the details of the sampling procedures for the National Health Interview Survey and the procedures used to estimate standard errors.

Nonsampling errors constitute all the other errors that surveys are heir to. Some nonsampling errors are essentially random—copying errors, for example—and they will tend to cancel out as the sample size increases, though they will increase the standard errors of the estimates. Other nonsampling errors, such as memory errors and systematic coding errors, will tend to cumulate and cannot be decreased just by increasing the size of the sample. Nonsampling errors can themselves be subdivided into nonresponse errors (people are left out of the frame, left out of the sample, or do not answer specific questions) and response or "measurement" errors (answers are obtained from respondents but they are in some sense "wrong"). For reviews of many aspects of nonsampling error in large-scale sample surveys, see Mosteller (1978) and Fienberg and Tanur (1983).

Most of the issues discussed in this article deal with response errors—an area where we believe that the cognitive sciences have knowledge that could assist survey researchers as they try to help respondents get the answers right.

#### References

- Adams, R.M., N.J. Smelser, and D. Treiman, eds. 1982. Behavioral and Social Science Research: A National Resource, Part I. Washington: National Academy Press.
- Fathi, D.C., J. Schooler, and E.F. Loftus. 1984. Moving Survey Problems into the Cognitive Psychology Laboratory. Proceedings of Social Statistics Section, American Statistical Association. (In press.)
- Fienberg, S.E., E.F. Loftus, and J.M. Tanur. 1985a. Recalling Pain and Other Symptoms. *Milbank Memorial Fund Quarterly/Health* and Society 63(3):582-97.

<sup>-----. 1985</sup>b. Cognitive Aspects of Health Surveys for Public Information and Policy. *Milbank Memorial Fund Quarterly/Health* and Society 63(3):598-614.

- Fienberg, S.E., and J.M. Tanur. 1983. Large-scale Social Surveys: Perspectives, Problems, and Prospects. *Behavioral Science* 28:135-53.
- Jabine, T.B. 1985. Reporting Chronic Conditions in the National Health Interview Survey: A Review of Findings from Evaluation Studies and Methodological Tests. (Unpublished.)
- Jabine, T.B., M. Straf, J.M. Tanur, and R. Tourangeau. 1984. Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington: National Academy Press.
- Kruskal, W., and F. Mosteller. 1980. Representative Sampling, IV: The History of a Concept in Statistics, 1895–1939. International Statistical Review 48(2):169–95.
- Lessler, J.T., and M.G. Sirken. 1985. Laboratory-based Research on the Cognitive Aspects of Survey Methodology. *Milbank Memorial Fund Quarterly/Health and Society* 63(3):565-81.
- Loftus, E.F., and D. Fathi. 1985. Retrieving Multiple Autobiographical Memories. Social Cognition. (In press.)
- Mosteller, F. 1978. Errors: Nonsampling Errors. In The International Encyclopedia of Statistics, ed. W.H. Kruskal and J.M. Tanur, 208– 29. New York: Free Press.
- National Center for Health Statistics. 1985. The National Health Interview Survey, Design 1973–1984 and Procedures 1975–1983. Vital and Health Statistics. Series 1, No. 18, (PHS) 85-1320. Washington.
- New York Times. 1984. Almost One in Five May Have Mental Disorder. October 3.

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, Ann Cartwright, Mary Grace Kovar, Judith Lessler, Dorothy Rice, Monroe Sirken, and several referees for helpful comments and suggestions on an earlier draft of the paper. The section entitled "Some cognitive dimensions of the National Health Interview Survey" draws heavily on the report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology organized by the committee on National Statistics of the National Research Council (see Jabine et al. 1984).

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