Historical Perspective

Research Synthesis Methods

Received 16 October 2014, Revised 30 October 2014, Accepted 08 November 2014 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1132

The meta-analytic big bang[‡]

William R. Shadish[a*†] and Jesse D. Lecy[b]

This article looks at the impact of meta-analysis and then explores why meta-analysis was developed at the time and by the scholars it did in the social sciences in the 1970s. For the first problem, impact, it examines the impact of meta-analysis using citation network analysis. The impact is seen in the sciences, arts and humanities, and on such contemporaneous developments as multilevel modeling, medical statistics, qualitative methods, program evaluation, and single-case design. Using a constrained snowball sample of citations, we highlight key articles that are either most highly cited or most central to the systematic review network. Then, the article examines why meta-analysis came to be in the 1970s in the social sciences through the work of Gene Glass, Robert Rosenthal, and Frank Schmidt, each of whom developed similar theories of meta-analysis at about the same time. The article ends by explaining how Simonton's chance configuration theory and Campbell's evolutionary epistemology can illuminate why meta-analysis occurred with these scholars when it did and not in medical sciences. Copyright © 2015 John Wiley & Sons, Ltd.

The big bang is a cosmological theory of how things started small but expanded outward until they are momentous. This article draws an analogy between the cosmological big bang (Figure 1) and the rapid growth of meta-analysis as an important tool in scholarly work. Meta-analysis started with a few articles published in the 1970s but expanded outward rapidly to have an important impact on science. In part, the analogy was inspired by graphical portrayals of the meta-analysis literature using citation network analysis (CNA), images that resemble the form of the cosmological big bang (Figure 2). More so, the analogy recognizes the historical importance of meta-analysis in science. One cannot take this analogy too seriously, of course; the cosmological big bang gave rise to the universe as we know it. Meta-analysis or any other scientific achievement pales by comparison.

Yet by any measure, meta-analysis has to be counted as one of the central methodological developments in science in the last 100 years. A recent Google search of the phrase 'meta-anal' yielded about 26 million hits, no surprise because meta-analysis is used in a great many branches of science, and even in the arts and humanities, albeit to a far lesser degree; the term is even referenced in the popular press. Meta-analysis is central to the evidence-based practice and policies movement in medicine and public health as a key method for summarizing what we know about what works, and it is slowly coming to play a similar role in education and psychology. In fields like psychology, it helped bolster the movement to report effect sizes and confidence intervals in scientific publications and to decrease the nearly monolithic emphasis on null-hypothesis significance testing that had mostly characterized the field in the 20th century. Meta-analysis gave concrete form to the notion that cumulative knowledge occurs in science through its unique form of quantitative synthesis of individual scientific results. As scientific developments go, then, meta-analysis did indeed spark its own big bang.

It is useful to speculate about why the meta-analytic big bang occurred in the 1970s and how it has unfolded since then. Thus, we focus primarily on the history of modern meta-analysis—the era that started around the time that Glass's (1976) article coined the phrase meta-analysis and outlined its essential characteristics. Other scholars have done histories that go back further in time (Bohlin, 2012; Chalmers, Hedges, and Cooper, 2002; Huberty, 2002; O'Rourke, 2007), reminding us of such early examples as Pearson's (1904a, 1904b) synthesis of correlations pertaining to the effectiveness of a typhoid vaccine. In contrast, the present article joins an emerging literature documenting how modern meta-analysis arose (e.g., Becker, 2007; Hunt, 1999; Cappelleri and Ingerick, 2014; Schmidt and Hunter, 2003). The article discusses what constitutes this meta-analytic big bang and reflects on

[a]*School of Social Sciences, Humanities and Arts, University of California, Merced, CA, USA*
[b]*420 Maxwell School of Syracuse University, Syracuse, NY, USA*
**Correspondence to: William R. Shadish, School of Social Sciences, Humanities and Arts, University of California, Merced, 5200 North Lake Rd, Merced, CA 95343, USA.*
[†]*E-mail: wshadish@ucmerced.edu*
[‡]*This article is based on the first author's presidential address to the Society for Research Synthesis Methodology, York, UK, July 1, 2014.*

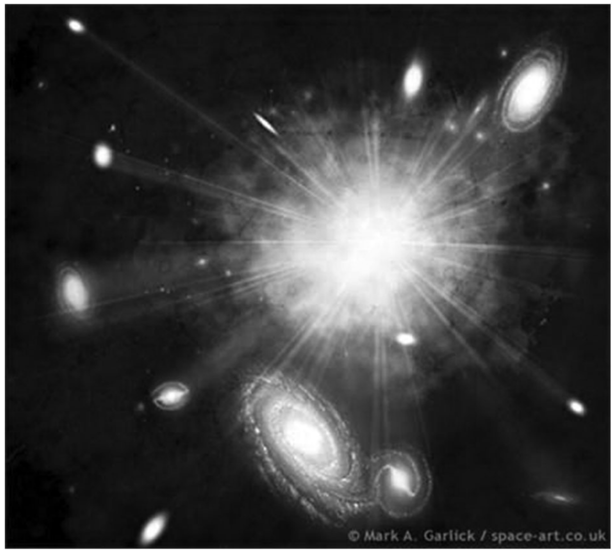

Figure 1. Artist rendition of the big bang (Reprinted by permission).

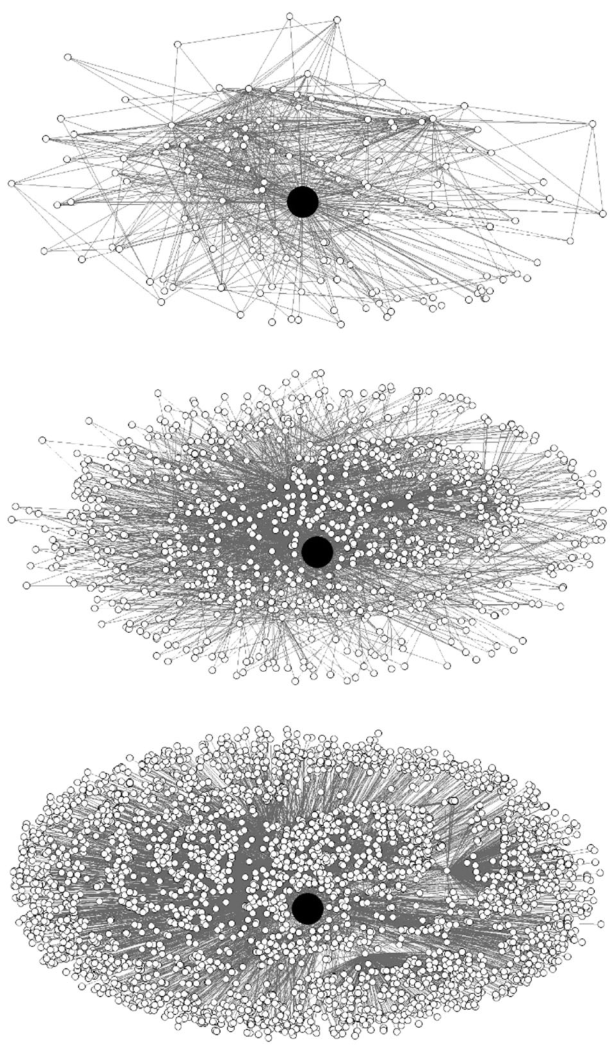

Figure 2. The meta-analytic big bang starting with only articles that cite Glass (1976) (the large black node in the center) and moving out through two levels of the citation network.

why it happened. In the former case, we will describe what happened using CNA, showing the huge impact that meta-analysis had across so many fields of scholarly work. In the latter case, we discuss why meta-analysis happened at the particular time that it did in the mid-1970s. To do so, we will rely heavily on the reflections of early founders of the field—Gene Glass, Robert Rosenthal, and Frank Schmidt—and then add our own observations to help systematize all this.

1. Citation network analysis and meta-analysis

Citations reflect, however imperfectly, the impact one scholarly work has on other scholarly works (Shadish, Tolliver, Gray, and Sen Gupta, 1995). In this paper, we survey the impact of meta-analysis by examining the networks of citations that emerged through citation patterns in the field. This article builds insight from a constrained snowball sample of citations that was generated with the seminal Glass (1976) article as the seed and two levels of data. The sampling was done using the R package CNA at a rate of 10% per level (Lecy and Beatty, 2012; Lecy and Moreda, 2013). Using this sample, we examine patterns of who cites the seed article, who cites those who cite the seed article (that is, searching two levels from the seed), and how all these works cite each other. By using this method, the citation practices of the scholarly community drive the collection of the sample, thus avoiding biases generated by peculiarities of search keywords or a scholar's specific views with regard to a domain of study.

This results in a picture of research clusters that have a common set of core citations (e.g., Harper and Peattie, 2011; Harris *et al*., 2011; Lecy *et al*., 2013). CNA uses Google Scholar rather than the Web of Science because Google Scholar relies on a more extensive index of citing sources (Noruzi, 2005), and we are interested in the broadest impacts of meta-analysis (e.g., even in the media), not just in the impact in scholarly journals indexed in the Web of Science. We used a sampling rate of 10% at each level, which implies the possibility that some important nodes in the network could be overlooked, although experience suggests that most key (highly cited) pieces will be present in the sample. This process generated a citation network of 16,863 publications in the network, which in order of frequency were from the social sciences ($N=6781$), medicine ($N=4447$), education ($N=2905$), research methods and statistics ($N=1901$), natural science ($N=809$), arts ($N=12$), humanities ($N=1$), and a handful of citations that were so incompletely described that they could not be categorized ($N=7$). Figure 2 shows the citation network of the 3601 nodes that have at least 300 citations (Figure 2 is the figure that inspired the big bang analogy), starting with only articles that cite Glass (1976) (the large black node in center) and moving out through two levels of the citation network. The size and complexity of the literature are apparent. Some contours of the research galaxy are visible with clusters apparent in the northeast and southeast quadrants; we will describe these clusters further shortly. This dense network visualization illustrates the rapid growth of the field of meta-analysis and the specialized branches of meta-analysis that formed in different disciplines and research methods.

1.1. Development of meta-analysis over time

Figure 3 illustrates the development of meta-analysis from 1975 to 2013. The networks represent the 410 publications from Figure 2 that were cited at least 1000 times (Figure 2 excludes 7 publications in natural sciences and humanities from this group, and excludes 3 publications dated prior to 1970). Each column shows time, and each row shows major areas of meta-analytic work—methods, medicine, education, and social sciences. Within a column, nodes are the same for each row except black nodes that represent articles published in the row area within the specific time frame. Nodes are not cumulative over time, so each column shows new nodes since the last time period within each area.

The figure suggests that the early meta-analytic work was mostly in methods, not surprising because the methods needed development before they could be applied topically. Medicine, in particular, was sparsely represented in the early years. Over time, work on meta-analytic methods continued to increase, meta-analytic work in both medicine and the social sciences increased, and such work in education increased but to a lesser degree.

1.2. Thinning the network

The sample of 16,863 references is too large to offer easy insight into the core structure of the field. Citations follow a power law (Newman, 2005), meaning that the vast majority or scientific publications are cited a small number of times or not at all and a few are cited thousands of times. Consequently, a small number of publications tend to comprise the skeletal structure of specific research fields. We can use this fact to judiciously filter articles and focus on the core publications in the network. Figure 4 contains only references that are cited at least 2000 times, a network of the top 113 publications in the sample (when multiple editions of a work exist, CNA uses citations to all of them but uses the date of the most recent in the label). The size of the node in the figure indicates centrality within this subgraph (node size does not represent total citation count because some

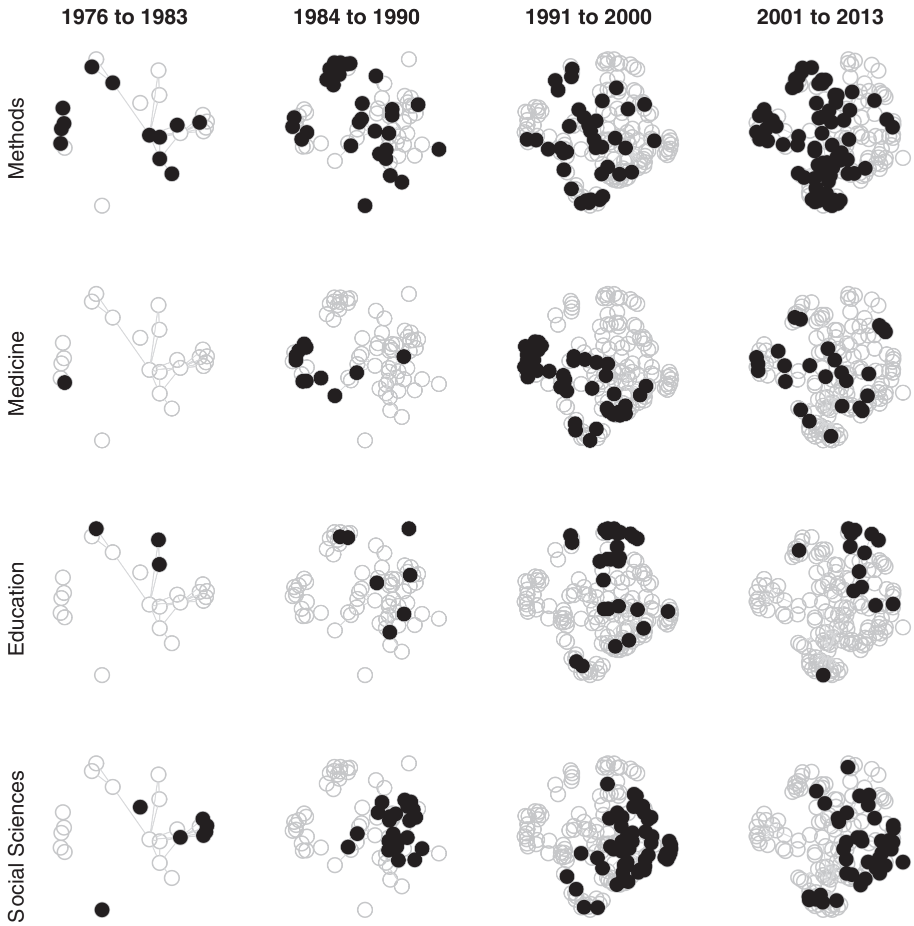

Figure 3. The development of meta-analysis over time. The networks represent the 410 publications from Figure 2 that were cited at least 1000 times. Each column shows time, and each row shows major areas of meta-analytic work. Within a column, nodes are the same for each row except black nodes that represent articles published in the row area within the specific time frame. Nodes are not cumulative over time, so each column shows new nodes since the last time period within each area. Number of nodes in the four columns are 23, 71, 161, and 155 from left to right. Number of black dots in the 16 cells from left to right by row are 12, 31, 38, 74; 1, 11, 42, 26; 3, 7, 28, 14; 7, 22, 53, 41.

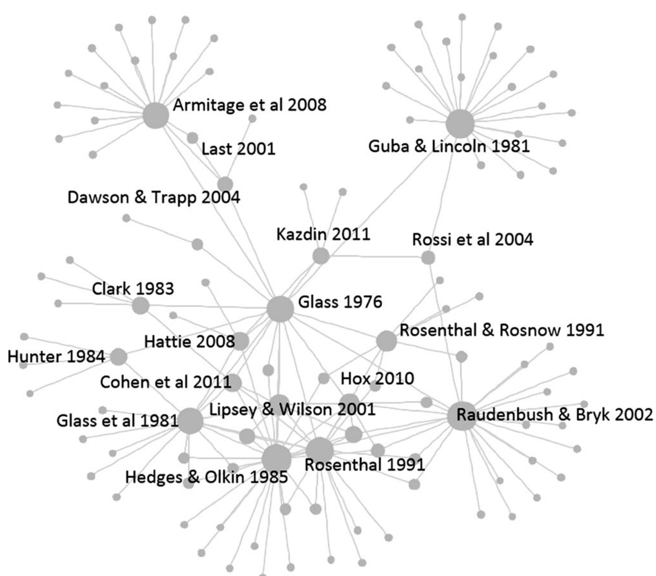

Figure 4. Thinned citation network of 113 nodes cited at least 2000 times each. The size of the node indicates its centrality in the network. Glass (1976) is at the center. Key meta-analytic works are to the south, southwest, and southeast of Glass. Three areas with mutual citations to meta-analysis in the network are multilevel modeling (southeast), medical methods and statistics (northwest), and qualitative methods (northeast).

highly cited works may not have many ties within this particular network). Glass (1976) is roughly in the middle of the graph. To the southwest of Glass (1976) are citations that the network analysis suggests are the meta-analytic canon, that is, the body of works that represents the main reference points for the field. Note in particular that Hedges and Olkin (1985) is the most central work in the subgraph (indicated by having the largest node). Other central works are Rosenthal (1991), Glass, McGaw, and Smith (1981), Lipsey and Wilson (2001), and Rosenthal and Rosnow (1991). Most of the smaller nodes to the south and west of the canon are applications of meta-analysis.

The meta-analytic canon exists within a larger scientific network in Figure 4 with important works that span meta-analysis and other fields. To the southeast of Glass is a set of nodes on multilevel modeling, with the highest centrality nodes being Raudenbush and Bryk (2002), Hox (2010), Snijders and Bosker (2011), and Goldstein (2011). The connection between multilevel modeling and meta-analysis is the strongest in the larger scientific network in Figure 4, with many co-citations. This is not surprising given that meta-analysis can be formulated as a multilevel modeling problem.

In the northwest quadrant of Figure 4 is a cluster of nodes from methods and statistics in medicine, reflecting the widespread use of meta-analysis in medical research. This is anchored by Armitage, Berry, and Matthews (2008), Dawson and Trapp (2004), and Last (2001). This probably reflects the fact that the introduction of meta-analysis to medicine had a profound impact on research methods in that field. In the northeast quadrant of Figure 4 is a cluster of nodes on qualitative research methods anchored by Guba and Lincoln (1981). The ties between meta-analysis and qualitative research are scant, with Guba and Lincoln (1981) using meta-analysis as one of the several foils for why they prefer qualitative over quantitative methods. Finally, close to Glass (1976) are central citations connecting meta-analysis to program evaluation (Rossi, Lipsey, and Freeman, 2004) and single-case design (Kazdin, 2011; often called N-of-1 studies in medicine). The latter are methods texts in areas where meta-analysis has found wide application.

1.3. The systematic review network

Figure 5 shows the citation network of the 469 references that represent systematic review methodology and statistics in meta-analysis. We coded these citations into various subfields. The upper part of the figure shows the central systematic review citations in the social sciences and education. In addition to the previously cited meta-analytic canon in Figure 4, central citations include Borenstein, Hedges, Higgins, and Rothstein (2011), Hedges (1981), and Light and Pillemer (1984). With a few exceptions (e.g., Hedges, 1981), all these citations are overviews of the field. The lower part of the figure contains the central meta-analytic citations from the medical and statistics literatures, including Higgins, Thompson, Deeks, and Altman (2003) on measuring heterogeneity, Egger, Davey Smith, Schneider, and Minder (1997) on publication bias, Stroup *et al*. (2000) on reporting standards for meta-analysis of observational studies in epidemiology, Moher *et al*. (1999) on the quality of reporting of meta-analyses standards for reporting meta-analyses of randomized trials, Liberati *et al*. (2009) on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) reporting standards for meta-analysis, Duval and Tweedie (2000) on the trim-and-fill method for publication bias, and Sacks, Berrier, Reitman, Ancona-Berk, and Chalmers (1987) that was central to the introduction of meta-analysis in medicine.

1.4. Highly cited works in systematic reviews

Table 1 shows the 14 most highly cited works in the systematic review network. These include many of the central works discussed in the previous paragraphs. However, the first two items in the table help illustrate the difference

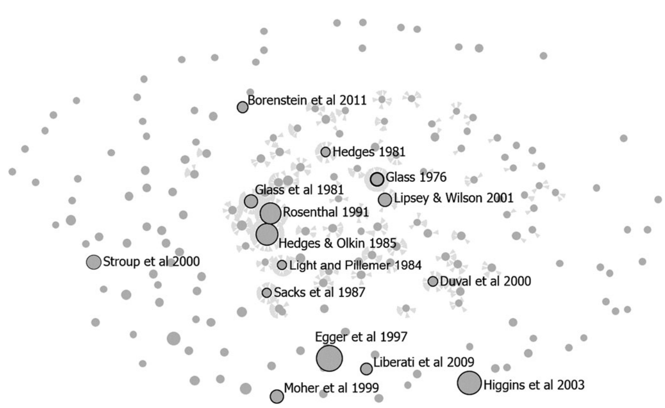

Figure 5. Citation network for 469 nodes about systematic review methods and statistics. Size of node indicates centrality in the network. The upper part of the network contains central works in the social sciences, and the lower part contains central works coming more from medicine and statistics.

Table 1. Highly cited works in the systematic review network.

Authors	Citations	Title (first few words for long titles)
DerSimonian and Laird, 1986	12,201	Meta-analysis in clinical trials
Baeza-Yates *et al*., 1999	11,926	Modern information retrieval
Egger *et al*., 1997	10,027	Bias in meta-analysis
Higgins *et al*., 2003	8867	Measuring inconsistency in meta-analysis
Hedges and Olkin, 1985	7952	Statistical methods for meta-analysis
Moher *et al*., 2009	5101	Preferred reporting items for systematic reviews
Schulz and Chalmers *et al*., 1995	4795	Empirical evidence of bias
Rosenthal, 1991	4536	Meta-analytic procedures for social research
Stroup *et al*., 2000	4367	Meta-analysis of observational studies in epidemiology
Glass, McGaw, and Smith, 1981	3693	Meta-analysis in social research
Glass, 1976	3656	Primary, secondary, and meta-analysis of research
Moher *et al*., 1999	3626	Improving the quality of reports of meta-analyses
Lipsey and Wilson, 2001	3484	Practical meta-analysis
Begg and Mazumdar, 1994	3351	Operating characteristics of a rank correlation test

between centrality in a graph and citation count in the Google Scholar database. The mostly highly cited work in the systematic review network is DerSimonian and Laird (1986), which describes a widely used statistical method; the second most highly cited is Baeza-Yates and Ribeiro-Neto (1999), a very general text on information retrieval methods that is largely unknown in the systematic review literature. Neither is central to the systematic review network in Figure 4 because most works in the network do not cite them, and they do not cite other key works in the network. They are only loosely integrated into the sample.

It is somewhat puzzling why the articles in the meta-analysis sample do not cite DerSimonian and Laird (1986) more often. DerSimonian and Laird (1986) presented a widely used random effects model and associated example of how to integrate medical clinical trials. More than 12,000 works that cite this piece are not heavily represented in the current sample. Perhaps, the snowball sample did not capture a full picture of the medical literature, although subsequent results tend to suggest that this is not the case. Perhaps, the fact that the DerSimonian and Laird model used only dichotomous outcomes as examples (e.g., survival and incidence of a condition) limited its perceived usefulness to similar medical studies using dichotomous outcomes and less so to the social sciences where dichotomous outcomes are more the exception than the rule. Readers may perhaps provide other suggestions.

1.5. Common topics in the systematic review network

We coded the 469 references in the systematic review network for the general topic that they addressed (Figure 6). From most to least frequently occurring, the topics were general systematic review theory, effect size, statistics, the empirical program, overviews, publication bias, guidelines and reporting, quality, qualitative reviews, heterogeneity, computer programs, retrieval, and network meta-analysis. Of course, the largest of these topics could have been further subdivided. For example, we could have coded separately such areas as cumulative meta-analysis, diagnostic and screening meta-analysis, and individual patient data meta-analysis. The data could also be coded for substantive topics such as psychotherapy and cardiovascular disease. In that sense, then, the raw frequencies in Figure 6 are not very informative because they would change with additional codes.

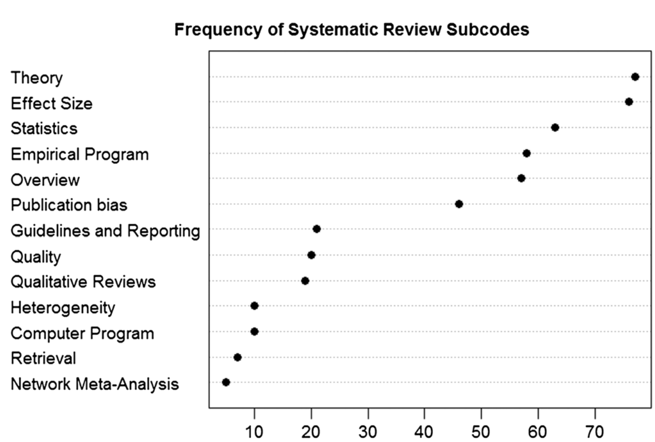

Figure 6. Frequency of systematic review topics.

More informative is Table 2, which shows the five most highly cited references for each topic. These works provide a useful reading list for anyone wanting to review past and current issues in systematic reviews. Most of the topics in Table 2 are self-explanatory, but two are worth comment. Studies categorized as the empirical program use meta-analytic methods to study statistical and methodological issues with empirical data from the pertinent literature. Examples from Table 2 include Schulz, Chalmers, Hayes, and Altman's (1995) classic empirical study of bias and quality in randomized controlled trials and Peterson's (1994) meta-analysis of Cronbach's alpha internal consistency reliability coefficient. Scores of these empirical studies exist in the present sample alone (which itself only used a 10% sampling rate), and they are an often-overlooked contribution of meta-analysis. Also of note is the literature on methods for synthesizing results from studies that used qualitative methods (e.g., Noblit and Hare, 1988). While some qualitative researchers espouse the belief that each case is unique and cannot be integrated with any other case, a large number of those researchers recognize both the scientific value and the policy relevance of finding credible ways to synthesize their work.

2. Why, when, and how did the meta-analytic big bang happen?

For good reasons, Glass (1976) traditionally obtains credit for the start of meta-analysis. Contemporaneously, however, two other psychologists were both doing similar work—Robert Rosenthal and Frank Schmidt. Glass tends to obtain special credit because he coined the phrase meta-analysis and he was the first to outline its key characteristics. Yet it is instructive to review the work of all three of these scholars to clarify how it came to pass that in the late 1970s all of them were working on a similar problem and all of them proposed solutions with a good deal of overlap. In this section, we will briefly review what Glass, Rosenthal, and Schmidt each proposed, discuss some of the conditions that facilitated their work, and place those matters into a larger context from the philosophy and social psychology of science so as to make sense of it as a coherent whole. We rely extensively on the reflective articles by Glass, Rosenthal, and Schmidt that appear in this special issue. Readers may find it helpful to examine those articles first.

2.1. The scholars

Gene V. Glass received his bachelor's degree in 1962 from the University of Nebraska, Lincoln, with a joint major in mathematics and German. He worked as a research assistant for education faculty member Robert Stake during his last undergraduate year. He then entered doctoral training in statistics, measurement, and experimental design in the School of Education at the University of Wisconsin, Madison, receiving his PhD in 1965. His primary motivating interest regarding meta-analysis was to find a way to synthesize hundreds of studies on the effectiveness of psychotherapy in a scientifically rigorous way.

His presidential address to the American Educational Research Association in 1975, later to become the seminal publication in meta-analysis (Glass, 1976), was entitled 'Primary, secondary, and meta-analysis of research'. It defined meta-analysis as the analysis of summary statistics from studies rather than the analysis of raw data. He proposed and subsequently implemented many of the defining characteristics of meta-analysis, seeing study-level data as the unit of analysis, using a variety of effect sizes (e.g., *d*, *r*, probits) appropriate to different outcome metrics, taking averages (albeit unweighted) of effect sizes, and coding and analyzing potential moderating study characteristics related to average effect sizes across studies rather than within studies.

Frank L. Schmidt received his bachelor's degree in psychology from Bellarmine College in Louisville, Kentucky, in 1966, and he then earned his doctorate in industrial psychology from Purdue University in 1970. His primary motivating interest was to understand whether apparently disparate test validity correlation coefficients truly did reflect unique and specific aspects of the testing situation or whether that heterogeneity was artifactual and masked an underlying generalizability. He initially approached the problem by converting correlation coefficients into *z*-values and then averaging them (Schmidt, Berner, and Hunter, 1973). His seminal work regarding meta-analysis was Schmidt and Hunter (1977), in which he coined the phrase validity generalization to refer to the procedures he developed for integrating study statistics, although he eventually also used the term meta-analysis to refer to those procedures (Hunter, Schmidt, and Jackson, 1982).

Robert Rosenthal earned his bachelor's degree in 1953 from the University of California, Los Angeles, in Psychology and then received his doctorate in clinical psychology in 1956 from the same school. Starting around 1960, Rosenthal began both publishing and accumulating studies done by other scholars of what today is called interpersonal expectancy effects. These are studies of how people's expectations can influence other people, for example, in teacher–student relations, researcher–subject interactions in psychology laboratories, doctor–patient relations, or manager–employee situations.

Facing a very large and accumulating literature, he became interested in finding quantitative ways to synthesize the results over all these studies. Influenced in part by his Harvard colleague Fred Mosteller (Mosteller

Table 2. Highly cited works on various systematic review topics.

Authors	Citations	Title (first few words for long titles)
Theory		
Glass, 1976	3656	Primary, secondary, and meta-analysis of research
Glass, 1977	864	Integrating findings: the meta-analysis of research
Mulrow, 1994	815	Rationale for systematic reviews
Oxman and Guyatt, 1988	485	Guidelines for reading systematic reviews
Schmidt, 1992	470	What do data really mean?
Effect size		
Parmar *et al.*, 1998	1237	Extracting summary statistics to perform meta-analysis…
Rosenthal *et al.*, 2000	925	Contrasts and effect sizes in behavioral research…
Rosenthal and Rubin, 1982a	891	A simple, general purpose display of magnitude…
Dunlap *et al.*, 1996	826	meta-analysis of experiments with matched groups…
Nakagawa and Cuthill, 2007	714	Effect size, confidence interval and statistical…
Statistics		
DerSimonian and Laird, 1986	12,201	Meta-analysis in clinical trials
Hedges and Olkin, 1985	7952	Statistical methods for meta-analysis
Hedges, 1981	1066	Distribution theory for Glass's estimator…
Thompson and Higgins, 2002	828	How should meta-regression analyses be…
Hedges and Vevea, 1998	800	Fixed and random effects models in meta-analysis
*Empirical program**		
Schulz *et al.*, 1995	4759	Empirical evidence of bias
Easterbrook *et al.*, 1991	1719	Publication bias in clinical research
Antman *et al.*, 1992	1428	A comparison of results of meta-analyses…
Peterson, 1994	1382	A meta-analysis of Cronbach's coefficient alpha
Wanous *et al.*, 1997	1329	Overall job satisfaction: how good are single-item…
Overviews		
Rosenthal, 1991	4536	Meta-analytic procedures for social research
Stroup *et al.*, 2000	4367	Meta-analysis of observational studies in epidemiology
Glass *et al.*, 1981	3693	Meta-analysis in social research
Lipsey and Wilson, 2001	3484	Practical meta-analysis
Borenstein *et al*, 2011	1773	Introduction to meta-analysis
Publication bias		
Egger *et al.*, 1997	10,027	Bias in meta-analysis detected by a simple graphical test
Begg and Mazumdar, 1994	3351	Operating characteristics of a rank correlation test…
Rosenthal, 1979	2587	The file drawer problem and tolerance for null results
Duval and Tweedie, 2000	1352	Trim and fill: a simple funnel plot-based method…
Dickersin, 1990	857	The existence of publication bias and risk factors…
Guidelines for conduct and reporting		
Moher *et al.*, 2009	5101	Preferred reporting items for systematic reviews…
Moher *et al.*, 1999	3626	Improving the quality of reports of meta-analyses…
Liberati *et al.*, 2009	2777	The PRISMA statement for reporting systematic…
Irwig *et al.*, 1994	724	Guidelines for meta-analyses of evaluating diagnostic…
Rosenthal, 1995	558	Writing meta-analytic reviews
Assessing quality		
Juni *et al.*, 1999	1241	The hazards of scoring the quality of clinical trials…
Chalmers *et al.*, 1981	1217	A method for assessing the quality of a randomized…
Verhagen *et al.*, 1998	983	The Delphi list: a criteria list for quality assessment…
Moher *et al.*, 1995	947	Assessing the quality of randomized controlled trials…
Oxman and Guyatt, 1991	599	Validation of an index of the quality of review articles
Qualitative reviewing		
Noblit and Hare, 1988	1171	Meta-ethnography: synthesizing qualitative studies
Popay *et al.*, 1998	580	Rationale and standard for the systematic review of…
Sandelowski *et al.*, 1997	431	Focus on qualitative methods: qualitative metasyn…
Paterson and Canam, 2001	391	Meta-study of qualitative health research: a practical…
Larsson, 1993	322	Case survey methodology: qualitative analysis of…
Heterogeneity		
Higgins *et al.*, 2003	8867	Measuring inconsistency in meta-analyses
Deeks *et al.*, 2001	1019	Statistical methods for examining heterogeneity…
Huedo-Medina *et al.*, 2006	552	Assessing heterogeneity in meta-analysis: *Q*-statistics…

(*Continues*)

Table 2. (Continued)

Authors	Citations	Title (first few words for long titles)
Hardy and Thompson, 1998	340	Detecting and describing heterogeneity in meta-analy…
Ioannidis *et al*., 2007	253	Uncertainty in heterogeneity estimates in meta-analysis…
Computer programs		
Viechtbauer, 2010	386	Conducting meta-analyses in R with the metafor…
Mullen, 2013	341	Advanced basic meta-analysis: version 1.10
Bax *et al*., 2006	315	Development and validation of MIX: comprehensive…
Mullen and Rosenthal, 1985	155	Basic meta-analysis: procedures and programs
Schwarzer, 1989	147	Meta-analysis programs
Retrieval		
Baeza-Yates *et al*., 1999	11,926	Modern information retrieval
Dickersin *et al*., 1994	1686	Identifying relevant studies for systematic reviews
Counsell, 1997	320	Formulating questions and locating primary studies…
Hopewell *et al*., 2002	103	A comparison of handsearching versus MEDLINE…
Hahn *et al*., 2002	78	MEDSYNDICATE: a natural language system for…
Network meta-analysis		
Lu and Ades, 2004	515	Combination of direct and indirect evidence…
Salanti *et al*., 2008	207	Evaluation of networks of randomized trials
Lu and Ades, 2006	166	Assessing evidence inconsistency in mixed treatment…
Dias *et al*., 2010	129	Checking consistency in mixed treatment comparison…
Hoaglin *et al*., 2011[†]	86	Conducting indirect-treatment –comparison…

*Studies that use meta-analytic methods to study statistical and methodological issues with empirical data from the pertinent literature.

[†]This is Part 2 of two-part piece. The first part is Jansen, Fleurence, Devine, Itzler, Barrett, Hawkins, Lee, Boersma, Annemans, and Cappelleri (2011).

and Bush, 1954), he calculated both *p*-values and effect sizes on each finding, initially preferring to combine *p*-values (e.g., Rosenthal, 1966), later combining effect sizes as well (Rosenthal, 1969, 1978, 1984; Rosenthal and Rosnow, 1975; Rosenthal and Rubin, 1978a, 1978b, 1989). Unlike Glass or Schmidt, he did not much focus on labeling his procedures, although he once called it quantitative assessment of research domains (Rosenthal, 1980). After reading Glass's work, he also used the term meta-analysis (e.g., Rosenthal and Rubin, 1982b; Rosenthal, 1984).

2.2. *The conditions that helped create meta-analysis*

As persons, Glass, Rosenthal, and Schmidt are exceptionally bright and talented scholars, and a purely psychological approach to understanding their contributions might focus on their personal characteristics (Gholson, Shadish, Neimeyer and Houts, 1989). Here, we take a more social approach, thinking of them as nodes in social networks that existed for years before they formulated their mature ideas about meta-analysis. In those networks, they occupied a niche where sometimes common and sometimes idiosyncratic conditions existed or converged. As a result of this diversity of influence, each produced a somewhat different approach to the common problem. In this section, we briefly discuss what some of those conditions were that influenced their eventual theory of meta-analysis.

2.2.1. A widely perceived and compelling problem. Each of these scholars was trying to solve a problem that was widely perceived to be compelling in the disciplines in which they worked—how to deal with the enormous proliferation of research after the end of World War II, especially in psychology and education in the United States, that made traditional review methods increasingly ineffective. Glass wanted to draw conclusions from over 500 studies of the effects of psychotherapy. Schmidt had to do the same for over 500 studies on the validity of psychological tests. Rosenthal had collected over 300 studies of interpersonal expectancy effects and wanted to synthesize conclusions from them.

The sheer number of studies meant that scholars could not rely on personal memory of the characteristics and outcomes of every study to reliably integrate them and relate them to each other. The usual cognitive biases that applied to all human beings made it increasingly unclear that human judgment could do better than statistical methods to integrate scientific results (Meehl, 1954). The flaws of relying on traditional null-hypothesis significance testing were becoming apparent (Meehl, 1978), as were the flaws of relying on the traditionally used vote-counting methods for summarizing results (Bush and Wang, 2009; Hedges and Olkin, 1980; Light and Smith, 1971). In this context, many scholars recognized that they needed some alternative to traditional literature review methods.

2.2.2. The decade of the 1970s was the right time in history. Karl Pearson (1904a, 1904b) tested the effects of a new typhoid vaccine by averaging correlation coefficients over a handful of studies. If he had called it meta-analysis, would that work have sparked the same big bang that Glass (1976) did? Arguably not, because it was not the right time in history given that scientists of that era were not faced with a problem of integrating hundreds of studies (although perhaps in physics they were; Hedges, 1987) and that the prior statistical development of effect sizes and methods for weighting studies using inverse variance methods did not yet exist. Indeed, statistical development of quantitative methods for literature reviews was not a respected specialty in mainstream statistics until the 1980s, much less in 1904.

Ideas about quantitative integration were merely hinted at in the early part of the 20th century, for example, when Sir Ronald Fisher said 'It sometimes happens that although few can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are lower than would have been obtained by chance' (1932, p. 99). The 1970s provided much more fertile ground for the development of a meta-analytic big bang. Much of the quantitative development necessary to meta-analysis had been done in statistics, scientists were faced with too many studies to handle with narrative methods, and a few statisticians like Larry Hedges took risks with their careers to make quantitative literature review methods part of mainstream statistics.

2.2.3. Controversy is motivating. Glass, Rosenthal, and Schmidt each faced substantive research problems of great import with sometimes powerful opposition to their findings. Glass engaged with a contingent of researchers led by the famous psychologist Hans Eysenck (1965) who concluded that psychotherapy had no more effect than a placebo and who responded to the Smith and Glass (1977) meta-analysis of psychotherapy studies by calling it an exercise in mega-silliness (Eysenck, 1978). Schmidt, whose findings suggested that validity coefficients were not situation-specific but rather were generalizable, thus challenged the members of a lucrative consulting industry of industrial/organizational psychologists whose livelihood depended in part on convincing clients to pay them to produce individualized validity coefficients for their specific context. Rosenthal's findings called into question the notion of how objective results from psychological research could be, which was a nontrivial challenge in a field that had hitherto prided itself on the objectivity of its methods (Rosenthal and Rubin, 1978b).

Of these three founders, only Glass explicitly cites the controversy with Eysenck as a central motivation to his work. However, it seems plausible that the controversial attention that Schmidt's and Rosenthal's work received motivated them to pursue that work with ever more rigor and thoroughness in order to ensure it was as beyond reproach as they could muster.

2.2.4. Intellectual risk taking. Glass, Schmidt, and Rosenthal were each trying to create a set of methods for research synthesis that no one had done before. Glass, for example, describes the trepidation he felt when first presenting the initial formulation of meta-analysis to his colleagues in the field of education. The risk of failure and its consequences for career development had to be salient to them. Yet such risk taking can have payoffs when such researchers pursue novel if imperfect answers to a problem. Two related quotes that capture this risk taking are from Joseph Lau, a physician and medical researcher who was central to the spread of meta-analysis in medicine, especially in the 1980s: 'In retrospect, I see [my lack of formal statistical training] … as a plus because this ignorance allowed my imagination to wander unencumbered' (Cappelleri and Ingerick, 2014, p. 3) and 'because some of the naive ideas of mine often led to somewhere!' (Cappelleri and Ingerick, 2014, p.13). With great risk can sometimes come great reward.

2.2.5. Luck favors the prepared mind. Glass, Rosenthal, and Schmidt did not come to their solutions at one time in one fell swoop. They had each been thinking for years, often as far back as graduate school, about issues related to what would eventually become meta-analysis. Glass was exposed to the work of William Hays, who introduced the notion of effect sizes to a wide psychological audience with his graduate-level psychology statistics textbook (Hays, 1963). Glass then asked his graduate school statistics teacher, George Box (son-in-law of Sir Ronald Fisher), what he thought of effect sizes, and Box's less than fully positive evaluation led Glass to study the topic further (Glass and Hakstian, 1969). Glass had read the work of Astin and Ross (1960) and Underwood (1957) who treated studies as the unit of analysis, as he would eventually propose to do himself in meta-analysis, and he had a long standing interest in the nature of generalization, a central intellectual underpinning of the notion that meta-analysis might show that the study outcomes reflect some common underlying features (Bracht and Glass, 1968).

In his own reflections, Frank Schmidt says that his graduate school instructors exposed him to quite divergent opinions about why test validity coefficients varied quite widely from each other in magnitude. One teacher in industrial/organizational psychology told Schmidt that these coefficients were heterogeneous because they really were situation-specific—each factory or office in each town with different employees all resulted in different test validities because test validities were situation-specific. But one instructor who specialized in psychometrics, Hugh Brogden, suggested that observed variability in coefficients was due to sampling error or measurement error. A fellow student in that classroom recently relayed anecdotally to the first author that Schmidt asked Brogden whether he should believe Brogden or the other instructor. Brogden is said to have replied that Schmidt had paid his tuition as a graduate student and that having heard two faculty members' opinions, Schmidt should now make up his own mind. Schmidt did, and his theory, more than any others, now takes measurement error into account.

Rosenthal had a long-standing interest in what it meant to say that one study replicated another (e.g., Rosenthal, 1966). This interest stemmed in part from colleagues in psychology who challenged his interpersonal expectancy findings by showing that many of them failed to replicate by conventional standards of null-hypothesis significance testing. This prompted Rosenthal to ponder whether replication meant both studies rejected the null hypothesis, or had similar-sized *p*-values, or something else. His eventual choice to combine *p*-values over studies came partly out of this more general background.

2.2.6. The effects of their particular social networks. Each of these scholars existed in a particular social network that influenced the development of their meta-analytic theories in at least three ways. One is that the networks exposed them to key technical ideas already discussed. Examples include Schmidt's exposure to Brogden and the possible roles of sampling and measurement error, Glass as taking graduate statistics from George Box leading to discussion of the role of effect sizes, and Rosenthal as having Fred Mosteller as a colleague to encourage Rosenthal to combine *p*-values.

A second role of the network was to provide moral support and encouragement to them at an early stage in their career when they were taking great risks. Glass explicitly cites Paul Meehl's encouragement as important, and Schmidt does the same for the encouragement he received from Lee Cronbach and Anne Anastasi. A third role of the network is that some people served to connect scholars not just to each other and to colleagues doing similar work but also to some larger implications of their ideas for psychology, education, and other fields as well. Glass and Schmidt both cite Lee Cronbach at Stanford in this regard, Cronbach being one of the giants of social science methodology in both education and psychology. In medicine, perhaps, Thomas Chalmers served a similar role in facilitating the spread of meta-analytic ideas and methods.

2.3. Putting it all together: context, trial-and-error, and luck

Out of this context, Glass, Rosenthal, and Schmidt each created a solution to the problem of quantitative syntheses of very large numbers of studies. Yet they did not create identical solutions. Within themselves over time, and between them, they tried and often discarded some elements of a solution and then tried others. Rosenthal's early work required some access to the raw data to proceed, and he quickly realized the limitations that posed for comprehensive reviews. So, he moved to combining study-level statistics that were reported in publications. Schmidt combined effect sizes, but they were mostly correlation coefficients that formed the bulk of validity coefficients. That left Glass to be the first to propose using different kinds of standardized effect sizes for different kinds of outcome metrics. Only Schmidt and Hunter (1977; Schmidt, Hunter, Pearlman, and Shane, 1979) weighted effect sizes at first, weighting by sample size. Rosenthal sometimes used a formal test for heterogeneity from Snedecor (1956) that he learned from Harvard colleague William Cochran (see also Rosenthal, 1984, p. 77), and Schmidt and colleagues used rules that are relevant to distinguishing error variability from total variability, closely related to formal statistical heterogeneity.

The general solution that came to be most widely accepted as the framework for meta-analysis included combining effect sizes, weighting larger studies more heavily, testing for heterogeneity, coding study characteristics, and then analyzing studies as the unit of analysis, embedding all these in a larger systematic review process that addressed not just statistics but also the research steps preceding and following analysis, all united under the label of meta-analysis. The power of a good label cannot be underestimated. Glass proposed the name that came to be widely accepted, meta-analysis. He attributes the inspiration to a colleague, philosopher Michael Scriven, who had just coined the phrase meta-evaluation as the evaluation of evaluation (Scriven, 1969). The title of Glass's (1976) paper, 'Primary, Secondary, and Meta-analysis of Research', also had attractive rhetorical properties of consonance (the recurrence of similar sounds in close proximity) and parallelism (similarity of structure in a series of related words) that made it memorable, and that implied a solution as general as all statistical analyses no matter what the substantive field is. One is reminded of psychologist Donald Campbell's gift for coining memorable phrases such as internal and external validity, and also quasi-experiments (Campbell and Stanley, 1963), phrases that have become so much a part of the scientific lexicon that their origins are often forgotten. By contrast, no other label for integrating studies was as successful.

None of these three theorists proposed all these features of modern meta-analysis. Glass probably came closest, but even Glass acknowledges how things could have been different, saying about Rosenthal: 'If Bob had just gone a little further in quantifying study characteristics and subjecting the whole business to regression analyses and what-not, and then thinking up a snappy name, it would be his name that came up every time the subject is research integration'.

So, the 'invention' of meta-analysis did not stop in the 1970s. After these theorists outlined a basic framework for meta-analysis, that framework has subsequently been elaborated over time both by these scholars themselves and by all the many scholars who came after, including topics such as publication bias, effect size calculation, network meta-analysis, retrieval methods, and the development of computer programs for doing meta-analytic work, as we saw in Table 2. If one can think of the development of meta-analysis as a scientific revolution in

Kuhnian (Kuhn, 1962) terms, then this subsequent development is what Kuhn (1962) called the prolonged normal science puzzle solving phase that constitutes the bulk of scientific work following such a revolution.

2.4. More formal conceptualizations of the development of meta-analysis

This description of how and why meta-analysis began when it did fits two well-known theories from the social psychology of science and the philosophy of science.

2.4.1. Simonton's chance configuration theory of creativity and discovery. The development of meta-analysis was clearly a creative act in response to a problem. Social psychologist Dean Simonton has spent much of his career studying the nature and causes of creative scientific achievement (e.g., Simonton, 2004). Simonton assumes that the scientist has already identified an important problem, and so Simonton focuses only on the creative problem solving process. We say more about this in the next section on evolutionary epistemology; in that epistemology, Simonton can be conceptualized as focusing only on the blind variation and selective retention part. In his chance configuration theory of scientific creativity (Simonton, 1988), Simonton says that creative scientific accomplishments like meta-analysis come together substantially by chance in three steps into a successful combination of elements and conditions. First, the researcher produces random permutations (combinations) of elements in an effort to solve a problem. Second, one permutation of these elements produces a stable configuration in the sense that the researcher believes that the elements are satisfactory to propose as a solution to the scientific community. Third, that configuration is then communicated through the usual scientific channels, and the community provides feedback and criticism about the adequacy of the solution for solving the problem at issue. Eventually, a refined configuration of the theory may win widespread support and is retained and transmitted by the community through books, journals, professional societies, and all the other methods available for the transmission of knowledge.

We see this process at play in the creation of meta-analysis. Glass, Rosenthal, and Schmidt each explored various combinations of elements. For example, Rosenthal began by exploring how to combine *p*-values and then *z*-transformations. He communicated his proposed solutions to the scientific community in journals and books. The scientific feedback process was critical of combining *p*-values, not rejecting it but pointing out its limitations. Rosenthal then changed his theory to combine effect sizes in addition to *p*-values.

Similarly, Schmidt's configuration included the validity generalization label, but the scientific community did not adopt that label outside the narrow test validation context in which it originated. So, Schmidt's theory eventually used the more widely accepted term meta-analysis for the applications of his method beyond the validity generalization context. Glass did not propose weighting effect sizes by a function of sample size and did not deal with the issue of independence of effect sizes within studies. Those lacunae were criticized, and the eventually accepted theory of meta-analysis dealt better with those issues.

Simonton's theory has a good deal of support in both empirical studies and computer simulations of the role that chance plays in creativity (e.g., Hong, 2013; Wuketits, 2001). Yet many scientists find it difficult to accept that chance plays a substantial role in their creative work. Simonton himself has explored other labels such as Darwinian, sightedness, and blind variation and selective retention (Simonton, 2010, 2011, 2013), and scientists seem to react better to terms like serendipity than to the word chance (Roberts, 1989). Simonton also acknowledges that much scientific work is intentionally more mundane than creative so that a search for really novel variations does not often come into play. For truly creative achievements like the creation of meta-analysis, however, the scientist has to engage in a search outside the usual space where solutions are typically found. In that context, Simonton suggests that chance plays a large role in ultimate creative success.

To illustrate the chanciness of the process that led to meta-analysis, consider what would have happened if Glass had never been exposed to Scriven's term meta-evaluation in the early 1970s (would Glass have created the term meta-analysis?), if Glass had not gone to the University of Wisconsin and taken statistics from Box after reading Hays (would Glass have pursued research on effect sizes at all?), if Glass had not seen the small amount of prior research that used studies as the unit of analysis (would Glass have had the idea of analyzing study statistics?), or if Glass had done his work in 1904 when Pearson did his work (would the contemporary scientific community see meta-analysis as solving a problem that was important to them at that time?).

Glass certainly did not pursue these elements at the times he did (e.g., in graduate school) in an intentional search for a solution to the problems giving rise to meta-analysis. Yet these elements eventually came together in a truly creative configuration one decade later when he actually was searching for a solution, and that configuration would not likely have occurred if it were not for all those prior accidents of history that placed Glass in the right place at the right time with the right set of elements to solve the problem. This is not, of course, to disparage the potential role of such factors as individual intelligence or hard work in creative achievement. Yet high intelligence and hard work abound among scientists who often never do truly creative work. It takes having that lucky combination of elements to develop a truly creative solution.

One is also reminded of Malcolm Gladwell's thesis in his book *Outliers* (Gladwell, 2008). Gladwell notes that society tends to attribute success to individual intelligence and ambition. Instead, he argues that factors such as geography, birth date, culture, family, access to resources, and the idiosyncrasies of the contacts made on life paths. He talks about Paul Allen and Bill Gates being high school students in a part of the country where the right combination of intellectual freedom and computational resources available to students allowed them to forge the Microsoft revolution and Disk Operating System. So, many of these factors are not much under an individual's control but occur by chance.

2.4.2. Campbell's evolutionary epistemology. At an even more general level, Simonton's theory cites as its inspiration the evolutionary epistemology of psychologist Donald Campbell (1960, 1974a, 1974b; see also Popper, 1972), particularly what Campbell referred to as blind variation and selective retention in both biological and epistemological evolutions. Campbell proposed that the development of knowledge in science follows a process analogous to the development of species as outlined in evolutionary biology by Charles Darwin. In evolutionary biology, random genetic mutations result in species changes. Successful mutations are those that improve the ability of the species to survive or to solve problems in its environment, and they allow the genetic mutations to be transmitted to subsequent generations during reproduction. Campbell suggested that successful knowledge develops in a similar way. Chance knowledge variations (blind variations in Simonton's terms) result in potential solutions to extant epistemological problems. Those solutions are evaluated in research, and the successful variations are then transmitted through journals, books, societies, education, and the other methods for the transmission of knowledge in science.

The evolution of meta-analysis fits this theory well. A challenging problem existed in the scientific community, that is, the size of the post-World War II research literature and the growing perception that traditional review methods were inadequate to the task. Researchers like Glass, Schmidt, and Rosenthal tried variations in new methods for addressing the problem; methods we have previously argued were substantially based on chance, luck, or serendipity. Evaluation of their proposed solutions by scientific colleagues was substantially (but not always entirely) positive, and these meta-analysts were able to respond compellingly to criticisms of their proposals. The resulting methods were rapidly disseminated through the usual scientific channels, becoming part of the accepted canon of science.

2.5. Why did a more general approach to the problem not arise first in medicine?

A natural question is why the development of a general theory of meta-analysis first occurred in the social sciences and not in medicine. After all, medical and related sciences have arguably dominated the last 20–30 years of statistical and methodological development in meta-analysis and have pioneered the use of meta-analysis in evidence-based practice through institutions like the Cochrane Collaboration. Those sciences also have substantially more financial resources with which to support research. So, why was medicine not the first to develop meta-analysis?

One answer may be that the epistemological problem in medicine was not as challenging as in the social sciences, so little pressure existed in medicine to develop the kind of broad solutions that emerged in the social sciences. Medical research focused on narrow questions and preferred homogenous data emerging from studies that were conducted reasonably similarly. At the limit, medical researchers focused on single studies of proprietary treatments being prepared for Food and Drug Administration review, where no other study might exist. The number of studies was also smaller because of the preference given to randomized experiments in medicine, compared with the social sciences that often included both randomized and nonrandomized experiments in the same meta-analysis. For such reasons, then, relatively few studies existed on each individual narrow questions, compared with the hundreds of studies that social scientists wanted to synthesize to answer their questions. Further, because medical research relies so heavily on dichotomous outcomes, some existing statistics like the Mantel–Haenszel method (Mantel and Haenszel, 1959) could satisfactorily summarize results over studies. So, medicine did not need a more general method capable of asking very general questions with heterogeneous data over hundreds of studies using diverse outcome metrics that require diverse effect size measures.

A good illustration is Stjernswärd's (1974, 2009) quantitative synthesis, often cited as an early example of meta-analysis in medicine. Stjernswärd's question was whether postoperative radiation therapy in early breast cancer patients increased mortality. It was not whether 'radiation works', which would be parallel to the very general question Glass asked about whether 'psychotherapy works'. Given the narrow question, the Stjernswärd study included just five controlled trials, not the more than 500 studies Glass wanted to synthesize to answer his very general question. Further, because all the outcomes were dichotomous, Stjernswärd could integrate them using the Mantel–Haenszel method; he did not need to use the wide array of effect size estimators (e.g., d, r, probit, and tobit) that Glass assembled to cope with the diverse outcome measures and metrics commonly used in the social sciences. Of course, this is not to say that the medical sciences were wrong to take this approach to integrating studies on narrow questions with dichotomous outcomes. Rather, it is merely to say that historical context created no pressure in medicine to create a more general meta-analytic theory that the social science context required.

3. Discussion

This paper is an internal (personalized) history of meta-analysis, that is, a history written by someone who has been intimately involved in meta-analysis since the early 1980s and who has historically written positively about the potential and accomplishments of that field (Shadish, 2007; Shadish, Cook, and Campbell, 2002). So, this article may be biased toward a positive slant on the field. Some scholars understandably view internal histories with skepticism because they risk underrepresenting more critical views (Danziger, 1994; Farr, 1996). For example, even today some scholars remain less sympathetic to the systematic review enterprise (e.g., Berk, 2007), and I have not represented their views much here.

Moreover, our focus on Gene Glass, Robert Rosenthal, and Frank Schmidt is not meant to minimize the contributions of other scholars whose work very quickly followed on the heels of the founders of the field. A good example is the work of Larry Hedges (e.g., 1980, 1982a, 1982b, 1982c and 1983; Hedges and Olkin, 1980, 1983a, 1983b, 1984), whose early and rapid systematization of the statistics of meta-analysis gave the field wider credibility and a firm foundation for future work. No wonder Figures 4 and 5 both suggest that Hedges and Olkin (1985) is perhaps the central work in the field. Yet Hedges and others like him (e.g., Harris Cooper, Richard Light, and Mark Lipsey) did not actually found the field of meta-analysis (nor, of course, do they claim to have done so), and this article is about the founders.

3.1. The future of the meta-analytic big bang

Glass, Rosenthal, and Schmidt leave us with much to ponder about the future of meta-analysis. Glass wonders if the very idea of using study statistics as the unit of scientific work will survive in an era in which raw data itself can be stored and then accessed by others for secondary analysis. Glass and Schmidt both see issues in the relative roles of main effects versus interactions in meta-analysis, with the study of main effects having perhaps overly dominated much meta-analytic work in the past. Schmidt continues to challenge the field to better take measurement error into account, given that a good argument can be made that we should be more interested in assessing constructs rather than operations. Rosenthal wonders if the field has become so formally statistical that it overlooks useful solutions that do not fall within the accepted canon, such as the exclusive use of formal heterogeneity testing rather than such simple procedures as displaying effect sizes with Tukey's five-point box-and-whiskers plot or full box plots with labels and data values for effect sizes that are far away from the box.

Nor are these scholars the only ones posing interesting problems for meta-analysis to address in the future. For example, Lau discusses the need to develop more accessible computer programs that can address all stages of the systematic review process, not just the statistics (Cappelleri and Ingerick, 2014). Relatively few meta-analysts have risen to Rubin's (1990) challenge that the fundamental purpose of the field should be to use data to model the likely outcome of a hypothetical ideal study rather than to summarize existing studies that are often more or less removed from the study we really would like to have conducted to answer the question (Madan, Chen, Aveyard, Wang, Yahaya, Munafo, Bauld, and Welton, 2014; Schmidt and Hunter, 2015; Shadish, Matt, Navarro, and Phillips, 2000; Vanhonacker, 1996; Welton, Caldwell, Adamopoulos, and Vedhara, 2009). Surely, then, meta-analysis does not seem likely to run short of problems to address in the near future.

To conclude by returning to the big bang analogy, however, cosmologists have asked whether the universe created by the big bang will expand forever or will eventually collapse in on itself. We might ask the same question of meta-analysis: will the meta-analytic universe keep expanding forever, spinning off new galaxies into ever more disciplines and new technical specialties like network meta-analysis? Or will it eventually contract or even collapse in on itself, for example, being replaced by syntheses of archived raw data across studies rather than summary statistics? Or do we live in the epistemological equivalent of a cosmological multiverse (Greene, 2011) in which other solutions that we have not even conceived already exist somewhere to solve the problem of synthesizing knowledge, solutions that will eventually make meta-analysis as we know it obsolete? None of us can know the answer to such questions. We can only know that we will each be a part of the answer by the decisions we make and the research we do in the field of systematic reviews.

Acknowledgement

This research was supported in part by a grant from the University of California Office of the President to the University of California Educational Evaluation Consortium

References

Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. 1992. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *Journal of the American Medical Association* **268**: 240–248.

Armitage P, Berry G, Matthews JNS. 2008. *Statistical Methods in Medical Research* (4th ed.). Wiley, New York.
Astin AW, Ross S 1960. Glutamic acid and human intelligence. *Psychological Bulletin* **57**: 429–34.
Baeza-Yates R Ribeiro-Neto B. 1999. *Modern Information Retrieval*. Addison Wesley, New York.
Bax L, Yu LM, Ikeda N, Tsuruta H, Moons KG. 2006. Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC Medical Research Methodology* **6**: 50.
Becker BJ. 2007. Multivariate meta-analysis: contributions of Ingram Olkin. *Statistical Science* **22**: 401–406. DOI: 10.1214/07-STS239
Begg CB, Mazumdar M. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**: 1088–1101.
Berk RA. 2007. Statistical inference and meta-analysis. *Journal of Experimental Criminology* **3**: 247–270.
Bohlin I. 2012. Formalizing syntheses of medical knowledge: the rise of meta-analysis and systematic reviews. *Perspectives on Science* **20**: 273–309.
Borenstein M, Hedges LV, Higgins JPT, Rothstein HR 2011. *Introduction to Meta-analysis*. Wiley, New York.
Bracht GH, Glass GV. 1968. The external validity of experiments. *American Educational Research Journal* **5**: 437–474.
Bush BJ, Wang MC. 2009. Vote-counting procedures in meta-analysis. In HM Cooper, LV Hedges, J.C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-analysis* (2nd Ed.) Russell Sage Foundation, New York; pp. 207–220.
Campbell DT. 1960. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review* **67**: 380–400. DOI: 10.1037/h0040373
Campbell DT. 1974a. Evolutionary epistemology. In *The Philosophy of Karl Popper*, Schlipp PA (Eds.). Open Court, La Salle, IL; pp. 413–463.
Campbell DT. 1974b. Unjustified variation and selective retention in scientific discovery. In *Studies in the Philosophy of Biology: Reduction and Related Problems*, Ayala F, Dobszhansky T (Eds.). Macmillan, London, UK: pp. 139–161.
Campbell DT, Stanley, JC. 1963. *Experimental and Quasi-Experimental Designs for Research*. Rand-McNally, Chicago.
Cappelleri JC, Ingerick M. 2014. A conversation with Joseph Lau. *Research Synthesis Methods* (published online 8 April 2014). DOI: 10.1002/jrsm.1116
Chalmers I, Hedges LV, Cooper H. 2002. A brief history of research synthesis. *Evaluation and the Health Professions* **25**: 12–37.
Chalmers TC, Smith Jr. H, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. 1981. A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials* **2**: 31–49.
Counsell C. 1997. Formulating questions and locating primary studies for inclusion in systematic reviews. *Annals of Internal Medicine* **127**: 380 –387.
Danziger K. 1994. *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge University Press, New York.
Dawson B, Trapp RG. 2004. *Basic and Clinical Biostatistics* (4th ed.). McGraw-Hill, New York.
Deeks JJ, Altman DG, Bradburn MJ. 2001. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context* (2nd ed.). BMJ Books, London.
DerSimonian R Laird N. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**: 177–188.
Dias S Welton NJ, Caldwell DM, Ades AE. 2010. Checking consistency in mixed treatment comparison. *Statistics in Medicine* **29**: 932–944.
Dickersin K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association* **263**: 1385–1389.
Dickersin K, Scherer R, Lefebvre, C. 1994. Identifying relevant studies for systematic reviews. *British Medical Journal* **309**: 1286–1291.
Dunlap WP, Cortina JM, Vaslow JB, Burke MJ 1996. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods* **2**: 170–177.
Duval S, Tweedie R. 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**: 455–463.
Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. 1991. Publication bias in clinical research. *Lancet* **337**: 1102.
Egger M, Davey Smith G, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **315**: 629–634.
Eysenck HJ. 1965. The effects of psychotherapy. *International Journal of Psychiatry* **1**: 97–178.
Eysenck HJ. 1978. An exercise in mega-silliness. *American Psychologist* **33**: 517.
Farr RM. 1996. *The Roots of Modern Social Psychology*. Blackwell Publishing, Oxford, UK: pp. 1872–1954.
Fisher RA. 1932. *Statistical Methods for Research Workers* (4th Ed.). Oliver & Boyd, London.
Gholson BG, Shadish WR, Neimeyer RA, Houts AC (Eds.). 1989. *Psychology of Science: Contributions to Metascience*. Cambridge University Press, Cambridge, England.
Gladwell M. 2008. *Outliers: The Story of Success*. Little, Brown and Company, New York.
Glass GV. 1976 Primary, secondary and meta-analysis of research. *Educational Researcher* **5**: 3–8.

Glass GV. 1977. Integrating findings: the meta-analysis of research. *Review of Research in Education* **5**: 351–379.
Glass GV, Hakstian AR. 1969. Measures of association in comparative experiments. *American Educational Research Journal* **6**: 403–414.
Glass GV, McGaw B, Smith ML. 1981. *Meta-analysis in Social Research*. Sage Publications, Thousand Oaks, California.
Goldstein H. 2011. *Multilevel Statistical Models* (4th ed.) Wiley, New York.
Greene B. 2011. *The Hidden Reality: Parallel Universes and the Deep Laws of the Cosmos*. Knopf, New York.
Guba EG, Lincoln YS. 1981. *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. Jossey-Bass, San Francisco, California.
Hahn U, Romacker M. 2002. MEDSYNDIKATE--a natural language system for the extraction of medical information from findings reports. *International Journal of Medical Informatics* **67**: 63–74.
Hardy RJ, Thompson SG. 1998. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* **17**: 841–856.
Harper G, Peattie K. 2011 Tracking the influence of the first special journal issue on 'green marketing': a citation network analysis. *Social Business* **1**: 239–61.
Harris JK, Beatty KE, Lecy JD, Cyr JM, Shapiro RM 2011. Mapping the multidisciplinary field of public health services and systems research. *American Journal of Preventive Medicine* **41**: 105–111.
Hays W. 1963. *Statistics for Psychologists*. Holt, New York.
Hedges LV. 1980. Unbiased estimation of effect size. *Evaluation in Education: International Progress* **4**: 25–27.
Hedges LV. 1981. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* **6**: 107–128.
Hedges LV. 1982a. Estimation of effect size from a series of independent experiments. *Psychological Bulletin* **92**: 490–499.
Hedges LV. 1982b. Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics* **7**: 119–137.
Hedges LV. 1982c. Fitting continuous models to effect size data. *Journal of Educational Statistics* **7**: 245–270.
Hedges LV. 1983. A random effects model for effect sizes. *Psychological Bulletin* **93**: 388–395.
Hedges LV. 1987. How hard is hard science, how soft is soft science: the empirical cumulativeness of research. *American Psychologist* **42**: 443–455.
Hedges LV, Olkin I. 1980. Vote counting methods in research synthesis. *Psychological Bulletin* **88**: 359–369.
Hedges LV, Olkin I. 1983a. Clustering estimates of effect magnitude from independent studies. *Psychological Bulletin* **93**: 563–573.
Hedges LV Olkin I. 1983b. Regression models in research synthesis. *The American Statistician* **37**: 137–140
Hedges LV, Olkin I. 1984. Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin* **96**: 573-580.
Hedges LV, Olkin I. 1985. *Statistical Methods for Meta-analysis*. Academic Press, Orlando, FL.
Hedges LV, Vevea JL. 1998. Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**: 486–504.
Higgins JP, Thompson SG, Deeks JJ, Altman DG 2003. Measuring inconsistency in meta-analyses. *British Medical Journal* **327**: 557–560.
Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, Boersma C, Thompson D, Larholt KM, Diaz M, Barrett A. 2011. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices—part 2. *Value in Health* **14**: 429–437.
Hong FT. 2013. Deciphering the enigma of human creativity: can a digital computer think? *Journal of Computer Science and Systems Biology* **6**: 228–261. 10.4172/jcsb.1000120.
Hopewell S, Clarke M, Lusher A, Lefebvre C, Westby M. 2002. A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials. *Statistics in Medicine* **21**: 1625–1634.
Hox JJ. 2010. *Multilevel Analysis: Techniques and Applications*. Routledge, New York.
Huberty CJ. 2002. A history of effect size indices. *Educational and Psychological Measurement* **62**: 227–240. DOI: 10.1177/0013164402062002002
Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. 2006. Assessing heterogeneity in meta-analysis: *Q* statistic or I^2index? *Psychological Methods* **11**: 193–206.
Hunt MM. 1999. *How Science Takes Stock: The Story of Meta-analysis*. Russell Sage Foundation, New York.
Hunter JE, Schmidt FL, Jackson GB. 1982. *Meta-analysis: Cumulating Research Findings Across Studies*. Sage Publications, Beverly Hills, CA.
Ioannidis JP, Patsopoulos NA, Evangelou E 2007. Uncertainty in heterogeneity estimates in meta-analyses. *British Medical Journal* **335**: 914–916.
Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. 1994. Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine* **120**: 667–676.
Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. 2011. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR task force on indirect treatment comparisons good research practices: part 1. *Value in Health* **14**: 417–428.

Juni P, Witschi A, Bloch R., Egger M. 1999. The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association*, **282**: 1054–1060.
Kazdin AE. 2011. *Single-Case Research Designs: Methods for Clinical and Applied Settings* (2nd ed.). Oxford University Press, New York.
Kuhn TS. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.
Larsson R. 1993. Case survey methodology: quantitative analysis of patterns across case studies. *The Academy of Management Journal* **36**: 1515–1546.
Last JM (eds.). 2001. *A Dictionary of Epidemiology* (4th ed.). Oxford University Press, New York.
Lecy JD, Beatty KE. 2012. Representative literature reviews using constrained snowball sampling and citation network analysis. Available at SSRN: http://ssrn.com/abstract=1992601 or 10.2139/ssrn.1992601 DOI: 10.2139/ssrn.1992601#_blank
Lecy JD, Mergel IA, Schmitz HP. 2013. Networks in public administration. *Public Management Review* **16**: 643–665. DOI: 10.1080/14719037.2012.743577
Lecy JD, Moreda D. 2013. cna: citation network analyzer. R package version 0.3-3.
Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* **6**: e1000100. DOI: 10.1371/journal.pmed.1000100
Light RJ, Pillemer DB. 1984. *Summing Up: The Science of Reviewing Research*. Harvard University Press, Cambridge, Massachusetts.
Light RJ, Smith PV. 1971. Accumulating evidence: procedures for resolving contradictions among different research studies. *Harvard Educational Review* **41**: 429–472.
Lipsey MW, Wilson DB. 2001. *Practical Meta-analysis*. Sage Publications, Thousand Oaks, California.
Lu G, Ades AE. 2004. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* **23**: 3105–3124.
Lu G, Ades AE. 2006. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* **101**: 447–459.
Madan J, Chen YF, Aveyard P, Wang D, Yahaya I, Munafo M, Bauld L, Welton N. 2014. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **177**: 295–314. DOI: 10.1111/rssa.12018
Mantel N, Haenszel W. 1959. Statistical aspects of the analysis of data from the retrospective analysis of disease. *Journal of the National Cancer Institute* **22**: 719–748. DOI: 10.1093/jnci/22.4.719 DOI: 10.1093%2Fjnci%2F22.4.719, PMID 13655060
Meehl PE. 1954. *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Univer. of Minnesota Press, Minneapolis.
Meehl PE. 1978. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology* **46**: 806–834.
Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. for the QUOROM group. 1999. Improving the quality of reporting of meta-analysis of randomized controlled trials: the QUOROM statement. *Lancet* **354**: 1896–1900.
Moher D, Jadad A, Nichol G, Penman M, Tugwell P, Walsh S. 1995. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* **16**: 62–73.
Mosteller FM, Bush RR. 1954. Selected quantitative techniques. In G. Lindzey (Ed). *Handbook of Social Psychology: Vol. 1. Theory and Method*. Addison-Wesley, Cambridge, MA: pp. 289–334.
Mullen B. 2013. *Advanced Basic Meta-analysis: Version 1.10*. Taylor and Francis, Hoboken, NJ: 2013.
Mullen B, Rosenthal R. 1985. *Basic Meta-analysis: Procedures and Programs*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
Mulrow CD. 1994. Rationale for systematic reviews. *British Medical Journal* **309**: 597–599.
Nakagawa S, Cuthill I. 2007 Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* **82**: 591–605.
Newman ME. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46** 323–351.
Noblit GW, Hare RD. 1988. *Meta-Ethnography: Synthesizing Qualitative Studies*. Sage Publications, London.
Noruzi A. 2005. Google Scholar: the new generation of citation indexes. *Libri: International Journal of Libraries and Information Services* **55**: 170–80.
O'Rourke K. 2007. A historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine* **100**: 579–582.
Oxman AD, Guyatt GH. 1988. Guidelines for reading literature reviews. *Canadian Medical Association Journal* **138**: 697–703.
Parmar MKB, Torri V, Stewart L. 1998. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* **17** 2815–2834.
Paterson B, Thorne S, Canam C, Jillings C 2001. *Meta-Study of Qualitative Health Research: A Practical Guide to Meta-analysis and Meta-Synthesis*. Sage, Thousand Oaks, CA.

Pearson K. 1904a. Antityphoid inoculation. *British Medical Journal* **2**: 1667–1668.
Pearson K. 1904b. Report on certain enteric fever inoculation statistics. *British Medical Journal* **2**: 1243–1246.
Peterson RA. 1994. A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research* **21**: 381–391.
Popay J, Rogers A, Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research* **8**: 341–351.
Popper KR. 1972. *Objective Knowledge, An Evolutionary Approach*. Oxford University Press, New York.
Raudenbush S, Bryk A. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Thousand Oaks, California.
Roberts RM. 1989. *Serendipity: Accidental Discoveries in Science*. Wiley, New York.
Rosenthal R. 1966. *Experimenter Effects in Behavioral Research*. Appleton-Century-Crofts, NY.
Rosenthal R. 1969. Interpersonal expectations. In *Artifact in Behavioral Research*. Rosenthal R, Rosnow RL (Eds). academic Press, New York: pp. 181–277.
Rosenthal R. 1978. Combining results of independent studies. *Psychological Bulletin* **85**: 185–193.
Rosenthal R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* **86**: 638–641.
Rosenthal R. 1980. *New Directions for Methodology of Social and Behavioral Science: Quantitative Assessment of Research Domains* (eds.). Jossey-Bass, San Francisco.
Rosenthal R. 1984. *Meta-Analytic Procedures for Social Research*. Sage, Beverly Hills, CA.
Rosenthal R. 1991. *Meta-Analytic Procedures for Social Research* (2nd ed.). Sage Publications, Thousand Oaks, California.
Rosenthal R. 1995. Writing meta-analytic reviews. *Psychological Bulletin* **118**: 183–192.
Rosenthal R, Rosnow RL. 1975. *The Volunteer Subject*. John Wiley, New York.
Rosenthal R, Rosnow R. 1991. *Essentials of Behavioural Research*. McGraw-Hill, New York.
Rosenthal R, Rosnow R, Rubin DB. 2000. *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach*. Cambridge University Press, New York.
Rosenthal R, Rubin DB. 1978a. Interpersonal expectancy effects: the first 345 studies. *Behavioral and Brain Sciences* **1** 377–386.
Rosenthal R, Rubin DB. 1978b. Issues in summarizing the first 345 studies of interpersonal expectancy effects. *Behavioral and Brain Sciences* **1**: 400–415.
Rosenthal R, Rubin DB. 1982a. A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* **74**: 166–169.
Rosenthal R, Rubin DB. 1982b. Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology* **74**: 708–712.
Rosenthal R, Rubin DB. 1989. Effect size estimation for one-sample multiple-choice-type data: design, analysis, and meta-analysis. *Psychological Bulletin* **106**: 332–337.
Rossi PH. Lipsey MW, Freeman HE. 2004. *Evaluation: A Systematic Approach* (7th ed.). Sage Publications, Thousand Oaks, California.
Rubin DBR. 1990. A new perspective. In *The Future of Meta-analysis*. Wachter KW, Straf ML (Eds.). Russell Sage Foundation, New York: pp. 155–165).
Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. 1987. Meta-analyses of randomized controlled trials. *New England Journal of Medicine* **316**: 450–455.
Salanti G, Higgins JP, Ades AE, Ioannidis JP. 2008. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* **17**: 279–301.
Sandelowski M, Docherty S, Emden C. 1997. Qualitative metasynthesis: issues and techniques. *Research in Nursing & Health* **20**: 365–371.
Schmidt FL. 1992. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist* **47**: 1173–1181.
Schmidt FL, Berner JG, Hunter JE. 1973. Racial differences in validity of employment tests: reality or illusion? *Journal of Applied Psychology* **58**: 5–9.
Schmidt FL, Hunter JE. 1977. Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology* **62**: 529–540.
Schmidt FL, Hunter JE. 2003. History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975 – 2001. In *Validity Generalization: A Critical Review*. Murphy KR (eds.). Lawrence Erlbaum Associates, Mahwah, New Jersey: pp. 31–65.
Schmidt FL, Hunter JE. 2015. *Methods of Meta-analysis: Correction Error and Bias and Research Findings*. Sage Publications, Thousand Oaks, California.
Schmidt FL, Hunter JE, Pearlman K., Shane GS. 1979. Further tests of the Schmidt-Hunter Bayesian validity generalization model. *Personnel Psychology* **32**: 257–281.
Schulz KF, Chalmers I, Hayes RJ, Altman DG. 1995. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* **273**: 408–412.
Schwarzer R. 1989. *Meta-analysis Programs*. Freie Universität, Berlin.
Scriven M. 1969. An introduction to meta-evaluation. *Educational Product Report* **2**: 36–38.

Simonton DK. 1988. *Scientific Genius: A Psychology of Science*. Cambridge University Press, Cambridge, UK: 1988.
Simonton DK. 2004. *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*. Cambridge University Press, Cambridge, UK.
Simonton DK 2010. Creative thought as blind-variation and selective-retention: combinatorial models of exceptional creativity. *Physics of Life Reviews* **7**: 156–179.
Simonton DK. 2011. Creativity and discovery as blind variation: Campbell's (1960) BVSR model after the half-century mark. *Review of General Psychology* **15**: 158–174.
Simonton DK. 2013. Creative thought as blind variation and selective retention: why creativity is inversely related to sightedness. *Journal of Theoretical and Philosophical Psychology* **33**: 253–266.
Shadish WR. 2007. A world without meta-analysis. *Journal of Experimental Criminology* **3**: 281–291.
Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton-Mifflin, Boston.
Shadish WR, Matt GE, Navarro AM, Phillips G 2000. The effects of psychological therapies under clinically representative conditions: a meta-analysis, *Psychological Bulletin* **126**: 512–529.
Shadish WR, Tolliver D, Gray M, Sen Gupta SK. 1995. Author judgments about works they cite: three studies from psychology journals. *Social Studies of Science* **25**: 477–498.
Smith ML, Glass GV. 1977. Meta-analysis of psychotherapy outcome studies, *American Psychologist* **32**: 752–60.
Snedecor GW 1956. *Statistical Methods* (5th ed). Iowa State College Press, Ames, Iowa.
Snijders TAB, Bosker RJ. 2011. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications, Thousand Oaks, California.
Stjernswärd J 1974. Decreased survival related to irradiation postoperatively in early operable breast cancer. *The Lancet* **2**: 1285–1286.
Stjernswärd J. 2009 Meta-analysis as a manifestation of 'bondförnuft' ('peasant sense'). *JLL Bulletin*: *Commentaries on the history of treatment evaluation* (www.jameslindlibrary.org). [Personal reflection]
Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *Journal of the American Medical Association* **283**: 2008–2012.
Thompson SG, Higgins JP. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* **21**: 1559–1573.
Underwood BJ 1957. Interference and forgetting. *Psychological Review* **64**: 49–60.
Vanhonacker WR. 1996. Meta-analysis and response surface extrapolation: a least squares approach. *The American Statistician* **50**: 4.
Verhagen AP, de Vet HCW, de Bie RA, Kessels AG, Boers M., Bouter LM, Knipschild PG. 1998. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology* **15**: 1235–1241.
Viechtbauer W 2010. metafor: meta-analysis package for R. R package version 1.4-0, URL http://CRAN.R-project.org/package=metafor.
Wanous JP, Reichers AE, Hudy MJ. 1997. Overall job satisfaction: how good are single-item measures? *Journal of Applied Psychology* **82**: 247–252.
Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K. 2009. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *American Journal of Epidemiology* **169**: 1158–65. DOI: 10.1093/aje/kwp014
Wuketits FM 2001. The philosophy of Donald T. Campbell: a short review and critical appraisal. *Biology and Philosophy* **16**: 171–188.

