

The Science of Self-Report

January 01, 1997

BETHESDA, MARYLAND-The accuracy and reliability of reports of one's own behavior and physical state are at the root of effective medical practice and valid research in health and psychology. To address this crucial element of research, the National Institutes of Health (NIH) held an informative conference here in November, "The Science of Self-report: Implications for Research & Practice," at which more than 500 researchers and policymakers learned about many of the critical limits of "self-report" as a research tool as well as some of the latest techniques to enhance its effectiveness.

Sponsored by the Office of Behavioral and Social Science Research (OBSSR), the symposium drew participants from virtually every area of health and medicine policy, practice, and research. The issue of self-report as a primary tool in research and as an unavoidable component in health care is of central concern to medical and social science researchers and medical and psychology practitioners, and many other scientists.

Drawing on the expertise of disciplines ranging from anthropology to sociology, the conference's 32 speakers and introducers featured 10 APS members, including the following NIH staff: Wendy Baldwin (NIH Office of Extramural Research deputy director), Norman Anderson (OBSSR director), Virginia Cain (OBSSR), Howard Kurtzman (National Institute of Mental Health), and Jaylan Turkkan (Program Co-Chair).

Value, Limits, and Improvements

"The issue we have to consider regarding self-report data is not that it should be replaced by external measurements but that we will always need self-report about many behaviors that are simply going to be unobservable by anyone else. We're going to need it because the interpretation of events may be important, and only the individual can provide those interpretations," said Baldwin in the opening session initiating the two-day conference. But assessing patient compliance with medical regimens and eliciting medical histories are just two of the particularly important areas in which self-report data is routinely, and perhaps blindly, accepted as reliable in many current medical contexts.

"Consequently, the effort should be placed on improving the self-report measures, as opposed to just looking for weaknesses or how they can be replaced by external measures," Baldwin emphasized in her comments that set the tone for the exceptionally practical conference. In fact, all speakers at the conference emphasized the invaluable nature of self-report measurements and called for a continual effort to improve their utility. "Where we have other validation, that's great! But we have a very important job ahead of us to make sure that we can learn why self-report either works well or doesn't, and when it works well, and when it doesn't," said Baldwin. Observational and experimental studies have shown that there are barriers to accuracy at every stage of the autobiographical report process—perception of the state of the self, encoding and storage of memory, understanding the question being asked, recalling the facts, and judging how and what to answer. And one intention of the conference was

to systematically review the documented problems across several research and medical contexts.

Reporting Symptoms and Physiology

Psychologist James Pennebaker, Southern Methodist University, presented data from studies on the ability to perceive one's own physical symptoms and other aspects of physiology such as heart rate.

"People are generally not good at this," he finds, "but there are interesting sex differences." In laboratory settings, men are better at perceiving their inner physiological states than are women, but the difference is largely erased when the studies are conducted in a more natural environment. This is because men and women emphasize different sources of information, when asked to define their internal states: Men rely more directly on internal bodily cues, while women rely more on situational cues. There is, of course, a lack of normal situational cues in the laboratory setting. One practical application of the skill of defining one's internal state is that diabetics must be trained to monitor their own blood glucose levels. Having instead to resort to chemical testing for glucose is often impractical.

Reporting Pain

Pain is not a simple sensory event and is not proportional to tissue damage, reported APS Member Francis Keefe, of Duke University Medical Center's Pain Management Program. In his discussion of the perception of pain, Keefe explained that pain is influenced by psychological, social and cultural factors, all of which act via a gating mechanism in the spinal cord, to influence the perception of pain. Also, the intensity of pain is separate from the degree of unpleasant affect associated with it, and this difference is reflected in pharmacology: While the drug fentanyl reduces the intensity of pain, diazepam reduces its unpleasantness.

Affect, in turn, modifies pain tolerance: A negative mood decreases tolerance for experimental pain in the cold pressor test, and affect at the time of pain influences the later recall of the intensity of the pain. Keefe says some pain specialists have advocated training patients with chronic pain (i.e., cancer patients) to be more emphatic and expressive in describing their pain to their doctors, in order to help ensure that adequate pain relief is prescribed. However, he says, many pain control techniques are effective because they influence affect and mood, more than they influence the intensity of the pain per se.

Reporting Data Through High-Tech "Diaries"

In his presentation on high-tech techniques to obtain self-report data, APS Member Saul Shiffman of the University of Pittsburgh's Department of Psychology indicated that written daily or weekly diaries have not proven themselves very good for accurate recording of simple objective events like smoking. In fact, people often fail even to accurately enter many simple events into their memory, let alone document them on paper. To avoid the problem, he describes the technique of Ecological Momentary Assessment (EMA). EMA requires the subject/patient to carry a custom-designed palm-top computer, which prompts him throughout the day to answer a question (e.g., "Are you smoking right now?"). The question is posed according to the desired sampling, which can be purely random over time or contingent upon various other behaviors (like drinking coffee). By avoiding recall completely, this method can provide a very revealing picture of the subject's pattern of behavior. It also generates great

quantities of data, but the analysis of that data poses unique and controversial statistical problems, because they do not fit into the standard definitions of repeated measures.

Reporting Temporal Frequencies of Behavior From Memory

Several presenters stressed the problems posed by aspects of the mechanisms of memory encoding and recall. Norman Bradburn of the National Opinion Research Center and the University of Chicago was the first of many speakers to note that remembering is very definitely a reconstructive task. It typically suffers from several distortions, including the bundling of events, and the tendency to “telescope” events. or bring them forward in the past when remembering.

Rounding errors are frequent when self-reported time intervals approach conventional discrete units of time (e.g., an hour, a week, a month, a year). Events six or eight days ago tend to be remembered as “one week” ago, and whatever the unit of time (e.g., an hour, a week, a month, a year). Events six or eight days ago tend to be remembered as “one week” ago, and whatever the unit of time appropriate to the interval, errors are made in whole unit chunks rather than in parts of units. “We are more likely to think in terms of three weeks, than 20 days,” said Bradburn. “Many people do not enumerate events, even when we might expect the question to lead them to do so. Rather, they estimate the number of events on the basis of some rule.”

Sex Differences Reporting Temporal Facts

And, just as many have thought, women *do* remember dates better than men. To help the respondent reconstruct the past, the interviewer or questionnaire should ask questions that are structured according to the way in which the events are likely to be encoded. Memories are rarely linked to calendar dates but rather to notable life events (e.g., graduation from college). Roger Tourangeau of the National Opinion Research Center further analyzed the distinction between questions designed to encourage estimation and questions designed to encourage recall of individual events.

Decompositional Approach

And, Geeta Menon of New York University’s Department of Marketing has analyzed the role of the decompositional question in eliciting recall of regular versus sporadic behaviors. Should we simply ask the open-ended question “How many times did you do X last week?” Or, should we ask the same thing using a decompositional approach? For example, “How many times did you do X while driving? While sitting at home? While working? ... ”

Menon’s research indicates that the open-ended question (“How many times did you do X in the last month?”) tends to encourage the subject to answer by referring to a “rule,” or an estimate of frequency. For regularly occurring behaviors this elicits accurate answers with the minimum of mental effort. For behaviors that are more sporadic, on the other hand, it is better to ask decompositional questions (i.e., to help the respondent by breaking the problem up into chunks). For irregular behaviors, a rule is less useful, and it is desirable to encourage the subject to recall each instance, using an enumeration strategy.

False and Forgotten Memories

Demonstrations that there can be both false negatives and false positives in memories of events that occurred long ago (or did not occur at all) have a particular relevance to the problem of sexual abuse of children. Speaking on the subject of false positives in memory, APS Fellow Elizabeth Loftus of the Psychology Department at the University of Washington presented findings—demonstrated in many experiments- that it is possible to create false memories. Such “memories” can be induced either by: (1) simply having the subject imagine a scenario vividly, and then later asking them to recount “memories” of similar events, or (2) by frankly telling a subject that a specific event happened and then reinforcing the associated “memory” by attempting to convince the subject of the authenticity of the event (e.g., by coaxing the subject with the question “Can’t you try to remember the time you got lost at the shopping mall?”).

People can import true memories from other events, thereby giving their false event memories seeming credibility, people can forget the source of a memory by wrongly attributing the memory of a fantasy to memory of a real event, and people make up completely unfounded facts, as well. The confidence one feels in the validity of one’s recall also has little correlation with its accuracy.

Linda Williams of the University of New Hampshire’s Family Research Laboratory has documented the other side of this issue, the false negative for a documented event. In these studies, children who were seen at hospitals for instances of sexual abuse were asked, many years later, to recall any such events. A substantial minority of the children, including those who had findings on physical exam that confirmed the abuse, failed to recall the instances. Interestingly, the forgetting was not correlated with the use of force or coercion by their abuser. The children were, however, more likely to forget abuse at the hands of individuals closest to them (i.e., in terms of familial relation, familiarity, or friendship).

Prolong the Pain

Psychologist Daniel Kahneman of the Woodrow Wilson School at Princeton University studies the memory of pain, as in painful medical procedures. Do we remember the quantity of pain as something like its intensity multiplied by its duration? Not at all. We remember an average of the moment of peak intensity and the pain at the end of the procedure. This has applications to colonoscopy, which is distinctly unpleasant, and for which one would like the subject to return for a repeat test every ten years. Strangely, Kahneman suggested, his research findings may mean that in order to make the long-term memory of the pain less severe, one should extend the time of the procedure, by keeping the colonoscope inserted, but not moving it. The pain is less for those last few minutes, even though we have added several minutes of diminished pain to the end of a painful experience.

Mood and Memory

APS Fellow John Kihlstrom of the Department of Psychology at Yale University took a logical and deductive approach to the problem of the influence of affect on memory. Although some experimenters have failed to find a link, he says, others have. There are some robust paradigms of mood-dependent memory. Because memory is reconstructive, not merely a readout of data, it is a cognitive task. Performance on other cognitive tasks is affected by mood, and so we should expect recall to be influenced by mood. For example, many mental patients report being abused in childhood. Is this a causal association or an example of preferential recall of mood congruent memories? What is needed to untangle this link, he says, are prospective studies.

Sensitive Topics

Nora Cate Shaeffer of the Department of Sociology at the University of Wisconsin-Madison addressed the problem of self-report in sensitive topics, such as sexual behavior or drug abuse. People will tend to present themselves in a positive light, sometimes to look good, and sometimes to “please” the researcher. The more serious an illegal behavior (e.g., the “harder” the drug), the less likely people are to report their recent use of it, while events in the distant past are less sensitive, and consequently are less likely to be concealed. Men tend to exaggerate their sexual histories, while women tend to understate them. But in any individual case, one doesn’t know how accurate a source is. Not only do people calculate the risk of revealing sensitive information (e.g., they may ask themselves “Will my spouse find out?” “Will the police find out?”), but they may even reinterpret the question, so as to allow themselves to answer evasively. (For example, a respondent may reason as follows: “Well, I did have that abortion, but I’m really not ‘the kind of person’ who would do that normally, so I’ll say “never.” Or, “This interviewer has a hell of a nerve; it’s none of his business, ergo I don’t feel dishonest lying about this.”)

Medical Compliance

Cynthia Rand of the Johns Hopkins University Asthma and Allergy Center discussed the problem of medical noncompliance. This generates a problem for research as well as practice. If everyone in a study takes half as many pills as they say they did, the FDA-approved and officially sanctioned dosage will be twice as high as the dosage that most people reported worked best. (Yes, this suggests that to avoid an overdose of medication, it may be best to be no more compliant than the average participant in the clinical trial that determined the proper dose!) What can be done to increase the honesty of responses? For starters, a physician’s question such as “You’re taking the pills the way I prescribed, aren’t you?” is not likely to uncover any problems with compliance. It is important to discuss the patient’s experience with the regimen in more detail, to reveal possible problems or hidden issues.

Ethics in Self-report

APS Fellow Donald Bersoff of the Villanova University School of Law addressed the knotty problems arising from ethical considerations in asking sensitive questions. If a subject reports self-destructive behavior, should the researcher intervene? Does that violate confidentiality and thereby compromise the autonomy of the subject?

Bersoff implores researchers to at least address these issues before beginning research studies. For example, before undertaking a study on the attitudes of teenagers toward dangerous behaviors, researchers should consider what they will do if they find out that a teenager is contemplating suicide, or is using heroin. “Have a plan, have a policy, discuss the pros and cons of breaking confidentiality before the issue comes up,” said Bersoff. “Too many researchers of sensitive topics don’t even think about what they will do, until they have in hand the information, and then they must agonize over their choices.”

Ethnic and Cultural Considerations

In many cases, the accuracy of a subject’s response depends on the understanding of the question. Spero

Manson of the Department of Psychiatry at the University of Colorado's School of Medicine has rewritten surveys specifically for Native American populations, and, with sensitivity to cross-cultural issues is able to raise the consistency of the scores very significantly. He cites one particular Indian culture in which it is considered very important never to give voice to certain negative thoughts; consequently, questions about suicidal ideation are either simply skipped by respondents at very high rates or are not answered frankly.

Efficient Screen for Depression

Ronald Kessler of Harvard Medical School's Department of Health Care Policy has been developing a short screening test for major depression. A psychiatrist asks questions until he knows the answers he is seeking, but screening tests must be designed for administration by non-specialists, with minimal preparation. Kessler's test, intended for screening large populations and subject to severe budget constraints, is an extreme version of this problem. The screen must not yield many false positives, it must be understandable by people of widely varying literacy and cultural backgrounds; 75% of the general population should score zero on the test, meaning that it is sensitive to only the serious cases. Interestingly, out of scores of possible questions, he has been able to narrow the survey to six very robust questions! They will be made available on the worldwide web at URLs <http://www.umich.edu/~icpe/> or www.umich.edu/~ncsum/.

If You Can't Beat Them, Join Them

Douglas Massey of the University of Pennsylvania's Population Studies Center presented a novel approach to securing sensitive or personal data in his presentation titled "When surveys fail," addressing the fact that many such research efforts simply demand that the researcher abandon the traditionally administered surveyor questionnaire. For a detailed study of undocumented workers from Mexico, for example, he has combined ethnography and surveys into an approach, called "ethnosurvey," in which anthropologists get to personally know the members of a Mexican town and then travel to a town in the United States where many of the workers go to work.

By demonstrating their involvement in the community and their knowledge of its members and worker's relatives, Massey and colleagues are able to establish trust, over a period of years, and to get answers about the laborers' experiences, documenting answers to non-standardized questions in an extensive data recording sheet. But, of course, even ethnosurveys are plagued by the same problems of faulty recall and encoding that researchers using more standard surveys encounter.

Practical Implications for Symptoms, Illness, & Health

Linking the findings from self-report research directly to medical practice, speaker Arthur Barsky of Harvard Medical School's Division of Psychiatry at Brigham and Women's Hospital pointed out that there is a very poor correlation between the patient's report of the seriousness of his symptoms, the medical findings of the presence of a pathological condition, and the patient's utilization of health care.

Why, then, given the flawed nature of self-report of symptoms, is history-taking so important in medical practice? Several speakers reaffirmed the dogma that history-taking must come first. The implication would seem to be that the real skill of history-taking is in the ability to get useful information about the

patient, despite the fact that his/her self-report is probably riddled with factual errors. As other speakers stated repeatedly during these two days, the respondent is always telling us something important. It just isn't always the answer to the question we thought we were asking!