

How to Review a Meta-analysis

Mark W. Russo, MD, MPH

Dr. Russo is Medical Director of Liver Transplantation in the Department of Medicine at the Carolinas Medical Center in Charlotte, North Carolina.

Address correspondence to:
Mark W. Russo, MD, MPH
Medical Director of Liver Transplantation
Carolinas Medical Center, Transplant
Center, 3rd floor Annex Building
1000 Blythe Boulevard,
Charlotte, NC 28203;
Tel: 704-355-6649; Fax: 704-355-7184;
E-mail: Mark.Russo@carolinashealthcare.org

Abstract: Meta-analysis is a systematic review of a focused topic in the literature that provides a quantitative estimate for the effect of a treatment intervention or exposure. The key to designing a high quality meta-analysis is to identify an area where the effect of the treatment or exposure is uncertain and where a relatively homogenous body of literature exists. The techniques used in meta-analysis provide a structured and standardized approach for analyzing prior findings in a specific topic in the literature. Meta-analysis findings may not only be quantitative but also may be qualitative and reveal the biases, strengths, and weaknesses of existing studies. The results of a meta-analysis can be used to form treatment recommendations or to provide guidance in the design of future clinical trials.

Meta-analysis provides a standardized approach for examining the existing literature on a specific, possibly controversial, issue to determine whether a conclusion can be reached regarding the effect of a treatment or exposure. Results from a meta-analysis can refute expert opinion or popular belief. For example, Nobel Laureate Linus Pauling lectured for many years on the benefits of vitamin C in the treatment and prevention of the common cold. However, several years ago, a meta-analysis of clinical trials examining this issue demonstrated that there is no clear benefit of high doses of vitamin C on the common cold.¹

The first applications of meta-analysis were made more than 30 years ago in the psychiatric literature.² Meta-analysis later made its appearance in the gastroenterology literature, with one of its first applications occurring in the assessment of the effectiveness of antisecretory drug dosing for duodenal ulcers.³ Since then, meta-analysis has been applied to most conditions in gastroenterology and hepatology, including inflammatory bowel disease, cirrhosis, irritable bowel syndrome, and colon cancer.⁴⁻⁸

If it is well conducted, the strength of a meta-analysis lies in its ability to combine the results from various small studies that may have been underpowered to detect a statistically significant difference in effect of an intervention. For instance, 8 studies of streptokinase suggested its effectiveness in treating patients present-

Keywords

Meta-analysis, statistic, bias

Table 1. Checklist for Meta-analysis

Study question	<ul style="list-style-type: none"> • Objectives clearly stated • Clinically relevant and focused study question included • Effectiveness of intervention not convincingly demonstrated in clinical trials
Literature search	<ul style="list-style-type: none"> • Comprehensive literature search conducted • Searched information sources listed (ie, PubMed, Cochrane database) • Terms used for electronic literature search provided • Reasonable limitations placed on search (ie, English language) • Manual search conducted through references of articles, abstracts • Attempts made at collecting unpublished data
Data abstraction	<ul style="list-style-type: none"> • Structured data abstraction form used • Number of authors (>2) who abstracted data given • Disagreements listed between authors and how they were resolved • Characteristics of studies listed (ie, sample size, patient demographics) • Inclusion and exclusion criteria provided for studies • Number of excluded studies and reasons for exclusion included
Evaluation of results	<ul style="list-style-type: none"> • Studies were combinable • Appropriate statistical methods used to combine results • Results displayed • Sensitivity analysis conducted
Evaluation for publication bias	<ul style="list-style-type: none"> • Publication bias addressed through evaluation methods such as funnel plot or sensitivity analysis
Applicability of results	<ul style="list-style-type: none"> • Results were generalizable
Funding source	<ul style="list-style-type: none"> • Funding source(s) stated • No conflict of interest seen

ing with myocardial infarction, yet only 3 of these studies reported statistically significant results.⁹ Nevertheless, the results of a meta-analysis combining data across all 8 studies concluded that streptokinase was associated with a statistically significant reduction in mortality.⁹

As with the planning of any study, the study design of a meta-analysis determines the validity of its results. The Quality of Reporting of Meta-analyses (QUOROM) statement was published to provide guidelines for conducting meta-analyses, with the goal of improving the quality of published meta-analyses of randomized trials.¹⁰ A checklist assessing the quality of a meta-analysis has also been developed by the QUOROM group and is available online (<http://www.consort-statement.org/QUOROM.pdf>).

This article focuses on the key areas that the reader should be aware of to determine whether a meta-analysis was properly designed. These areas include the development of the study question; methods of literature search; data abstraction; proper use of statistical methods; evalu-

ation of results; evaluation for publication bias; sensitivity analysis; and applicability of results. A checklist for reviewing a meta-analysis is shown in Table 1. By utilizing a standardized approach for critiquing a meta-analysis, the internal validity of the analysis can be determined.

Development of the Study Question

The objectives of a meta-analysis and the question being addressed must be explicitly stated and may include primary and secondary objectives. The question at the focus of a meta-analysis should not have already been answered satisfactorily by the results of multiple well-conducted randomized trials. The more focused the question is, the more likely the study group will be homogenous. If the subjects across the studies are different, combining data from these studies is not appropriate. For example, a hypothetical meta-analysis on the effect of *Helicobacter pylori* eradication for reducing the risk of ulcer disease may not be useful or interesting because several studies

have already demonstrated a benefit in the eradication of *H. pylori* when ulcer disease is present. Furthermore, the study population would likely include patients with both gastric and duodenal ulcers, making the population heterogeneous. On the other hand, a meta-analysis comparing the efficacy of different antibiotic treatment regimens to eradicate *H. pylori* in patients with duodenal ulcers may be more relevant and would constitute a more homogenous study population.

The primary objective of a meta-analysis may not be solely to determine the effectiveness of an intervention. Results from a meta-analysis may be used to determine the appropriate sample size of a future trial, develop data for economic studies such as cost-effectiveness analyses, or demonstrate the association between an exposure and disease. Frequently, the results of a meta-analysis are used to highlight the weaknesses of previous studies and to recommend how to improve the design of future trials.

Literature Search

One of the first steps when reviewing a meta-analysis is to determine whether the authors conducted a comprehensive search for clinical trials and other types of studies, some of which may be unpublished, related to the research question. The information sources that were searched should be provided. Literature searches can include computerized and manual searches, which involve reviewing the references of an article's "ancestor search," as well as searching through abstracts, typically over the preceding 5 years. The most frequently used online resources for literature searches include PubMed, Cochrane Database, and Cancerlit. The Cochrane Collaboration was founded in 1993 and produces the Cochrane database of systematic reviews, which has generated more than 2,500 systematic reviews and meta-analyses (<http://www.cochrane.org/reviews/index.htm>). Reviews from the Cochrane database are typically of high quality and provide a helpful resource for those interested in performing or reviewing meta-analyses.¹¹ PubMed was developed by the US National Library of Medicine and includes over 17 million citations dating back to the 1950s. Cancerlit is produced by the US National Cancer Institute and is a database consisting of more than 1 million citations from over 4,000 sources dating back to 1963.

At least two reviewers should search sources for articles relevant to the meta-analysis, and the keywords used in the online searches should be provided in the article. Many authors include only full-length papers because abstracts do not always provide enough information to score the paper. The number of studies that were included and excluded should also be provided, as well as the reasons for exclusion.

Data Abstraction

Data abstraction is one of the most important steps in conducting a meta-analysis, and the methods of data abstraction that were used by the authors should be described in detail. In high-quality meta-analyses, a standardized data abstraction form is developed and utilized by the authors and may be provided in the paper as a figure. The reader of a meta-analysis should be provided with enough information to determine whether the studies that were included were appropriate for a combined analysis.

Two or more authors of a meta-analysis should abstract information from studies independently. It should be stated whether the reviewers were blinded to the authors and institution of the studies undergoing review. The results from the data abstraction are compared only after completing the review of the articles. The article should state any discrepancies between authors and how the discrepancies were resolved.

Results should be collected only from separate sets of patients, and the authors should be careful to avoid studies that published the same subjects or overlapping groups of subjects that appeared in different studies under duplicate publications. Raw numbers, in addition to risk ratios, should be recorded. Results from intention-to-treat analyses should be reported, when possible. The gold standard of data abstraction in a meta-analysis is to include patient level data from the studies combined in the meta-analysis, which usually requires contacting the authors of the original studies.¹² Obtaining patient level data may reveal differences among the trials that otherwise would not have been detected.

A quality score for each study included in a meta-analysis may be useful to ensure that better studies receive more weight. More than 20 instruments have been identified for the assessment for quality in both randomized clinical trials as well as meta-analyses of prospective cohort studies.¹³ Results can vary by the type of quality instrument, and a sensitivity analysis may need to be performed to determine the impact of the quality score on the results.¹³ As with data abstraction, two reviewers should score the quality of the studies using the same quality instrument, and results from the quality assessment should be compared. Agreement among the reviewers should be reported, and differences in quality scores should be reconciled through discussion.

As with clinical trials, inclusion and exclusion criteria for the studies included in the meta-analysis need to be well defined and established beforehand.¹⁴ One goal of inclusion and exclusion criteria is to create a homogenous study population for the meta-analysis.¹² The rationale for choosing the criteria should be stated, as it may not be apparent to the reader. Inclusion criteria may be based

on study design, sample size, and characteristics of the subject. Examples of exclusion criteria include studies not published in English or as full-length manuscripts. It has been reported that meta-analyses that restrict studies by language overestimate treatment effect only by 2%.¹⁰ The number of studies excluded from the meta-analysis and the reasons for the exclusions should also be provided.

Statistical Techniques

When determining whether a meta-analysis was properly performed, the statistical techniques used to combine the data are not as important as the methods used to determine whether the results from the studies should have been combined. If the data across the studies should not have been combined in the first place because their populations or designs were heterogeneous, statistical methods will not be able to correct these mistakes.

Two commonly used statistical methods for combining data include the Mantel-Haenszel method, which is based on the fixed effects theory, and the DerSimonian-Laird method, which is based on the random effects theory.¹⁵ One of the goals of these methods is to provide a summary statistic of an intervention's effect or exposure, as well as a confidence interval. The fixed effects model examines whether the treatment produced a benefit in the studies that were conducted. In contrast, the random effects model assumes that the studies included in the meta-analysis are a random sample of a hypothetical population of studies. The summary statistic is typically reported as a risk ratio, but it can be reported as a rate difference, person-time data, or percentage.

Arguments can be made for using either the fixed effects or random effects models, and sometimes results from both models are included. The random effects model provides a more conservative estimate of the combined data, with a wider confidence interval, and the summary statistic is less likely to be significant. The Mantel-Haenszel method can be applied to odds ratios, rate ratios, and risk ratios, whereas the DerSimonian-Laird method can be applied to ratios, as well as rate differences and incidence density (ie, person-time data).

The statistical test for homogeneity, which is also referred to as the test for heterogeneity, is frequently misused and misinterpreted as a test to validate whether the studies were similar and appropriate (ie, homogenous) to combine. The test may complement the results from data abstraction, supporting the interpretation that the studies were homogeneous and appropriate to combine. The test for homogeneity investigates the hypothesis that the size of the effect is equal in all included studies. $P < .1$ is considered to be a conservative estimate. If the test for homogeneity is significant, calculating a combined

estimate may not be appropriate. If this is the case, the reviewer should re-examine the studies included in the analysis for substantial differences among study designs or characteristics of subjects.

Evaluating the Results

Data abstraction results should be clearly presented in order for the reader to determine whether the included studies should have been combined in the first place. The meta-analysis should provide a table outlining the features of the studies, such as the characteristics of subjects, study design, sample size, and intervention, including the dose and durations of any drugs. Substantial differences in the study design or patient populations signify heterogeneity and suggest that the data from the studies should not have been combined.¹² For example, a meta-analysis was conducted on the risk of malignancy in patients with inflammatory bowel disease who were taking immunosuppressants.¹⁶ The patients had either ulcerative colitis or Crohn's disease and were taking azathioprine, 6-mercaptopurine, methotrexate, tacrolimus, or cyclosporine. Due to the differences in patient populations and types of treatment among the studies, the results from these studies should not have been combined.

The typical graphic displaying meta-analysis data is a Forest plot, in which the point estimate for the risk ratio is represented by a square or circle and the confidence interval for each study is represented by a horizontal line. The size of the circle or square corresponds to the weight of the study in the meta-analysis, with larger shapes given to studies with larger sample sizes or data of better quality or both. The 95% confidence interval is represented by a horizontal line except for the summary statistic, which can be shown by a diamond, the length of which represents the confidence interval.

Sensitivity analysis is an evaluation method employed when there is uncertainty in one or more variables included in the model or when determining whether the conclusions of the analysis are robust when a range of estimates is used. A sensitivity analysis is usually included in a meta-analysis because of uncertainty regarding the effectiveness or safety of an intervention. The values at the extremes of the 95% confidence intervals for risk estimates of key variables or areas with the most uncertainty can be included in additional modeling to determine the stability of the conclusions. For example, in a meta-analysis I conducted with colleagues on the efficacy and safety of transjugular intrahepatic portosystemic shunt (TIPS), the rates of new or worsening encephalopathy ranged from 17% to 60%.¹⁷ This range was incorporated into a sensitivity analysis to report the best and worst case scenarios for encephalopathy post-TIPS.

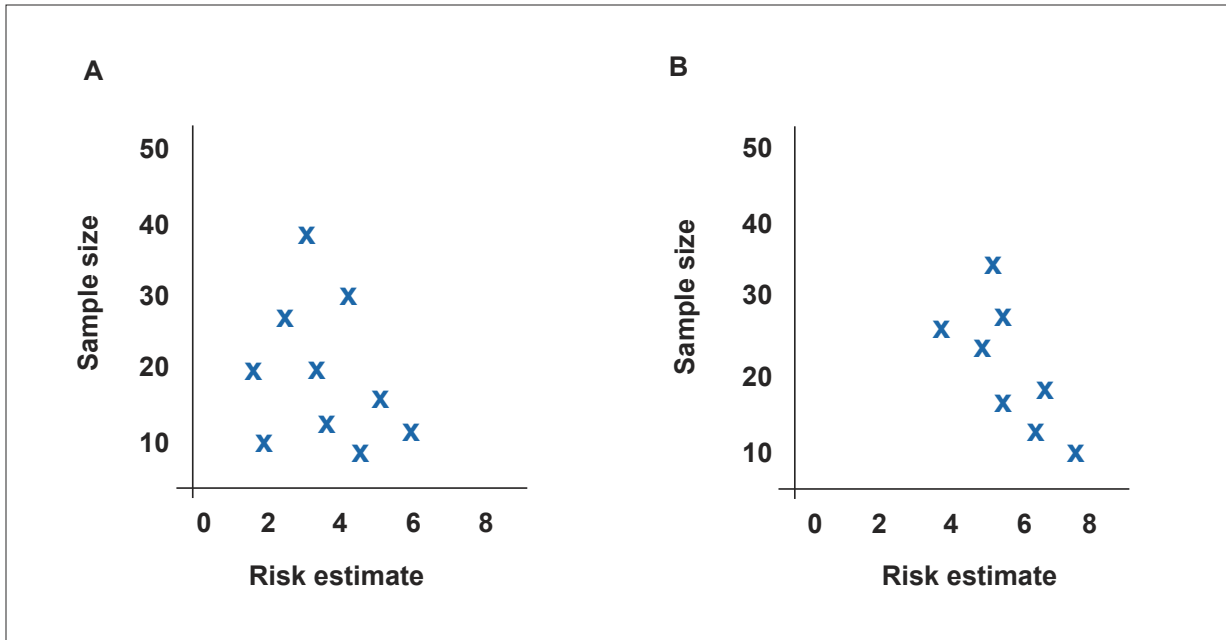


Figure 1. Example of funnel plot demonstrating no publication bias where the estimated true risk is 4. Risk estimates are evenly distributed around true risk (A). Example of funnel plot demonstrating publication bias where the estimated true risk is 4. Risk estimates cluster in lower right-hand corner, indicating that small studies with positive results are more likely to be published (B).

Assessing for Publication Bias

Meta-analyses are subject to publication bias because studies with negative results are less likely to be published and, therefore, results from meta-analyses may overstate a treatment effect. One strategy to minimize publication bias is to contact well-known investigators in the field of interest to discover whether they have conducted a negative study that remains unpublished. With the development of the National Library of Medicine's clinical trial registry (www.clinicaltrials.gov), researchers conducting meta-analyses have better opportunities to identify trials that are unpublished. Publication bias may lead to the overestimation of a treatment effect by up to 12%.¹⁰

A funnel plot can visually reveal the presence of a publication bias.¹⁸ A funnel plot is a graphic representation in which the size of the study on the y axis is plotted against the measure of effect on the x axis. Sampling error decreases as sample size increases and, therefore, larger studies should provide more precise estimates of the true treatment effect. In the absence of publication bias, smaller studies are scattered evenly around the base of the funnel (Figure 1A). In the presence of publication bias, small studies cluster around high-risk estimates with no or few small studies in the area of low-risk estimates

(Figure 1B). For example, a study reviewing the literature on the association between Barrett esophagus and esophageal carcinoma nicely demonstrated the presence of publication bias using funnel diagrams.¹⁹ Another method employed to address publication bias is a sensitivity analysis to determine the number of negative trials required to convert a statistically significant combined difference into a nonsignificant difference. Examples of these statistical methods to address publication bias include regression analysis, file-drawer analysis (failsafe N), and trim and fill analysis.¹⁸

Applicability of Results

The results of a meta-analysis, even if they are statistically significant, must have utility in clinical practice or constitute a message for researchers in the planning of future studies. The results must have external validity or generalizability and must impact the care of an individual patient. In addition, the studies included in the meta-analysis should include patient populations that are typically seen in clinical practice. There should be a balance between finding studies that are similar and appropriate to combine without becoming too focused, in order to avoid a study population that is too narrow.

Meta-analysis Beyond Randomized, Clinical Trials

Although randomized clinical trials are usually the focus of a meta-analysis, the same methodology used for randomized trials can be applied to cohort studies.²⁰ The Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group has proposed a checklist for conducting meta-analyses of prospective studies,²¹ which is similar to the checklist for randomized trials proposed by QUOROM and the checklist shown in Table 1.

There may be settings in which randomized data are not available, such as the association between a risk factor and cancer. For example, combined data from 21 prospective studies in a meta-analysis demonstrated a significant association between body mass index and pancreatic cancer.²² The dangers of combining results from cohort studies is that bias is more likely to be introduced in cohort studies than in randomized trials and the study populations among cohort studies are more likely to be heterogeneous. Nevertheless, if there are multiple cohort studies in an area of interest and few or no randomized clinical trials, then the results from a meta-analysis may emphasize the need for one or more randomized trials and provide recommendations for optimal study design.

Summary

Meta-analysis can be a powerful tool to combine results from studies with similar design and patient populations that are too small or underpowered individually to demonstrate a statistically significant association. As with clinical trials, having an appropriate study question and design are essential when performing a meta-analysis to ensure that there is internal validity and that the results are clinically meaningful. Heterogeneity among studies in study designs or patient populations is one of the most common flaws in meta-analyses. Heterogeneity can be avoided by thoughtful data abstraction performed by two or more authors who use a standardized data abstraction form. By applying a systematic approach to meta-analysis, many of the pitfalls can be avoided.

References

- Douglas RM, Hemila H, D'Souza R, et al. Vitamin C for preventing and treating the common cold. *Cochrane Database Syst Rev.* 2004;4:CD000980.
- Smith ML, Glass GV. Meta-analysis of psychotherapy outcomes studies. *Am Psychol.* 1977;32:752-760.
- Jones DB, Howden CW, Burget DW, et al. Acid suppression in duodenal ulcer: a meta-analysis to define optimal dosing with antisecretory drugs. *Gut.* 1987;28:1120-1127.
- Triester SL, Leighton JA, Leontiadis GI, et al. A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with non-stricturing small bowel Crohn's disease. *Am J Gastroenterol.* 2006;101:954-964.
- Su C, Lichtenstein GR, Krok K, et al. A meta-analysis of the placebo rates of remission and response in clinical trials of active Crohn's disease. *Gastroenterology.* 2004;126:1257-1269.
- Shi J, Wu C, Lin Y, et al. Long-term effects of mid-dose ursodeoxycholic acid in primary biliary cirrhosis: a meta-analysis of randomized controlled trials. *Am J Gastroenterol.* 2006;101:1529-1538.
- Moskal A, Norat T, Ferrari P, Riboli E. Alcohol intake and colorectal cancer risk: a dose-response meta-analysis of published cohort studies. *Int J Cancer.* 2007;120:664-671.
- Quartero AO, Meineche-Schmidt V, Muris J, et al. Bulking agents, antispasmodic and antidepressant medication for the treatment of irritable bowel syndrome. *Cochrane Database Syst Rev.* 2005;2:CD003460.
- Stampfer MJ, Goldhaber SZ, Yusuf S, et al. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. *N Engl J Med.* 1982;307:1180-1182.
- Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet.* 1999;354:1896-1900.
- Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA.* 1998;280:278-280.
- Chalmers TC, Buyse M. Meta-analysis. In: Chalmers TC, Blum A, Buyse M, et al, eds. *Data Analysis for Clinical Medicine: The Quantitative Approach to Patient Care.* Rome, Italy: International University Press; 1988:75-84.
- Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054-1060.
- Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ.* 1997;315:672-675.
- Petitti DB. Statistical Methods in Meta-analysis. *Meta-analysis, Decision Analysis, and Cost-effectiveness Analysis.* New York, NY: Oxford University Press; 1994:90-114.
- Masunaga Y, Ohno K, Ogawa R, et al. Meta-analysis of risk of malignancy with immunosuppressive drugs in inflammatory bowel disease. *Ann Pharmacother.* 2007;41:21-28.
- Russo MW, Sood A, Jacobson IM, Brown RS Jr. Transjugular intrahepatic portosystemic shunt for refractory ascites: an analysis of the literature on efficacy, morbidity, and mortality. *Am J Gastroenterol.* 2003;98:2521-2527.
- Rothstein HR, Sutton AJ, Borenstein M, eds. Publication bias in meta-analysis. In: *Publication Bias in Meta-analysis: Prevention, Assessment, and Adjustments.* Chichester, UK: John Wiley & Sons, Ltd; 2005:1-7.
- Shaheen NJ, Crosby MA, Bozymski EM, Sandler RS. Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology.* 2000;119:333-338.
- Russo MW, Goldsweig CD, Jacobson IM, Brown RS Jr. Interferon monotherapy for dialysis patients with chronic hepatitis C: an analysis of the literature on efficacy and safety. *Am J Gastroenterol.* 2003;98:1610-1615.
- Stroup DF, Berlin JA, Morton S, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA.* 2000;283:2008-2012.
- Larsson SC, Orsini N, Wolk A. Body mass index and pancreatic cancer risk: a meta-analysis of prospective studies. *Int J Cancer.* 2007;120:1993-1998.