Formalizing Syntheses of Medical Knowledge: The Rise of Meta-Analysis and Systematic Reviews

Ingemar Bohlin

Perspectives on Science, Volume 20, Number 3, Fall 2012, pp. 273-309 (Article)

Published by The MIT Press

Formalizing Syntheses of Medical Knowledge: The Rise of Meta-Analysis and Systematic Reviews

Ingemar Bohlin
Department of Sociology and Work Science, University of Gothenburg, Sweden

Originally developed by social scientists, meta-analysis is now one of the central methods of evidence-based medicine. This paper offers an account of the emergence of meta-analysis in social science in the 1970s, its adoption in medicine in the 80s, and the birth of the closely related format of systematic reviews in the early 1990s. The paper investigates the extent to which medical meta-analysis relied on previous work by social scientists, as well as the manner in which systematic reviews grew out of these developments. Throughout the exposition, attention is paid to the formalization of the procedures involved in synthesizing research.

In recent years, the provision of health care has undergone profound changes in many countries. Since the early 1990s, the phrase *evidence-based medicine* has been a frequently used watchword for such reform. The essence of this concept is that reliable knowledge about the effects of drugs and other medical interventions be produced, systematized, and translated into clinical recommendations. The enterprise involves a double movement, beginning and ending at the level of individuals: data on patient samples are processed into general facts, which are again translated into formats designed to help practitioners make judicious decisions concerning individual patients. In the movement from general knowledge to individual cases, clinical practice guidelines constitute a key tool, while two methods are central to the preceding movement, the construction of an evidence base. One is the randomized clinical trial, frequently referred to as the "gold standard" of evidence-based medicine. The other is meta-analysis, a technique for combining the results of separate studies address-

The research on which this article is based was funded by the Swedish Research Council. I wish to thank Iain Chalmers and Trisha Greenhalgh, both of whom have offered detailed and very constructive comments on a draft version of the paper.

Perspectives on Science 2012, vol. 20, no. 3

ing similar topics. Both methods have been instrumental in establishing the concept of evidence-based medicine, as well as in extending it beyond the health care sector. There is much to suggest that, in the process in which evidence-based medicine has evolved into what is often referred to as the "evidence-based practice and policy movement," or simply the "evidence movement," meta-analysis, along with the closely related format of systematic reviews, has been at least as important as randomized experiments.

The literature on meta-analysis includes two comprehensive accounts of its history. One is a semi-popular book by Morton Hunt (1997), a freelance science writer; the other is an article by Iain Chalmers, Harris Cooper, and Larry Hedges (2002), three long-time protagonists of this methodology. In addition, a number of brief historical accounts have appeared as integral parts of discussions on the proper application of the method, both in medicine and in the social sciences, where it first came into wide use (such accounts include Ellenberg 1988; Jenicek 1989; Chalmers, Hetherington, Elbourne et al. 1989; Olkin 1990; Dickersin and Berlin 1992; Cooper and Hedges 1994; and Egger, Smith, and O'Rourke 2001). These construals of the emergence of meta-analysis and systematic reviews all instantiate what Andy Pickering (1984) has dubbed the "scientist's account." They are structured, that is, by an evaluative stance on the application of the techniques, each idea discussed being assessed for its contribution to the currently accepted form of this methodology. This is in contrast to the *empirical* literature on practices associated with evidence-based medicine and the evidence movement more broadly, which, given their significance, is still quite limited.[1] In a monograph on the emergence of evidence-based medicine a few years ago, sociologist Jeanne Daly (2005) devoted a chapter to meta-analysis and systematic reviews. To the best of my knowledge, Daly's has remained the only account of the history of these methods put forward by an author clearly outside the circle of actors.

This article describes the emergence of meta-analysis and its development into a methodology which has proved relatively easy to transfer to new domains of practice and policy. My reconstruction shares some common ground with several of the accounts mentioned above, especially the more comprehensive ones. In terms both of theoretical perspective and empirical focus, however, this interpretation departs from previous accounts. The perspective adopted is symmetrical, i.e., I do not commit my-

1. The shortage of empirical research on evidence-based practices was pointed out in Mykhalovskiy and Weir 2004. Though many valuable studies have since appeared, key methods and procedures remain poorly understood.

self to actors' views of the scientific merit of the techniques discussed. Empirically, three themes will be explored, none of which has received the attention I believe it deserves. The first concerns actors' efforts at formalizing the procedures involved in synthesizing research. The second concerns the extent to which the form of meta-analysis which was established in clinical research during the 1980s relied on techniques previously developed by social scientists. The third concerns the circumstances in which the concept of systematic review was invented, and the nature of its relation to meta-analysis. While the first theme pertains to essential arguments and practical means deployed to establish meta-analysis and systematic reviews across contexts, the second and third themes concern historical connections between specific methodological formats. The overall aim of the paper is to shed new light on the emergence of these formats.

The article is organized as follows. The first section below describes the introduction of meta-analytic techniques in American social science in the 1970s. The second section is an account of how meta-analysis became widely adopted in clinical medicine in the 1980s, and the controversy that its popularity engendered. The third section clarifies the relation between the developments addressed in the two previous sections. Then follows a discussion of the role of guidelines in attempts to enhance the status of meta-analysis in clinical research in the latter half of the 1980s. The fifth section is devoted to the manner in which systematic reviews grew out of meta-analysis. In the concluding section the main points of this historical reconstruction are summarized.

The Emergence of Meta-Analysis in the Social Sciences

By the second half of the 1960s, the volume of research published in the social sciences had reached such a level that findings relating to many issues were difficult to summarize. For instance, a plethora of findings had been reported concerning the way in which beliefs, personal orientations, and other characteristics of American students were affected by their experience of attending college. The inconclusiveness of those findings prompted Kenneth Feldman and Theodore Newcomb, the latter having long contributed to this field, to take stock of what was known. In 1969 they published *The Impact of College on Students,* a two-volume attempt at "integration" of the results of hundreds of reports having appeared from the mid 1920s up to 1967. While their analysis and conclusions were offered in the first volume, the second comprised tables setting out data extracted from the studies covered. This is Feldman and Newcomb's rationale for publishing that material (1969b, p. iii):

> These summary tables are, in a very real sense, the "raw" data upon which many of our generalizations in Volume One are built. Since we lay no claims to infallibility, we hope that our readers will use these tables to decide the validity of the inferences we have drawn from them.[2]

Implicit in the notion that Feldman and Newcomb's theoretical conclusions were based on empirical data, the accuracy of which could be assessed by other investigators, was an analogy between their work and the primary publications reviewed. Such an analogy had previously been drawn by Conyers Herring, a physicist calling for greater efforts at "creative synthesis" of research in his own field (1968, pp. 27, 30). Stressing the "genuine intellectual challenge" of synthesizing a multitude of independent findings, Herring (1968, p. 30) remarked that

> After all, science consists in the creation of simplicity out of the complexity of nature, and it is scarcely less of a feat to create new simplicity out of the complexity of the literature.

In a 1971 article addressing methodological and theoretical aspects of the kind of project he had been pursuing with Theodore Newcomb, Kenneth Feldman made this analogy even more explicit. "Systematically reviewing and integrating" the literature of some specialty, he suggested, "may be considered a type of research in its own right—one using a characteristic set of research techniques and methods" (1971, p. 86). This notion, that what is now often referred to as *research synthesis* is an activity paralleling primary research and involving similar methodological decisions, has subsequently been established as axiomatic. The task of synthesizing current research on any given topic, champions of meta-analysis and systematic reviews have emphasized again and again, must not be taken any less seriously than conducting primary studies. Research synthesis, according to its proponents, must be recognized as a scientific specialty, and hence needs a well-defined methodology.

Though meta-analysis, research synthesis and cognate terms had not been invented at the time, Feldman's 1971 article clearly instantiates this position. Reviews are instrumental in cumulating knowledge, he argued, and one of the reasons why cumulative progress was not a conspicuous feature of the development of the behavioral sciences might well be that commonly adopted reviewing practices were inadequate. Too often, reviews of some specific body of literature did "little more than string together short summaries of selected articles" (1971, p. 89). In reviews of

2. A similar passage appears in Feldman and Newcomb 1969a, p. 2.

that kind, Feldman maintained, no "integration" of the literature is effected. Especially where the number of studies covered is considerable, tables like those appearing in the second volume of *The Impact of College on Students* ought to serve a key function, since

> The amount of information available in the typically tens or hundreds of research reports in an area being integrated would be overwhelming unless the integrator develops a systematic schema for indexing, coding, and retrieving this information. (Feldman 1971, p. 88)

Within a few years of the publication of this paper, substantial efforts were devoted to developing "systematic schemata" of the kind it recommended, along with other tools enabling those engaged in what Feldman termed integration of research findings to extract and process information from individual studies in a highly structured manner.

Like Feldman and Newcomb, the pioneers of the new techniques were American social scientists, predominantly from educational research and psychology with a background in statistics. The major concept unifying the techniques developed by this community was that of meta-analysis, introduced in 1976 by Gene Glass. With his colleague and then wife Mary Lee Smith, Glass conducted two large-scale reviews in the 1970s, one on the effects of psychotherapy (Smith and Glass 1977), the other on the relationship between class size and pupils' achievements in elementary and secondary school (Glass and Smith 1979).[3] While preparing the first of these reviews for publication, Glass (1976) went on to address conceptual issues involved in such work, coining the term "meta-analysis" to distinguish it from secondary analysis. The latter was an established term for analyses of raw data performed subsequently, by researchers not involved in the studies in which the data were produced, either to test the validity of the conclusions drawn or to ask new questions. Meta-analysis, by contrast, did not rely on access to primary data.[4] Instead, the raw data used for the kind of analysis proposed by Glass were the findings reported in large numbers of studies. In cases like the one concerning the effects of psychotherapy, where no consensus had emerged even though a multitude of studies had been reported, extracting "the knowledge that lies untapped in completed research studies," as Glass (1976, p. 4) put it, seemed far more important than carrying out yet another investigation.

3. A few years later, a monograph on each subject appeared; Smith et al 1980; Glass et al 1982.

4. In meta-analysis of clinical research, using primary data has subsequently become common practice; see Clarke and Stewart 2001.

For the purposes of this paper, the technical details of the statistical procedures recommended by Glass are far less important than his strictures on the mode of reviewing literature that was dominant in the social sciences at the time. This genre, in which no statistics of any kind is employed, remains the standard form of reviewing in many humanities and social scientific disciplines. Dubbing such reviews "traditional," "discursive," and "narrative," Glass (1976, pp. 4, 6) denounced their "dizzying lists" of verbal summaries of research reports, and contrasted their "literary exposition" with "quantitative rigor." In a paper published the following year, this criticism was reiterated and expanded. The "narrative, rhetorical integration" effected through "chronologically arranged verbal descriptions," Glass maintained (1977, pp. 351–52), may well have been appropriate in psychology and educational research during the 1940s and 50s, when the number of studies covered in a review rarely exceeded a dozen or two. Given the volume of research having appeared by the late 1960s, however, it was imperative that a new form of reviewing be devised.

The US communities of educational research and psychology were quick to endorse the concept of meta-analysis. In the late 1970s and early 80s, a number of publications reiterated, varied, and enlarged upon the position taken by Glass and his co-authors, contrasting quantitative methods for integrating research with the traditional, narrative or "literary" form of reviewing. In these publications, two crucial arguments were offered, both of which had been stated by Kenneth Feldman in 1971.

The first crucial point was that conducting research reported in primary publications and reviewing such research should be regarded as parallel activities. Due to the increasing volume of social scientific research output, the need for reviews summarizing the primary literature had grown dramatically. The importance of reviews in defining what was known about any number of subjects tended to be far greater than that of individual studies. Given their role in certifying knowledge, there was no reason why the standards by which research reviews were produced should be less rigorous than those applied in the studies integrated. In the words of Harris Cooper and Robert Rosenthal (1980, p. 442), two leading early proponents of meta-analysis:

> Because literature reviews have such great information-gatekeeping potential, it is crucial that we apply standard, replicable, and rigorous criteria to them. These criteria should be at least as demanding as those that we require for primary data handling.[5]

5. Similar arguments are offered in Cooper 1979 and 1982.

The reference to replicability in this passage is far from gratuitous. Just as the objectivity of findings produced by scientists is commonly taken to be underwritten by the replicability of their observations and experiments, the reliability of research reviews may be assumed to be measurable by the degree to which analysts independently reviewing the same set of studies identify the same patterns. Only if reviewing is considered a form of research analogous with that reported in the primary literature does such a requirement make sense.

The second key point concerns the practical means by which replicability and objectivity can be achieved. A recurrent criticism of traditional or narrative reviews in publications championing meta-analysis is that their interpretations and conclusions, to a large extent, are contingent on the specific perspective and the individual judgments of the reviewer. What Feldman called "systematic schemata" are an example of resources by which this subjective element of reviewing can be curtailed. That traditional or narrative reviews were frequently referred to as "informal" (in Glass et al 1981, pp. 14 and 18, for instance) was not a coincidence. The techniques developed by Glass and other proponents of meta-analysis from the mid 1970s onward constitute *formal tools.* The function of these tools is to discipline the interpretations of those engaged in combining results from individual studies, enabling them to draw inferences that can be validated intersubjectively.

In 1980, an article appeared that did much to convince social scientists that reviewing practices prevalent in their field were unsound. Gregg Jackson, author of this article, had randomly sampled 36 recent "integrative reviews" from leading journals in psychology, educational research, and sociology. Exploring a number of aspects of the interpretive and analytic procedures adopted by the reviewers, Jackson (1980) drew attention to many alleged weaknesses. The chief target of his criticism was the lack of explicitly stated methodological procedures. Only four of the reviews, for instance, reported indexes or bibliographies that had been used to identify primary studies, and only seven indicated whether the full set of studies located, or just a subset, were analyzed. In short, Jackson's empirical study of reviewing practices lent strong support to the view that explicit, shared methodological principles for the integration of research findings were needed. Within a few years, several handbooks offering such criteria appeared (Glass et al 1981; Hunter et al 1982; Light and Pillemer 1984; Cooper 1984; Rosenthal 1984; Hedges and Olkin 1985).

The Introduction of Meta-Analysis in Medical Research

The mid 1980s saw the publication of several programmatic articles in which the technique of meta-analysis was introduced to audiences within

the health care sector.[6] These papers offered overviews of the new methodology developed by social scientists, focusing on what might be gained by adopting it in medicine and allied disciplines. Its use in the health care domain was recommended, on the apparent assumption that such analyses were not already being conducted. At the time these programmatic articles appeared, however, meta-analysis was already being launched as a new approach in medical research.[7]

A major role in demonstrating what meta-analysis could offer the biomedical sciences was played by a group of clinicians and statisticians whose leading figures, Salim Yusuf, Rory Collins and Richard Peto, were all based in Oxford in the early 1980s. By the time the Oxford group, as they were often called, started conducting meta-analyses, randomization had been established as a crucial element of medical research, and the forms of bias inherent in observational studies were recognized. Hence a central principle of the meta-analyses of the Oxford group was that only data from randomized controlled trials (RCTs) should be included.[8]

Where a medical condition affects a large number of people, effects of treatments that are not dramatic, and thus may fail to be detected in small randomized trials, may yet be large enough to save a significant number of lives. The prospect of discovering such moderate beneficial effects of medical interventions constitutes a major reason for combining data from multiple controlled trials. In 1982 the Oxford group published its first meta-analysis, assessing the effect of intravenous infusion of streptokinase in patients with acute myocardial infarction (Stampfer et al 1982). In this review, which is frequently cited as a classic example of meta-analysis, eight studies were included. The results of these trials, which involved from 167 up to 730 patients, were inconclusive: while five of the studies

6. Such papers include Ottenbacher and Petersen 1984a, Einarson et al 1985, and Curlette and Cannella 1985. The communities primarily addressed by these authors were, respectively, those of clinical pediatrics, pharmacy, and nursing research.

7. Designating this category of papers programmatic is less than fully justified, because some of them did refer to meta-analyses that had been carried out in the biomedical area. Einarson et al. 1985 cite one such case: Coffman 1981, a review of studies in gerontology. Similarly, the literature covered by Curlette and Cannella 1985 includes a few meta-analyses of studies addressing patient education programs and the value of psychological interventions during hospital stays.

It is also worth mentioning that some of these authors themselves applied the method they were promoting. Ottenbacher and Petersen 1984a is a companion paper of Ottenbacher and Petersen 1984b, a meta-analysis of studies of vestibular stimulation, and an illustrative example of meta-analysis of pharmaceutical studies is offered in Einarson et al. 1985.

8. Though the RCT is widely regarded as a superior study design in clinical research, the epistemic virtues of randomization are disputed, even in medicine. See, for instance, Worrall 2002.

indicated that the treatment reduced mortality and three found that it increased the risk of death, conventional levels of significance (a P value of 0.05 or less) were reached in only two of the trials. Combining the results, however, Yusuf and his coauthors found that they indicated that intravenous streptokinase therapy reduced mortality by 20%, and this finding was statistically significant.

Three years later, the same group published a new, extended meta-analysis (Yusuf et al 1985). This time, 20 randomized trials of intravenous streptokinase were included, some of which had been overlooked in the previous review. Results from these trials were combined with those of 4 trials of intravenous urokinase, another fibrinolytic agent. Again, before being subjected to meta-analysis, the data were inconclusive. In 16 studies, only 5 of which achieved statistical significance, the treatment was found to reduce mortality, while 8 trials yielded point estimates suggesting a negative (but statistically nonsignificant) effect. Aggregating the data, however, Yusuf and his coworkers found that the intravenous fibrinolytic agents had the highly significant ($P<0.001$) effect of conferring a 22% reduction in risk of death.

These and other meta-analyses set forth by the Oxford group were greeted in many quarters as valuable contributions to medical knowledge, but the methodology employed was disputed by a number of commentators, and even those who were in favor of the new approach expressed serious reservations. The circumstances surrounding certain decisions concerning treatments for breast cancer, in particular, fuelled controversy. Breast cancer is a common affliction, and many trials of various therapies have been performed. In September 1985, a consensus conference addressing the effectiveness of available treatments was convened by the National Institutes of Health in the US. The conclusions reached at this conference relied, to a considerable extent, on an unpublished meta-analysis conducted by Richard Peto, Rory Collins, and other researchers associated with the Oxford group.[9] The basis for the recommendations issued was widely disputed, and partly in response to this, a workshop was organized in Bethesda, Maryland, in May 1986. The aim of this meeting was to allow proponents and critics of meta-analysis to exchange views on the merits and limitations of the method.[10]

As Susan Ellenberg, one of the conveners of the meeting, subsequently noted (1988, p. 479), the controversy over the legitimacy of meta-analysis mainly concerned the acceptable degree of heterogeneity of studies in-

9. National Institutes of Health Consensus Development Conference Statement 1985. Several years later, this meta-analysis remained unpublished; see Ellenberg 1988, p. 473.

10. The proceedings were published as Yusuf et al 1987.

cluded in a review. A number of statistical problems were considered during the workshop, but the thorniest questions were of a non-statistical nature. "While methods exist to combine studies," as one of the speakers at the workshop remarked, "adequate procedures for deciding what to combine are not yet . . . developed" (DeMets 1987, p. 347). There are two aspects to this issue: one concerns inter-trial variation with respect to treatments and patients, the other pertains to quality.

Concerns about the quality of the studies from which data were culled were raised at an early stage in the history of meta-analysis. Soon after Glass launched his concept, Hans Eysenck (1978) disputed, in a scathing attack, the wisdom of including material drawn from poorly designed investigations when attempting to synthesize existing knowledge on any given subject. As meta-analysis became a widely used method in clinical research in the 1980s, considerable attention was devoted to that issue. Most of those who engaged in the discussion shared Eysenck's apprehension, but the difficulty was to devise reliable criteria for determining whether or not a clinical trial was well enough designed. Thomas Chalmers, who, besides the Oxford group, was the main protagonist of meta-analysis of biomedical research in this period, took the view that such criteria were available. Key aspects of the methodology adopted in study trials, such as the randomization procedures, the blinding of patients and observers, and the statistical analyses performed, Chalmers and his coauthors argued, should always be assessed when conducting a meta-analysis (Sacks et al 1987, p. 453). Though there was agreement that such components were essential, however, many researchers warned that the subjective element of assessments of quality could hardly be eliminated (Naylor 1988, p. 893; Jenicek 1989, p. 38). This view formed the rationale of the practice adopted by the Oxford group, who, rather than excluding or weighting studies on the basis of assessments of methodological rigor, included all randomized trials addressing a given subject. In a discussion section at the 1986 workshop in Bethesda (Yusuf et al 1987, p. 230), Richard Peto defended that practice:

> You need an objective rule as to what goes in and what doesn't go in. The rule that seems to be the most objective is to include studies only if they were properly randomized. I think if you try to adopt any other rule, you can get into impossible difficulties.

The second aspect of the issue of what studies should be included in a meta-analysis has often been called the "apples and oranges" problem. The pertinence of this metaphor is that when calculating some average property of objects that may be regarded as falling into separate categories, it may not be obvious what entity is being measured. As no two clinical tri-

als are ever congruous in every respect, researchers planning to conduct a meta-analysis have to make a decision as to the point beyond which the degree of diversity in the set of studies selected would prevent the exercise from yielding useful information. When planning their 1985 analysis of trials of fibrinolytic therapy in acute myocardial infarction, for instance, Yusuf and his coauthors apparently decided that calculating the aggregate effect of intravenous infusion of two different agents, streptokinase and urokinase, fell within such a limit. In their previous paper on streptokinase, on the other hand, inter-trial variation was considered a problem. Because two of the eight trials selected for that review were regarded as differing from the others with respect to study design, two versions of an aggregate effect of the therapy were calculated, one based on all the studies and one excluding the two deviant investigations Stampfer et al 1982, p. 1181).[11]

In a presentation given at the 1986 meeting, Peto (1987, pp. 233–34) recognized that the heterogeneity of trials addressing similar questions presents a difficulty to any attempt at synthesizing results. However, drawing on a distinction between the *size* and *direction* of therapeutic effects, he suggested that heterogeneity provides an argument in favor of meta-analysis rather than against it. While the magnitude of a therapeutic effect often varies across subgroups of patients, what benefits one subset, so this argument goes, is rarely harmful in another. Hence the great advantage of the methodology of meta-analysis is that it allows the direction of effects to be detected. Determining with quantitative accuracy the effect in some specific population of patients is not the aim of this kind of analysis. The point, instead, is to establish whether or not a given treatment has a positive effect; if it does, physicians may then prescribe it to patients belonging to a variety of subgroups.[12]

The major claim made by Thomas Chalmers, the Oxford group, and others integrating findings from clinical trials by meta-analytic procedures in the 1980s, was that these techniques enabled them to detect average effects which did not emerge in individual studies. The rule of thumb suggested by Peto in the passage quoted above similarly stressed the significance of identifying average effects. While there have been exceptions over the years, in retrospect, this simple formula seems to have been quite useful. This is not to imply, however, that the ability to detect average effects is the only presumed advantage of meta-analysis; bringing a se-

11. Only because the smaller sample yielded an even greater positive effect, it seems, did the authors base their conclusions on all eight studies.

12. The same point had previously been made in Yusuf et al 1985, pp. 359–60. Besides reporting another very successful meta-analysis, that paper includes the fullest discussion of methodological issues offered by the Oxford group.

ries of similar studies together also affords an opportunity to explain diverging findings. Where both favorable and adverse effects of some treatments have been demonstrated, this may be due to differences across settings, categories of recipients, or with respect to the research designs employed. Such patterns, underlying conflicting outcomes, may remain invisible until individual studies are compared in a systematic manner. The capacity of meta-analysis to yield information of this kind was stressed by social scientists from an early stage (Light 1984; Light and Pillemer 1984, esp. pp. 86–103).

Returning to the 1986 meeting, not all participants were persuaded that Peto or other practitioners of meta-analysis had found satisfactory solutions to the twin problems posed by the heterogeneity and methodological deficiencies of individual studies. This is how Robert Wittes (1987, p. 274), one of the most outspoken critics, described the implications for the technique of meta-analysis, as practiced by the Oxford group and others, of methodological imperfections in the trials included:

> The concept of data pooling is deeply troubling to many investigators. Trained as they have been on the primary importance of data quality and of meticulous attention to detail in the conduct of a trial, they have an understandable aversion to the prospect of combining carefully acquired data from trials with excellent quality control with data of less certain provenance.

Wittes also addressed the problem of heterogeneity, emphasizing the diversity commonly exhibited by studies addressing similar topics in regard of therapies as well as of the patients selected. The drugs being studied are often used in various combinations, and even where there is uniformity in that respect, they are often administered by different routes and in greatly varying doses. As for patients, in many cases the methods of diagnosing the disease targeted differ significantly, as do other components of the patient entry criteria. Wittes's conclusion was that where a series of well designed RCTs existed in which both the intervention studied and the patient selection criteria employed were the same, a meta-analysis was to be welcomed, owing to the increased sample size achieved. Where the trials from which data are combined are dissimilar, however, and where they have not been consistently carried out in a rigorous manner, a meta-analysis would add nothing of value. In a word, there is

> no reason to assume that the pooling of RCTs can bail us out of the uncertainty that will accompany any situation where excellent clinical trials, the cornerstone of therapy evaluation, do not exist. (Wittes 1987, p. 276)

Another speaker, Larry Norton, rejected the new approach in even stronger terms. Where the number of patients included in available clinical trials is too small to allow reliable conclusions to be drawn concerning the effect of some treatment regimen, the options are either to launch a large trial or to combine through meta-analysis the results of a series of small ones. Norton was strongly in favor of the first option, branding the second alternative "dangerous" (1987, p. 335). The "technical difficulties, methodological controversies [and] mathematical quandaries" aired at the meeting suggested to Norton that the field was "in an early state of development even on fundamental matters." Hence he warned (1987, p. 333) that

> If we are not sure of our analytical techniques, if these techniques are still in evolution, we must at least be sure that the results of our analyses are recognized for what they are: interesting science, maybe even useful science, but not solid fact upon which to base life and death clinical decisions.

Despite the objections of skeptics like Wittes and Norton, the late 1980s and early 90s witnessed a surge of interest in the methodology they denounced. A few months after the Bethesda workshop was held, 86 meta-analyses had appeared in the English-language medical literature, according to one estimate (Sacks et al. 1987). A few years later, a search for meta-analyses in the Medline database produced 21 citations for 1986, and then a sharp rise year by year, 431 publications found for 1991 (Dickersin and Berlin 1992, p. 155). The interest in meta-analyses has continued to increase, and this study design is now cited more frequently than any other in clinical research (Patsopoulos, Analatos, and Ioannidis 2005). "Few methodological advances," it has been remarked,

> have been implemented so quickly and so extensively as this family of techniques and procedures designed to provide quantitative syntheses of entire areas of research endeavor. Further, few advances (outside of the development of the randomized clinical trial) have captured the imagination and attention of so many scientists. (Bausell et al 1995, p. 239)

The Provenance and Scope of Meta-Analysis

The claim that the technique of meta-analysis was implemented very quickly presupposes that at the time it was established, the method was a recent invention. That would have been the case had the methodology of combining results from separate studies that came into wide use in clinical research during the latter half of the 1980s, been an outgrowth of the

technique developed in educational research and psychology a few years earlier. The fact that the method currently used to aggregate numerical data from independent trials of drugs and other healthcare interventions goes under the name coined by Glass in 1976 does not imply, however, that everybody agrees that the historical roots of the procedure can be traced to the social sciences.

The complexity of the historical issues involved may be illustrated by the label under which the Oxford group presented their calculations of aggregate results of RCTs. Though the analyses carried out by Yusuf, Peto, and their associates are now routinely referred to as meta-analyses and were often classified as such at the time they appeared, their authors consistently designated them "overviews." The autonomy vis-à-vis previous work in the social sciences suggested by that term was equally clear in the references cited. Excepting one single citation (Rosenthal 1978, a technical paper), the publications of the Oxford group carry no references to the social science literature on meta-analysis. Instead, the key source adduced in support of the methodology employed was a paper by Nathan Mantel and William Haenszel (1959), proposing a statistical procedure for calculating relative risks in retrospective epidemiological studies.

Almost all the speakers at the Bethesda workshop of 1986 disregarded previous efforts by social scientists in the same manner. The chief exception was a group of authors led by Thomas Chalmers. Not only did Chalmers and his coauthors identify the methodology debated at the meeting as meta-analysis, they also referred to the article in which Glass had introduced the concept, as well as to several of the textbooks on meta-analysis recently published by social scientists (Chalmers et al 1987). In a discussion section at the meeting, Richard Peto and Thomas Chalmers differed as to whether "overview" or "meta-analysis" was the more precise and understandable term (Yusuf et al 1987, pp. 229–30). No consensus was reached, and in many texts from this period the two denominations are used interchangeably. As instantiated by the passages from a presentation by Robert Wittes quoted above, a third term was "pooling" of data. The terminological disagreement was settled in 1989, for most practical purposes, when the National Library of Medicine added meta-analysis as a new medical subject heading.[13] Quantitative estimates of the aggregate effect of RCTs are still occasionally referred to as overviews, but with diminishing frequency.

As the divergent citation practices adopted by Thomas Chalmers and the Oxford group suggest, those who used the appellation "meta-analysis"

13. Four years later, moreover, meta-analysis was given the status of publication type. Dickersin and Berlin 1992, p. 154; Bausell et al 1995, p. 245.

were more prone to regard the technique as relying on insights gained by social scientists than those who preferred the terms "overview" or "pooling." Terminological preferences, however, did not reflect unequivocally perceptions of the provenance of the method. The main reason there is no such straightforward correspondence is that procedures differing significantly from the technique introduced by Glass are sometimes subsumed under the term he coined. So, for instance, Stephen Thacker (1988, p. 1686), in an article addressing the increasing use of quantitative techniques in reviews of clinical trials, proposed that meta-analysis be defined as "any systematic method that uses statistical analyses" for calculating an overall result by combining data from independent studies. The approaches apparently subsumed under Thacker's definition included "vote counting," one of the procedures that the method launched by Glass had been designed to supplant.[14] The authors of one of the commentaries on meta-analysis cited in the previous section sided with Glass in this matter. Estimating the number of meta-analyses published up to the early 1990s, Barker Bausell and his coauthors (Bausell et al 1995, p. 242) expressly excluded reviews in which findings on a given topic were summarized by means of vote-counting techniques.

Pooling of data was contested in a similar manner. As mentioned earlier, data pooling was sometimes used as an alternative name for meta-analysis. It is not surprising, therefore, that Thacker's catch-all definition of meta-analysis included pooling alongside vote-counting and the procedure recommended by Glass.[15] If it was less than crystal clear what meta-analysis referred to, however, that was true of pooling, too. The notion of data pooling was inclusive enough, it appears, that some of the practices categorized under this heading were considered far less acceptable than others. Hence, Keith O'Rourke and Allan Detsky, in a commentary (1989, p. 1022), contrasted "appropriate meta-analysis" with "simple statistical pooling of the data."

The term meta-analysis, then, has been used both in a narrow and in an exceedingly liberal sense. Any construal of the origin and diffusion of the method depends on the scope of the definition of "meta-analysis" em-

14. In two other, similar papers, too, vote counting was classified as a form of meta-analysis, though in one case with the proviso that it was the crudest type. Louis et al. 1985, p. 2; Oxman and Guyatt 1988, p. 701. At the 1986 workshop, moreover, Richard Light, who in 1971 had coauthored a paper in which vote counting was discarded, was prepared to accept a definition of meta-analysis at least as inclusive as the one offered by Thacker. In a discussion section, Light took the view that meta-analysis might refer to "any type of quantitative summary of data." Light and Smith 1971; Yusuf et al 1987, p. 229.

15. The form of analysis proposed in Light and Smith 1971 as an alternative to vote counting was included, too; Thacker 1988, p. 1868.

ployed. One does not have to look far, for instance, to find indications that a narrower definition would have produced a different picture of the early history of the methodology than that conveyed by some of the estimates, summarized in the preceding section, of the number of meta-analyses published. Thomas Chalmers and his coauthors identified 86 meta-analyses in the English-language medical literature up to October 1986 (Sacks et al 1987). In a subsequent paper, the Chalmers group reported a comparison between meta-analyses addressing the same treatments (Chalmers et al 1987). Selecting replicate analyses among the 86 publications previously located and adding an unpublished one that they had themselves conducted, they ended up with 46 papers synthesizing data on the efficacy of 20 different treatments. The collation of information from this sample of meta-analyses included categorizing the statistical methods employed. In the great majority of cases, it was found that one of three methods was used: the technique introduced by Mantel and Haenszel in 1959, the methodology developed by Glass and his colleagues, or "crude pooling." The number of analyses categorized as falling either under the latter rubric or under the heading of "unknown" statistical methods made up almost one third of the sample.[16] Were one to adopt the position of O'Rourke and Detsky, then, setting "simple statistical pooling" apart from exercises in meta-analysis proper, the body of literature representing the early history of meta-analysis in clinical research would have been considerably smaller than the one examined by Chalmers and his coauthors. If this is true for the 46 publications offering replicate analyses, then, very likely, it is true for the 86 meta-analyses previously identified for the period up to October 1986 as well.

One of the other estimates of the volume of published meta-analyses cited above was put forward by Kay Dickersin, Karen Higgins, and Curtis Meinert. Reproducing almost exactly the figure reported by the Chalmers group, these authors identified 83 meta-analyses in clinical research through 1987. In addition, the two estimates agree fairly well concerning the pre-1980 period. Chalmers and associates found 17 meta-analyses published prior to 1980, four of which had appeared before 1970, while Dickersin and her coauthors (1990, p. 57) identified 19 that were published prior to 1979. Both groups of authors, moreover, cite a 1955 paper as providing the first meta-analysis reported in the medical literature. An observation made by Dickersin, Higgins, and Meinert, however,

16. Chalmers et al 1987, pp. 736, 738 (table IV). To be precise, 19 out of 59 meta-analyses were categorized as relying either on "crude pooling" or on "unknown" methods. That 59 separate meta-analyses were being discussed though the number of publications selected was 46 may appear incongruous, but this was due to some papers addressing more than one treatment and a few using more than one method.

suggests that the figure just mentioned may exaggerate the number of medical meta-analyses conducted before the mid 1970s. In all but one of the ten analyses that had appeared by 1975, they remark, "crude" statistical methods were applied. The first meta-analysis using what these authors (1990, p. 57) referred to as a "formal" technique—in this case, the Mantel-Haenszel procedure—was reported in 1974. Again it is evident that, provided one does not count instances of data pooling by means less sophisticated than the ones introduced by Glass and by Mantel and Haenszel, the number of meta-analyses reported in the medical literature prior to the methodological advances in quantitative social science set in motion by Glass, may have been small indeed. The categorization undertaken by Dickersin, Higgins, and Meinert suggests that before 1976, one single paper (Stjernswärd 1974) had been published in which the results of clinical trials were combined in a manner for which the term meta-analysis would be appropriate.

Conversely, given a far wider definition of meta-analysis, a long prehistory of the methodology emerges. Prominent among those who have been cited as early pioneers of the field is Karl Pearson. In a 1904 paper, Pearson assessed the evidence on a new vaccine against typhoid by collecting data from several tests and calculating average effects of the inoculations. This review has been described as "the earliest example we have found of what we would now call a meta-analysis" (Shadish and Haddock 1994, p. 262; cf. Cooper and Hedges 1994, p. 5; Egger et al 2001, p. 8; and Chalmers et al 2002, p. 14). According to Harris Cooper (1998, p. 107), a leading advocate of meta-analysis in the social sciences from the late 1970s on, the method was hatched by three research teams who, more or less independently, "rediscovered and reinvented" the technique introduced by Pearson in 1904.[17] Other works cited as having proposed meta-analytic procedures long before current techniques were developed include monographs on statistics published by Ronald A. Fisher and Leonard H. C. Tippett in the early 1930s (Olkin 1990, pp. 4–5; Cooper and Hedges 1994, pp. 5–6). The first textbook of meta-analysis, it has even been suggested (Egger et al 2001, pp. 6–7), was a treatise on astronomical observation published by George Biddell Airy in 1861.

Owing to the lack of an agreed definition of meta-analysis, then, accounts of the origin of the field may appear confusingly discrepant. Such confusion may be reduced considerably, if not dispelled entirely, by means

17. Besides the team led by Glass, the reinventors of the method, on this account, were Robert Rosenthal and Donald B. Rubin, and John E. Hunter, Frank L. Schmidt, and Ronda Hunter. Three of the textbooks on meta-analysis in the social sciences that appeared in the early 1980s were authored by members of these groups: Glass et al 1981; Hunter et al 1982; and Rosenthal 1984.

of a distinction between the introduction of statistical procedures subsequently employed in meta-analysis, on the one hand, and the actual use of those tools in the conduct of meta-analysis on the other. Dickersin, Higgins, and Meinert (1990, p. 54), relying on that distinction, offer a list of key statistical innovations, beginning with the 1932 edition of a standard work by Ronald A. Fisher.[18] Clearly, statistical techniques that were to become essential components of meta-analysis were worked out decades before the methodology was established as a specialty in the social sciences and then gained the same status in medicine. Irrespective of how the work of Pearson, for instance, is perceived, there is little doubt that formal methods for integrating data from separate clinical trials were devised independently of developments occurring in psychology and educational research. The Mantel-Haenszel procedure, in particular, has been important in this regard. As mentioned above, a 1974 paper has been cited as reporting the first analysis in which that procedure was used to aggregate data across clinical studies. In the first meta-analysis carried out by Thomas Chalmers, appearing three years later (Chalmers et al 1977), the Mantel-Haenszel procedure was again applied. Only in 1982, however, on being informed that he was to receive an award from the Evaluation Research Society "for his meta-analyses," did Chalmers become aware of the obvious parallels between the form of analysis he had invented himself and methodological advances previously made in the social sciences.[19] As for the Oxford group, they appear to have been familiar with Chalmers's work from the outset. In an interview (Anonymous 1993, p. 3), Yusuf Salim has suggested that what he and Richard Peto set out to do was refine, in the light of criticism that had been leveled against it, the method introduced by Chalmers.[20]

To some extent, then, the introduction of meta-analysis in medicine seems to have been unrelated to developments in the social sciences. Yet there are strong indications that the emergence of this methodology in clinical research depended, to a considerable degree, on the work of social scientists. When the first handful of textbooks on meta-analysis were pub-

18. No contribution by Karl Pearson is included in this list. For a discussion stressing the limitations of Pearson's form of statistical analysis, as compared to meta-analysis as developed by Glass, see Einarson et al 1985, p. 1958.

19. This account is offered in a retrospective interview with Chalmers; Hunt 1997, p. 82. Repeating the same story in another interview, Chalmers added the following comment: "And this is how I learned what I was doing. Apparently I had been doing meta-analyses for years and hadn't known it." Daly 2005, p. 156.

20. Objections to Chalmers's 1977 analysis had been raised by Goldman and Feinstein 1979. That paper is cited in Stampfer et al 1982, the first meta-analysis of the Oxford group, along with a 1978 article by Chalmers and associates; no reference to Chalmers et al 1977 is offered.

lished by Glass and colleagues of his, during the first half of the 1980s, the number of meta-analyses reported in the medical literature was still very small. In the mid 1980s, as mentioned earlier, several articles of a programmatic nature appeared. These papers summarized the development of meta-analysis in the social sciences, assessing the merits of the method and concluding that it was promising enough that it ought to be introduced in clinical research. As the number of meta-analyses addressing medical topics rose sharply in this period, discussions of that kind were quickly superseded by articles weighing the strengths and limitations of analyses actually performed. Publications by social scientists were frequently cited in such papers, and most commentators agree that far more than the term meta-analysis was assimilated in the process. Hence Milos Jenicek, author of the first textbook on meta-analysis in medicine (1987), as well as of several subsequent books on evidence-based medicine, has called the method "a result of a migration from psychology and education into the health sciences" (1989, p. 35). Equally clear, but more elaborate, is the following formulation by Susan Ellenberg (1988, pp. 472–73):

> Formal meta-analysis as a methodologic tool is relatively new to scientific research. While a thorough review of the medical literature of the last 50 years would undoubtedly yield a few examples of quantitatively oriented research reviews, the popularity of this approach to research synthesis clearly originated within the social science community in the mid 1970s.

Guidelines, Replicability, and Reliability

As mentioned above, the parallel between empirical studies and analyses integrating data across studies constituted a central theme in the arguments of social scientists promoting meta-analysis in the late 1970s and early 80s. In the debate on meta-analysis in medical research that soon followed, the same parallel was drawn from an early stage on. "A meta-analysis," it is maintained in what may be regarded as the first overview of this kind of analysis in the health care sector (Louis et al 1985, p. 1), "is to a primary research study as a primary research study is to its study subjects."[21] Given the prevalence of the idea that the accuracy and reliability

21. By an "overview" I understand a review of recent developments which is not programmatic in the sense discussed above; see n. 7. Louis, Fineberg and Mosteller reviewed meta-analyses reported in areas related to public health. Like many related reviews appearing in the following years, theirs addressed the medical literature, whereas the programmatic texts cited earlier summarized work done in the social sciences, recommending researchers within medicine to adopt similar techniques. Prior to the publication of the overview by Louis and associates, methodological problems of meta-analyses of clinical re-

of scientific observations and experiments rest on their replicability, one would expect this notion of a parallelism to be linked to the requirement that meta-analyses be reproducible. As mentioned earlier, such a link was evident in publications by social scientists advocating meta-analysis in the late 1970s and early 80s. As the method came into wide use in clinical research a few years later, the link was equally visible in this area. Some commentators were quite optimistic, suggesting that replicating meta-analyses in medicine ought to be an easy matter (see, for instance, O'Rourke and Detsky 1989, p. 1023), while others took a far more cautious view. Whatever position was taken, and whether or not explicit reference was made to replicability, the degree in which the results of any given meta-analysis were corroborated by independent investigations was a key issue in debates over the strengths and limitations of the method in this period—and has remained so.

Now, if meta-analyses are considered scientific accomplishments, there is reason to regard them as a form of observational, retrospective investigation rather than as experimental studies. This seems to have been well understood.[22] The implication is that one should expect the conclusions of meta-analyses to be no less biased than the results of observational studies. An obvious way to address this difficulty is to pay close attention to criteria by which the studies to be included in meta-analyses are selected. This, then, is why both advocates and critics of meta-analysis attached considerable weight to such criteria, as was evident, for instance, in the discussions of the 1986 meeting in Bethesda. In a review of meta-analysis in public health (Louis et al 1985, p. 2), the problem was described thus:

> Unlike the investigator in a prospective primary study, the meta-analyst has no control, except through selection of papers, over what "treatments" have been applied or how "subjects" are assigned to them.

A viable solution to the problem alluded to here would have to comprise more than inclusion criteria, however. Thomas Chalmers and his co-authors stated the general problem succinctly, and drew a firm conclusion, in their contribution to the Bethesda workshop. Meta-analysis, they argued (Chalmers et al 1987, p. 315),

---

search had been discussed in a paper which also, to some extent, served the function of reviewing the emerging area of medical meta-analysis. See Yeaton and Wortman 1984.

22. But see Gelber and Goldhirsch 1986, p. 1696, for what appears to be a deviant view. "An overview, per se," these authors suggested, using a term which in some quarters has been preferred to meta-analysis, "is a clinical trial."

> can achieve the status of a scientific discipline only if strenuous efforts are made to overcome its major drawback as a reproducible process, that it is retrospectively conducted and thus fraught with the dangers of bias.

Those who wished to contribute to the establishment of meta-analysis as a scientific method in clinical research in the late 1980s, then, had to find ways of minimizing bias. The challenge was to devise procedures that could demonstrably reproduce results already reported. In order to arrive at such procedures, champions of meta-analysis had to propose and seek acceptance for criteria on which specific methodological decisions could be based. In some cases, relatively elaborate criteria formulated for this purpose were offered as comprehensive sets, commonly referred to as guidelines (chief examples are Sacks et al 1987; L'Abbé et al 1987; and Teagarden 1989).[23] Again, a precedent had been set in the social scientific literature, both for the idea of setting guidelines and for specific items in the lists proposed.[24]

There was substantial overlap between different guidelines set out in the late 1980s. Besides directions as to what statistical techniques are appropriate for various purposes, essential points addressed search strategies, inclusion and exclusion criteria, and procedures for the extraction of data. The most fundamental instruction provided in these guidelines, however, concerned protocols. It is a commonplace that prior to any empirical investigation, the research questions to be answered must be decided, along with the type of data to be collected and the analytic methods to be used. In some areas, including clinical research, statements on these matters are generally referred to as protocols. Insisting that research protocols be set up before any observations are made is a means of preventing "data dredging," "post hoc hypothesis formation," and "statistical massage," phrases which all refer to attempts to identify, in a set of data gathered for a different purpose, some pattern worth reporting. Insofar as meta-analysis is considered a form of science, processing in such an inductive manner data pooled from a series of studies ought to be judged just as unacceptable. Put differently, the injunction that a protocol be set up prior to the conduct of any meta-analysis is a corollary of the claim that the method is scientific.

23. These guidelines must not be confused with clinical practice guidelines, a ubiquitous component of evidence-based medicine. Guidelines of the kind discussed in this paper concern procedures adopted in meta-analysis, not in clinical practice.

24. The first author explicitly to propose methods for conducting meta-analysis in the form of guidelines, as far as I know, was Harris Cooper; see Cooper 1982.

As mentioned in a previous section, participants at the Bethesda workshop disagreed as to whether, and if so, how, the quality of studies to be included in meta-analysis should be assessed. One of the positions argued was that the only reliable criterion available was whether or not the randomization process had been conducted properly; the contrary position was that other aspects of study design, too, ought to be taken into account. The disagreement was visible in guidelines published in the same period, as were certain assumptions shared by both parties to this dispute. Everyone agreed that eliminating the subjective element of judgments concerning the quality of investigations was very difficult. In some cases, accordingly, guidelines recommended that assessments of study quality be performed by two independent observers whose disagreements would then be resolved in a consensus meeting. The same procedure, with observers being blinded to parts of the material, was recommended for the extraction of data from the studies selected (L'Abbé et al 1987, p. 227; Sacks et al 1987, p. 452).[25] For those making the detailed decisions that this form of data gathering involves, coding forms constituted an indispensable resource.[26]

It was by mobilizing consensus concerning the need to observe criteria like the ones summarized above, then, that proponents of meta-analysis tried to gain acceptance for the claim that their method was scientific. The criteria on which these efforts centered were often referred to as rules. In a previous section, for instance, I quoted a comment made by Richard Peto at the 1986 workshop, to the effect that "an objective rule" is needed for decisions as to which clinical trials to include in meta-analyses. Peto's solution, as we have seen, was that randomization provides the only sound basis for such a rule (Peto, comment in discussion, in Yusuf et al 1987, p. 230). The fact that investigators failed to reach an agreement based either on Peto's solution or on any other inclusion criterion proposed was recurrently deplored in this period, and commentators sometimes expressed this regret in terms of a lack of general rules. "At present we have no accepted rules" for this kind of decision, the author of one of the sets of guidelines cited above typically remarked (Teagarden 1989, p. 277).

Broad agreement on the appropriateness of a set of rigorous, fairly detailed rules for the conduct of meta-analysis was the only way in which systematic and reproducible procedures, and hence reliable results, could

25. A later study has suggested that blinding is not necessary in meta-analysis; see Berlin 1997.

26. For explicit references to coding schemes or data abstract forms, see, for instance, L'Abbé et al 1987, 226; and Teagarden 1989, pp. 275–76. For ethnographic observations of the use of data abstracting templates in systematic reviewing, see Moreira 2007, p. 188 ff.

be ensured. The key resource employed in the efforts to guarantee the rigorousness of meta-analytic procedures was the analogy with methodological principles observed by scientists in general. Like scientific investigations of any kind, the pivotal argument went, meta-analyses must be carried out in compliance with a set of agreed guidelines or rules. The rules and considerations specific to any empirical study should be stated in a protocol and, again as in any scientific investigation, set out in publications reporting the research. On the one hand, information about the methods used is a prerequisite for the replicability of any experiment or observation. On the other hand, it is by assessing methods that readers are able to decide whether the results reported should be taken seriously. Given the crucial parallel invoked again and again by champions of meta-analysis, both points apply equally to meta-analysis and to primary research.

From Meta-Analysis to Systematic Reviews

The concept of "systematic review" is a more recent invention than meta-analysis. The term is currently used to denote a manner of synthesizing research findings which is widely employed both within and beyond the health care sector, even in areas where meta-analysis is regarded with skepticism. The procedures by which systematic reviews are conducted are commonly assumed to have been extrapolated directly from meta-analysis as applied in clinical research in the 1980s. As we shall see, that view is misleading; the route from medical meta-analysis to systematic reviews was not straightforward. In three other respects, however, I find myself in agreement with common views of the rise of systematic reviewing. First, the concept emerged from meta-analytic techniques. Second, the transition occurred in a medical research context; it is from medicine that this form of research synthesis has subsequently expanded into other areas. Third, a widely cited article published in 1987 by Cynthia Mulrow was instrumental in setting in motion the process through which the format of systematic reviews was established.

A few years earlier, Mulrow had co-authored a review of research addressing the effects of digitalis therapy (Mulrow et al 1984). The manner in which the studies selected for that review were evaluated was very clear. Deriving from an authoritative publication a list of criteria by which the quality of clinical trials may be determined, the authors specified whether and in what degree the respective studies were found to satisfy each of the requirements. In her influential 1987 article, Mulrow again used an explicit set of criteria to assess the quality of a certain category of publications. This time the material selected for scrutiny consisted of those review articles, 50 in all, that had appeared in four leading medical journals

between June 1985 and June 1986. The scholarly standard of these reviews, Mulrow concluded, was disappointingly low. Because the paper in which she reported her findings was itself a review, two levels of reviewing were involved. The key assumption made, informing Mulrow's form of analysis on the one hand, yielding a lucidly argued verdict on the material assessed on the other, was that summaries of existing work should be based on predefined, explicitly stated criteria.

That assumption, as readers of the present paper will immediately recognize, is fundamental to meta-analysis. Hence it is far from surprising that the criteria on which Mulrow's evaluation of the review articles was based were derived from the literature on meta-analysis. Interestingly, however, there is no hint in Mulrow's paper that considerable efforts to establish the method were under way in clinical medicine. By 1987, many meta-analyses of medical research had been reported, the first overviews of the emerging specialty had appeared (Yeaton and Wortman 1984; Louis et al 1985), and that year saw the publication both of the proceedings of the Bethesda workshop (Yusuf et al 1987) and of two sets of guidelines for the conduct of meta-analysis of clinical trials (Sacks et al 1987; L'Abbé et al 1987). Mulrow did not mention these developments. Instead, she presented the criteria employed in her evaluation as being adapted from guidelines for "information synthesis" (1987, 485). Closely related to the concept of meta-analysis, this is one of the neologisms conceived in the methodological debates that Glass's 1976 article triggered.[27]

More specifically, Mulrow adduced three sources for her criteria. The first was a monograph by Richard Light and David Pillemer (1984), one of the first books on meta-analysis published in the social sciences. The second was a paper proposing guidelines for the synthesis of research on health education and promotion (Mullen and Ramirez 1986). This area, addressing the effectiveness of educational programs for patients undergoing or about to undergo various forms of medical treatment, represents an extension of educational research into the health care sector. The third source adduced by Mulrow was an unpublished manual describing a set of procedures for assessing the empirical basis of knowledge claims made in the medical literature (Policy Research Incorporated 1978).[28]

27. Related terms include data synthesis, research integration and research synthesis.

28. This manual stems from a project addressing the provision of medical information that was carried out by a corporation based in Baltimore. Though the reference offered by Mulrow does not coincide fully with the title I have tracked down, this appears to be the manual she used. I am grateful to Peter Goldschmidt, co-investigator in the Medical Practice Information Demonstration Project and one of the authors of the manual, for making a copy available to me. The roots of this project lay outside meta-analysis, but later, having become familiar with that method, Goldschmidt proposed "information synthesis" as an

While guidelines and checklists are offered in all three sources, they differ in other respects. The unpublished manual focuses on the medical literature, but makes no reference to meta-analysis. The two other texts, by contrast, argue the merits of summarizing research findings by meta-analytic procedures, but pay little attention to their application in clinical research. Though the volume by Light and Pillemer has been a very successful textbook of meta-analysis, the term itself, in fact, is consistently avoided in it. Nor did Mulrow come across meta-analysis of clinical trials through her sample of reviews; the method was not used in any of the 50 articles.

Mulrow's apparent unawareness that meta-analytic techniques were already being employed in clinical medicine may have encouraged her to take a step back from meta-analysis to more fundamental issues. The emergence of systematic reviews represents precisely such a step. Seeking out guidelines for the conduct of meta-analysis proposed outside clinical research, where she was herself active, Mulrow was able to take a message back to her own field.[29] The message did not concern the details of statistical procedures. The significant point made, instead, was that reviews themselves constitute a form of research and thus should conform, to the same extent as primary research, to methodological criteria defined prior to the investigation. Mulrow set up eight fairly simple criteria, such as whether the purpose of the review was clearly stated, whether the procedures adopted to identify potential material for the review were specified, and whether the basis on which studies were selected from the material yielded by the search procedures was described. Demonstrating that such criteria were routinely violated in reviews published in the top journals was a highly effective way of advocating the basic idea that standard methodological principles must be heeded in reviews, too. Some years earlier, as mentioned in a previous section, Gregg Jackson (1980) had ad-

---

alternative, cognate approach. By 1987, this term had been in use for several years at the Veterans Administration in Washington, D.C., where Goldschmidt served as director of the Health Services Research and Development Service. Goldschmidt 1986; Hedrick and Inui 1986, pp. 867, 880; Marcus et al 1987.

29. Asked in retrospect how she arrived at the position argued in her 1987 paper, Mulrow has offered an account which supports this interpretation. In a recent piece by Edward Huth, former editor of *Annals of Internal Medicine,* where Mulrow's paper was published, she is quoted as recalling having had the impression that overviews of various medical topics were often based on opinion rather than evidence. Hence, she recollects, "I began to look for literature on reviews—and found much good work in the social science field. I applied that work to thinking about reviews published in medical journals and (voilà)—the *Annals* article." Huth 2008. This does not imply that Mulrow had never come across meta-analysis of clinical trials. In fact, in the piece just cited, she is quoted as recalling attending a presentation of that form of analysis given by Richard Peto in the early 1980s.

dressed reviews published in social science journals in a similar empirical manner; Mulrow's 1987 paper may be considered a highly successful repeat performance in medicine.[30] Crucially, however, she did not emphasize the function that quantitative techniques may serve in synthesizing research. Finding that quantitative procedures were used in only three of the 50 reviews, Mulrow recommended more frequent use of such techniques, but this was a minor feature of her argument.[31]

This subsumption of quantitative methods under more general principles is the essential element in the transition from meta-analysis to systematic reviews. For someone not conversant with meta-analysis as practiced in medicine it was not possible to make this step in a fully explicit manner, at least not in a clinical context, and in Mulrow's paper the point is not unequivocal. The following year, however, it was clearly stated by Andrew Oxman and Gordon Guyatt, who were part of a team based at McMaster University, Canada, which, a few years later, was to launch the concept of evidence-based medicine (Guyatt 1991; Evidence-Based Medicine Working Group 1992). Oxman and Guyatt, who were well informed about meta-analysis as a method for aggregating clinical trial data, offered a set of guidelines for assessing the quality of medical research reviews. At the end of a section addressing appropriate ways of statistically combining data from separate studies, they made the decisive step in the transition from meta-analysis to systematic reviews crystal clear:

> It is important to remember that all the other guidelines we have discussed still apply whether or not the authors of a review have used meta-analysis. (Oxman and Guyatt 1988, p. 702)[32]

Whereas meta-analysis was widely debated in the 1980s, both in medicine and in the social sciences, the use of the concept of information synthesis was infrequent (and has remained so). Hence it may seem odd that Mulrow, not even mentioning meta-analysis, suggested that her criteria for evaluating reviews had been adapted from guidelines for information synthesis. That term, however, in itself reflects the very position concern-

30. No reference to Jackson's paper was offered in Mulrow's article.

31. The quantitative procedures used in these three reviews can hardly be regarded as examples of meta-analysis, and no reference to that concept was made either in these articles or by Mulrow. It was not long, however, before Mulrow started referring to meta-analysis of clinical research. In a 1988 paper proposing a revised, simpler set of criteria, the textbook by Light and Pillemer was again adduced as a source, but this time along with two overviews of meta-analysis as applied in medicine; Mulrow et al 1988. One of the best early overviews of meta-analysis of clinical research was published by Stephen Thacker, co-author of this paper, later that year.

32. The following year, the point was made in an equally unmistakable manner by two other authors; O'Rourke and Detsky 1989, p. 1023.

ing quantitative methods that constitutes the decisive step in the transition from meta-analysis to systematic reviews. When introducing this concept, Peter Goldschmidt suggested that an information synthesis "may include a meta-analysis, the statistical manipulation of findings from multiple research studies" (1986, p. 220; italics removed). Information synthesis, in other words, was proposed as a more inclusive format, addressing more general issues involved in the reviewing of research. In another publication that year, discussing methods of summarizing research on health education, Goldschmidt's concept was adopted to allow qualitative as well as quantitative aspects of synthesizing research to be addressed (Mullen and Ramirez 1986, p. 204).[33] That article was cited by Mulrow as one of the sources for the eight criteria by which she assessed her sample of research reviews, and it appears to have been from its authors that she picked up the concept of information synthesis. On closer inspection, then, Mulrow's use of that term, far from being an historical oddity, was fully consonant with her stance on quantitative reviewing techniques.[34]

As mentioned earlier, Thomas Chalmers and others worked hard to secure for meta-analysis the standing of a scientific discipline. Extending the parallelism argument to research reviews in general, as proposed by Cynthia Mulrow, Andrew Oxman, and Gordon Guyatt, meant suggesting that the wider genre, too, ought to attain such status. It was entirely appropriate, therefore, for the two latter authors to publish an article titled "The Science of Reviewing Research" (Oxman and Guyatt 1993). The scientific character of reviewing had previously been assumed by social scientists involved in the meta-analytic revolution of the late 1970s and early 80s; the title of Oxman and Guyatt's paper, in fact, reiterated the subtitle of the 1984 volume by Light and Pillemer. Not everyone agreed, however, that the parallelism argument ought to be extrapolated from meta-analysis to literature reviews in general. Commenting on this paper by Oxman and Guyatt, Chalmers challenged that idea:

> If one assumes that a good meta-analysis includes all of the things that you've pointed out are necessary for doing a review . . .—is there any place for a systematic review without the quantitative

33. These authors cited an earlier version of Goldschmidt's paper.

34. Incidentally, there is a reference to information synthesis in the brief editorial in which Gordon Guyatt introduced the concept of evidence-based medicine, too. Cited as one of the new skills clinicians need to master, along with literature retrieval and critical appraisal of the studies identified, information synthesis seems to be employed as a general term under which meta-analysis might be subsumed. In the full-blown call for evidence-based medicine which soon followed, meta-analysis was mentioned instead, as

> analysis? Is there any place for a review of data in the literature where the data have not been statistically analyzed? Should not all qualitative reviews that do not analyze the data as data be replaced by adequately done meta-analyses? (Chalmers, comment in discussion, in Oxman and Guyatt 1993, p. 134)

To this Oxman replied that there are cases where the data produced in well-conducted clinical trials are too meager to justify a meta-analysis. Even in the absence of adequate quantitative data, however, research bearing on specific clinical questions should be identified and synthesized according to rigorous, explicit rules. Put differently, meta-analytic procedures form part of a larger enterprise in which quantitative techniques are of secondary importance. This has since become the dominant view (see, for instance, Chalmers and Altman 1995, p. viii; Egger et al 2001, p. 5; Clarke 2006, p. 117).

Now, if literature reviews are to be conducted by formalized procedures whether meta-analyses are included or not, then a term is needed to distinguish reviews which satisfy this requirement from those that do not. As indicated above, the appeal of "information synthesis" was quite limited. Instead, from the early 1990s on, the term *systematic review* has been adopted for this purpose. For all the effectiveness of Mulrow's 1987 article in convincing practitioners and users of medical research that more reliable routines for reviewing the literature were needed, and though it is frequently cited as having introduced the format of systematic review, that term does not appear in it. Mulrow certainly stressed the need for systematic procedures, and suggested that her paper "presents a systematic review method" (1987, p. 487). Yet it was not until an Oxford unit called the Cochrane Centre began operating in the autumn of 1992 that "systematic review" was introduced as a label for the reviewing format then being established (Chalmers et al 1992).[35] The first director of this center was Iain Chalmers (no relation of Thomas Chalmers), a British obstetrician. Some fifteen years earlier, he had started collecting randomized clini-

---

one of two important new methods of clinical research, the randomized clinical trial being the other. Guyatt 1991; Evidence-Based Medicine Working Group 1992, p. 2420.

35. Earlier publications in which this term appears include Cochrane 1989, after whose author the above-mentioned center was named, as well as Chalmers 1991, pp. 136, 141. In neither case, however, is the term used in the programmatic sense of Chalmers et al 1992, an editorial announcing the new center. Clearly, the phrase had been used occasionally long before acquiring its current status. Thus a Medline search for the period up to 1991 produces 85 hits for the text words "systematic review," the earliest one deriving from the 1954 volume of a German journal. Cf. Petticrew and Roberts 2006, pp. 16–19, where many other early examples are given.

cal trials bearing on the effect of perinatal care, in order to create an overview of reliable evidence in this area (Chalmers 1991, p. 135; Daly 2005, p. 161). The project gradually assumed impressive proportions, and by the end of the 1980s a massive two-volume work was completed (Chalmers et al (eds.) 1989). Many of the 89 chapters included had several authors, and numerous meta-analyses of trial data were reported, Chalmers and many of his collaborators being well versed in this methodology. Having overseen the publication of these reviews, Chalmers realized how valuable continuously updated versions might be. It was decided that the groups responsible for the reviews should be maintained and enlarged, and updated versions were distributed by floppy disk every six months (Chalmers 1991, p. 137).

In October 1993, a year after the opening of the first Cochrane Centre, an international network was founded, comprising national and regional centers as well as collaborative review groups. Governmental agencies conducting, maintaining and disseminating systematic reviews have since been set up in many countries, but this network, the Cochrane Collaboration, is by far the world's most important organization carrying out these tasks. More than 28,000 people are active in the network in over 100 countries, and more than 4500 systematic reviews of research on a wide range of health-related issues can currently be accessed via the website of the organization (www.cochrane.org). Through the activities of the Cochrane Collaboration, often referred to as a cornerstone of evidence-based medicine, systematic reviews have been established as a major format of bringing research findings together, within and beyond biomedicine.

Conclusion

Meta-analysis was conceived in the mid 1970s, in response to the need for reliable summaries of large numbers of studies in American quantitative social science. By the mid 80s, a handful of textbooks having appeared, the method was fairly well established in this field. A few years later, the use of similar techniques in clinical research increased rapidly, from a very low level. Then, in the early 90s, the basic principles of meta-analysis were extended into systematic reviews, a format which, frequently but not necessarily, includes statistical analyses of the data. The methodology of meta-analysis and systematic reviews is a central component of evidence-based medicine, which was launched in the same period. Over the last decade or so, as the concepts of evidence-based practice and policy have become prevalent far beyond medicine, formal methods of synthesizing research have been established in social care, education, and other areas. Given that systematic reviews may include qualitative as well as quanti-

tative studies, there appears to be no limit to the transferability of this format.[36]

In this paper, three themes have been pursued in order to shed light on the emergence of formal methods of synthesizing research. The first theme concerns two closely related aspects of the strategy employed by proponents of these methods. On the one hand, what may be called the parallelism assumption is fundamental to the entire enterprise. It is by projecting on to reviews of the research literature a certain image of common scientific practice that advocates of meta-analysis and systematic reviews have defined the essential requirements of research synthesis. Indeed, to make the parallelism crystal clear, this kind of analysis is sometimes designated "secondary research."[37] On the other hand, it is by means of guidelines and other formal tools that this parallelism, and hence the rigorousness of the procedures involved in research synthesis, can be achieved.

The second theme concerns the extent to which meta-analytic techniques developed by social scientists in the 1970s informed that practice of combining data from independent studies which grew common in medicine in the latter half of the 80s. Because definitions of meta-analysis diverge widely, so do accounts of its origin. Meta-analysis as applied in clinical research is often assumed to owe little to the work of social scientists. The Oxford group, whose syntheses of data from separate clinical trials did much to convince the medical community that this new form of analysis was valuable, even indispensable, adopted a designation for the method and a practice of referencing the literature, both of which had the effect of minimizing the social scientific background. This was in contrast to Thomas Chalmers, an equally important protagonist of medical meta-analysis who stressed the utility of techniques pioneered by social scientists. As demonstrated above, there is much to suggest that medical researchers engaging in meta-analysis in fact drew substantially on those techniques.[38]

36. In educational research, however, a strong case has been made for the claim that beyond a certain point, the formalization of reviewing practices results in distortion of the findings synthesized. I am currently preparing an article addressing this controversy, under the working title, "Not Beyond this Point: Understanding Epistemological Barriers to the Synthesis of Educational Research."

37. This term appears in Gregg Jackson's article (1980, p. 445), but without being emphasized. The earliest instance of its current usage of which I am aware is Chalmers et al 1993, pp. 412–13.

38. In this disagreement between main actors in the field under study, I am thus siding with Thomas Chalmers against the Oxford group. Here the limits of a symmetrical perspective are evident: where my analyst's concerns coincide with those of actors, the distinction between topics and resources breaks down and symmetry must be relinquished. For an illuminating discussion of this problem, see Collins 1995, pp. 298–99.

The third theme concerns the nature of the historical link between meta-analysis and systematic reviews. The concept of systematic review was introduced in the early 1990s, as a label for the analyses provided by the Cochrane Collaboration. In an article published by Cynthia Mulrow a few years earlier, key arguments justifying this new form of research review were offered, but contrary to what is often assumed, this is not where the term systematic review was coined. Mulrow was not well versed in the use of meta-analysis in clinical research and did not stress the need for quantitative analyses of the data. Drawing on core principles underlying the development of meta-analytic techniques in social science, however, Mulrow made a strong case for the need to formalize the procedures by which research findings are synthesized. The historical development from meta-analysis as introduced by social scientists to medical meta-analysis and on to the format of systematic reviews, as applied within and beyond the health care sector, is characterized by a considerable degree of continuity. At the core of this continuity is the reliance on formal tools legitimized by the assumed parallelism between primary studies and research synthesis.

References

Anonymous. 1993. "Cardiologist Salim Yusuf Gets to the Heart of Meta-Analysis." *Science Watch* 4 (8): 3–4.

Bausell, R. Barker, Yu-Fang Li, Meei-Ling Gau, and Karen L. Soeken. 1995. "The Growth of Meta-Analytic Literature from 1980 to 1993." *Evaluation and the Health Professions* 18: 238–51.

Berlin, Jesse A. 1997. "Does Blinding of Readers Affect the Results of Meta-Analyses?" *Lancet* 350: 185–86.

Chalmers, Iain. 1991. "Improving the Quality and Dissemination of Reviews of Clinical Research." Pp. 127–46 in *The Future of Medical Journals.* Edited by Stephen P. Lock. London: BMJ.

Chalmers, Iain and Doug Altman 1995. Foreword. Pp. vii–ix in *Systematic Reviews.* Edited by Iain Chalmers and Doug Altman. London: BMJ.

Chalmers, Iain, Harris Cooper, and Larry V. Hedges. 2002. "A Brief History of Research Synthesis." *Evaluation and the Health Professions* 25: 12–37.

Chalmers, Iain, Kay Dickersin, and Thomas C. Chalmers. 1992. "Getting to Grips with Archie Cochrane's Agenda." *British Medical Journal* 305: 786–88.

Chalmers, Iain, Murray Enkin, and Marc J. N. C. Keirse (eds.). 1989. *Effective Care in Pregnancy and Childbirth.* Oxford: Oxford University Press, 2 vols.

Chalmers, Iain, Murray Enkin, and Marc J. N. C. Keirse. 1993. "Pre-

paring and Updating Systematic Reviews of Randomized Controlled Trials of Health Care." The Milbank Quarterly 71: 411–37.

Chalmers, Iain, Jini Hetherington, Diana Elbourne, M. J. N. C. Kierse, and M. Enkin. 1989. Materials and Methods Used in Synthesizing Evidence to Evaluate the Effects of Care During Pregnancy and Childbirth. Pp. 39–65 in *Effective Care in Pregnancy and Childbirth.* Edited by I. Chalmers, M. Enkin, and M. J. N. C. Keirse. Oxford: Oxford University Press, Vol I.

Chalmers, Thomas C., Jayne Berrier, Henry S. Sacks, Howard Levin, Dinah Reitman, and Raguraman Nagalingham. 1987. "Meta-Analysis of Clinical Trials as a Scientific Discipline. II: Replicate Variability and Comparison of Studies that Agree and Disagree." *Statistics In Medicine* 6: 733–44.

Chalmers, Thomas C., Howard Levin, Henry S. Sacks, Dinah Reitman, Jayne Berrier, and Raguraman Nagalingam. 1987. "Meta-Analysis of Clinical Trials as a Scientific Discipline. I: Control of Bias and Comparison with Large Co-Operative Trials." *Statistics In Medicine* 6: 315–25.

Chalmers, Thomas C., Raymond J. Matta, Harry Smith Jr., and Anne-Marie Kunzler. 1977. "Evidence Favoring the Use of Anticoagulants in the Hospital Phase of Acute Myocardial Infarction." *New England Journal of Medicine* 297: 1091–96.

Clarke, Mike. 2006. The Cochrane Collaboration. Pp. 115–23 in *Public Health Evidence: Tackling Health Inequalities.* Edited by A. Killoran, C. Swann, and M. P. Kelly, Oxford: Oxford University Press.

Clarke, Michael J. and Lesley A. Stewart. 2001. Obtaining Individual Patient Data from Randomised Controlled Trials. Pp. 109–21 in *Systematic Reviews in Health Care: Meta-Analysis in Context,* 2nd Edn. Edited by M. Egger, G .D. Smith, and D. G. Altman, London: BMJ.

Coffman, Thomas L. 1981. "Relocation and Survival of Institutionalized Aged: A Re-examination of the Evidence." *The Gerontologist* 21: 483–500.

Collins, Harry. 1995. Science Studies and Machine Intelligence. Pp. 286–301 in *Handbook of Science and Technology Studies.* Edited by Shiela Jasanoff, Gerald E. Markle, James C. Petersen, and Trevor J. Pinch London: Sage.

Cooper, Harris M. 1979. "Statistically Combining Independent Studies: A Meta-Analysis of Sex Differences in Conformity Research." *Journal of Personality and Social Psychology* 37: 131–46.

Cooper, Harris M. 1982. "Scientific Guidelines for Conducting Integrative Research Reviews." *Review of Educational Research* 52: 291–302.

Cooper, Harris M. 1984. *The Integrative Research Review: A Systematic Approach.* Beverly Hills, CA: Sage.

Cooper, Harris 1998. *Synthesizing Research: A Guide for Literature Reviews,* 3rd Edn. Thousand Oaks, CA: Sage.

Cooper, Harris and Larry V. Hedges. 1994. Research Synthesis as a Scientific Enterprise. Pp. 3–14 in *The Handbook of Research Synthesis.* Edited by Harris Cooper and Larry V. Hedges . New York: Russell Sage Foundation.

Cooper, Harris M. and Robert Rosenthal 1980. ":Statistical Versus Traditional Procedures for Summarizing Research Findings." *Psychological Bulletin* 87: 442–49.

Curlette, William L. and Kathleen Silva Cannella. 1985. "Going Beyond the Narrative Summarization of Research Findings: The Meta-Analysis Approach." *Research in Nursing and Health* 8: 293–301

Daly, Jeanne. 2005. *Evidence-Based Medicine and the Search for a Science of Clinical Care.* Berkeley: University of California Press; New York: Milbank Memorial Fund.

DeMets, David L. 1987. "Methods for Combining Randomized Clinical Trials: Strengths and Limitations." *Statistics in Medicine* 6: 341–48.

Dickersin, Kay and Jesse A. Berlin. 1992. "Meta-Analysis: State of the Science." *Epidemiologic Reviews* 14: 154–76.

Dickersin, Kay, Karen Higgins, and Curtis L. Meinert. 1990. "Identification of Meta-Analyses: The Need for Standard Terminology." *Controlled Clinical Trials* 11: 52–66.

Egger, Matthias, George Davey Smith, and Keith O'Rourke. 2001. Rationale, Potentials, and Promise of Systematic Reviews. Pp. 3–19 in *Systematic Reviews in Health Care: Meta-Analysis in Context,* 2nd Edn. Edited by M. Egger, G. D. Smith, and D. G. Altman. London: BMJ.

Einarson, Thomas R., William F. McGhan, J. Lyle Bootman, and Darrell L. Sabers. 1985. "Meta-Analysis: Quantitative Integration of Independent Research Results." *American Journal of Hospital Pharmacy* 42: 1957–64.

Ellenberg, Susan S. 1988. "Meta-Analysis: The Quantitative Approach to Research Review." *Seminars in Oncology* 15: 472–81.

Evidence-Based Medicine Working Group. 1992. "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine." *Journal of the American Medical Association* 268: 2420–25.

Eysenck, Hans J. 1978. "An Exercise in Mega-Silliness." *American Psychologist* 33: 517.

Feldman, Kenneth A. 1971. "Using the Work of Others: Some Observations on Reviewing and Integrating." *Sociology of Education* 44: 86–102.

Feldman, Kenneth A. and Theodore M. Newcomb. 1969a. *The Impact of College on Students. Vol I: An Analysis of Four Decades of Research.* San Francisco: Jossey-Bass.

Feldman, Kenneth A. & Theodore M. Newcomb 1969b. *The Impact of College on Students. Vol II: Summary Tables.* San Francisco, Cal.: Jossey-Bass.

Gelber, Richard D. and Aron Goldhirsch. 1986. "The Concept of an Overview of Cancer Clinical Trials with Special Emphasis on Early Breast Cancer." *Journal of Clinical Oncology* 4: 1696–1703.

Glass, Gene V. 1976. "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher* 5(10): 3–8.

Glass, Gene V. 1977. "Integrating Findings: The Meta-Analysis of Research." *Review of Research in Education* 5: 351–79.

Glass, Gene V., Barry McGaw, and Mary Lee Smith. 1981. *Meta-Analysis in Social Research.* Beverly Hills, CA: Sage.

Glass, Gene V. and Mary Lee Smith. 1979. "Meta-Analysis of Research on the Relationship of Class Size and Achievement." *Educational Evaluation and Policy Analysis* 1: 2–16.

Glass, Gene V., Leonard S. Cahen, Mary Lee Smith, and Nikola N. Filby. 1982. *School Class Size: Research and Policy.* Beverly Hills, CA: Sage.

Goldman, Lee and Alvan R. Feinstein. 1979. "Anticoagulants and Myocardial Infarction: The Problem of Pooling, Drowning, and Floating." *Annals of Internal Medicine* 90: 92–94.

Goldschmidt, Peter G. 1986. "Information Synthesis: A Practical Guide." *Health Services Research* 21: 215–37.

Guyatt, Gordon H. 1991. "Evidence-Based Medicine." *ACP Journal Club,* A-16 (supplement 2 to *Annals of Internal Medicine* 114).

Hedges, Larry V. and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis.* Orlando, FL.: Academic Press.

Hedrick, Susan C. and Thomas S. Inui. 1986. "The Effectiveness and Cost of Home Care: An Information Synthesis." *Health Services Research* 20: 851–80.

Herring, Conyers 1968. "Distill or Drown: The Need for Reviews." *Physics Today* 21(9): 27–33.

Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-Analysis.* New York: Russell Sage Foundation.

Hunter, John E., Frank L. Schmidt, and Gregg B. Jackson. 1982. *Meta-Analysis: Cumulating Research Findings across Studies.* Beverly Hills, CA: Sage.

Huth, Edward J. 2008. "The Move Toward Setting Standards for the Content of Medical Review Articles." James Lind Library. http://www.jameslindlibrary.org/illustrating/articles/the-move-toward-setting-scientific-standards-for-the-content-of (accessed March 6, 2012)

Jackson, Gregg B. 1980. "Methods for Integrative Reviews." *Review of Educational Research* 50: 438–460.

Jenicek, Milos. 1987. *Meta-analyse en Médecine: Evaluation et Synthèse de l'Information Clinique et Épidémiologique.* St Hyacinthe, Québec: Edisem.

Jenicek, Milos. 1989. "Meta-Analysis in Medicine: Where We Are and Where We Want to Go." *Journal of Clinical Epidemiology* 42: 35–44.

L'Abbé, Kristan A., Allan S. Detsky, and Keith O'Rourke. 1987. "Meta-Analysis in Clinical Research." *Annals of Internal Medicine* 107: 224–33.

Light, Richard J. 1984. "Six Evaluation Issues That Synthesis Can Resolve Better Than Single Studies." *New Directions for Program Evaluation* 24: 57–73.

Light, Richard J. and David B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research.* Cambridge, MA: Harvard University Press.

Light, Richard J. and Paul V. Smith. 1971. "Accumulating Evidence: Procedures for Resolving Contradictions among Different Research Studies." *Harvard Educational Review* 41: 429–71.

Louis, Thomas A., Harvey V. Fineberg, and Frederick Mosteller. 1985. "Findings for Public Health From Meta-Analysis." *Annual Review of Public Health* 6: 1–20.

Mantel, Nathan and William Haenszel. 1959. "Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease." *Journal of the National Cancer Institute* 22: 719–48.

Marcus, Susan H., Prakash L. Grover, and Dennis A. Revicki. 1987. "The Method of Information Synthesis and its Use in the Assessment of Health Care Technology." *International Journal of Technology Assessment in Health Care* 3: 497–508.

Moreira, Tiago. 2007. "Entangled Evidence: Knowledge Making in Systematic Reviews in Healthcare." *Sociology of Health & Illness* 29: 180–97.

Mullen, Patricia D. and Gilbert Ramirez. 1986. "Information Synthesis and Meta-Analysis." *Advances in Health Education and Promotion* 2: 201–39.

Mulrow, Cynthia D. 1987. "The Medical Review Article: State of the Science." *Annals of Internal Medicine* 106: 485–88.

Mulrow, Cynthia D., John R. Feussner, and Ramon Velez. 1984. "Reevaluation of Digitalis Efficacy: New Light on an Old Leaf." *Annals of Internal Medicine* 101: 113–17.

Mulrow, Cynthia D., Stephen B. Thacker, and Jacqueline A. Pugh. 1988. "A Proposal for More Informative Abstracts of Review Articles." *Annals of Internal Medicine* 108: 613–15.

Mykhalovskiy, Eric and Lorna Weir. 2004. "The Problem of Evidence-Based Medicine: Directions for Social Science." *Social Science & Medicine* 59: 1059–69.

National Institutes of Health Consensus Development Conference State-

ment. 1985. "Adjuvant Chemotherapy for Breast Cancer." *Journal of the American Medical Association* 254: 3461–63.

Naylor, C. David. 1988. "Two Cheers for Meta-Analysis: Problems and Opportunities in Aggregating Results of Clinical Trials." *Canadian Medical Association Journal* 138: 891–95.

Norton, Larry. 1987. "Commentary." *Statistics in Medicine* 6: 333–35.

Olkin, Ingram. 1990. History and Goals. Pp. 3–10 in *The Future of Meta-Analysis.* Edited by K. W. Wachter and M. L. Straf. New York: Russell Sage Foundation.

O'Rourke, Keith and Allan S. Detsky. 1989. "Meta-Analysis in Medical Research: Strong Encouragement for Higher Quality in Individual Research Efforts." *Journal of Clinical Epidemiology* 42: 1021–24.

Ottenbacher, Kenneth J. and Paul Petersen. 1984a. "Quantitative Reviewing of Medical Literature: An Approach to Synthesizing Research Results in Clinical Pediatrics." *Clinical Pediatrics* 23: 423–27.

Ottenbacher, Kenneth J. and Paul Petersen. 1984b. "The Efficacy of Vestibular Stimulation as a Form of Specific Sensory Enrichment: Quantitative Review of the Literature." *Clinical Pediatrics* 23: 428–33.

Oxman, Andrew D. and Gordon H. Guyatt. 1988. "Guidelines for Reading Literature Reviews." *Canadian Medical Association Journal* 138: 697–703.

Oxman, Andrew D. and Gordon H. Guyatt. 1993. "The Science of Reviewing Research." *Annals of the New York Academy of Sciences* 703: 125–34.

Patsopoulos, Nikolaos A., Apostolos A. Analatos, and John P. A. Ioannidis. 2005. "Relative Citation Impact of Various Study Designs in the Health Sciences." *Journal of the American Medical Association* 293: 2362–66.

Peto, Richard. 1987. "Why Do We Need Systematic Overviews of Randomized Trials?" *Statistics in Medicine* 6: 233–40.

Petticrew, Mark and Helen Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide.* Malden, MA: Blackwell.

Pickering, Andy. 1984. "Against Putting the Phenomena First: The Discovery of the Weak Neutral Current." *Studies in History and Philosophy of Science* 15: 85–117.

Policy Research Incorporated. 1978. *Validation Procedures Manual.* Baltimore, MD: Policy Research Inc.

Rosenthal, Robert. 1978. "Combining Results of Independent Studies." *Psychological Bulletin* 85: 185–93.

Rosenthal, Robert. 1984. *Meta-Analytic Procedures for Social Research.* Beverly Hills, CA: Sage.

Sacks, Henry S., Jayne Berrier, Dinah Reitman, V. A. Ancona-Berk, and

T. C. Chalmers. 1987. "Meta-Analyses of Randomized Controlled Trials." *New England Journal of Medicine* 316: 450–55.

Shadish, William R. and C. Keith Haddock. 1994. Combining Estimates of Effect Size. Pp. 261–81 in *The Handbook of Research Synthesis.* Edited by H. Cooper and L. V. Hedges. New York: Russell Sage Foundation, 261–81.

Smith, Mary Lee and Gene V. Glass. 1977. "Meta-Analysis of Psychotherapy Outcome Studies." *American Psychologist* 32: 752–60.

Smith, Mary Lee, Gene V. Glass, and Thomas I. Miller. 1980. *The Benefits of Psychotherapy.* Baltimore, MD: Johns Hopkins University Press.

Stampfer, Meir J., Samuel Z. Goldhaber, Salim Yusuf, R. Peto, and C. H. Hennekens . 1982. "Effect of Intravenous Streptokinase on Acute Myocardial Infarction: Pooled Results from Randomized Trials." *New England Journal of Medicine* 307: 1180–82.

Stjernswärd, Jan. 1974. "Decreased Survival Related to Irradiation Postoperatively in Early Operable Breast Cancer." *Lancet* 2: 1285–86.

Teagarden, J. Russell. 1989. "Meta-Analysis: Whither Narrative Review?" *Pharmacotherapy* 9: 274–81.

Thacker, Stephen B. 1988. "Meta-Analysis: A Quantitative Approach to Research Integration." *Journal of the American Medical Association* 259: 1685–89.

Wittes, Robert E. 1987. "Problems in the Medical Interpretation of Overviews." *Statistics in Medicine* 6: 269–76.

Worrall, John. 2002. "What Evidence in Evidence-Based Medicine?" *Philosophy of Science* 69: S316–30.

Yeaton, William H. and Paul W. Wortman. 1984. "Evaluation Issues in Medical Research Synthesis." *New Directions for Program Evaluation* 24: 43–56.

Yusuf, Salim, Rory Collins, Richard Peto, M. Furberg, M. J. Stampfer, S. Z. Goldhaber, and C. H. Hennekens . 1985. "Intravenous and Intracoronary Fibrinolytic Therapy in Acute Myocardial Infarction: Overview of Results on Mortality, Reinfarction and Side-Effects from 33 Randomized Controlled Trials." *European Heart Journal* 6: 556–85.

Yusuf, Salim, Richard Peto, John Lewis, R. Collins, and P. Sleight. 1985. "Beta Blockade During and After Myocardial Infarction: An Overview of the Randomized Trials." *Progress in Cardiovascular Diseases* 27: 335–71.

Yusuf, Salim, Robert Simon, and Susan S. Ellenberg (eds.). 1987. "Proceedings of the Workshop on Methodologic Issues in Overviews of Randomized Clinical Trials." *Statistics in Medicine* 6: 217–410.