

The ANNALS of the American Academy of Political and Social Science

<http://ann.sagepub.com>

Meeting the Challenges of Evidence-Based Policy: The Campbell Collaboration


ANTHONY PETROSINO, ROBERT F. BORUCH, HALUK SOYDAN, LORNA DUGGAN and JULIO SANCHEZ-MECA

The ANNALS of the American Academy of Political and Social Science 2001; 578; 14

DOI: 10.1177/0002716201578001002

The online version of this article can be found at:
<http://ann.sagepub.com/cgi/content/abstract/578/1/14>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

On behalf of:



American Academy of Political and Social Science

Additional services and information for *The ANNALS of the American Academy of Political and Social Science* can be found at:

Email Alerts: <http://ann.sagepub.com/cgi/alerts>

Subscriptions: <http://ann.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations (this article cites 21 articles hosted on the SAGE Journals Online and HighWire Press platforms):
<http://ann.sagepub.com/cgi/content/refs/578/1/14>

Meeting the Challenges of Evidence-Based Policy: The Campbell Collaboration

By ANTHONY PETROSINO, ROBERT F. BORUCH, HALUK SOYDAN,
LORNA DUGGAN, and JULIO SANCHEZ-MECA

ABSTRACT: Evidence-based policy has much to recommend it, but it also faces significant challenges. These challenges reside not only in the dilemmas faced by policy makers but also in the quality of the evaluation evidence. Some of these problems are most effectively addressed by rigorous syntheses of the literature known as systematic reviews. Other problems remain, including the range of quality in systematic reviews and their general failure to be updated in light of new evidence or disseminated beyond the research community. Based on the precedent established in health care by the international Cochrane Collaboration, the newly formed Campbell Collaboration will prepare, maintain, and make accessible systematic reviews of research on the effects of social and educational interventions. Through mechanisms such as rigorous quality control, electronic publication, and worldwide coverage of the literature, the Campbell Collaboration seeks to meet challenges posed by evidence-based policy.

Anthony Petrosino is a research fellow at the Center of Evaluation, American Academy of Arts and Sciences and coordinator for the Campbell Crime and Justice Group.

Robert F. Boruch is University Trustee Professor at the University of Pennsylvania Graduate School of Education and cochair of the Campbell Collaboration Steering Group.

Haluk Soydan is the director of the Center for Evaluation at the Center for Evaluation, Swedish National Board of Health and Welfare and a cochair of the Campbell Collaboration Steering Group.

Lorna Duggan is a forensic psychiatrist in the United Kingdom and reviewer for the Cochrane Collaboration.

Julio Sanchez-Meca is a psychology professor at the University of Murcia in Spain and the director of the Unit for Meta-Analysis.

NOTE: We appreciate the assistance of Maureen Matkovich and colleagues at the National Criminal Justice Reference Service who supplied the data in Figure 1. The first author's work was supported by grants from the Mellon Foundation to the Center for Evaluation and from the Home Office to Cambridge University. Thanks to Brandon C. Welsh for his comments on earlier drafts of the article.

DONALD Campbell (1969) was an influential psychologist who wrote persuasively about the need for governments to take evaluation evidence into account in decisions about social programs. He also recognized, however, the limitations of the evidence-based approach and the fact that government officials would be faced with a number of political dilemmas that confined their use of research. The limits of evidence-based policy and practice, however, reside not only in the political pressures faced by decision makers when implementing laws and administrative directives or determining budgets; they also reside in problems with the research evidence.

Questions such as, What works to reduce crime in communities? are not easily answered. The studies that bear on these questions are often scattered across different fields and written in different languages, are sometimes disseminated in obscure or inaccessible outlets, and can be of such questionable quality that interpretation is risky at best. How can policy and practice be informed, if not persuaded, by such a fragmented knowledge base comprising evaluative studies that range in quality? Which study, or set of studies, if any at all, ought to be used to influence policy? What methods ought to be used to appraise and analyze a set of separate studies bearing on the same question? And how can the findings be disseminated in such a way that the very people Donald Campbell cared about—the decision makers in government and elsewhere—receive findings from these analyses that

they trust were not the product of advocacy?

Donald Campbell unfortunately did not live long enough to bear witness to the creation of the international collaboration named in his honor that ambitiously attempts to address some of the challenges posed by evidence-based policy. The Campbell Collaboration was created to prepare, update, and disseminate systematic reviews of evidence on what works relevant to social and educational intervention (see <http://campbell.gse.upenn.edu>). The target audience will include decision makers at all levels of government, practitioners, citizens, media, and researchers.

This article begins with a discussion of the rationale for the Campbell Collaboration. We then describe the precedent established by the Cochrane Collaboration in health care. This is followed by an overview of the advent and early progress of the Campbell Collaboration. We conclude with the promise of the Campbell Collaboration in meeting the challenges posed by evidence-based policy.

RATIONALE

Surge of interest in evidence-based policy

There are many influences on decisions or beliefs about what ought to be done to address problems like crime, illiteracy, and unemployment. Influential factors include ideology, politics, costs, ethics, social background, clinical experience, expert opinion, and anecdote (for example,

Lipton 1992). The evidence-based approach stresses moving beyond these factors to also consider the results of scientific studies. Although few writers have articulated the deterministic view that the term “evidence-based” suggests, it is clear that the vast majority of writers argue that decision makers need to—at the very least—be aware of the research evidence that bears on policies under consideration (for example, Davies, Nutley, and Smith 2000). Certainly the implicit or explicit goal of research-funding agencies has always been to influence policy through science (Weiss and Petrosino 1999), and there have always been individuals who have articulated the need for an evidence-based approach (for example, Fischer 1978). But there has been a surge of interest, particularly in the 1990s, in arguments for research-, science-, or evidence-based policy (for example, Amann 2000; Boruch, Petrosino, and Chalmers 1999; Nutley and Davies 1999; Wiles 2001).

One indirect gauge of this surge is the amount of academic writing on the topic. For example, in Sherman’s (1999) argument for evidence-based policing, decisions about where to target police strategies would be based on epidemiological data about the nature and scope of problems. The kinds of interventions employed, and how long they were kept in place, would be guided by careful evaluative studies, preferably randomized field trials. Cullen and Gendreau (2000) and MacKenzie (2000) are among those who made similar arguments about correctional treatment. Davies (1999), Fitz-

Gibbon (1999), MacDonald (1999), and Sheldon and Chilvers (2000), among others, articulated views about evidence-based education and social welfare.

A more persuasive indicator that evidence-based policy is having some impact is initiatives undertaken by governments since the late 1990s. Whether due to growing pragmatism or pressures for accountability on how public funds are spent, the evidence-based approach is beginning to take root. For example, the United Kingdom is promoting evidence-based policy in medicine and the social sectors vigorously (for example, Davies, Nutley, and Smith 2000; Wiles 2001). In 1997, its Labour government was elected using the slogan, “What counts is what works” (Davies, Nutley, and Smith 2000). The 1998 U.K. Crime Reduction Programme was greatly influenced by both the University of Maryland report to Congress on crime prevention (Sherman et al. 1997) and the Home Office’s own syntheses (Nuttall, Goldblatt, and Lewis 1998). In Sweden, the National Board of Health and Welfare was commissioned by the government to draft a program for advancing knowledge in the social services to ensure they are evidence based (National Board of Health and Welfare 2001).

In the United States, the Government Performance and Review Act of 1993 was implemented to hold federal agencies responsible for identifying measurable objectives and reaching them. This has led to the development of performance indicators to assess whether there is value

added by agencies. The 1998 reauthorization of the Safe and Drug Free Schools and Communities Act required that programs funded under the law be research based. The news media now commonly ask why police are not using research-based eyewitness identification techniques (Gawande 2001) or why schools use ineffective drug prevention programs (for example, Cohn 2001).

All of these signs seem to indicate more than a passing interest in evidence-based policy. As Boruch (1997) noted, different policy questions require different types of scientific evidence. To implement the most effective interventions to ameliorate problems, careful evaluations are needed. An evidence-based approach to what works therefore requires that these evaluations be gathered, appraised, and analyzed and that the results be made accessible to influence relevant decisions whenever appropriate and possible.

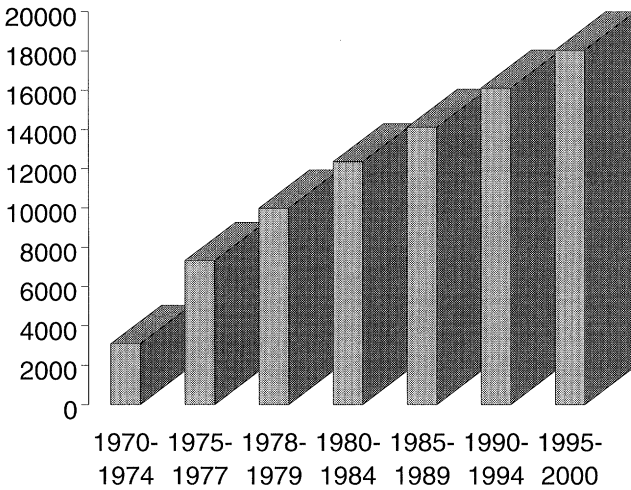
Challenges to evidence-based policy: Evaluation studies

If evidence-based policy requires that we cull prior evaluation studies, researchers face significant challenges in doing so. For one, the relevant evaluations are not tidily reported in a single source that we can consult. Instead they are scattered across different academic fields. For example, medical, psychological, educational, and economic researchers more routinely include crime measures as dependent variables in their studies (for example, Greenberg and Shroder 1997). These evaluations are as relevant as those reported in justice journals.

Coinciding with fragmentation, evaluation studies are not regularly published in academic journals or in outlets that are readily accessible. Instead a large percentage of evaluative research resides in what Sechrest, White, and Brown (1979) called the fugitive literature. These are government reports, dissertations and master's theses, conference papers, technical documents, and other literature that is difficult to obtain. Lipsey (1992), in his review of delinquency prevention and treatment studies, found 4 in 10 were reported in this literature. Although some may argue that unpublished studies are of lesser quality because they have not been subjected to blind peer review as journal articles are, this is an empirical question worthy of investigation. Such an assertion, at the very least, ignores the high-quality evaluations done by private research firms. Evaluators in such entities do not have organizational incentives to publish in peer-reviewed journals, as professors or university-based researchers do.

Relevant studies are not reported solely within the confines of the United States or other English-speaking nations. Recently the Kellogg Foundation supported an international project that has identified more than 30 national evaluation societies, including those in Brazil, Ghana, Korea, Sri Lanka, Thailand, and Zimbabwe (see <http://home.wmis.net/~russon/icce/eorg.htm>). One argument is that it is not important to consider evaluations conducted outside of one's jurisdiction because the cultural context will be very different. This is

FIGURE 1
**CUMULATIVE GROWTH OF EVALUATION STUDIES:
 NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE DATABASE**



SOURCE: The National Criminal Justice Reference Service (www.ncjrs.org).

another assertion worthy of empirical test. Harlen (1997) noted that many in education believe evaluative studies and findings from different jurisdictions are not relevant to each other. This ignores the reality that interventions are widely disseminated across jurisdiction without concern for context. For example, the officer-led drug prevention program known as D.A.R.E. (Drug Abuse Resistance Education) is now in 44 nations (Weiss and Petrosino 1999). Harlen (1997) suggested that we must investigate the role of context across these evaluations. This is most effectively done through rigorous research reviews. Such reviews, however, are difficult with international literature without translation capabilities.

Another challenge to gathering evaluative studies is that there is no finite time by which the production of this evidence stops. Research, including evaluation, is cumulatively increasing (Boruch, Petrosino, and Chalmers 1999). An example is provided in Figure 1. Consider the cumulative growth of studies indexed either as evaluation or as evaluative study by the National Criminal Justice Reference Service for its database of abstracts. The data underscore the challenge faced in coping with the burgeoning evaluation literature.

It would be good to identify and acquire all relevant evaluations and keep abreast of new studies as they become available. It would be even better if all evaluations were of

similar methodological quality and came to the same conclusion about the effectiveness of the intervention. Unfortunately not all evaluations are created equal. The results across studies of the same intervention will often differ, and sometimes those differences will be related to the quality of the methods used (see Weisburd, Lum, and Petrosino 2001). This highlights the importance of appraising evaluation studies for methodological quality.

Challenges to evidence-based policy: Reviewing methods

But what is the best way to draw upon these existing evaluation studies to understand what works and develop evidence-based policy? Certainly, relying on one or a few studies when others are available is very risky because it ignores evidence. For example, relying on one study if five relevant studies have been completed means that we ignore 80 percent of the evidence (Cook et al. 1992). It is true that the one study we pick may be representative of all the other studies, but as mentioned previously, studies in an area often conflict rather than converge. Evaluation studies themselves are part of a sampling distribution and may differ because of chance probability. Until we do a reasonable job of collecting those other studies, an assertion of convergence based on an inadequate sampling of studies is unsupported.

Criminologists have generally understood the problem of drawing conclusions from incomplete evidence and have a half century's experience in conducting broad surveys of the literature to identify relevant

evaluations (for example, Bailey 1966; Kirby 1954; Lipton, Martinson, and Wilks 1975; Logan 1972; Witmer and Tufts 1954). Although a few of these earlier syntheses were remarkably exhaustive, the science of reviewing that developed in the 1970s focused attention on the methods used in reviews of research.

Methods for analyzing separate but similar studies have a century of experience (Chalmers, Hedges, and Cooper in press), but it was not until the 1970s that reviews became scrutinized like primary reports of surveys and experiments. This was ironic, as some of the most influential and widely cited articles across fields were literature reviews (Chalmers, Hedges, and Cooper in press). Beginning in the 1970s, not only were the traditional reviews of evaluations under attack, but the modern statistical foundation for meta-analysis or quantitative analysis of study results was also being developed (for example, Glass, McGaw, and Smith 1981; Hedges and Olkin 1985). Research confirmed that traditional reviews, in which researchers make relative judgments about what works by using some unknown and inexplicit process of reasoning, were fraught with potential for bias (Cooper and Hedges 1994). Quinsey (1983) underscored how such bias could affect conclusions about research following his review of research on sex offender treatment effects: "The difference in recidivism across these studies is truly remarkable; clearly by selectively contemplating the various studies, one can conclude anything one wants" (101).

One major problem noted with regard to traditional reviews was their lack of explicitness about the methods used, such as why certain studies were included, the search methods used, and how the studies were analyzed. This includes the criteria used to judge whether an intervention was effective or not. Consider the debate over the conclusions in the Lipton, Martinson, and Wilks (1975) summary of more than 200 correctional program evaluations, briskly reported first by Martinson (1974). Despite finding that nearly half of the evaluations reported in Martinson's article had at least one statistically significant finding in favor of treatment, his overall conclusions were gloomy about the prospects of correctional intervention. The criterion for success was not readily apparent, but it must have been strict (Palmer 1975).

These earlier reviews, like Martinson's (1974), were also problematic because they seemed to rely on statistical significance as the criterion for judging whether an intervention was successful. This later proved to be problematic, as research showed that statistical significance is the function not only of the size of the treatment effect but of methodological factors such as sample size (for example, Lipsey 1990). For example, large and meaningful effects reported in studies with small samples would be statistically insignificant; the investigator and traditional reviewer would consider the finding evidence that treatment did not succeed. Given that most social science research uses small samples, moderate and important intervention

effects have often been interpreted as statistically insignificant and therefore as treatment failures.

Systematic reviews

Evidence-based policy requires overcoming these and other problems with the evaluation studies and methods for reviewing them. There is consensus among those who advocate evidence-based policy that systematic reviews are an important tool in this process (Davies 1999; Nutley, Davies, and Tilley 2000). In systematic reviews, researchers attempt to gather relevant evaluative studies, critically appraise them, and come to judgments about what works using explicit, transparent, state-of-the-art methods. Systematic reviews will include detail about each stage of the decision process, including the question that guided the review, the criteria for studies to be included, and the methods used to search for and screen evaluation reports. It will also detail how analyses were done and how conclusions were reached.

The foremost advantage of systematic reviews is that when done well and with full integrity, they provide the most reliable and comprehensive statement about what works. Such a final statement, after sifting through the available research, may be, "We know little or nothing—proceed with caution." This can guide funding agencies and researchers toward an agenda for a new generation of evaluation studies. This can also include feedback to funding agencies where additional process, implementation, and theory-

driven studies would be critical to implement.

Systematic reviews, therefore, are reviews in which rigorous methods are employed regardless of whether meta-analysis is undertaken to summarize, analyze, and combine study findings. When meta-analysis is used, however, estimates of the average impact across studies, as well as how much variation there is and why, can be provided. Meta-analysis can generate clues as to why some programs are more effective in some settings and others are not. Meta-analysis is also critical in ruling out the play of chance when combining results (Hedges and Olkin 1985).

Systematic reviews have other byproducts. They can reconcile differences between studies. Because each study document is scrutinized, systematic reviews can underscore deficiencies in report writing and lead to better systems for collecting data required by reviewers. Reviews also ensure that relevant evaluations—which may have been ignored and long forgotten—are eternally utilized. It is satisfying to investigators to find their studies still considered 20 years or more after completion (Petrosino forthcoming).

Systematic reviews have been influential. This is understandable, as Weiss (1978) predicted that policy makers would find good syntheses of research compelling because they would reconcile conflicting studies when possible and provide a comprehensive resource for their aides to consult. Hunt (1997) discussed how the results from meta-analysis contradicted conclusions in earlier traditional reviews, in areas such as

psychotherapy, class size, and school funding. Palmer (1994) noted that meta-analyses like Lipsey's (1992) helped to counter the prevailing pessimism about the efficacy of correctional treatment generated by earlier reviews.

Challenges to evidence-based policy: Current systematic reviews

There seems to be growing convergence among researchers and others that systematic reviews are a critical tool for evidence-based policy (for example, Nutley, Davies, and Tilley 2000). This is reflected in the decision by the United Kingdom's most prestigious social science funding agency, the Economic and Social Research Council, to support an evidence-based policy and practice initiative featuring systematic reviews (see <http://www.esrc.ac.uk/EBPesrcUKcentre.htm>). On closer scrutiny, however, we find that there are some challenges to the use of systematic reviews in evidence-based policy as they are currently done.

One problem is that there is often a lack of transparency in the review process. Completed syntheses are generally submitted to peer-reviewed journals, long after the research question has been determined and the methods selected. Except for rare occasions in which reviewers submit a grant proposal for funding, researchers do not a priori describe why they are doing the review and what methods they will employ. Without transparent processes from beginning to end, ex post facto decisions that can influence a review and slant it knowingly or

unknowingly toward one conclusion or another are possible. This is especially important in persuading policy makers who want to be sure that research is not the product of slick advocacy.

Because there is no uniform quality control process, systematic reviews, like evaluations, range on a continuum of quality. In some cases, the quality is due to the methods employed. Some reviewers may use meta-analytic methods but inadequately describe their decision process. Other reviews may detail exhaustive search processes but then use questionable methods for analysis. It is difficult for even the discerning reader to know how trustworthy the findings are from a particular review. In other cases, the quality of the review is due to the way it is reported. Sometimes the nature of the outlet dictates how explicit and transparent the reviewers can be. Some reviews, particularly those prepared for academic print journals with concerns about page lengths, are briskly written. Dissertations and technical reports are usually very detailed but are less accessible to readers. Systematic reviews may all contain Materials and Methods sections, but the level of detail in each may vary depending on the dissemination outlet.

Policy makers and other interested users of research have a wide range of questions about what works to which they want answers. Although funding agencies will sponsor reviews at times to meet these information needs, reviews are generally conducted because of the

interests of individual researchers. For example, in criminology, offender treatment has been a controversial and popular topic and the target of most systematic review activity (Petrosino 2000). Other important areas for review such as police training, services for crime victims, court backlog interventions, and so on have been inadequately covered. Evidence-based policy requires that evaluations in these areas be synthesized, even if they are less relevant to longstanding criminological debates.

Even if reviews did cover many more questions than they currently do, they are often not disseminated in such ways that decision makers and the public can get them. All too often, reviews are published by academics in peer-reviewed journals, outlets that are not consulted by policy makers. In fact, decision makers often get their information about research from news media, which can selectively cover only a few of the thousands of evaluative studies relevant to crime and justice reported each year (Weiss and Singer 1988). Tyden (1996) wrote that publishing an academic paper to disseminate to policy makers was akin to shooting it over a wall, blindly, into water. The path to utilization by decision makers was haphazard at best.

To examine dissemination further, we analyzed the citations for 302 meta-analyses reported by Lipsey and Wilson (1993) of psychological and educational treatment studies. Nearly two-thirds listed in the reference section were published in academic journals. These were scattered across 93 journals during the years

covered (1977-1991). Only the *Review of Educational Research* published an average of one review or more per year. Unless researchers were using other unknown mechanisms such as oral briefings and internal memos to communicate to decision makers, it seems very unlikely that this evidence got into the hands of anyone other than research specialists working in narrow areas.

Most systematic reviews also tend to be one-off exercises, conducted only as funding, interest, or time permits. Rarely are they updated to take into account new studies that are relevant to the review, a challenge that is more significant given the cumulative growth of evaluation reports highlighted in Figure 1. Yet years may go by before an investigator pursues funding to update an existing review. The methodology and statistical foundation for meta-analysis is still rapidly evolving, with improved techniques and new software being developed to solve data problems. It is rare to find reviews that take into account these new techniques, conducting analyses to determine if results using different methods converge.

Some reviewers publish in print journals, an inefficient method for disseminating reviews. Because print journals find them too costly, cases in which reviewers take into account cogent criticisms by others and conduct reanalysis are rarely reported. Unlike medical journals, criminological journals do not have a strong tradition in routinely printing letters to the editor that respond to criticisms with additional analyses.

Some journals also have lengthy lag times between submission and publication, delaying the dissemination of evidence even further.

THE COCHRANE COLLABORATION

Are there ways of overcoming challenges to using systematic reviews in evidence-based policy? A precedent for doing so was established in the health care field. Archie Cochrane was a noted epidemiologist who wrote persuasively about the need for medical practitioners to take scientific evidence into account in their practice. Cochrane (1972) lamented the fact that although randomized trials had shown some practices to be effective and others harmful, clinical practitioners and medical schools were ignoring the information. He later (Cochrane 1979) wondered why the medical sciences had not yet organized all relevant trials into subspecialties so that decision makers could take such evidence into account. A protégé of Cochrane, an obstetrician turned researcher named Iain Chalmers, soon identified and reviewed randomized trials relevant to childbirth and prenatal interventions (see www.cochrane.org).

In the early 1990s, the U.K. National Health Service (NHS), under the direction of Sir Michael Peckham, initiated the Research and Development Programme with the goal of establishing an evidence-based resource for health care. Because of the success of their earlier project on childbirth and pregnancy studies, Chalmers and his colleagues

were asked to extend this effort to all areas of health care intervention. The U.K. Cochrane Centre was established, with core funding from the NHS, to begin the work. It soon became clear that the amount of work far surpassed the capacity of one center or one nation to take into account. In 1993, in honor of his mentor, Chalmers and his colleagues launched the international Cochrane Collaboration to "help people make well-informed decisions about healthcare by preparing, maintaining and promoting the accessibility of systematic reviews of the effects of healthcare interventions." In just 8 years, the Cochrane Collaboration has been able to organize thousands of individuals worldwide to contribute to its work. Much more information about the Cochrane Collaboration can be found at its Web site, www.cochrane.org. But the Cochrane Collaboration, in a very brief time, established a number of mechanisms to address challenges to using systematic reviews in evidence-based health care policy.

For example, collaborative review groups (CRGs) are responsible for the core work of systematic reviewing. CRGs are international networks of individuals interested in particular health areas such as breast cancer, epilepsy, injuries, and stroke. Each CRG has an editorial board, generally comprising persons with scientific or practical expertise in the area, who are responsible for quality control of protocols (plans) and completed drafts of reviews.

It is useful to examine how a Cochrane review is prepared. First, individuals approach the CRG in

which the intended topic area seems appropriate. Once a title for the proposed review is agreed on, it is circulated to ensure that no other similar reviews are being prepared by reviewers from other CRGs. Reducing overlap and duplication is a crucial goal for the Cochrane Collaboration, as scarce resources must be used judiciously. Reviews are needed in so many areas of health care that wide coverage is a priority. Once the title is agreed on, the reviewers must then submit a protocol for the review. The protocol is a detailed plan that spells out a priori the question to be answered, the background to the issue, and the methods to be employed. The protocol then goes through a round or two of criticism by the CRG editors. It must conform to a certain template to facilitate electronic publication using the Cochrane Collaboration's software, *Review Manager*, or *RevMan*. Once the protocol is approved, it is published in the next edition of the quarterly electronic publication, the *Cochrane Library*, and made available to all subscribers for comment and criticism. The editorial board must decide which criticisms should be taken into account.

The reviewers then prepare the review according to the protocol. Although deviation from the plan is sometimes necessary, the protocol forces a prospective, transparent process. Post hoc changes are readily detected, and analyses can be done to determine if they altered findings. After the reviewers conduct the review and write up a draft, it too is submitted to the CRG editorial board. Once the review draft is

completed, the editors critique it again. It is also sent to external readers, including researchers as well as practitioners and patients. This round of criticism is designed to improve the methods in the review and to ensure that the final review is written as accessibly as possible to a nonresearch audience, including health care patients, providers, and citizens. For each completed Cochrane review, the Cochrane Consumer Network crafts a one-page synopsis written accessibly for patients and other consumers and posts the synopses at its Web site (see <http://www.cochrane.org/cochrane/consumer.htm>). Once the review is approved, it is published in the next issue of the *Cochrane Library* and again made available for external criticism by subscribers. Again the editors and the reviewers have to determine which of these criticisms ought to be taken into account in a subsequent review update. Cochrane reviews must be updated every 2 years, to take into account new studies meeting eligibility criteria.

Another important mechanism for the Cochrane Collaboration is the methods groups. These are international networks of individuals who conduct systematic reviews focused on the methods used in systematic reviews and primary studies. For example, a methods group might collect all systematic reviews in which randomized trials are compared to nonrandomized trials. In this review, they would seek to determine if there is a consistent relationship between the reporting of random assignment and results. Their objective is to

make sure that decisions in reviews, such as setting eligibility criteria, be informed as much as possible by evidence. The Cochrane Collaboration is also facilitated by 15 centers around the world; they promote the interests of the collaboration within host countries, train people in doing systematic reviews, and identify potential collaborators and end users. Finally Cochrane fields and networks, such as the Cochrane Consumer Network, focus on dimensions of health care other than health problems and work to ensure that their priorities are reflected in systematic reviews.

The main product of the Cochrane Collaboration is the *Cochrane Library*. This electronic publication is updated quarterly and made available via the World Wide Web or through CD-ROMs mailed to subscribers. The January 2001 issue contained 1000 completed reviews and 832 protocols (or plans for a review) in one central location using the same format. Wolf (2000) noted that the uniformity allows the reader to understand and find all of the necessary information in each review, facilitating training and use. Another important feature of the *Cochrane Library* is the *Cochrane Controlled Trials Register (CCTR)*. The *CCTR* has more than a quarter million citations to randomized trials relevant to health care, an important resource in assisting reviewers find studies so they can prepare and maintain their reviews.

Empirical studies have reported that Cochrane syntheses are more rigorous than non-Cochrane systematic reviews and meta-analyses pub-

lished in medical journals. For example, Jadad and his colleagues (1998) found that Cochrane reviews provided more detail, were more likely to test for methodological effects, were less likely to be restricted by language barriers, and were updated more than print journal reviews. The *Cochrane Library* is quickly becoming recognized as the best single source of evidence on the effectiveness of health care interventions (Egger and Davey-Smith 1998). Reviews by the Cochrane Collaboration are frequently used to generate and support guidelines by government agencies such as the National Institute for Clinical Excellence (for example, see Chalmers, Hedges, and Cooper in press). In 1999, the U.S. National Institutes of Health made the *Cochrane Library* available to all 16000 of its employees. It is now accessible by all doctors in Brazil, the U.K. NHS, and all U.K. universities (Mark Starr, personal communication, 2001). Finally the queen recognized Iain Chalmers for his efforts with the United Kingdom's greatest honor: knighthood!

Thus the Cochrane Collaboration has been able to meet many of the challenges posed by evidence-based policy in health care. By requiring detailed protocols, the Cochrane Collaboration addresses the lack of transparency in most systematic reviews of research. Through rigorous quality control, they produce commendable reviews. By publishing electronically, dissemination is quickened, and the ability to update and correct the reviews in light of new evidence is realized. By provid-

ing an unbiased, single source for evidence and producing reviews, abstracts, and synopses for different audiences, they facilitate utilization.

THE CAMPBELL COLLABORATION

With the success of the Cochrane Collaboration, the same type of organization was soon suggested for reviewing social and educational evaluations. Adrian Smith (1996), president of the Royal Statistical Society, issued a challenge when he said,

As ordinary citizens . . . we are, through the media, confronted daily with controversy and debate across a whole spectrum of public policy issues. Obvious topical examples include education—what does work in the classroom?—and penal policy—what is effective in reducing reoffending? Perhaps there is an opportunity . . . to launch a campaign directed at developing analogues to the *Cochrane Collaboration*, to provide suitable evidence bases in other areas besides medicine [emphasis added]. (378)

A number of individuals across different fields and professions organized and met to determine how best to meet this challenge. Several exploratory meetings were held during 1999, including two headed by the School of Public Policy at University College—London, and one organized in Stockholm by the National Board of Health and Welfare. These meetings, which included researchers and members of the policy and practice communities, provided evidence that the development of an infrastructure

similar to Cochrane’s for social and educational intervention including criminal justice should be vigorously pursued (Davies, Petrosino, and Chalmers 1999; www.ucl.ac.uk/spp/publications/campbell.htm).

Early days and progress

The Campbell Collaboration was officially inaugurated in February 2000 at a meeting in Philadelphia, with more than 80 individuals from 12 nations participating. The Campbell Collaboration was founded on nine principles developed first by the Cochrane Collaboration (see Table 1). At the February 2000 inaugural meeting, it was agreed that the headquarters (secretariat) should reside at the University of Pennsylvania. An international eight-member steering group was officially designated to guide its early development.

The first three Campbell coordinating groups (similar to Cochrane’s CRGs) were created to facilitate systematic reviews in their areas: education, social welfare, and crime and justice. The Campbell Education Coordinating Group is focused on developing protocols and reviews in the following critical areas: truancy, mathematics learning, science learning, information technology learning, work-related learning and transferable skills, assessment and learning, comprehensive school reform, school leadership and management, professional education, and economics and education. The Campbell Social Welfare Coordinating Group has also organized itself into several areas: social work, transportation, housing, social casework with certain ethnic clientele, child welfare, and

TABLE 1
PRINCIPLES OF THE
CAMPBELL COLLABORATION

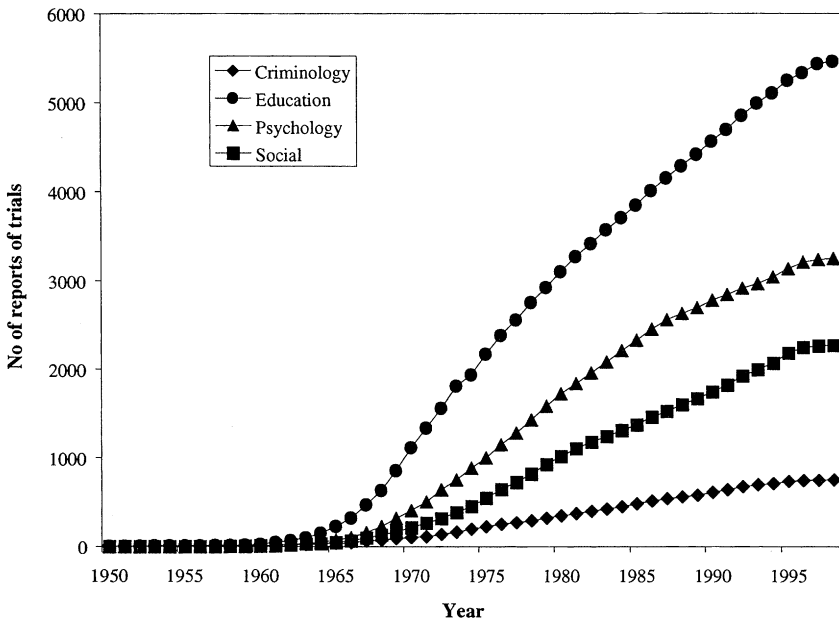
Collaborating by fostering open communication, cooperation, and transparency
Building on the enthusiasm of individuals by involving and supporting people of different skills and backgrounds
Avoiding unnecessary duplication by coordinating and maximizing economy of effort
Minimizing bias by maximizing scientific rigor, assuring broad participation, and avoiding conflicts of interest
Keeping current by ensuring that systematic reviews are kept up to date through incorporation of new evidence
Ensuring relevance by promoting reviews that use outcomes that matter to people making choices
Promoting access by widely disseminating the collaboration’s products and taking advantage of strategic alliances
Ensuring quality by inviting critical comment, applying advances in methodology, and developing systems for quality improvement
Continuing to renew by updating reviews, editorial processes, and key functions and by engaging new collaborators

SOURCE: C2 Steering Group (2001).

employment programs within the welfare system. The early progress of the Campbell Crime and Justice Coordinating Group is described elsewhere in this issue (see Farrington and Petrosino 2001).

The Campbell Collaboration Methods Group was developed to increase the precision of Campbell reviews by conducting reviews to investigate the role of methodological and statistical procedures used in systematic reviews, as well as characteristics in original studies (see <http://web.missouri.edu/~c2method>). Three methods subgroups were created during the past year, including statistics, quasi-experiments, and

FIGURE 2
C2-SPECTR



SOURCE: Petrosino et al. (2000).

process and implementation subgroups. In conjunction with the Campbell secretariat and the coordinating groups, the methods group has taken the lead in developing a preliminary quality control process based on the Cochrane model (see Appendix A). The Campbell Communication and Dissemination Group will develop best practice in translating results to a variety of end users, including policy makers, practitioners, media, and the public.

To facilitate the work of reviewers, the *Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR)* is in development. As

Figure 2 shows, preliminary work toward *C2-SPECTR* has already identified more than 10000 citations to randomized or possibly randomized trials (Petrosino et al. 2000), and this has now been augmented with new additions to the data file. Like the *CCTR* in health care, *C2-SPECTR* should serve as a productive resource and facilitate preparation and maintenance of reviews. Plans to build a complimentary database of nonrandomized evaluations are being discussed.

During 2000, Campbell and Cochrane groups mutually participated in the NHS Wider Public Health Project (see www.york.ac.uk/

crd/publications/wphp.htm), an attempt to collate evidence from systematic reviews relevant to the United Kingdom's intended health policies. As the NHS now considers education, social welfare, and criminal justice directly or indirectly related to public health, Campbell groups were involved (for example, Petrosino 2000). Several hundred systematic—or possibly systematic—reviews were identified in these areas and can now be used to help us map the terrain and identify target areas where high-quality reviews are needed.

Funding from the Ministry of Social Affairs of Denmark has been acquired to establish a Campbell Center in Copenhagen to facilitate efforts in the Nordic region. Resources have also been secured to create the Meta-Analysis Unit at the University of Murcia in Spain (see <http://www.um.es/sip/unidadma1.html>), the first step in developing a Campbell Center for Mediterranean nations. Objectives of these centers include facilitating reviews through training, identifying end users and collaborators, and promoting dissemination and utilization.

These are just a few of the many developments in the early days of the Campbell Collaboration. Although the collaboration anticipates creating an electronic publication that will make available *C2-SPECTR* and other helpful resources, its critical product will be high-quality systematic reviews. For the Campbell Collaboration to achieve the kind of success Cochrane has obtained in the health care field, it will have to ensure that these reviews are as

unbiased and technically sound as possible. Appendix B provides a preliminary model of stages in a Campbell review.

CONCLUSION

Donald Campbell articulated an evidence-based approach before the methods of systematic reviewing and meta-analysis became staples of empirical inquiry. Still we think he would be pleased with the efforts of hundreds of individuals who are working worldwide to advance the international collaboration that bears his name.

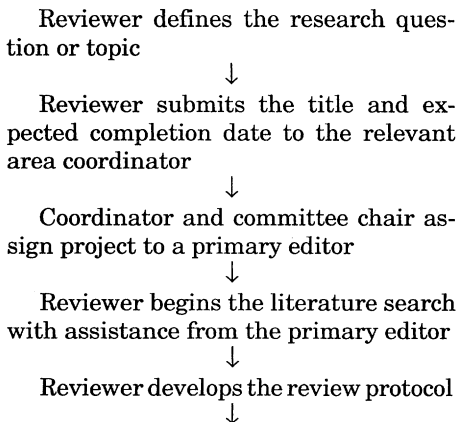
The challenges of evidence-based policy are many. We think the Campbell Collaboration will help to meet some of them. For example, with rigorous quality control and protocols, the Campbell Collaboration will attempt to produce the same level of transparency with unbiased reviews for which the Cochrane Collaboration is lauded. Through electronic publication of reviews in a single source (that is, *Campbell Library*), it will attempt to extend beyond communicating with researchers and facilitate dissemination and utilization by decision makers and ordinary citizens. Maintaining reviews, taking into account evidence worldwide, and preparing reviews using the best science available should enhance the use of Campbell reviews.

Systematic reviews are certainly an important tool for evidence-based policy, but like Campbell before us, we do not wish to zealously oversell scientific evidence. Systematic reviews will not resolve all enduring political and academic conflicts, nor

will they often—if at all—provide neat and tidy prescriptions to decision makers on what they ought to do. In their proper role in evidence-based policy, they will enlighten by explicitly revealing what is known from scientific evidence and what is not. They will also generate more questions to be resolved. One unanticipated benefit in the short life of the Campbell Collaboration is the sustained forum it provides for discussions about evaluation design and how to engage people from around the world.

Criminology is a noble profession because it aims to reduce the misery stemming from crime and injustice. To the extent that this collaboration can fulfill Don Campbell's vision of assisting people in making well-informed decisions, it will help criminologists stay true to criminology's original and noble intent.

APPENDIX A A FLOW CHART OF THE STEPS IN THE CAMPBELL REVIEW QUALITY CONTROL PROCESS



Editorial team (including a methodologist and external reviewer[s]) reviews the protocol



Editorial team submits the protocol to the Campbell Database



Reviewer completes the literature searching, quality assessment of primary studies, data extraction, and analysis



Reviewer writes and submits a draft review to primary editor



Reviewer receives and incorporates feedback from the editorial team



Reviewer submits a second draft



Primary editor obtains external reviews of second draft



Reviewer incorporates feedback from external reviewer



Reviewer submits final review



Review is published

SOURCE: C2 Steering Group (2001).

APPENDIX B STAGES OF A CAMPBELL SYSTEMATIC REVIEW

1. Formulate review questions
2. Define inclusion and exclusion criteria
 - Participants
 - Interventions and comparisons
 - Outcomes
 - Study designs and methodological quality
3. Locate studies; develop search strategy considering the following sources:
 - *C2-SPECTR*
 - Electronic databases and trials registers not covered by *C2-SPECTR*
 - Checking of reference lists

- Hand searching of key journals
 - Personal communication with experts in the field
4. Select studies
 - Have eligibility checked by more than one observer
 - Develop strategy to resolve disagreements
 - Keep log of excluded studies, with reasons for exclusions
 5. Assess study quality
 - Consider assessment by more than one observer
 - Use simple checklists rather than quality scales
 - Assess handling of attrition
 - Consider blinding assessors to authors, institutions, and journals
 - Assess randomization and power
 6. Extract data
 - Design and pilot data extraction form
 - Consider data extraction by more than one extractor
 - Consider blinding of extractors to authors, institutions, and journals
 7. Analyze and present results
 - Tabulate results from individual studies
 - Examine plots
 - Explore possible sources of heterogeneity
 - Consider meta-analysis of all trials or subgroups of trials
 - Perform sensitivity analyses, examine funnel plots
 - Make list of excluded studies available to interested readers
 - Examine process/implementation of interventions
 8. Interpret results
 - Consider limitations, including publication and related biases
 - Consider strength of evidence
 - Consider applicability
 - Consider statistical power
 - Consider economic implications
 - Consider implications for future research

SOURCE: C2 Steering Group (2001).

References

- Amann, Ron. 2000. Foreword. In *What Works? Evidence-Based Policy and Practice in Public Services*, ed. Huw T. O. Davies, Sandra Nutley, and Peter C. Smith. Bristol, UK: Policy Press.
- Bailey, William C. 1966. Correctional Outcome: An Evaluation of 100 Reports. *Journal of Criminal Law, Criminology and Police Science* 57:153-60.
- Boruch, Robert F. 1997. *Randomized Experiments for Planning and Evaluation: A Practical Guide*. Thousand Oaks, CA: Sage.
- Boruch, Robert F., Anthony Petrosino, and Iain Chalmers. 1999. The Campbell Collaboration: A Proposal for Multinational, Continuous and Systematic Reviews of Evidence. In *Proceedings of the International Meeting on Systematic Reviews of the Effects of Social and Educational Interventions July 15-16*, ed. Philip Davies, Anthony Petrosino, and Iain Chalmers. London: University College-London, School of Public Policy.
- Campbell, Donald T. 1969. Reforms as Experiments. *American Psychologist* 24:409-29.
- Chalmers, Iain, Larry V. Hedges, and Harris Cooper. In press. A Brief History of Research Synthesis. *Evaluation & the Health Professions*.
- Cochrane, Archie L. 1972. *Effectiveness and Efficiency. Random Reflections on Health Services*. London: Nuffield Provincial Hospitals Trust.
- . 1979. 1931-1971: A Critical Review, with Particular Reference to the Medical Profession. In *Medicines for the Year 2000*. London: Office of Health Economics.
- Cohn, Jason. 2001. Drug Education: The Triumph of Bad Science. *Rolling Stone* 869:41-42, 96.
- Cook, Thomas D., Harris Cooper, David S. Cordray, Heidi Hartmann, Larry V. Hedges, Richard J. Light, Thomas A.

- Louis, and Frederick Mosteller, eds. 1992. *Meta-Analysis for Explanation*. New York: Russell Sage.
- Cooper, Harris C. and Larry V. Hedges, eds. 1994. *The Handbook of Research Synthesis*. New York: Russell Sage.
- C2 Steering Group. 2001. *C2: Concept, Status, Plans. Overarching Grant Proposal*. Philadelphia: Campbell Collaboration Secretariat.
- Cullen, Francis T. and Paul Gendreau. 2000. Assessing Correctional Rehabilitation: Policy, Practice, and Prospects. In *Policies, Processes, and Decisions of the Criminal Justice System, Criminal Justice 2000*, vol. 3, ed. Julie Horney. Washington, DC: National Institute of Justice.
- Davies, Huw T. O., Sandra M. Nutley, and Peter C. Smith, eds. 2000. *What Works? Evidence-Based Policy in Public Services*. Bristol, UK: Policy Press.
- Davies, Phillip. 1999. What Is Evidence-Based Education? *British Journal of Educational Studies* 47:108-21.
- Davies, Philip, Anthony Petrosino, and Iain Chalmers, eds. 1999. *Proceedings of the International Meeting on Systematic Reviews of the Effects of Social and Educational Interventions, July 15-16*. London: University College-London, School of Public Policy.
- Egger, Matthias and George Davey-Smith. 1998. Bias in Location and Selection of Studies. *British Medical Journal* 316:61-66.
- Farrington, David P. and Anthony Petrosino. 2001. The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science* 578:35-49.
- Fischer, Joel. 1978. Does Anything Work? *Journal of Social Service Research* 1:215-43.
- Fitz-Gibbon, Carol. 1999. Education: The High Potential Not Yet Realized. *Public Money & Management* 19:33-40.
- Gawande, Atul. 2001. Under Suspicion. The Fugitive Science of Criminal Justice. *New Yorker*, 8 Jan., 50-53.
- Glass, Gene V., Barry McGaw, and Mary L. Smith. 1981. *Meta-Analysis in Social Research*. London: Sage.
- Greenberg, David and Mark Shroder. 1997. *The Digest of Social Experiments*. 2d ed. Washington, DC: Urban Institute Press.
- Harlen, Wynne. 1997. Educational Research and Educational Reform. In *The Role of Research in Mature Educational Systems*, ed. Seamus Hegarty. Slough, UK: National Foundation for Educational Research.
- Hedges, Larry V. and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. New York: Academic Press.
- Hunt, Morton. 1997. *The Story of Meta-Analysis*. New York: Russell Sage.
- Jadad, Alejandro R., Deborah J. Cook, Alison Jones, Terry P. Klassen, Peter Tugwell, Michael Moher, and David Moher. 1998. Methodology and Reports of Systematic Reviews and Meta-Analyses: A Comparison of Cochrane Reviews with Articles Published in Paper-Based Journals. *Journal of the American Medical Association* 280:278-80.
- Kirby, Bernard C. 1954. Measuring Effects of Criminals and Delinquents. *Sociology and Social Research* 38:368-74.
- Lipsey, Mark W. 1990. *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.
- . 1992. Juvenile Delinquency Treatment: A Meta-Analytic Inquiry into the Variability of Effects. In *Meta-Analysis for Explanation*, ed. Thomas D. Cook, Harris Cooper, David S. Cordray, Heidi Hartmann, Larry V. Hedges, Richard J. Light, Thomas A. Louis, and Frederick Mosteller. New York: Russell Sage.
- Lipsey, Mark W. and David B. Wilson. 1993. The Efficacy of Psychological,

- Educational and Behavioral Treatment: Confirmation from Meta-Analysis. *American Psychologist* 48:1181-209.
- Lipton, Douglas S. 1992. How to Maximize Utilization of Evaluation Research by Policymakers. *Annals of the American Academy of Political and Social Science* 521:175-88.
- Lipton, Douglas S., Robert Martinson, and Judith Wilks. 1975. *The Effectiveness of Correctional Treatment*. New York: Praeger.
- Logan, Charles H. 1972. Evaluation Research in Crime and Delinquency: A Reappraisal. *Journal of Criminal Law, Criminology and Police Science* 63:378-87.
- MacDonald, Geraldine. 1999. Evidence-Based Social Care: Wheels Off the Runway? *Public Money & Management* 19:25-32.
- MacKenzie, Doris Layton. 2000. Evidence-Based Corrections: Identifying What Works. *Crime & Delinquency* 46:457-71.
- Martinson, Robert. 1974. What Works? Questions and Answers About Prison Reform. *Public Interest* 10:22-54.
- National Board of Health and Welfare. 2001. *A Programme of National Support for the Advancement of Knowledge in the Social Services*. Stockholm: Author.
- Nutley, Sandra and Huw T. O. Davies. 1999. The Fall and Rise of Evidence in Criminal Justice. *Public Money & Management* 19:47-54.
- Nutley, Sandra, Huw T. O. Davies, and Nick Tilley. 2000. Letter to the Editor. *Public Money & Management* 20:3-6.
- Nuttall, Christopher, Peter Goldblatt, and Chris Lewis, eds. 1998. *Reducing Offending: An Assessment of Research Evidence on Ways of Dealing with Offending Behaviour*. London: Home Office.
- Palmer, Ted. 1975. Martinson Revisited. *Journal of Research in Crime and Delinquency* 12:133-52.
- . 1994. *A Profile of Correctional Effectiveness and New Directions for Research*. Albany: State University of New York Press.
- Petrosino, Anthony. 2000. Crime, Drugs and Alcohol. In *Evidence from Systematic Reviews of Research Relevant to Implementing the "Wider Public Health" Agenda*, Contributors to the Cochrane Collaboration and the Campbell Collaboration. York, UK: NHS Centre for Reviews and Dissemination.
- . Forthcoming. *What Works to Reduce Offending? A Systematic Review of 300 Randomized Field Trials in Crime Reduction*. New York: Oxford University Press.
- Petrosino, Anthony, Robert F. Boruch, Cath Rounding, Steve McDonald, and Iain Chalmers. 2000. The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to Facilitate the Preparation and Maintenance of Systematic Reviews of Social and Educational Interventions. *Evaluation Research in Education* 14:293-307.
- Quinsey, Vernon L. 1983. Prediction of Recidivism and the Evaluation of Treatment Programs for Sexual Offenders. In *Sexual Aggression and the Law*, ed. Simon N. Verdun-Jones and A. A. Keltner. Burnaby, Canada: Simon Fraser University Criminology Research Centre.
- Sechrest, Lee, Susan O. White, and Elizabeth D. Brown, eds. 1979. *The Rehabilitation of Criminal Offenders: Problems and Prospects*. Washington, DC: National Academy of Sciences.
- Sheldon, Brian and Roy Chilvers. 2000. *Evidence-Based Social Care: A Study of Prospects and Problems*. Dorset, UK: Russell House.

- Sherman, Lawrence W. 1999. Evidence-Based Policing. In *Ideas in American Policing*. Washington, DC: Police Foundation.
- Sherman, Lawrence W., Denise C. Gottfredson, Doris Layton MacKenzie, John E. Eck, Peter Reuter, and Shawn D. Bushway. 1997. *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Smith, Adrian. 1996. Mad Cows and Ecstasy: Chance and Choice in an Evidence-Based Society. *Journal of the Royal Statistical Society* 159:367-83.
- Tyden, Thomas. 1996. The Contribution of Longitudinal Studies for Understanding Science Communication and Research Utilization. *Science Communication* 18:29.
- Weisburd, David, Cynthia M. Lum, and Anthony Petrosino. 2001. Does Research Design Affect Study Outcomes? Findings from the Maryland Report Criminal Justice Sample. *Annals of the American Academy of Political and Social Science* 578:50-70.
- Weiss, Carol Hirschon. 1978. Improving the Linkage Between Social Research and Public Policy. In *Knowledge and Policy: The Uncertain Connection*, ed. L. E. Lynn. Washington, DC: National Academy of Sciences.
- Weiss, Carol Hirschon and Anthony Petrosino. 1999. Improving the Use of Research: The Case of D.A.R.E. Unpublished grant proposal to the Robert Wood Johnson Foundation, Substance Abuse Policy Research Program. Cambridge, MA: Harvard Graduate School of Education.
- Weiss, Carol Hirschon and Eleanor Singer. 1988. *The Reporting of Social Science in the National Media*. New York: Russell Sage.
- Wiles, Paul. 2001. Criminology in the 21st Century: Public Good or Private Interest? Paper presented at the meeting of the Australian and New Zealand Society of Criminology, University of Melbourne, Australia, Feb.
- Witmer, Helen L. and Edith Tufts. 1954. *The Effectiveness of Delinquency Prevention Programs*. Washington, DC: U.S. Department of Health, Education and Welfare, Social Security Administration, Children's Bureau.
- Wolf, Frederic M. 2000. Lessons to be Learned from Evidence-Based Medicine: Practice and Promise of Evidence-Based Medicine and Evidence-Based Education. *Medical Teacher* 22:251-59.