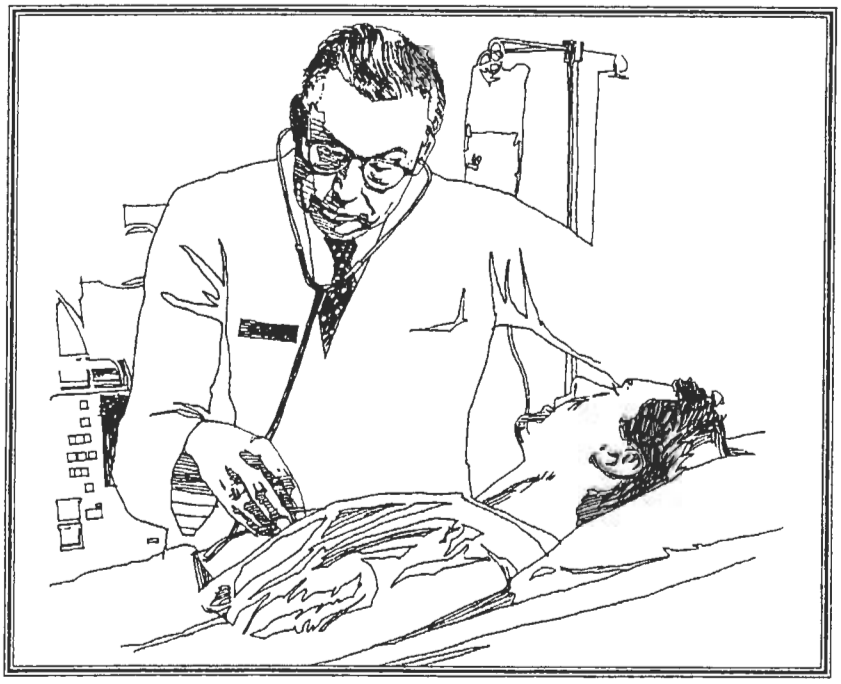
Conference Summary

Effective Dissemination of Clinical and Health Information



AHCPR
Agency for Health Care Policy and Research

Bonnach

Conference Summary

Effective Dissemination of Clinical and Health Information

Editors:
Lee Sechrest
Thomas E. Backer
Everett M. Rogers
Timothy F. Campbell
Mary L. Grady

U.S. Department of Health and Human Services
Public Health Service
Agency for Health Care Policy and Research
Rockville, MD 20852

AHCPR Pub. No. 95-0015
December 1994

Foreword

The Agency for Health Care Policy and Research (AHCPR) is charged with enhancing the quality, appropriateness, and effectiveness of health care services through a broad program of scientific research, clinical practice guideline development, data development, and dissemination of research findings and guidelines. To accomplish this ambitious mission, AHCPR must first identify what works and is cost effective in health care, and then disseminate this information to a number of key audiences. Physicians and other health care practitioners, who need to incorporate this information into their day-to-day practice, are one important audience. Others include patients, policymakers, educators, individual consumers and large-scale purchasers of care, and managers of health care systems. Given this diversity of audiences and information needs and the goal of building tighter linkages between actual practice and the underlying scientific evidence, effective dissemination is an essential component of AHCPR's mission.

Unfortunately, most traditional methods of disseminating information do not guarantee the actual use of the information. While it is clear that the disseminated information must be perceived as credible, easily understood, and readily available when needed, it is not clear which techniques, individually or in combination, are central for effective dissemination to specific audiences. As a starting point, AHCPR awarded a grant to Dr. Lee Sechrest at the University of Arizona to convene a conference that would define the field of dissemination in relation to health care practices, provide initial guidance about the dissemination methods most likely to be fruitful, and identify areas in need of further study.

This conference was constructed around three major issues:

(1) **Multiplicity of audiences.** AHCPR is required to disseminate information to the full range of health care providers, the health care business community, and patients and the general public. These three major audiences and their subgroups, pose quite different challenges with respect to dissemination methods and their results.

(2) **Characteristics of effective messages.** Some ways of structuring messages are more effective than others. Information must be tailored so that it reaches and can be carefully considered by members of the target audience. It must be presented in ways that enhance its credibility, while making it easily understood and persuasive. To the extent that scientific findings challenge long-standing practice or beliefs, these considerations are even more important.

(3) **Channels of communication.** Channels by which messages are communicated differ in terms of availability, technical complexity, cost, appeal, and other variables. Some channels have a long history of use, while others are just beginning to be developed and studied.

For each session at the conference, the principal speaker was requested to prepare and present an overall paper on a topic relevant to the focus of the session. This presentation was followed by three shorter papers addressing more specific issues or describing the particular experiences or knowledge of the speakers.

One session was devoted to each of the three audiences of practitioners, administrators, and consumers. One session considered the characteristics of effective messages, and one session each dealt with more conventional and more innovative channels of communication.

The papers in this volume bring to bear the expertise from both academic disciplines and the public and private sectors to the complex task of disseminating health and clinical information. This book is intended to further AHCPR's mission of providing information to policymakers, consumers, and the health care industry to foster improved decisionmaking and health outcomes. We hope that the document will serve as a springboard for further dialogue on what constitutes effective dissemination and how to achieve it.

Clifton R. Gaus, Sc.D.
Administrator
Agency for Health Care Policy and Research

Acknowledgments

Planning for the conference that is the basis for this volume began more than a year prior to the conference. Ultimately, a proposal for support of the conference was approved and funded. A grant (No. HSO6806) was awarded to Lee Sechrest at the University of Arizona.

Gratitude is owed in large measure to Margaret VanAmringe, former Director of AHCPR's Center for Research Dissemination and Liaison, and to her staff for support for the idea of having the conference and for making its occurrence possible. In addition to Margaret VanAmringe, many planning meetings were attended by Jane E. Linkletter, Allan J. Lazar, and Terry E. Shannon, all of whom made many helpful suggestions and facilitated arrangements for the conference.

General guidance in planning for the conference was provided by members of the Evaluation Group for Analysis of Data at the University of Arizona, and different members of that group performed a variety of services important success of the conference. Special appreciation is due Visvanathan, who served as the administrative assist Lee Sechrest during most of the period of planning f conference as well as during its aftermath and in the p ration of this volume. Karen Olson also provided va help along the way and during the conference.

We wish to thank all those persons who prepared p for presentation at the conference for their excellent e the nature of which will be evident in the contents c volume. We are also appreciative of those who attend conference and who participated in so many important in discussions, both formal and informal. No confere successful without a fine audience, and ours was exce

Finally, we wish to thank Ann R. Feild and Myra I for their editorial assistance.

Boruch

Contents

Contents (continued)

p. 543 ?

Chapter 4

On "Good, Certain, and Easy Government:" The Policy Use of Statistical Data and Reports[1]

Robert Boruch, Ph.D., and Erling Boe, Ph.D.
Graduate School of Education, University of Pennsylvania
Philadelphia, Pennsylvania

The object of this chapter is to understand something about why, how, and to what effect we seek information in sample surveys and controlled experiments designed to inform policy. It seems sensible to consider five fundamental questions:

1. Why should we do research on data use?

2. Why is our understanding of data use so limited?

3. What kinds of research on the use of data have been done?

4. What have we learned from such work?

5. What are the major problems in research on data use?

In this chapter, the policy uses of administrative records are excluded. They are considered elsewhere by others, including Alvey (1983) and Kilss and Alvey (1984).

Why Should We Do Research on the Use of Data?

The origins of contemporary efforts to exploit statistical data in the interest of public policy lie in the 17th century. John Graunt's *Natural and Political Observations Made Upon the Bills of Mortality* (1662; 1939), for instance, reflects the self-consciousness of those who sought to understand not only the data at hand, but how to exploit it with integrity.

Graunt's concluding chapter poses a question related to the one posed here:

> It may be now asked, "to what purpose tends all this laborious bustling and groping? To know the number of . . . people, fighting men, teeming women, what years are fruitful, what proportions neglect the Orders . . ." (p. 96).

Graunt makes no bones in his first response to his own question:

> To this I might answer in general by saying that those who cannot apprehend the reason of these enquiries are unfit to trouble themselves to ask them (p. 97).

His second answer places him among many contemporary scholars:

> . . . it is much pleasure in deducing so many abstract and unexpected inferences (p. 97).

His third response to his own question is politic:

> . . . the foundation of this honest and harmless policy is to understand the land and the hands of the territory to the governed according to all their intrinsic and accidental differences . . . by the knowledge whereof trade and government may be made more certain and regular . . . so as trade might not be hoped for where it is impossible . . . [all] necessary to good, certain, and easy government . . . (pp. 98–100).

Graunt's later remarks make it clear that he thinks it is in government's interest to pay attention to statistics. However, he is not at all convinced that there is any reason for disclosing the data to the general public.

Despite such early efforts, the history of statistical social research in the interest of public policy as a formal and sustained interest of government is embarrassingly brief. The problems encountered in collecting high-quality evidence, of course, are not trivial. That statistical data will invariably lead to "good, certain, and easy government" is not at all clear and some cynics would find in the announcement a utopian charm. Nor is it evident that government will always

[1] This research was developed with the support of National Science Foundation grants to the Social Science Research Council and the Center for Statistics and Probability, Northwestern University, during the 1980s and U.S. Department of Education grants to the University of Pennsylvania. An earlier, imperfect version of this chapter was presented at the annual meeting of the American Statistical Association, August 13–17, 1984, Philadelphia, in a symposium chaired by Richard Savage and organized by Bruce Spencer. Related work on the topic has been reported in Boruch, Wortman, and Cordray (1981) and the National Academy of Sciences Committee on National Statistics (1984). The mistakes in the authors' earlier thinking are recognized in this version.

share this view that statistical understanding is desirable, much less helpful, in making decisions.

One purpose of inquiry into the use of data for policy is a scholarly one. Knowing when, how, and why data are helpful; inventing theory to guide an understanding of the process; and generating evidence that illuminates misperceptions in use, neglect, and abuse of data seem worth serious intellectual attention. This purpose is legitimate.

A second purpose of research on data use by policymakers is proactive. Learning about the conditions under which data use can be enhanced and about the procedures that foster use seems important. Good data are a scarce and often expensive resource. To the extent that we learn to increase the likelihood, frequency, and depth of use, we drive down the costs of data. Poor data are in ample supply and learning how to ensure their identification and abandonment seems no less important.

A third purpose is ruthlessly defensive. Cuts in Federal statistical budgets and in related applied research budgets over the past two decades have been cyclic but notable. Chelimsky (1984), the U.S. House of Representatives Committee on Government Operations (1982), and the U.S. GAO (1984) reported declines of 18 to 20 percent. The cuts, both proposed and actual, were often met by opposition from data producers, of course, and from a fragmented user group. Systematic data on data use arguably would have prevented cuts that were uninformed; they could also have helped to justify the cuts that were warranted.

The opposition to cuts came from, among others, policymakers. These included Senator Paul Sarbanes and some of his colleagues on the Senate Joint Economic Committee; the President's Council of Economic Advisors, notably Chairman Michael Boskin; and the administration, through President Bush's Building a Better America program, which gave improvement of statistics a high priority (Wallman, 1989). Former Commerce Secretary Elliot Richardson (1990) was a strong advocate of the statistics programs.

Some Presidents of the United States have chosen to attend to the Science Advisor to the President. They have chosen to institutionalize such advice through, for instance, the President's Science Advisory Committee (PSAC), as Eisenhower did in 1957. Both the Science Advisor and the PSAC have been vehicles for transmitting quantitative information and translating it at the policy level.

Other Presidents have chosen to downgrade if not ignore such counsel. The PSAC was terminated by President Nixon in 1973. The Science Advisor under Nixon was not among the strongest. The reasons for these occurrences are arguably important for science policy and the Nation, and understanding the reasons is a legitimate aim of research (Golden, 1988).

The achievement of the ultimate goal, improving public policy through the use of data, is difficult to verify. It is a challenge to determine whether data have been used to this end and to learn how to use data better. This goal is a fourth purpose of research on the use of data.

Why Is Our Understanding of Data Use So Limited?

Suppose we again turn to Graunt (1662; 1939):

> Now having (I know not by what accident) engaged my thoughts upon the *Bills of Mortality*—and proceeding to "observations" which I happened to make (for I designed them not)...finding some Truths and not commonly- believed Opinions...I proceeded further to consider what benefit the knowledge...would bring to the World... (unpaginated Dedicatory).

His "consideration" of the benefits of his analysis and of the data could easily qualify him for the role of science advisor to the Crown or the Presidency. Such consideration has not always been encouraged.

Some reasons for inattention to the topic of data use have origins in the history of professional statistics. For instance, the Latin motto of the British Royal Statistical Society was, until recently, "Aliis exterendum." The literal translation is "Let others thrash it out." (This comes from Cochran, 1976, and he doesn't pursue the point.) The line of thinking is that it is "others" who demand or should demand the data. Why then shouldn't *they* worry about its use, instead of the statistician? It is perhaps partly a consequence of this attitude that nowhere in the great textbooks on sampling or on experiments do we find serious, direct attention to data use or misuse in the policy arena or anywhere else.

Our ignorance about the use of information and our inattention may also reflect a blind side of social science disciplines. Nobel laureate Theodore Schultz (1982), for instance, regretted a decade ago that members of his own discipline, economists, do no cost-benefit analyses of their work. More generally, economists often do not have systematic data on the consequences of data use because they do not collect data on use. A major exception is the zeal with which scholars subscribe to citation counts as an index of academic productivity. These counts are a proxy measure of the use of scholarly information by scholars, of course. Such counts are useful in evaluating basic research, but it is the use of research by policymakers that concerns us here.

Another reason for inattention to data use is that research on data use does not fit neatly into any single academic discipline. Rather, it is and perhaps must be multidisciplinary. This situation, of course, engenders problems of differences in scientific language and academic tribal customs, problems that are tractable but do take time to resolve.

It was startling, for example, to discover at a 1984 Social Science Research Council conference that sociologists who were doing research on the use of social research and statisticians who were producing research data had great difficulty in determining what one group could do for the other. Data on data use would seem to have implications for the technical design of surveys, for example. Statisticians could educe a few of them, but not without considerable effort. That conference, which I chaired, was a bust.

A final difficulty is access to policymakers. Researchers on data use have to reach policymakers to assay reasons for data use and nonuse. There has been less "insider" research than is necessary. It is only recently, for example, that we have had good insider postmortems on the rise and fall of the PSAC (Golden, 1988).

What Kinds of Research on the Use of Data Have Been Done?

How have we tried to understand the use of statistical data and research in the policy arena? Four kinds of approaches are easy to identify:

- Investigative.
- Descriptive.
- Prospective/experimental.
- Management.

All are analytical in the sense that the individuals who produce the work try to understand. There is delightful variety in each of these (admittedly pedestrian) categories, and this variety is considered in what follows.

Consider first the investigative approach. It will surprise some to learn that the U.S. Congress has been an important vehicle for understanding the use of statistical data and research. The U.S. House of Representatives Committee on Government Operations (1967) produced a fine series of papers on the topic. The Committee's members included individuals who are still in public service, such as Senator Robert Dole; people who served well, such as Representative Donald Rumsfeld; and individuals who were regrettably greedy by Internal Revenue Service standards, such as Representative Cornelius Gallagher. The Committee's consultants and staff deserve more recognition than they received.

The Committee's report puts the readers of this chapter in good company. Among those who thought hard about how to produce good data and encourage their use were individuals who took as an important task the making of law and policy, such as Senator Daniel Moynihan. The group included scholar-executives whose efforts helped to pave the way for policy research of diverse kinds in the United States, such as Henry Riecken. Those who contributed included young researchers such as Carol Weiss, who continues to make distinctive contributions on the use of research, and Don Price, Anatol Rapoport, Herb Simon, Ralph Tyler, Dael Wolfe, and Wassily Leontif.

The second category is descriptive social science based on sample surveys to determine which policymakers use what kind of data, when, and why. One may, for instance, identify a population of managers and decisionmakers: ask them whether they use social research, statistical data, or evaluations, and probe to understand their answers.

The work by Boyer and Langbein (1991), Carol Weiss (1977), Nate Caplan (1977, 1979), Robert Rich (1981), and others falls into this category. Work has been undertaken in a variety of settings; for example, applied research on mental health and education. It is reported in new refereed journals with which some readers may be unfamiliar—*Knowledge: Creation, Diffusion, Utilization* (published by Sage) and *New Directions for Program Evaluation* (published by Jossey-Bass)—and in a more familiar one, the *American Statistician.*

A third category of research includes prospective studies of ways to enhance use. Randomized field experiments, for instance, address such questions as how data use is influenced by kind of information, provider, time, and form. Fairweather, Sanders, and Tornatsky (1974), for example, worked to understand how to best introduce innovative treatment programs to over 300 Veterans' Administration hospitals in the United States. Stevens and Tornatsky (1980) undertook randomized experiments to discover how introducing the idea of one kind of data, notably evaluation data, could be enhanced in human services agencies by group versus private consultation and personal versus telephone consultation.

A fourth class of work involves management case studies of one or more organizations and their efforts to use and enhance the use of data. Examples include Sebring and Boruch's (1983b) study of some 20 divisions of the Illinois State Department of Education and their cooperative use of High School and Beyond longitudinal data. Martin Bulmer (1983) and others focused on royal commissions in the United Kingdom and Federal commissions in the United States to understand how these groups capitalize on social research generally. Lois Wise (1983) did a fascinating case study of the National Institute of Justice policy on data access and the institutional mechanisms that enhance or impede the repeated use of a class of data sets.

Within each category, there is great variety, of course. Consider, for instance, the survey approach. One may begin with a population of policymakers, as others have done, or one may identify a population of data sets or reports, then undertake interviews with potential and actual users of the data or reports. For example, our report to the Congress (Boruch, Cordray, Pion, and Leviton, 1981) on the use of program evaluation reports targeted the universe of reports produced by the Department of Education over a 3-year period. Similarly, Sherman and Hamilton (1984) focused on police departments' use of the Police Foundation's Minneapolis domestic violence experiment. Alkin and Lewy (1983) examined the use of Israel's Van Leer study of educational equity by the Knesset, the Ministry of Education and Culture, teachers, journalists, and others with influence on policy (see also U.S. GAO [1984]).

Another approach is to begin with a law or bill and undertake surveys to discover what kinds of data were and were not used in debate and in the bill's construction. This approach is not common. Indeed, some of the best known work on lawmaking, such as Eric Redman's *Dance of Legislation*, gives one the distinct impression that the only data used in constructing Federal bills concern the site of the membership of lobbying groups.

What Have We Learned From Such Work?

Let us summarize a few of the things that researchers have learned in their efforts to understand the use and interpretation of data.

Definitions and distinctions are crucial. In some work, for example, it has been helpful to define use of data, or research more generally, in terms of "serious consideration," "decisions," and "impact" (Leviton and Boruch, 1983). Other researchers, such as Carol Weiss (1977), have depended on functional definitions to establish use, for example, in improving, terminating, or enhancing programs. Robert Merton urged us in 1944–45 to attend to use along the entire continuum of decisions, with points representing alternatives under discussion, the actual choice among alternatives, implementation of the choice, and evaluation (U.S. House of Representatives, 1967).

Information may be seriously considered, for instance, but has no influence on specific decisions. Use in decisions made at one level of government may have no impact on law or regulations constructed at another level. How we define use,

then, influences what we can ultimately say about it. Definition, or lack of it, accounts for gratuitous squabbles over what is useful and what is not.

Similarly, classes of actual and potential users of data and how these classes change over time need to be nailed down. The classes include those who are formally acknowledged as policymakers, such as elected officials and senior executives. They may also include outsiders; for instance, interest groups with advocacy roles. They may include inside-outsiders: the President's Science Advisor, National Academy of Sciences advisory group members, and consultants to the Office of Management and Budget. They also can include those in the service of legislatures, notably legislative staff and audit agencies.

Conscientiousness and caution are warranted in talking about "decisionmakers." We recognize that decisions are often a matter of consensus and even more often involve iteration among levels of government, individuals, and committees. The process of decisionmaking is not easily traceable (see Lindblom and Cohen, 1979; Majchrzak and Stepanich, 1984).

Recognizing how statistics are woven into applied research, especially evaluations of projects, policies, and programs, is important. For example, almost all of the major studies undertaken by the U.S. Department of Education to determine program needs and how well programs are implemented and to estimate program effects are based heavily on statistical data. The same is true of studies undertaken under the Department of Labor's Knowledge Development Plan for the Youth Employment and Demonstration Projects Act of 1977.

Qualitative information plays a role in these studies, to be sure. However, just because such studies are labeled as "implementation analysis," "evaluation of outcomes," or "knowledge development" rather than as "statistics" does not mean we should ignore them. We do the statistical community a disservice by failing to recognize the statistical character of such studies. We are also likely to deceive ourselves about the use of data if we confine our attention to visible large-scale surveys such as the Panel Study of Income Dynamics, the National Assessment of Educational Progress (NAEP), and so forth.

The matter is also complex because most good studies address mixtures of questions that demand mixtures of data. For example, the National Institute of Education's Study of Compensatory Education Evaluation, one of the most useful of such efforts, involved mounting and synthesizing the results of some 35 projects to address numerous policy questions (Hill, 1980; Leviton and Boruch, 1984). A large-scale field experiment may obtain good quantitative data on a variety of response variables. If the experiment does not also entail qualitative and quantitative data on the implementation and character of the program and control conditions, it is unlikely to yield results useful to policy.

Special features of the phenomenon of data use can be measured—crudely, at least—although measurement is difficult. We may speculate that the quality of information, frequency of its use, and its accessibility are important in theory. Experience with a variety of interview protocols and survey instruments suggests that these traits are indeed reliably measurable.

Such measures can be expensive and time consuming, however. Moreover, concepts such as quality of data are complex, involving timeliness, relevance, accuracy, specificity, and quantity. Still, the perceptions of respondents seem to hang together; that is, "quality" often turns out to be a well-defined factor (in the sense of factor analysis) that helps to account for covariance among respondent reports about characteristics of useful data (O'Reilly, 1982).

Direct routine observation of data use is desirable but difficult. It would be expensive to make on-site visits to 1,000 school districts, for instance, to establish the extent to which such districts rely on the research and development (R&D) that is sponsored by the Federal Government. Consequently, one may be forced to rely on the low-budget approach—self-reports by school superintendents.

This approach has been taken by the U.S. Department of Education, among others (National Center for Education Statistics [NCES], 1990). A recent report was produced through the Fast Response Survey system (conducted by Westat, Inc.) at a fraction of the cost of direct observation. A high response rate (95 percent) was obtained from a probability sample of 1,000 districts. The report recognizes the limits of relying on self-reports, but it does not take the next step—to suggest side studies that would help establish the validity of such reports. One can imagine a parallel study of a small sample to dig more deeply into specific products and the school superintendents' uses of them.

Data and evaluations are indeed used. Verifiable evidence is fragmented but available in the form of surveys, cases, and experiments. Consider, for example, a recent study of Federal evaluation reports on educational programs, reports based to varying degrees on statistical data (Leviton and Boruch, 1983). We examined 14 evaluations of massive Title I programs in Compensatory Education, Upward Bound, Services to Neglected and Delinquent Youth, and others. Defining the use of an evaluation report as "serious consideration" or better, we found verifiable evidence for about 160 distinctive contributions in management or policy, including law. Adding instances that could not be corroborated increased the number by about 20 percent. Of all these contributions, 37 percent involved serious consideration alone and 63 percent involved verifiable decisions; 76 percent of those that involved decisions also involved an impact on laws, regulations, or administrative directives.

Recent work by NCES (1990) shows that organizational variety in the use of R&D is substantial. Of the Nation's 15,000 school districts, it was estimated that nearly 10 percent had not heard of any major Federal programs that allow districts to capitalize on Government R&D. About 40 percent of the districts recognized all four of the Government's major programs. Of the districts that did not recognize any federally sponsored resource for education, about half also failed to recognize any other resources. Regardless of recognition, over 20 percent of the districts surveyed did not report receiving R&D resources from any source since 1987 (p. 19).

At least some of the uses of policy-relevant data stem from the data's origins. Applied research and evaluations of a statistical kind are mission-oriented to begin with. At their best, they involve considerable front-end efforts to establish constituencies for each of the questions that they address (Hill, 1980; Leviton and Boruch, 1984; Rog, 1983). Other uses of more generalized (omnibus) studies also demand considerable planning to yoke a subset of survey topics with prob-

able policy issues; for example, Coleman, Bartot, Lewin-Epstein, and Olson's (1979) planning for High School and Beyond longitudinal surveys.

There are some important subtleties. One may argue, for instance, that estimates of frequency of data use are biased downward, relative to a true value, for a variety of reasons. Many data uses are buried in discussion of broad issues, and they are often forgotten. Further, turnover in groups of users and the biases that may prevail among respondents who are or were users can complicate matters. Claims that a report was provided to and used by a commission may be as easy to find as claims that the report was useless if the topic is controversial. Finally, we know that hard information is only one of a number of factors that are seriously considered in decisions. It should not surprise us when we have difficulty in establishing that a particular data set, research report, or evaluation has indeed been used.

It is also clear that the population of actual users is often a special subset of the population of potential users. For instance, the Police Foundation (Sherman and Hamilton, 1984) executed a remarkable randomized field experiment on alternative ways for police to handle domestic violence. The innovative treatment—arrest of one or more participants—differed from conventional police policy; that is, referral or discussion on site. Arrest, in Minneapolis at least, worked in the sense that it reduced subsequent violence. The Foundation later surveyed over 150 large police departments and determined that about 30 were able to remember the results of the experiment correctly. Six departments said the study influenced policy. For about 12 departments, the findings supported action already taken.

A third subtlety concerns the normal filtering systems that affect the use of data. For example, National Longitudinal Studies and High School and Beyond data have been used in academic reports by manpower experts (O'Neill, 1983). Those reports have been augmented through further data analysis by the Congressional Budget Office. The results of the Congressional Budget Office reports, in turn, are filtered and given serious attention that leads to decisions and perhaps recommendations by the National Academy of Sciences Committee on Youth Employment. These recommendations may then lead to changes in law, agency regulations, or policy.

Information-filtering processes are important. Indeed, they are an essential element of any institutional system that seeks to enhance the use of research findings; for example, the American Federation of Teachers' education research and dissemination program (Rauth, Biles, Billups, and others, 1983). The complex nature of the filtering process makes it very difficult to track and document data use. Audit trails are often absent or imperfect.

Ensuring the trustworthiness of the data and relevance to policy is crucial. The issues vary and options have been developed. Some matters are delicate. Consider, for instance, a statistical data agency that collects high-quality data relevant to policy. Should such an agency—NCES, for example—be responsible for the public release of the data, independent of operating agencies whose policies may be illuminated by the data? The independent release is a guarantee of integrity. Or should an operating agency, such as one of the Department of Education's program delivery components, release the data to the public, along with interpretation? The component agency, one may argue, knows more about the data's policy uses than does the statistical agency.

The wisdom of Congress, in the case of NCES, puts priority on the agency's independence: NCES is authorized to publicly disclose the data. This authority prevents, for example, data suppression by an operating agency whose interests may be threatened; it prevents release of data in company with analyses that are debatable at best and untrustworthy at worst.

In an effort to satisfy the policymaking agency's needs, NCES has created some options. The NAEP, for instance, was issued publicly in two steps. First, NCES released the raw data accompanied by explanations (in press conferences) of quality, technical detail, and descriptive meaning. Second, the data were released hours later by a policy group that provided background and policy-related explanations. This group, one that had advised on the design of the survey, had received and reviewed the data a little while earlier. The data's integrity and credibility are preserved by the first action; its policy relevance is preserved by the second.

A second and similarly delicate matter concerns the extent to which the statistical agency "analyzes" the data. To the extent that analyses are deep and directly relevant to specific volatile policy, the statistical agency may be at risk of lower credibility. Observers, including policymakers, may see a policy bias even where there is none. To the extent that the agency refrains from analyzing data, however, the agency's staff will be handicapped in understanding the data and how to do better surveys. To further complicate matters, it must be recognized that the choice of which data to collect and present and the choice of even simple cross-tabulations based on race, say, is "analysis" at a low level and in disguise, yet the statistical agency must make these choices.

Consider one of the options developed at NCES and at the U.S. Census Bureau. Each agency has developed a specialized research report system that involves staff members' production of analyses that are relevant to policy. A report issued under this system is independent of the public release of the data but cannot precede the release. Further, it is explicitly denominated as an individual's work, without institutional endorsement of findings. Finally, before publication the report is reviewed by peers (i.e., individuals who understand the topic at issue) in the interest of enhancing scholarship and quality. This tactic arguably enhances the caliber of staff work and the data product without putting the agency into an overt policy analysis posture.

For the record, the NCES experience in developing options is a fine subject for the attention of able scholars and bureaucrats. The quality of the NCES experience is reflected in a variety of documents; an excellent working document that deserves attention is the *NCES Role in Policy-Relevant Analysis of Data* (1991).

Staying close to the policymaking process seems crucial; maintaining distance from the policymaker seems important too. Staying close is important partly because the need for timeliness of reports, a factor that heavily influences use, cannot be managed easily at a distance. Maintaining distance seems important for the reputation and credibility of the source, although the importance of this factor is not clear (Boyer and Langbein, 1991).

How does one stay close and simultaneously maintain distance? The answer is with delicacy and with a keen sense of options.

Consider, for example, that until the 1970s, GAO's work consisted almost entirely of that which GAO chose to do. In the absence of serious involvement by Congress in the choice of the work, it is arguable that little of GAO's product was used then by policymakers in the legislative branch. A decade ago, for instance, legislative staff appear to have been dissatisfied with GAO's performance (see Boyer and Langbein, 1991).

According to former Assistant Comptroller General Harry Havens (1990), from 80 to 100 percent of GAO resources are now being dedicated to congressional requests and studies mandated by statute, more than doubling the rate in 1977. Each year, GAO testifies more than 200 times and initiates more than 1,000 projects. GAO now formally trains its senior managers to testify and provides experience in the activity to third-level area management.

Havens (1990) recognizes the downside of being this useful to the policymaker: "The close working relationship that has emerged between the GAO and the Congressional committees . . . could pose significant risks to credibility" (p. 22). Congressional committee members or staff may choose to frame requests that ignore important features of the problem at hand; indeed, they have done so at times. They may apply pressure to produce specific answers. They may pose questions that simply cannot be answered reasonably.

For GAO, maintaining distance is based on explicit institutional goals: independence, objectivity, and accuracy. The protection of these goals lies partly in strategic planning and involves identifying the broad issues to be explored and some priorities with both committee staff and agency officials. Within GAO's divisions as well as at the Comptroller General level, a major task is educating Congress broadly about what GAO does and what GAO's standards of evidence are. This microlevel education is important for the protection of the agency and the public, as well as Congress. That is, dialog between GAO and Congress reduces the likelihood that stupid or egregious questions will be put forward.

Given that there is agreement within GAO on at least two levels (Division as well as Senior Executive) that the issue is important, the agency's integrity is protected by a system that routinely requires that questions such as the following be addressed (again, we depend on Havens, 1990, p. 24):

- Is the right question being asked?
- Does GAO have an efficient and reliable way of developing the data needed to answer the question?
- Is the proposed analytic approach likely to yield a valid result?
- Are the right people working on the assignment?
- Will the results be available in time to meet the anticipated need?

Consider next the PSAC. The reasons for PSAC's demise under the Nixon administration are said to have been an explosive increase in the diversity of PSAC targets, divergence between the advice and policy positions, and limited access. The statute governing the Office of Science and Technology Policy (OSTP), which was created in 1976, was developed by Congress to foster access, in that the director is an agency head; integrity, in that Congress has oversight over some OSTP activity; and competence, in that the agency has staff. Richard Atkinson, former director of the National Science Foundation (NSF), maintains that "despite the evident craftsmanship of the OSTP Act, Congress has been unable to guarantee that the scientific implications of national policy issues will receive sustained attention . . . by the President . . . and senior policy officials" (1988, p. 13).

In education at the national level, the Planning and Evaluation Service at the Department of Education, for example, has the primary authority for monitoring, assessing, and evaluating service programs sponsored by the Federal Government. Much of the work is initiated at the behest of the Under Secretary or Secretary of Education and in conjunction with the various divisions of the Department. In the past, the Planning and Evaluation Service's attempts to remain independent of the program units resulted in "failures" to consider program concerns adequately and diminished the program's "ownership" of studies (Ginsburg, 1991, p. 4). The Service's current arrangements are based on joint agreements with program units. The latter do have veto power; however, disputes are resolved at the Deputy Secretary level.

Congress is not the primary audience for Department of Education reports. Nonetheless, the Planning and Evaluation Service has been a sufficiently important source of data that Congress has increased its requests of the unit. In fiscal year 1990–91, for example, about 25 percent of the projects given in the Secretary's evaluation agenda were statutorily mandated by Congress (NCES, 1990). For instance, Public Law 100–297 asked that the unit examine the consequences of alternative formulas for allocating chapter 1 (compensatory education) monies to States and school districts. The integrity of this process is supported partly by the separation of powers between the executive and legislative branches of Government.

Some intuition that certain variables are related to information use is supported by evidence. In particular, use is linearly related to blocks of ostensibly pertinent variables. However, there is a lot of unexplained variance.

For example, O'Reilly's (1982) regression analysis suggests that the frequency of use of many kinds of information is a linear function of the prospective users' access to and quality of the data, the context and task complexity, and the background of the data user. The adjusted R^2 for such equations, however, is in the range of 0.30 to 0.40, with individuals as the unit of analysis. An independent study of 100 congressional staffers by Boyer and Langbein (1991) produced regressions of the use of evaluative reports in health policy debates on factors such as timeliness, credibility of methods, and others in which the R^2 ranged from 0.50 to 0.72. The importance of timeliness here is reflected in policy analysts' views of how to ensure that data are useful (e.g., Ginsburg, 1991). The absence of a detractor (i.e., a critic of the evaluation) turns out to be a fine predictor of use both by members of Congress and their staffs, but clarity's importance is more evident for the members.

Lest this seem obvious to the reader, remember that at least some managers subscribe to the idea that data need not conform to statistical standards unless it has to be defended publicly. That is, decisionmakers can and do make some decent decisions based on poor data. Thus, the relationship between data use and data quality is imperfect.

A further complication is that the equation's coefficients vary depending on type of information used. Education is negatively related to use of files, for example, and positively related to use of outside consultants in some settings (e.g., O'Reilly, 1982).

Another complication is that people are sensitive to the different ways in which similar material is presented. Physicians, clinical psychologists, and others with similar training are more likely to attend to clinical case studies in reaching understanding or decisions. The cases themselves may, of course, have been sampled opportunistically from a statistically orthodox probability sample. Those with legal training may be less disposed toward both clinical cases and statistics. They may be content to rely on summaries or on the discussion on which the data are based (for instance, see Weiss, 1977 and 1983).

Focusing solely on the use of data sets or reports is misleading. The data production methods and structure are themselves useful products and ought to be taken into account. For example, in Sherman and Hamilton's (1984) survey of the users of results of the Minneapolis Domestic Violence Experiment, the authors remark that it is doubtful that any police department 5 years earlier would have been willing to engage in controlled randomized experiments. Once the Minneapolis experiment was completed and a report was issued, however, about 60 percent of knowledgeable departments reported a willingness to engage in randomized field experiments in the interest of discovering what works. Engaging professionals in the idea of randomized experiments for field testing new ideas is difficult. The Minneapolis data generated interest in the method of its collection. That interest helps to position the research community to capitalize more frequently on better methods.

Similarly, the techniques of doing surveys are an important byproduct of large-scale surveys. NAEP, for example, stimulated some States to learn how to augment the basic NAEP sample in the interest of obtaining better statewide estimates of student achievement. Further, items from NAEP tests are used in regional and some local districts because the NAEP technology seemed sensible to extend from National to State and regional levels of Government (Sebring and Boruch, 1983a).

Data-use enhancement policies and systems are important and warrant special study. There is considerable variety in such systems. Consider contemporary illustrations from both the private and public sector.

Since 1987, the Rockefeller Foundation has explicitly yoked theory, research, and action in its funding policies. The Equal Opportunity division, in particular, supports the Urban Institute and others in an effort that includes original data collection, analysis, and a remarkably vigorous system for distribution and personal presentation of reports in public policy forums (Gibson, Ricketts, and Rodriguez, 1991). The work is designed to enhance understanding of persistent poverty, to educe its implications, and to bring the results into policy forums. The Foundation's indices of the grantee's performance include the confirmable use of the Urban Institute's published work by policy-relevant groups, notably the GAO, the Office of Management and Budget, the Domestic Policy Council, and others. The indices include news coverage, especially by the printed media. The Urban Institute has been consistent in maintaining logs on all such matters and in pursuing an active series of colloquiums, seminars, and so forth in the Federal arena.

Needless to say, a Federal or State foundation would find it difficult to sponsor such an effort. One of the main goals of the relevant Federal agencies and the Rockefeller Foundation is, however, well met by the Foundation's initiative: having existing data well analyzed and discussed in the policy development effort.

In the Federal statistical community, the idea of enhancing data use during the 1980s hinged on the user-producer relationship. According to Wallman (1989), the discussions of the period resulted in a recognition that strengthening the user-producer relationship falls equally on those working inside Government and those outside, the creation of programs such as the American Statistical Association/National Science Foundation Agency Fellows and the Census Bureau's Annual Research Conference to help bring together insiders and outsiders, and a recognition that the environment is unstable. The organization that Wallman has headed, the Council of Professional Associations on Federal Statistics (COPAFS), is itself a product of that activity. COPAFS' interest in enhancing data use in the policy arena is a part of a much broader mission, to contribute "to developments that affect the integrity, quality, utility, and accessibility of Federal statistical resources" (p. 251).

In both the applied and basic research communities, the idea of data sharing has become important and is a legitimate element of data enhancement policy. Data produced by one research group are made available to others in the interest of verification, reanalysis, and better exploitation of the data by diverse communities. The production of public-use data tapes by the Census Bureau, NCES, and others constitutes an early form of sharing. More recently, the pool of shareable data has been expanded to include microrecords produced in evaluations of programs and projects and basic research records. The expansion helps to enhance the integrity of the social scientific enterprise partly by ensuring that checks on data and analysis are possible. Examples and a broad handling of the topic in biomedical, epidemiological, and social research are given in Cordray, Pion, and Boruch (1991).

The Federal agencies that have adopted formal policies on data sharing have done so partly in the interest of enhancing or ensuring the integrity of the statistical research enterprise and partly in the interest of enhancing data use. They include agencies that led the way during the early 1980s in framing policy on data disclosure, notably the National Science Foundation and the National Institute of Justice (Garner, 1981), and more recent entries to the field, such as the National Heart, Lung, and Blood Institute (Cordray, Pion, and Boruch, 1991). At least one Federal agency, the Agency for Health Care Policy and Research, has taken as a major mission the enhancement of existing data bases and their use. Part of the encouragement for this data sharing comes from the National Academy of Sciences (USA) Committee on National Statistics (1984).

At the project level, a great deal of promise in large-scale work lies with satellite or piggybacking policy—policy that directs our attention away from unitary Federal efforts and toward multistate and multiagency uses of data. By "piggybacking," we mean augmenting samples or questionnaires in special ways and exploiting the infrastructure creatively

to produce more useful information; for example, by conducting randomized field experiments to test projects in conjunction with ongoing longitudinal studies.

Piggybacking has often been provoked more by fiscal distress than by formal interest in data-use enhancement. The costs of major surveys, for instance, must now be shared because the resources of individual sponsors are scarcer.

It is obvious that data generated for national probability samples are often perfectly useless, on account of sampling variation, for decisions and understanding at State and local levels. Piggybacking (in the sense of augmenting) samples in the National Longitudinal Studies, High School and Beyond, and other studies, has made longitudinal and cross-sectional studies more useful at State and local levels. Recent meetings of the Council of Chief State School Officers, for example, suggest that a leadership role can be played by State commissioners of education in this area, partly because of public pressures for evidence on academic excellence and partly because resources at the State level can be allocated to augmentation of samples or questions.

Within some States, piggybacking has led to the creation of systems that dramatically enhance the apparent usefulness of augmented samples. In Illinois, for example, a new system designed by Sally Pancrazio makes it easier to understand High School and Beyond data, to translate the data into digestible forms for 20 subagencies, and to facilitate the committee process of educing the implications of the data (Sebring and Boruch, 1983b).

The most ambitious of piggybacking approaches recognize that large-scale surveys are passive and data from such surveys will often be equivocal. That is, imputing cause and effect from observational surveys will be difficult, if not impossible. Why then do we not adjoin randomized side experiments to capitalize on infrastructure and obtain less equivocal data on interim-to-long-term policy questions? The idea seems sensible and indeed has been explored, but its time seems not yet to have come for Federal, State, or regional efforts.

The option of designing multiple studies, in contrast to a single massive study, deserves more attention. The topic seems not to have been considered deeply in the literature on statistical policy. It has received some attention in evaluation policy.

The issue is whether, in the interest of enhancing the likelihood that data will be used, launching multiple studies would be a better tactic than mounting a single large study. One can argue, for instance, that audiences for research may find multiple smaller studies more persuasive or credible than a single massive effort, and that the results of multiple small studies might be more likely to be used in making a decision than would the results of a single large study.

The word "might" is used here because evidence on the credibility of single versus multiple trials is weak. Informally, we have observed that beyond a certain sample size a major study's credibility does not improve much for the lay public. For instance, a study based on a sample of 500 might be just as persuasive to a legislative body as one based on a sample of 1,500 or 2,000. Furthermore, three independent studies, each based on a sample of, say, 500, may be far more persuasive to these decisionmakers, and thus more likely to be used, than one study based on a sample of 1,500. We take this to be a plausible assumption partly because members of Congress do elicit information on many studies in hearings.

A related matter has been addressed by the Planning and Evaluation Service of the Department of Education. Criticisms of the Department's evaluations during the 1980s focused on the weaknesses of massive one-time-only studies. Recent efforts to invigorate and improve the Department's evaluation portfolio emphasize the use of multiple smaller studies, sometimes in conjunction with a much larger one, to address a range of issues that are of interest to the Department's components, Congress, and the Office of Management and Budget, and to facilitate independent corroboration of results (Ginsburg, 1991).

There are other arguments for multiple studies, to be sure. For the research manager, multiple studies may be warranted because they can be mounted sequentially and so exploit a small staff over a longer period of time rather than a large staff in one shot. Moreover, multiple independent studies may be more likely than a single study to generate good data. Managing several studies, however, is often more difficult than managing a single big one, or at least this is what we are told by managers.

What Are the Major Problems in Research on Data Use?

The conceptual and theoretical problems are severe. What do we mean, indeed, by use of data? Or use of research? Or of reports? The meaning adopted in some current research includes informing decisions, identifying problems, and enhancing understanding. These meanings are not mutually exclusive, of course. When use occurs at all, it occurs in different ways that overlap. The definition is imperfect in that it is simplistic.

It is also imperfect because there is no coherent theory on the topic of use of information. We have rules of thumb and informed hunches, but no formal causal models of data use in the policy arena seem to have been articulated well, and few have been tested. Small theories have been explored and they invite more thinking (e.g., Caplan, 1979, on types of data and the audience for use; Stevens and Tornatsky, 1980, on who uses what kind of information).

Without formal theory, let us present an unpatented questionary theory that has, at times, guided the way we think about data use in the policy arena:

- Do the potential users *know* about the data or report?
- Do they *understand* the data?
- Do they have the *capacity* (resources) to take action on the basis of the data?
- Are they *willing* to take action?

A nice feature of this questionary theory is that it focuses on variables (stages) that have a clear temporal relationship and therefore invites a probabilistic model. Consider, for instance, a fairly recent NCES (1990) survey of school districts' use of R&D. At the first stage (first question), only about 65 percent of the districts even knew about federally sponsored R&D centers. Of these, roughly half had received and, we

assume, understood a product or service. For the next two stages no data were elicited, but if we assume the same conditional probabilities for each (e.g., 0.50), it is easy to determine that any given R&D product will be used by, at most, 10 percent of the sample.

The "theory" (here we retreat, with chagrin, to quotation marks) invites one to think about such odds, and how to change them in various policy contents. Do the Department of Health and Human Services rates at any stage differ much from the 10 percent found by NCES (1990)? Why or why not? How do the odds change over time? Or do they? What are the implications?

This questionary theory, however, can be criticized as being anti-intellectual at best. We believe that *why* should be asked in conjunction with each question. Determining *why* policymakers know about the data is no less important than determining *whether* they know. Often, GAO reports that surface in congressional committees other than the committee that helped to generate the report are learned about from the press, which implies that GAO must be able to respond to this medium.

The community of bureaucrat-scholars may also object to this questionary theory. The charge that the theory is insufficient is fairly leveled. An additional question may then be adjoined to each basic question: *How* can the question be answered more frequently in the affirmative if the object is to ensure that good information is used? The question of *how* falls properly into the bureaucrat-scholars' ambit.

The practical problems of interviewing individuals and reviewing documents require considerable effort to solve. Corroboration is also difficult. For example, we have asked congressional staffers early in an interview about the use of evaluations and have been told, "We don't use evaluations." Ten minutes into the conversation, the staffers have told us incidentally that certain reports (and presumably the data they are based on) are used to create agendas for congressional hearings and to frame the questions that are addressed to witnesses. With more probing of those who inform congressional staffers (e.g., individuals at the Congressional Budget Office), we find again that data have been used (at times) even though the congressional staffer may be unaware of the fact. To compound the difficulty, studies of the use of data usually must deal with decisionmakers who are busy, transient, not disinterested in the political implications of data, and so on.

Obtaining verifiable evidence on use of data is difficult. Even deciding what corroboration means can be difficult. In Feviton and Boruch (1983, 1984), for instance, we defined corroboration as statements of independent informants; for example, a program manager and a congressional staffer. These informants are not truly independent observers of use in that each has a vested interest and, because the interests are either antagonistic or cordial, they must communicate. When are two sources really independent in such work? What do we do about quasi-independence? Once the ground rules are constructed, we may be able to exploit more powerful devices for estimating frequency of data use (e.g., dual-system estimation).

The relationship between costs of data and benefits of data use, and the way to marry the two, is not entirely clear. Part of the ambiguity here lies in estimating costs, part lies in the uncertainty of connections between data and use, and part lies in the fact that decisions must, perforce, be based on imperfect information.

One challenge to contemporary work, for instance, is assaying the costs of data, the data's benefits, and the way that data quality affects or is affected by the uses to which the data are put. We often hear, for example, that randomized field experiments for evaluating new programs are very expensive. Yet information on expenditures in most evaluations, including controlled field experiments, is usually sparse. This appears to be as true now as it was nearly 20 years ago when Riecken, Boruch, Campbell, and others (1974) directed attention to the topic. More generally, there exist no special accounting conventions and no substantial body of statistical or social science literature on costs of data. Cost-benefit analyses will be difficult or impossible unless we become more conscientious in such reporting.

It seems sensible to expect scholarly journals to take some responsibility in this arena. Why, for example, do policy or research journals not publish the costs of producing the data on which a journal article is based? The routine use of appendices (or paid subscription notes) containing such information would help immeasurably in learning about costs and in building cost-benefit models whose parameters are sensible.

A remarkable exception to the lack of cost information, in a narrow sense, is the General Social Survey undertaken by the National Opinion Research Center (1990) and sponsored by the National Science Foundation and other foundations. The survey, which has been conducted almost every year since 1972, generates data on American society on the basis of national probability samples. The target includes attitudes (e.g., religion, politics) and behaviors (e.g., risky behavior) that inform discussion of trends, scholarly models of society and its structure, and cross-national comparisons.

Each round in the General Social Survey costs about $1 million, petty cash to the Defense Department, perhaps, but an important piece of the National Science Foundation's budget. To better understand the costs of this survey (and eventually cost-benefit ratios), General Social Survey directors have been conscientious in tracking them. Since 1977, the field costs have increased by about 4 percent per year in constant dollars, a rate that is well below the rates in other industries. The sheer quantity of data collected each year has increased, from 230,000 elements in 1972 to 1,120,000 elements in 1990, bringing the cost per element from $0.22 to $0.14.

Consider next the idea that the connection between data and their use is uncertain. Once articulated, the notion is obvious. However, there also exist some remarkable exceptions to this uncertainty. The exceptions that are easiest to identify involve yoking the data to an administrative mechanism that requires the data's use. Resource allocation and revenue-sharing laws that depend on census data for their execution typify the genre.

Far less clear but no less important are those cases in which data are collected primarily to inform decisions, such as in evaluations and other applied social research. The data's use is often not guaranteed—every bill reenacts the perils of Pauline on the ice floes. The policy link may be weak, mixed in character, and not well predicted; the data's use in the scholarly arena often has no immediate quantifiable benefit to society.

A formal approach to a different but related family of problems, is given in a tantalizing but difficult paper by Spencer

and Moses (1990). They define vague data use as settings in which the precise role of the data in a decision cannot be known until after the data are collected. The probability of a correct decision is construed as a function of vagueness (nonoptimality) in the decision rule, cost of data, and the true value of the threshold parameter for the decision. The question of how much to invest in the data has a complicated answer in that high levels of vagueness generally imply lower optimal expenditures, but increased expenditures can be warranted in certain conditions despite extreme vagueness.

Consider now the good statistician's conscientious focus on quality of data. To put the matter bluntly, quality has no clear empirical, conceptual, or theoretical linkage to the actual uses to which data are put. The statistician's efforts to improve quality are important, to be sure, but the efforts often displace attention to use. This displacement is unlikely to serve science or policy well in the long run, when we need to understand when mediocre data will suffice, when decisions are robust against poor data, when poor data yield the same estimates of parameters that good data do, and when high-quality data are desirable, feasible, and useful. There usually is considerable tension between the need for timeliness in reports and the need for credible methodology when the latter may impede timeliness (Boyer and Langbein, 1991).

Consider an example, one taken from *Probability and Statistical Inference in Ancient and Medieval Jewish Literature* (Rabinovitz, 1975). It is a slender volume and does not sell well; nonetheless, the parables are instructive.

In the 13th century (or thereabouts), rabbis debated the issue of how one ought to draw a sample to judge the value of an olive crop. This sample estimate was to be used in determining a tithe for the temple. Should one, they asked, merely hire an apprentice to grab a handful of olives from the nearest basket? Or should one develop a high-quality sample design, hire a survey agency to draw the sample, and obtain the estimate?

The debate lasted a couple of decades, partly because it bore on profound questions. It anticipated contemporary discussions about "quick and dirty" versus "slow and clean" surveys, about investing in imperfect versus defensible evidence, and so forth. How often are decisions correct and the data wrong, and in what senses?

The resolution for the rabbis was as follows. In this matter, they said, the investment one should make in addressing the question—how much one should spend on the data—must depend on the origins of the demand for the information. If the demand is rabbinic in its origin, merely a matter of bureaucracy and manmade rule, then an apprentice will do. If, on the other hand, the demand is biblical in its origin, if God wants the answer, one had better invest in the higher-caliber evidence.

This rule of thumb is not a bad one, and other such rules need to be developed, but they could be difficult to apply. Other ways of thinking about the matter need attention. Little has been done, for instance, to formalize a way to make decisions that are robust against imperfect data. How do we frame such decisions to build contingencies into the system? How do we better characterize the costs of bad decisions and those who bear the costs?

Melnick (1990) returns some of the burden to the analyst. He reminds us that outdated data must often be used, for example, in allocating Federal funds to school districts on the basis of the number of children living in an impoverished family. Only the decennial census produces such data, and it is at least 10 years old at its oldest point of application. Similarly, data that are accurate at the national level are often assumed by decisionmakers to apply to subnational decisions. Melnick's plea is that the statistician learn to generate estimates of error due to time lag and lower levels of aggregation. He implies that integrity in the statistical community requires that such an attempt be made. He places no related burden on the data user.

Summary

The practice of seriously questioning the value of policy-relevant data can be traced at least to the 17th century, notably to John Graunt's treatise on the bias of mortality in Great Britain. Serious research on the use of data for policy purposes is more recent. The current robustness of such research is attributable partly to self-defense (i.e., attempts to avoid budget cuts in statistics and evaluation programs) and partly to interest in learning how to enhance the usefulness and utility of high-quality data.

The research on data use can be characterized as investigative, as in the case of formal congressional inquiries. It is often descriptive and analytic in that it is based on surveys of the users of data and reports, especially policymakers and public servants, and the statistical modeling of factors that predict use. It is prospective and scientific, at times. Full-blown randomized field experiments have been run to determine how to enhance data use in particular settings, and case studies of government systems for producing statistics and evaluative reports have been carried out. Among the lessons are the following.

Definitions of use of statistical data and reports are crucial. It is easy to distort an understanding of use by failing to define use and to include types of use. The definition of the target is no less important. Statistics are often woven into text, sources go unacknowledged, and so forth. This makes tracking the use of particular products difficult.

Direct observation of use is desirable but difficult. Surveys based on self-reports are imperfect but are a frequent device for estimating use.

There is no question that statistical data and evaluative reports are often used to frame problems, illuminate them, develop options, and identify possible solutions. There are exceptions: a small-to-moderate fraction of potential users exhibits a sturdy ignorance of or indifference to some reports. However, claims that cannot be corroborated, about either use or nonuse, ought to be suspect.

Trustworthiness of the data and credibility of the institution are crucial to the use of data but are often a delicate matter in policy environments. How the agency that produces data (and may analyze it) can stay close to the user, in the interest of enhancing use, and distant from the policymaker, whose interests are often not entirely dispassionate (and should not be), is often unclear. The strategies used by the Census Bureau, the GAO, and the Department of Education, vary. All are instructive. They entail repeated internal reviews, involving competing stakeholders in the production of data; the use of clear professional standards; and serious consideration of how the disclosure of information by policy entities versus statistical entities affects the credibility of the data.

The contemporary research confirms some expectations about factors that influence use of data and evaluative reports in the policy arena—timeliness matters. So does credibility of the source, a factor that is related to quality of the product, a third important factor that is itself related to the absence of detractors of the data. Other important factors appear to depend on the nature of the audience.

Focusing solely on the use of a particular data set or report is a bit misleading. The methods of producing the data, that is, procedures for the design of probability sample surveys and controlled field experiments, are also used by data users. That is, through using data, data users learn about better methods and become more willing to consider or even demand them.

Policies that enhance data use are important and can be creative. The Rockefeller Foundation's program on persistent poverty, for instance, specifically yokes the grant recipient's evaluative policy research to the recipient's active dissemination of the research in policy arenas. In the Federal arena, the emergence of data-sharing as a desirable function, the production of user-friendly data files and evaluative reports, and the use of active dissemination systems have become important. The development of dissemination policies based partly on criteria such as timeliness and credibility is no less important.

The idea of multiple studies rather than single large studies in the evaluation arena is worth attention. To the extent that multiple independent studies are more credible (as in the case of replications) or meet the demands of different users (e.g., some stress finances; others may focus on outcomes or program implementation), the idea arguably leads to better use of a particular product. As a matter of practice and policy, the idea has often (not always) been adopted by, among others, the GAO and the Department of Education.

There are remarkable challenges in doing research on the use of data and evaluative reports. Some challenges bear on the construction of better, more durable theory about why, how, and when data are used and by whom. Some bear on methods. Tracking the actual use of data rather than relying on reported use is hard. Learning how to track more easily is desirable. There are challenges also in learning about the actual costs (rather than proposal budgets) of data and reports and benefits of their use, so that we can compute and make decisions on the basis of cost-benefit or cost-effectiveness ratios.

Such work merits serious attention. It has in fact received attention from arguably able people—Florence Nightingale, Mark Twain, Charles Dickens, Abraham Lincoln—in their thinking about statistical information and evaluation reports. Current practice arguably goes well beyond the work of these historical figures. Understanding how to get beyond current practice is a fine challenge for those who are interested in understanding and improving society by improving the use and usefulness of statistical data.

References

Alkin, M.C. and Lewy, A. (1983). *Impact of a major national evaluation study: Israel's Van Lee report*. Los Angeles: University of California, Los Angeles, Center for the Study of Evaluation.

Alvey, W. (Ed.) (1983). *Statistics of income and related administrative record research: 1983*. Washington, DC: Department of the Treasury, Internal Revenue Service, Statistics of Income Division.

Atkinson, R.C. (1988). Science advice at the cabinet level. In Golden, W.T. (Ed.), *Science and technology advice* (pp. 11–15). New York: Pergamon.

Boruch, R.F., Cordray, D.S., Pion, G.M., and Leviton, L. (1981). A mandated appraisal of evaluation practices: Digest of recommendations to the Congress and the Department of Education. *Educational Researcher*, 10(4), 10–13.

Boruch, R.F., Wortman, P.M., and Cordray, D. (Eds.) (1981). *Reanalyzing program evaluations: Policies and practices for secondary analysis of social and educational programs*. San Francisco, CA: Jossey-Bass.

Boyer, J.F. and Langbein, L.I. (1991). Factors influencing the use of health information research in Congress. *Evaluation Review*, 15(5), 507–532.

Braskamp, L.A. and Brown, R.D. (Eds.) (1980). Utilization of evaluative information. *New Directions for Program Evaluation*, No. 5.

Bulmer, M. (Ed.). (1983). Social science and policy-making: The use of research by governmental commissions. *American Behavioral Scientist*, 26(5), 558–568.

Caplan, N. (1977). A minimal set of conditions necessary for the utilization of social science knowledge in policy formulation at the national level. In Weiss, C.H. (Ed.), *Using social research in public policy-making*. Lexington, MA: Lexington Books.

Caplan, N. (1979). The two communities theory and knowledge utilization. *American Behavioral Scientist*, 22(3), 459–470.

Chelimsky, E. (1984). *Program evaluation and the use of extent data*. Unpublished paper, U.S. General Accounting Office, Washington, DC.

Cochran, W.G. (1976). Early development of techniques in comparative experimentation. In Owen, D.B. (Ed.), *On the history of statistics and probability* (pp. 1–26). New York: Marcel Dekker.

Coleman, J., Bartot, V., Lewin-Epstein, N., and Olson, L. (1979). *Policy issues and research design: Report to the National Center for Education Statistics* (U.S.D.E. Contract No. 300–78–0208). Chicago, IL: National Opinion Research Center.

Cordray, D.S., Pion, G.M., and Boruch, R.F. (1991). Sharing research data: With whom? When? How much? In *Data management in biomedical research: Report of a workshop* (pp. 39–85). Washington, DC: U.S. Department of Health and Human Services.

Fairweather, G.W., Sanders, D., and Tornatsky, L. (1974). *Creating change in mental health organizations*. New York: Pergamon.

Garner, J. (1981). National Institute of Justice: Access and secondary analysis. In Boruch, R.F., Wortman, P.M., and Cordray, D.S. (Eds.), *Reanalyzing program evaluations* (pp. 43–49). San Francisco, CA: Jossey-Bass.

Gibson, J., Ricketts, E., and Rodriguez, A. (1991). *An assessment of Equal Opportunity's Persistent Poverty Program: Report to the board of trustees*. New York: The Rockefeller Foundation.

Ginsburg, A. (1991). *Reinvigorating program evaluation at the U.S. Department of Education*. Washington, DC: Office of the Under Secretary for Policy and Planning, Planning and Evaluation Service.

Golden, W.T. (Ed.) (1988). *Science and technology advice to the President, Congress, and judiciary*. New York: Pergamon Press.

Graunt, J. (1939). *Natural and political observations made upon the bills of mortality* (Wilcox, W., Ed.). Baltimore, MD: Johns Hopkins University Press. (Original work published 1662.)

Havens, H.S. (1990). *The evolution of the General Accounting Office from voucher audits to program evaluations* (GAO/OP–2-HP). Washington, DC: U.S. General Accounting Office.

Hill, P. (1980). Evaluating education programs for federal policymakers: Lessons from the NIE Compensatory Education Study. In Pincus, J. (Ed.), *Educational evaluation in the public policy setting*. Santa Monica, CA: RAND Corp.

Kilss, B. and Alvey, W. (Eds.) (1984). *Statistical uses of administrative records: Recent research and present prospects*. Washington, DC: Internal Revenue Service, Statistics of Income Division.

Leviton, L.C. and Boruch, R.F. (1983). Contribution of evaluation to education programs and policy. *Evaluation Review*, 7, 563–598.

Leviton, L.C. and Boruch, R.F. (1984). Why the compensatory education evaluation was useful. *Journal of Policy Analysis and Management*, 3(2), 299–305.

Lindblom, C.E. and Cohen, D.K. (1979). *Usable knowledge: Social science and social problem-solving*. New Haven, CT: Yale University Press.

Majchrzak, A. and Stepanich, M.E. (1984). *Public sector organizational decisionmaking: An analysis of information use*. Lafayette, IN: Krannert Graduate School of Management, Purdue University.

Melnick, D. (1990). Government and statistics in time and space. In *Proceedings of the annual research conference: Bureau of the Census* (pp. 381–386). Washington, DC: U.S. Department of Commerce, Bureau of the Census.

National Academy of Sciences, Committee on National Statistics. (1984). *Sharing research data*. Washington, DC: National Research Council/National Academy of Sciences.

National Center for Education Statistics (1990). *Use of educational research and development resources by public school districts: Survey report* (NCES–90–084). Washington, DC: National Center for Education Statistics.

National Center for Education Statistics (1991). *NCES role in policy-relevant analysis of data*. Unpublished report to the NCES Advisory Council on Educational Statistics.

National Opinion Research Center (1990). *The NORC general social survey: Questions and answers*. Chicago, IL: National Opinion Research Center.

O'Neill, J. (1983). Some relevant policy uses of the national longitudinal surveys. *Social Indicators Newsletter*, 18, 5–10.

O'Reilly, C.A. (1982). Variations in decision-makers' use of information sources: The impact of quality and accessibility of information. *Academy of Management Journal*, 25(4), 756–771.

Rabinovitz, N.L. (1975). *Probability and statistical inference in ancient and medieval Jewish literature*. Toronto: University of Toronto Press.

Rauth, M., Biles, B., Billups, K., and Veitch, S. (1983). *Executive summary: American Federation of Teachers educational research and dissemination program* (NIE-G–81–0021). Washington, DC: American Federation of Teachers.

Rich, R.F. (1981). *Social science information and public policymaking*. San Francisco, CA: Jossey-Bass.

Richardson, E.L. (1990). Social choices and the democratic process: The need to stretch the limits of statistical capacity. In *Proceedings of the annual research conference: Bureau of the Census* (pp. 375–378). Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Riecken, H.W., Boruch, R.F., Campbell, D.T., Caplan, N., Glennan, T.K., and Pratt, J. (1974). *Social experimentation*. New York: Academic Press.

Rog, D.J. (1983). *Evaluation planning: Can it ensure utilization*? Paper presented at the annual meeting of the Evaluation Research Society, Chicago.

Schultz, T.W. (1982). Distortions of economic research. In Kruskal, W.H. (Ed.), *The social sciences: Their nature and uses* (pp. 121–134). Chicago, IL: University of Chicago Press.

Sebring, P.A. and Boruch, R.F. (1983a). How is the National Assessment of Educational Progress used: Results of an exploratory study. *Educational Measurement: Issues and Practice*, 2(1), 16–20.

Sebring, P.A. and Boruch, R.F. (1983b). *The Illinois case: Uses of High School and Beyond at the state level*. Springfield, IL: Illinois State Board of Education.

Sherman, L.W. and Hamilton, E. (1984). *The impact of the Minneapolis domestic violence experiment: Wave I findings*. Washington, DC: Police Foundation.

Spencer, B.D. and Moses, L.E. (1990). Needed data expenditure for an ambiguous decision problem. *Journal of the American Statistical Association*, 85(412), 1099–1104.

Stevens, W.F. and Tornatsky, L. (1980). The dissemination of evaluation: An experiment. *Evaluation Review*, 4(3), 339–354.

U.S. General Accounting Office (1984). *Status of the statistical community after sustaining budget reductions* (IMTEC–84–17). Washington, DC: U.S. General Accounting Office.

U.S. House of Representatives, Committee on Post Office and Civil Service, Subcommittee on Census and Population. (1982). *Hearings: Impact of budget cuts on federal statistical programs* (Serial No. 97–41, 97th Congress, 2d session). Washington, DC: U.S. Government Printing Office.

U.S. House of Representatives, Committee on Government Operations (1967). *The use of social science research in Federal domestic programs: Part III—The relation of private social scientists to Federal programs on national social problems* (No. 72–113). Washington, DC: U.S. Government Printing Office.

U.S. House of Representatives, Committee on Government Operations (1982). *Reorganization and budget cutbacks may jeopardize the future of the Nation's statistical system* (House Report No. 97–901, 97th Congress, 2d session). Washington, DC : U.S. Government Printing Office (99–247–0).

Wallman, K. (1989). The user producer dialogue: Quality counts. In *Proceedings of the fifth annual research conference* (pp. 251–254). Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Weiss, C.H. (1977). Improving the linkage between social research and public policy. In Lynn, L.E. (Ed.), *Using social research in public policy-making*. Lexington, MA: Lexington Books.

Weiss, C.H. (1983). Toward the future of stakeholder approaches in evaluation. In Bryk, A.S. (Ed.), *New directions for program evaluation: Stakeholder-based evaluation* (No. 17). San Francisco, CA: Jossey-Bass.

Weiss, C.H. and Bucuvalas, M.J. (1977). The challenge of social research to decision-making. In Weiss, C.H. (Ed.), *Using social research in public policy-making*. Lexington, MA: Lexington Books.

Weiss, C.H. and Bucuvalas, M.J. (1980). Truth tests and utility tests: Decision-makers' frames of reference for social science research. *American Sociological Review*, 45, 302–313.

Wise, L. (1983). *Existing data resources of the National Institute of Justice: Report to the Director*. Washington, DC: National Institute of Justice, Office of the Director.