

TUTORIAL

Meta-analysis: Key features, potentials and misunderstandings

Olaf M. Dekkers MD, PhD, MA, MSc 

Department of Clinical Epidemiology and
Department of Endocrinology, Leiden
University Medical Center, Leiden,
The Netherlands

Correspondence

Olaf M. Dekkers, Department of Clinical
Epidemiology, Leiden University Medical
Center, Leiden, The Netherlands.
Email: o.m.dekkers@lumc.nl

Abstract

A meta-analysis consists of a systematic approach to combine different studies in one design. Preferably, a protocol is written and published spelling out the research question, eligibility criteria, risk of bias assessment, and statistical approach. Included studies are likely to display some diversity regarding populations, calendar period, or treatment settings. Such diversity should be considered when deciding whether to combine (some) studies in a formal meta-analysis. Statistically, the fixed effect model assumes that all studies estimate the same underlying true effect. This assumption is relaxed in a random effects model and given the expected study diversity a random effects approach will often be more realistic. In the absence of statistical heterogeneity, fixed and random effects models give identical estimates. Meta-analyses are especially useful to provide a broader scope of the literature; they should carefully explore sources of between study heterogeneity and may show a treatment effect or an exposure–outcome association where individual studies are not powered. However, its validity largely depends on the validity of included studies.

KEYWORDS

heterogeneity, meta-analysis, protocol, statistical analysis, tutorial

Essentials

- A systematic review aims to appraise and synthesize the available evidence addressing a specific research question; a meta-analysis is a statistical summary of the results from relevant studies.
- A meta-analysis will provide a non-valid answer if included studies are not valid. Judgment of validity of individual studies is thus crucial.
- When deciding whether to perform a formal meta-analysis, study diversity and statistical heterogeneity should be considered.

1 | INTRODUCTION

In 2007, a meta-analysis showed an increased cardiovascular risk in patients using rosiglitazone, an anti-diabetic drug.¹ Here, a meta-analytic approach displayed its full potential: to show a side-effect for which individual studies were not powered. Similarly, the increased cardiovascular risk for rofecoxib was shown in a meta-analysis.²

The rosiglitazone paper was firmly debated; for example, the statistical approach was criticized.³ This highlights that meta-analyses are not immune to criticism, despite being conceived as high-level evidence. Choices in design or analysis of a review can be debated or criticized by readers, underlining the need for transparent reporting.

This tutorial discusses central features of meta-analyses, including potential misconceptions. A glossary (Table 1) provides an explanation of the methodological terms used; displayed terms are shown

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. *Research and Practice in Thrombosis and Haemostasis* published by Wiley Periodicals, Inc on behalf of International Society on Thrombosis and Haemostasis.

TABLE 1 Glossary with short explanation of technical terms used in meta-analyses

Term	Explanation
Cochrane's Q-test	Statistical test that examines the null-hypothesis that all studies have the same true effect. ¹³ A significant <i>P</i> value provides evidence of statistical heterogeneity. The test is based on the deviations of study estimates from the overall mean
Fixed effect model	Statistical method to obtain a weighted average of study estimates. Studies are weighted according the inverse of the variance, meaning that larger studies bear more weight. The fixed effect model assumes that included studies estimate the same underlying true effect
Forest plot	Graphical display of effect estimates of individual studies, often presented with a weighted estimate. Forest plots display studies' effect estimates and 95% confidence intervals, the weight the studies get in the meta-analysis (shown as box and/or percentage) and the overall weighted estimate with a 95% confidence interval
Funnel plot	A funnel plot is a graphical display plotting effect estimates against sample size or inverse of the variance. The idea behind a funnel plot is that study effects scatter around a mean effect, but that smaller studies can deviate more from this mean. Publication bias may be considered if smaller studies show on average a more positive effect than larger studies. Smaller studies are more prone to only getting published if the result is positive, large trials tend to get published anyway. There are statistical techniques to judge whether these small studies show a different effect compared to larger studies ¹⁸
I ² statistic	Measure to quantify the amount of heterogeneity between studies that cannot be explained by chance. It is quantified as a percentage between 0 and 100; as a general rule low, moderate, and high heterogeneity can be assigned to I ² values of 25%, 50%, and 75% ¹³
Individual Patient Data (IPD) meta-Analysis	In standard meta-analyses the individual study is the unit of analysis. In an IPD meta-analysis the researchers have access to data at the level of individual patients from different studies. This is especially useful to harmonize endpoints and perform analyses in prespecified subgroups
Meta-regression	Statistical technique to relate study characteristics to effect estimates. ¹⁹ For example the association between treatment duration and treatment effect for depression was studied in a meta-regression framework ²⁰
Network meta-analysis	A network meta-analysis allows the comparison of more than two groups. ²¹ This can be a useful approach when more than two treatment options exist for the same indication. An example is a comparison of the thrombosis risk of different oral contraceptives ²²
Random effects model	Statistical method to obtain a weighted average of study estimates. In contrast to a fixed effect model, a random effects model assumes that studies have different underlying true effects. The combined effect in a random analysis is an estimation of the mean of these underlying true effects. Technically, the random effects model takes the between-study variation into account
Subgroup analysis	Restricting the statistical analysis to a group of studies based on a specific characteristic. For example, an analysis can be restricted to randomized studies, studies with low risk of bias, or studies performed in children. Subgroup analyses can be used as a way to explore heterogeneity

in the main text in italics. In Table 2, ten potential misunderstandings in meta-analyses are shown.

2 | RESEARCH QUESTION AND STUDY PROTOCOL

Similarly to other study designs, systematic reviews start with a research question to be specified in terms of population, intervention or risk factor, or control group and outcome(s). Defining the research question is a balancing act: if very narrow, the review may end up with few studies only (for example: effect of 40 mg simvastatin on recurrent thromboembolism in 60- to 70-year-old males); if too broad, a meaningful overview may be cumbersome (for example: effect of inflammation on coagulation). A rough idea on the number of publications on a particular topic may guide framing of the research question.

A related point is that within a systematic review framework, researchers are dependent on how studies are performed and reported. If the effect of wine on coagulation is assessed in a review,

some studies may report on the association between coagulation and alcohol in general, without providing data for wine-drinkers only. Such considerations should be taken into account up front when deciding on inclusion criteria. For the wine-coagulation example, no formal rule will decide whether broad alcohol categories provide meaningful information to the review. This decision is up to the researchers, but arguments should be provided in the protocol and the paper that clarify decisions made.

It is advised to write and publish a protocol specifying details of design and analyses.⁴ The advantage prespecifying a protocol is that decisions will be made independently of study results.

3 | SEARCHING STUDIES AND EXTRACTING DATA

Many electronic databases can be used for searching studies; Embase, Medline, and Cochrane Library are the most well-known. More than one database should be searched, given the incomplete

TABLE 2 Ten potential misunderstandings in meta-analyses

Potential misunderstanding	Background
A meta-analysis is an objective procedure	Every meta-analysis is characterized by decisions regarding research question, eligibility criteria, risk of bias analysis, and statistical approach. These decisions should be reasonable and transparently reported. Probably, no single best and ultimately objective procedure exists. For this reason, different meta-analyses on the same topic may come to different conclusions
A meta-analysis provides the highest level of evidence	A meta-analysis is generally considered to provide high-level evidence. However, the validity of a meta-analysis depends largely on the validity of included studies ("garbage in—garbage out"); a meta-analytic design is thus not a guarantee for highest level evidence
Study quality is synonymous with risk of bias	Study quality is about the question whether a study has been optimally performed; risk of bias relates to threats of validity. A study can be high quality but still have a high risk of bias for certain bias domains. An example is a comparison between two surgical techniques; even if the study is optimally performed, it cannot, by design, be blinded
A risk of bias analysis resolves the bias	A risk of bias analysis mainly displays this bias risk; such a display does not resolve it, although a sensitivity analysis restricted to low risk of bias studies can be considered
Random effects models solve heterogeneity	Random effects models allow that different studies have a different underlying true effect; the random effects model thus does not explain, solve, or even remove heterogeneity
Assuming homogeneity between studies when the statistical test fails to show heterogeneity	In the presence of few studies only, tests for heterogeneity have low power; the presence of a nonsignificant test does thus not provide strong evidence for true homogeneity between studies. This is especially the case if the review includes <10 studies
Present the I^2 statistic as if it was a test	The I^2 statistic is formally not a test that can reject a null hypothesis. It provides a quantitative measure of the heterogeneity between studies beyond chance ¹³
Assuming funnel plot symmetry when the statistical test fails to show heterogeneity	In the presence of few studies only, the test for heterogeneity has low power; the absence of a nonsignificant test does thus not provide strong evidence for symmetry
Funnel plot asymmetry proves publication bias	Funnel plot asymmetry means that smaller studies show on average a different effect compared to larger studies; one explanation is publication bias, other explanations are effect modification and chance
Meta-analyses "speak for themselves"	Even meta-analyses need an interpretation. ¹⁶ Such an interpretation pertains to questions on validity, heterogeneity, and clinical relevance. For example, a recent review concluded that low-molecular-weight heparin lowered the risk of venous thromboembolism in patients with lower-limb immobilization. ¹⁸ The translation of this review to clinical practice requires a discussion whether it is clinically relevant to reduce thrombosis found by routine ultrasound screening, which was the way to assess the endpoint in most included papers

coverage of single databases.⁵ As writing database-specific search strings requires specific bibliographical knowledge, a search should be developed in collaboration with an information specialist. To optimize the search process, key articles should be provided to the information specialist, as they may contain clues to keywords and indexing. Additionally, references of key papers can be checked.

Inclusion of papers follows from the eligibility criteria. Data extraction should be done by two researchers independently. This reduces the error rate, but it also may help to discuss choices to be made.⁶ For example, a paper may present effects of wine drinking

on coagulation factors using different sets of adjusted confounders. Researchers have to decide which effect estimate to extract for a formal meta-analysis.

4 | RISK OF BIAS ANALYSIS

A meta-analysis will provide a nonvalid answer if included studies are not valid. The judgment of validity of individual studies is referred to as risk of bias assessment. For randomized trials risk of bias

assessment is standardized, and elements to be judged are concealment of allocation, blinding (participants, personnel and outcome assessors), selective outcome reporting, and incomplete outcome data.⁷ For these domains, researchers judge included studies for their risk of bias, which is reported at the study level. For example, an unblinded study is judged “high risk of bias” with respect to blinding. The full risk of bias analysis is preferably tabulated to facilitate an overview and to provide an overall idea on the validity of included studies. Although we are actually judging risk of bias, there is evidence that risk of bias is related to actual bias. For example, unblinded studies show on average a more positive effect than adequately blinded studies.⁸ This was shown in detail for a study assessing the effect of rosiglitazone on cardiovascular endpoints, where the unblinded design was the likely cause of case-forms being filled in more often in favor of the study drug.⁹ As risk of bias depends on reporting, some elements cannot always be judged; for example, loss-to-follow up (potentially introducing selection bias¹⁰) is often poorly reported, especially in observational studies.

The crucial distinction between randomized and observational studies is the potential for confounding and judging the risk of confounding is key for reviews of observational studies. Confounding means that compared groups are different with respect to important prognostic factors. In the context of interventions, this is called confounding by indication. It is a decision of the researchers what is considered a sufficient set of confounders to be adjusted for, and what statistical techniques are considered adequate. The mere statement that a study suffers from confounding because it is observational is too simplified. Moreover, confounding is a matter of degree. For the association between a vegetarian diet and thrombosis, the confounding will be almost intractable, whereas it has been shown that side effects can reliably be assessed in observational studies, as confounding is only marginally an issue.^{11,12} Importantly, standard statistical techniques (including propensity scores) cannot fully adjust for unmeasured confounding.

For observational studies, two other bias domains are also important. Misclassification can be of either risk factor or intervention under study, the outcome, or both. Selection bias refers to the bias introduced by selection mechanisms in studies. This type of bias requires methodological expertise and is often difficult to detect. Guidance exists for risk of bias assessment for observational studies on therapeutic interventions.¹⁰

Three approaches can be considered to incorporate results from a risk of bias analysis.

1. Researchers can restrict their meta-analysis to studies with low risk of bias. This is the approach for many meta-analyses on therapeutic interventions where observational studies are not eligible.
2. If risk of bias is considered too high, researchers may want to abstain from statistical combining the results.
3. Exploration (by *meta-regression* or *subgroup analyses*), this approach tries to answer the question whether risk of bias influences reported effect estimates.

5 | DIVERSITY AND HETEROGENEITY

By design, a systematic review includes studies from different patient populations. Likely, these different populations will display diversity, for example with regard to clinical characteristics, study period, or health care facilities. Some diversity is thus inevitable. Researchers should display such between-diversity as it will facilitate a judgment whether included papers “tell a similar story.”

Statistical assessment of heterogeneity only considers the heterogeneity of the quantitative estimates. Such statistical measures (*Cochrane’s Q-test* and *I² statistic*) approach the question of whether differences in effect estimates are beyond chance. Statistical tests are not powered to detect heterogeneity when the review includes fewer than 10 papers.¹³ Study diversity may translate into statistical heterogeneity, but this does not need to be the case. Researchers should thus, when thinking about heterogeneity and deciding whether to perform a formal meta-analysis, take into account both the clinical judgment and the statistical verdict.

6 | STATISTICAL ANALYSES

From a statistical point of view, meta-analyses are fairly simple: the pooled estimate is a weighted average of effect estimates of individual studies. The weighting is according to the inverse of the variance, which means that larger studies get more weight. A *forest plot* is a graphical display of a meta-analysis’ results.

Researchers have two basic statistical options to perform a meta-analysis: a fixed and a random effects model. The *fixed effect model* assumes that all studies have the same underlying true effect; this assumption is rigid (Do we really have certainty that all studies only differ due to chance?). A *random effects model* relaxes this assumption and does not assume that all studies have the same underlying true effect.

When comparing the two models, smaller studies get relatively more weight in a random model, and the confidence intervals are wider in a random effects model. This is shown in Figure 1, a graphical display of both models. What is the correct model? There is no final answer,¹⁴ although it is often realistic to assume some underlying heterogeneity and start with a random model. In case of absence of statistical heterogeneity, the two models give identical results.

Mostly, fixed and random effects models give very similar pooled estimates, the main difference being the wider confidence interval for the random model. There is an exception to this rule, when on average effects from smaller studies are different from effects in larger studies, as in Figure 1. As expected, the confidence interval from the fixed effect model is smaller, but there is also clear difference in pooled estimate.

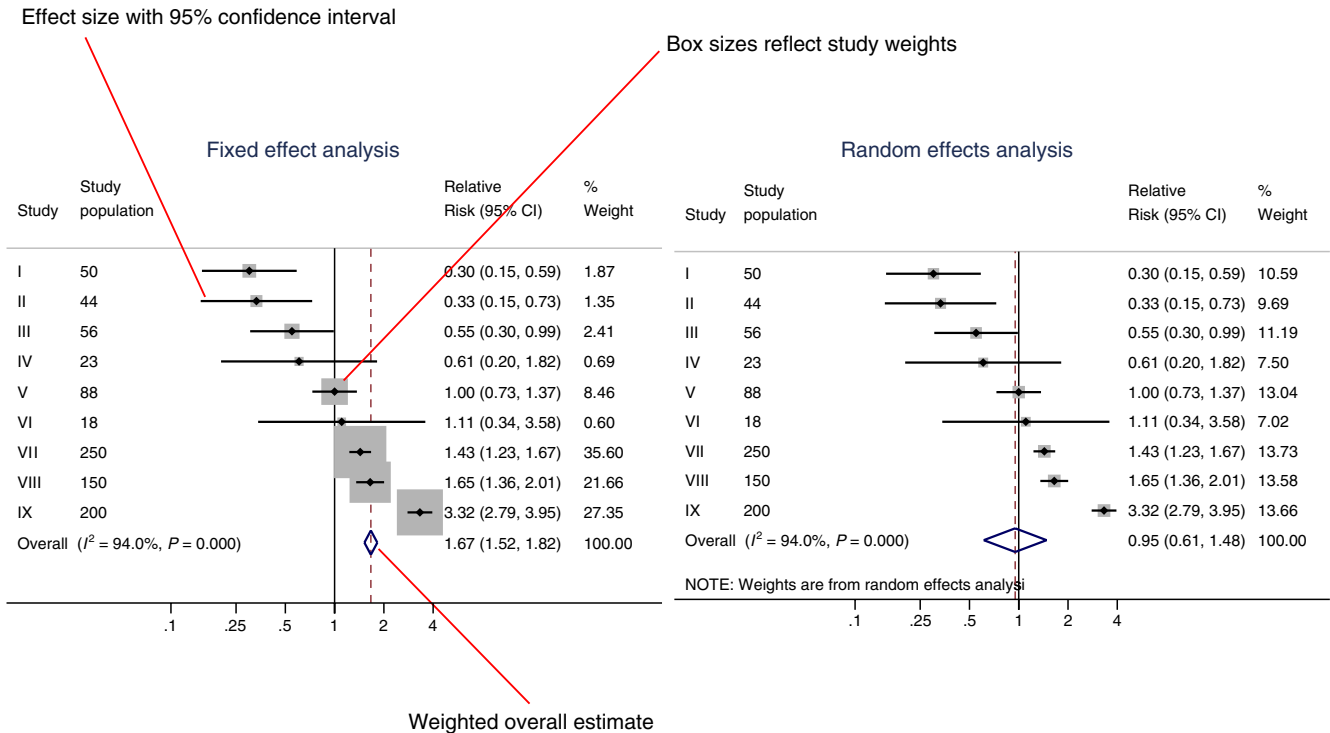


FIGURE 1 Graphical display of a fixed and random effects model. Forest plot showing two different statistical meta-analytic approaches for the same set of (fictional) studies: a fixed effect model (left) and a random effects model (right). In the forest plot effect estimates of individual studies, the study weights, a weighted overall effect and measures of heterogeneity (I^2 statistic and a P value for the heterogeneity test) are shown. CI, confidence interval

7 | PUBLICATION BIAS

Risk of bias refers to bias at the level of individual studies; publication bias distorts the overall picture. Publication bias occurs when studies with statistically significant positive effects are more likely to get published. There are many reasons for negative studies remaining unpublished, such as lower motivation of authors to finalize or submit negative studies, and unwillingness of journals to publish “uninteresting results.” Publication bias will often result in a too-positive picture of an intervention. This was shown for antidepressants, where published papers showed a 50% greater treatment effect, compared to unpublished papers on the same drug.¹⁵ Although often publication bias is considered a problem of meta-analyses, it is clearly a broader problem the moment a systematic overview is used to inform doctors, patients, and policymakers.

A *funnel plot* can facilitate the judgment whether publication bias is an issue. Publication bias may be considered if smaller studies show on average a more positive effect than larger studies.

8 | ANSWERING THE RESEARCH QUESTION

A systematic review provides an optimal opportunity to place studies in a broader context.¹⁶ Sometimes the interpretation

of a meta-analysis is straightforward, when all studies give the same picture. This was the case in studies on the association between acromegaly and mortality, where all studies showed a slightly increased risk, which became significant in a meta-analysis.¹⁷

But in other cases the interpretation is more difficult, for example when one or two large trials show effects not directly comparable to the weighted average of a much larger number of trials. Researchers should then carefully balance the arguments for a decision: are the two large trials less likely to be biased, or is the weighted estimate closer to the truth?

In summary, meta-analyses are especially useful to provide a broader scope of the literature; they should carefully explore sources of between study heterogeneity, and may show a treatment effect or an exposure–outcome association where individual studies are not powered. However, its validity largely depends on validity of included studies.

RELATIONSHIP DISCLOSURES

The author has nothing to disclose.

ORCID

Olaf M. Dekkers <http://orcid.org/0000-0002-1333-7580>

REFERENCES

1. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007;356:2457-71.
2. Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet*. 2004;364:2021-9.
3. Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. *Ann Intern Med*. 2007;147:578-81.
4. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350:g7647.
5. Lemeshow AR, Blum RE, Berlin JA, Stoto MA, Colditz GA. Searching one or two databases was insufficient for meta-analysis of observational studies. *J Clin Epidemiol*. 2005;58:867-73.
6. Tendal B, Higgins JP, Juni P, et al. Disagreements in meta-analyses using outcomes measured on continuous or rating scales: observer agreement study. *BMJ*. 2009;339:b3128.
7. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
8. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42-6.
9. Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. *JAMA*. 2010;304:793-4.
10. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
11. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med*. 2011;8:e1001026.
12. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet*. 2004;363:1728-31.
13. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-60.
14. Mueller M, D'Addario M, Egger M, et al. Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC Med Res Methodol*. 2018;18:44.
15. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358:252-60.
16. Blair A, Burg J, Foran J, et al. Guidelines for application of meta-analysis in environmental epidemiology. ISLI Risk Science Institute. *Regul Toxicol Pharmacol*. 1995;22:189-97.
17. Dekkers OM, Biermasz NR, Pereira AM, Romijn JA, Vandenbroucke JP. Mortality in acromegaly: a metaanalysis. *J Clin Endocrinol Metab*. 2008;93:61-7.
18. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.
19. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21:1559-73.
20. Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. *BMJ*. 2001;322:763-7.
21. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med*. 2013;159:130-7.
22. Stegeman BH, de Bastos M, Rosendaal FR, et al. Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis. *BMJ*. 2013;347:f5298.

How to cite this article: Dekkers OM. Meta-analysis: Key features, potentials and misunderstandings. *Res Pract Thromb Haemost*. 2018;2:658-663. <https://doi.org/10.1002/rth2.12153>