Limited Usefulness of Meta-Analysis for Informing Patient Care

John Concato^{a, b} Ralph I. Horwitz^c

^aCenter for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA; ^bDepartment of Medicine, Yale University School of Medicine, New Haven, CT, USA; ^cDepartment of Medicine, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA

Keywords

Meta-analysis · Evidence-based medicine · Medicine-based evidence · Conflict of interest

Psychotherapy and Psychosomatics has been a strong voice in challenging current dogma regarding meta-analysis, as reflected in statements such as “Other journals believe in pseudo-objectivity, obtained by increasingly complicated and cumbersome procedures, of systematic reviews and meta-analyses, where the presence of an author with clinical familiarity regarding the topic is an option” [1]. For readers interested in a broader discussion of relevant issues, the corresponding controversy can be found in numerous venues, including a series of articles published in the *Journal of Clinical Epidemiology*, with titles such as “Evidence-based medicine has been hijacked: a report to David Sackett” [2], “Evidence-based medicine was bound to fail: a report to Alvan Feinstein” [3], “Hijacked evidence-based medicine: stay the course and throw the pirates overboard” [4], and “Why evidence-based medicine failed in patient care and medicine-based evidence will succeed” [5]. In this commentary, we focus on – and question – the current practice of conducting and reporting meta-analyses for clinicians.

We contend that the veneration given to meta-analysis overlooks serious problems with the technique, even when done according to current standards. This report is not an exhaustive review of the conduct and reporting of meta-analysis, examples of which are available elsewhere [6, 7], and it also does not comprehensively address all applications of the methodology (e.g., regulatory aspects [8]). Rather, the focus of this article is on selected, under-recognized limitations of meta-analysis. We present examples and discuss conceptual issues in categories of internal conduct, external influences, and attention to clinical issues; other considerations are also noted. The take-home message is that harm can occur if practicing clinicians are expected to embrace the results of meta-analyses, and umbrella meta-analyses, that place more emphasis on mathematical models than on clinical relevance and applicability.

Representative Examples

Internal Conduct

Pertinent considerations involved in the intellectual contradictions surrounding meta-analysis are mentioned in a paper [9] reviewing published, often-cited meta-anal-

yses focusing on the benefits of antidepressant medications. In summary, evidence in support of medication efficacy is stronger for severe depression than it is for mild cases of depression, but meta-analyses on the topic are not concordant. Among various categories of problems potentially explaining this discrepancy, the authors identify issues with “the randomized controlled trial designs, the study samples, the psychometric scales, the methods of meta-analysis, the interpretation of results, and reporting of conflicts of interest” [9]. Although proponents of meta-analysis often comment that the conduct of meta-analysis can and should be better [4], the degrees of freedom available to meta-analysts, in terms of decision-making, is an appropriate topic for discussion. Specifically, when done by different investigators, meta-analyses on the same topic are vulnerable to obtaining disparate results and conclusions.

The aforementioned paper on treating depression addresses this issue directly, finding that “different methods (of meta-analysis) give different results and similar results seem to entertain a variety of interpretations” [9]. Additional details regarding how methods can differ are provided in the subsequent section of this report, but it should also be noted that the problem can even extend to meta-analyses of the same data. When provided with the same objectives, resources, and individual participant-level data on recombinant human bone morphogenetic protein, two groups conducting separate meta-analyses produced results having “notable differences” [10].

External Influences

Evaluations of meta-analyses and their underlying trials usually mention potential conflicts of interest. Efforts to minimize such conflicts – or at least report and try managing them – are usually reflected in publications as disclosure notices. The focus in this context is typically on investigator authors with links to industry, but academic investigators are not immune. For example, a meta-analysis of mindfulness-based interventions [11] was later retracted [12] for reasons including authors’ potential financial gain.

Beyond financial gain, one viewpoint emphasized how the influence of academic departments, universities, research institutions, other sponsors, and journals are relevant, and points out how “the prospect of fame can be even more seductive than fortune” [13]. A similar concern points out that meta-analysts “may hope to confirm what their studies had previously shown, or because they’ve [previously] suggested certain policies” [14]. Importantly, and unfortunately, given that multiple studies

can be combined in different ways, a meta-analysis can be more subject to external influences than any single investigation.

Clinical Issues

An example from the medical field involving hepatitis C provides an opportunity to emphasize how meta-analyses can falter by favoring nonclinical considerations over clinical relevance. In brief, a meta-analysis of direct-acting antiviral medications (DAAs) for the treatment of hepatitis C was conducted by the Cochrane Collaboration [15]. When assessing data on 2,996 participants from 11 randomized trials that included mortality as an outcome, the investigators found “no evidence of a difference when assessing the risk of all-cause mortality” [15].

Notably, and reflecting one of the aforementioned decisions that meta-analysts must make, the 11 trials involved were a subset of 138 relevant trials identified – based on a decision to look at mortality, rather than shorter-term outcomes such as sustained viral response (SVR), as the endpoint of interest. As emphasized by a commentary on the topic, however, “a large body of accumulated evidence indicates associations between SVR and improvements in liver function, fibrosis, extrahepatic outcomes of cirrhosis-related complications, and all-cause mortality” [16]. Another report pointed out that the purpose of the corresponding clinical trials was to evaluate the virological efficacy of new DAAs; such trials “were not designed to assess mortality” [17]. The controversy is based on judgments regarding whether data from longer-term trials are desirable as per the Cochrane viewpoint, or whether the evidence is sufficient based on other forms of knowledge, including clinical knowledge, in this situation.

One opinion noted that “the Cochrane review appears to assume that SVR, which has been used for the last 25 years as the main surrogate marker to assess the success of anti-HCV therapy, is fundamentally unreliable, as it has never been validated in a formal RCT. This implies that one should also repeal the current knowledge about clinical management of [chronic hepatitis C]” [18]. Indeed, given the impressive virological results, equipoise to directly test whether DAAs reduce mortality in a randomized trial is no longer considered to exist. Attempting to perform such a trial now could repeat the tragedy of decades ago, when enthusiasm for randomization led to mortality in the control arms of arguably unnecessary trials, given that ample but nonrandomized evidence in support of ECMO therapy was available but was not considered adequate [19].

Conceptual Issues

The scenarios described in the examples above are not mutually exclusive. Each example was used to help clarify and emphasize a specific point, yet each also involved numerous methodological options, potential conflicts regarding how those options were handled, and considerable versus little emphasis placed on clinical issues. For discussion purposes in this commentary, and now extending a conceptual perspective, we again consider categories of internal conduct, external influences, and attention to clinical issues.

Conduct of the Analysis Itself

Although described as involving well-documented methodologies [6], and despite guidelines for appropriate reporting [7], researchers encounter numerous decision points that allow for different decisions to be made when conducting a meta-analysis. These decisions include, but are not limited to, which studies to include, whether to account for publication bias, whether to exclude any studies based on an assessment of quality, which outcome to select as the primary endpoint, the number of relevant secondary endpoints, how to manage co-therapies, and various statistical options including how to address heterogeneity of studies. Another potential decision, although not always available for an investigator to choose, involves combining study-level or participant-level data, with the latter approach considered more rigorous. Overall, these many degrees of freedom, so to speak, will always have a component of subjectivity, even if intentional manipulation is not involved.

Influences on Analyses

A range of external influences potentially affecting meta-analyses are related to the numerous decision points just discussed, and such influences can be considered a specific concern. The complexities involved are beyond the scope of this paper, but two comments will be offered. First, and as mentioned, fame can equal or surpass financial gain as a motivating factor, meaning that all meta-analyses should be viewed with healthy skepticism. Although not addressing meta-analysis directly, the results of a report on human behavior “do not suggest that people are necessarily directionally biased (to either overestimate or under-estimate control [over potential bias]), but they are nevertheless inaccurate in their estimation of personal control” [20]. If so, managing conflicts regarding meta-analysis is even more challenging than currently believed.

Second, the distinction between industry and academia has blurred in recent years, including public-private partnerships wherein patent rights are released from the government to investigators and institutions. The interests of these parties can be divergent, further complicating assessments of external influences on meta-analysis. Overall, a simple dichotomy of pharma-based versus university-based research was never the case, and it is less applicable today.

Attention to Clinical Detail

Perhaps the most glaring limitation of meta-analysis is the lack of attention to clinical detail. As mentioned in a more general context of evidence generation, “clinically relevant issues of who and where were the patients, what and why were the treatments, and when and how were the outcomes assessed” [21] should be assessed in the conduct of clinical research. In meta-analysis, however, such issues are usually secondary to considerations of, for example, how to assess heterogeneity from a mathematical perspective.

Although this topic has many facets, “how the outcome is measured” can be a prominent feature of the disagreement. “Hard” endpoints of randomized trials, especially mortality as an outcome, should be comparable across studies if sufficient attention to detail is applied. Such consistency may be more challenging, however, when different criteria or instruments are used for outcomes towards the “soft” end of the spectrum. Examples include quality of life, depression, or pain scales, with various instruments available from which to choose to represent the same construct – and even different ways to apply the same instrument (e.g., abbreviated vs. full format).

Another, under-appreciated phenomenon related to outcomes of randomized trials can have a substantial impact on the usefulness of meta-analyses. Most often reflecting different participant characteristics (or perhaps secondary therapies) across different trials, the outcome rates of placebo arms for studies of the same intervention-outcome association examined in meta-analysis can vary considerably. A study [22] focusing on this issue found that placebo arms in trials of beta-blockers versus placebo and heart failure/sudden death had an approximate five-fold difference in outcomes (2.2 vs. 10%); placebo arms in trials of angiotensin receptor blockers versus placebo and all-cause mortality had an approximate ten-fold difference in outcome (2.6 vs. 29%). Combining such studies is problematic – what, exactly, is a clinician supposed to do with the “average” result of trials that included such disparate outcomes involving placebo arms?

This problem is evident even in informal encounters with the medical literature. After novel treatments were developed recently for idiopathic pulmonary fibrosis, for example, three trials involving two new agents were compared [23]. The primary outcome, a pulmonary function measure, worsened in all treatment arms, but less worsening over time represents evidence of a clinical benefit. Surprisingly, one trial's treatment group had a greater decline (i.e., worse outcome) than the placebo group of the two other studies [23]. This scenario likely reflects differences in study populations – so on what scientific basis would a future meta-analysis combine such studies?

Specific Considerations

“Next-Generation” Meta-Analysis

Umbrella meta-analyses are studies that evaluate more than one treatment comparison for the management of the same disease or condition [24, 25]. As noted by proponents of the method, “umbrella reviews are limited by the amount, quality and comprehensiveness of available information in primary studies” [24], similar to conventional meta-analysis. Yet, the corresponding increased complexity of clinical issues goes largely unrecognized. For example, an umbrella meta-analysis on “systemic treatment for advanced breast cancer included data from 45 different direct comparisons, each of which could have been a separate traditional meta-analysis” [24]. Overshadowing the virtuous claims of “objective, quantitative integration of the evidence,” the scientific credibility of umbrella meta-analysis is undermined by cautionary statements about “data syntheses require caution in the presence of incoherence,” and “one has to critically consider whether the data can be extrapolated to individual patients and settings” [24]. An assumption that participants, interventions, and outcomes in these multiple trials would be similar enough to be compared meaningfully is clinically unrealistic.

Network meta-analysis extends the umbrella concept by examining “all treatments for a given condition or disease and all possible comparisons between them” [25]. Unfortunately, using this approach further extends the problem of unscientific and clinically muddled comparisons. Prospective meta-analysis, described as “the design of multiple trials with the explicit, predefined purpose to, when completed, combine them in a meta-analysis” [25] theoretically would have fewer limitations – but, for reasons discussed in the next section, not if such analyses “encompass trials that use different designs and test a va-

riety of interventions, comparisons, settings and populations” [25]. On a related point, recent discussions [26, 27] suggest that too much trust is placed in individual RCTs themselves and the average treatment effect they generate, with corresponding implications for meta-analyses. As one of the viewpoints that was emphasized, “unless we are prepared to make assumptions, and to say what we know, [...] the credibility of the RCT does us very little good” [26].

Faulty Conceptual Leap

A fundamental problem with all meta-analyses involves an inherent conceptual leap imbedded in the design. Consider this statement, published over 30 years ago, as part of the justification for embracing meta-analyses: “In the absence of trials of [very large] size, however, some useful estimates of the likely effects [...] may still be obtained by combining the information (irrespective of the results) from all the randomised controlled trials, as is done when the results from individual centers are combined in multicentre trials. Such an overview does not, of course, implicitly assume that the selection criteria, treatment regimens, or definitions of outcome are similar in different trials, for they are not: it assumes merely that patients in one randomised trial can be compared unbiasedly with other patients in the same study” [28].

This assertion reveals an often-ignored assumption of meta-analyses: that combining data from multiple trials in a meta-analysis is equivalent to combining data from each site in a multicenter trial. However, this assumption may not be correct. The multicenter RCT employs a common protocol, as a logical extension of the single-center RCT, thereby minimizing the risk of extensive clinical heterogeneity that is so often observed in meta-analyses. The common protocol requires investigators to use identical criteria across sites for enrolling and classifying patients in the trial, for administering the main treatment and potential ancillary treatments, and for the surveillance and detection of outcome events. In contrast, a meta-analysis is designed with an inherent disregard for these methodological requirements, and with disregard as well for the implications regarding how the result can be used by physicians who are making clinical decisions for a patient at hand.

Meta-Analysis and Magic

Lastly, an informal aspect of meta-analysis involves the idea of getting “something for nothing” when combining studies using statistical procedures, as reflected in a mid-1990s characterization of the technique as “statisti-

cal alchemy for the 21st century” [29]. Recognizing that the comment (as well as our paper) does not examine specific statistical approaches, a similar theme resurfaced in a recent paper that alluded to “biology versus psychology in the era of statistical magic” [30]. The corresponding topic encapsulated many of the issues discussed in this report. In brief, the condition involved was female hypoactive sexual desire disorder (HSDD), and meta-analyses reached conflicting conclusions regarding the effectiveness of pharmacological versus psychological interventions [30]. Phrases such as “all successful trials should meet 10 precepts” [31] reflect the aspect of internal conduct – who decides on what standards are applicable? Regarding conflicts of interest, ideology regarding a biological versus a psychological perspective was mentioned as a specific source of potential bias and conflict of interest – extending beyond financial gain or fame as a motivating factor.

Selecting relevant comparison groups was also controversial, as was what represents a minimal difference in outcome that is meaningful (notwithstanding questions about the disorder itself [32]) – reflecting clinical issues warranting attention. In the HSDD context, reaching a conclusion based on a heterogeneous compilation of only four studies was the basis for invoking the phrase “statistical magic”. As stated more than 20 years earlier, a “prime scientific flaw of meta-analysis is its focus on the big process of aggregation, but not on the crucial small processes that produced the basic evidence” [29]. Umbrella and network meta-analyses further amplify this problem and should be discouraged for most clinical purposes.

Current Status and Looking Forward

Although meta-analyses are often placed at the top of hierarchies of evidence, an insightful criticism from more than 20 years ago anticipated the current problems of meta-analyses. Specifically, suggestions made at that time included conducting meta-analyses that go beyond the usu-

al focus on methodologic decisions and potential conflicts of interest, to involve knowledgeable clinicians who can “describe the crucial scientific (as well as humanistic) importance of clinical severity and clinical outcomes” [29], and who can also ensure that appropriate clinical taxonomies are applied.

As a simple rule of thumb [21], more consideration needs to be given to patient populations, characteristics of the intervention, and suitability of the outcome (all from a clinical perspective), and commensurate with attention to statistical considerations. Instead, and as has been noted recently, the current practice of conducting meta-analysis and (more so) umbrella meta-analysis has limited clinical utility: “When the evidence points clearly in one direction, there is little need for meta-analysis. When it doesn’t, a meta-analysis is unlikely to give the final answer” [14].

Conclusions

The limitations of meta-analysis are representative of the larger issue of how the accomplishments and popularity of evidence-based medicine are obscuring fundamental problems. Meta-analyses, including umbrella meta-analyses, are very prone to disparate conclusions when evaluating the same treatment-outcome association, and such variability is made worse by a variety of potential conflicts of interest. Even if these challenges were addressed, however, meta-analyses require attention to clinical detail on a scale comparable to the attention currently given to statistical procedures.

Disclosure Statement

The authors have no financial conflicts of interest to declare. This article reflects the views of the authors and should not be construed to represent the views or policies of the FDA.

References

- 1 Fava GA, Cosci F, Berrocal Montiel C, Tomba E. The Journal’s Mission. *Psychother Psychosom*. 2019;88(1):1–4.
- 2 Ioannidis JP. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol*. 2016 May;73:82–6.
- 3 Fava GA. Evidence-based medicine was bound to fail: a report to Alvan Feinstein. *J Clin Epidemiol*. 2017 Apr;84:3–7.
- 4 Ioannidis JP. Hijacked evidence-based medicine: stay the course and throw the pirates overboard. *J Clin Epidemiol*. 2017 Apr;84:11–3.
- 5 Horwitz RI, Singer BH. Why evidence-based medicine failed in patient care and medicine-based evidence will succeed. *J Clin Epidemiol*. 2017 Apr;84:14–7.
- 6 Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nature*. 2018 Mar;555(7695):175–82.
- 7 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009 Aug;151(4):264–9.

- 8 Food and Drug Administration. Meta-Analyses of Randomized Controlled Clinical Trials to Evaluate the Safety of Human Drugs or Biological Products Guidance for Industry [Internet]. 2018 [cited 2019 Aug 4]. Available from: <https://www.fda.gov/media/117976/download>.
- 9 Fountoulakis KN, Samara MT, Siamouli M. Burning issues in the meta-analysis of pharmaceutical trials for depression. *J Psychopharmacol*. 2014 Feb;28(2):106–17.
- 10 Low J, Ross JS, Ritchie JD, Gross CP, Lehman R, Lin H, et al. Comparison of two independent systematic reviews of trials of recombinant human bone morphogenetic protein-2 (rhBMP-2): the Yale Open Data Access Medtronic Project. *Syst Rev*. 2017 Feb;6(1):28–28.
- 11 Gotink RA, Chu P, Busschbach JJ, Benson H, Fricchione GL, Hunink MG. Standardised mindfulness-based interventions in healthcare: an overview of systematic reviews and meta-analyses of RCTs. *PLoS One*. 2015 Apr;10(4):e0124344.
- 12 PLOS ONE Editors. Retraction: Standardised Mindfulness-Based Interventions in Healthcare: An Overview of Systematic Reviews and Meta-Analyses of RCTs. *PLoS One*. 2019 Apr;14(4):e0215608.
- 13 Cappola AR, FitzGerald GA. Confluence, Not Conflict of Interest: Name Change Necessary. *JAMA*. 2015 Nov;314(17):1791–2.
- 14 de Vrieze J. The metawars. *Science*. 2018 Sep;361(6408):1184–8.
- 15 Jakobsen JC, Nielsen EE, Feinberg J, Katakam KK, Fobian K, Hauser G, et al. Direct-acting antivirals for chronic hepatitis C. *Cochrane Database Syst Rev*. 2017 Jun;6:CD012143.
- 16 The Lancet Gastroenterology Hepatology. Clinical effects of antivirals for hepatitis C: context is critical. *Lancet Gastroenterol Hepatol*. 2017 Aug;2(8):539.
- 17 Hawkes N. Are new hepatitis C drugs all they're cracked up to be? *BMJ*. 2017 Jun;357:j2961.
- 18 Ippolito G, Zumla A, Lanini S. Is there sufficient evidence to repeal three decades of clinical research on chronic hepatitis C? *Clin Microbiol Infect*. 2018 Apr;24(4):328–31.
- 19 Wolfson PJ. The development and use of extracorporeal membrane oxygenation in neonates. *Ann Thorac Surg*. 2003 Dec;76(6):S2224–9.
- 20 Gino F, Sharek Z, Moore DA. Keeping the illusion of control under control: Ceilings, floors, and imperfect calibration. *Organ Behav Hum Decis Process*. 2011 Dec;114(2):104–14.
- 21 Concato J. Is it time for medicine-based evidence? *JAMA*. 2012 Apr;307(15):1641–3.
- 22 Jane-wit D, Horwitz RI, Concato J. Variation in results from randomized, controlled trials: stochastic or systematic? *J Clin Epidemiol*. 2010 Jan;63(1):56–63.
- 23 Hunninghake GM. A new hope for idiopathic pulmonary fibrosis. *N Engl J Med*. 2014 May;370(22):2142–3.
- 24 Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ*. 2009 Oct;181(8):488–93.
- 25 Ioannidis J. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *Br J Sports Med*. 2017 Oct;51(20):1456–8.
- 26 Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018 Aug;210:2–21.
- 27 Horwitz RI, Singer B. Introduction. What works? And for whom? *Soc Sci Med*. 2018 Aug;210:22–5.
- 28 Collins R, Yusuf S, Peto R. Overview of randomised trials of diuretics in pregnancy. *Br Med J (Clin Res Ed)*. 1985 Jan;290(6461):17–23.
- 29 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol*. 1995 Jan;48(1):71–9.
- 30 Balon R, Segraves RT. Which Emperor Has New Clothes? Biology Versus Psychology in the Era of Statistical Magic. *J Sex Marital Ther*. 2017 Apr;43(3):201–5.
- 31 Pyke RE, Clayton AH. Psychological Treatment Trials for Hypoactive Sexual Desire Disorder: A Sexual Medicine Critique and Perspective. *J Sex Med*. 2015 Dec;12(12):2451–8.
- 32 Meixel A, Yancher E, Fugh-Berman A. Hypoactive sexual desire disorder: inventing a disease to sell low libido. *J Med Ethics*. 2015 Oct;41(10):859–62.