Meta-analysis: a critical appraisal of the methodology, benefits and drawbacks

ABSTRACT

Meta-analysis has become an integral part of evidence-based decision-making processes and is being increasingly used in medical and non-medical disciplines. Aggregate data or summary statistics continue to be the mainstay of meta-analysis and are used by many professional societies to support clinical practice guidelines. Meta-analyses synthesize the summary statistics from independent trials by pooling them to estimate the underlying common effect size. The results represent the highest level of evidence but only if the chosen studies are of high quality and the selection criteria are fully satisfied. It is important to address the issues of defining an explicit and relevant question, exhaustively searching for the totality of evidence, meticulous and unbiased data transfer or extraction, assessment of between study heterogeneity and the use of appropriate statistical methods for estimating summary effect measures. This article reviews the methodology, benefits and drawbacks of performing a meta-analysis.

In the 21st century, scientists and policy makers rely on an evidence-based decision-making process to guide them towards early interventions, better treatment methods and structured guidelines which work efficiently and reliably to produce the best possible outcomes. The validity of evidence, and hence the decisions based on this, relies on the quality of the data, the methodology used to extract them, the robustness and totality of the evidence and use of relevant methods to analyse them. Hence, decision makers need to be aware of factors that impact the quality of evidence along with any shortcomings in the process of gathering and processing such evidence. The use of meta-analysis as part of systematic review enables inclusion of analysis of quantitative data from independent trials in decision-making processes (Khan et al, 2016). However, problems arise in meta-analysis. These include:

- Regressions are often non-linear
- Effects are often multivariate rather than univariate
- Coverage can be restricted
- Bad studies may be included
- The data summarized may not be homogeneous
- Grouping different causal factors may lead to meaningless estimates of effects
- The theory-directed approach may obscure discrepancies (Eysenck, 1994).

This article covers a range of issues relating to the meta-analysis, and critically investigates its drawbacks and benefits.

Professor Shahjahan Khan, Professor of Statistics, School of Agricultural, Computing and Environmental Sciences, International Centre for Applied Climate Sciences and Centre for Health Sciences Research, University of Southern Queensland, Toowoomba, Queensland, Australia

Ms Breda Memon, Research and Clinical Nurse, South East Queensland Surgery and Sunnybank Obesity Centre, McCullough Centre, Sunnybank, Queensland, Australia

Professor Muhammed A Memon, Consultant Surgeon, Mayne Medical School, School of Medicine, University of Queensland, Brisbane, Queensland, Australia

Correspondence to: Professor MA Memon (mmemon@yahoo.com)

Prelude to meta-analyses: systematic reviews

A systematic review is the process of searching, gathering and investigating the relevant literature on a specific topic of interest. Ideally the results of meta-analyses should be rigorous, comprehensive, transparent, free from bias and reproducible. Khan et al (2003) describe five steps in performing a systematic review to ensure its objectivity:

1. Framing questions for a review
2. Identifying relevant work
3. Assessing the quality of studies
4. Summarizing the evidence
5. Interpreting the findings.

To ensure that the evidence is of highest quality, various researchers tried to devise processes, criteria and protocols to prevent biases and design flaws to improve the quality of reviews and evidence. The Quality of Reporting of Meta-analyses (QUOROM) addresses standards for improving the quality of meta-analysis of clinical randomized controlled trials (Moher et al, 1999). The Consolidated Standards of Reporting Trials (CONSORT) (Moher et al, 2010) encompass various initiatives to deal with the problems arising from inadequate reporting of randomized controlled trials. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al, 2009) is an evidence-based minimum set of items for reporting in systematic review and meta-analysis. The authors of PRISMA later introduced the PRISMA protocols (Moher et al, 2015; Shamseer et al, 2015). The Meta-analysis Of Observational Studies in Epidemiology (MOOSE) was proposed by Stroup et al (2000) which contains a checklist of specifications for reporting meta-analysis of observational studies.

Meta-analysis

Glass (1976) introduce meta-analysis as a statistical procedure to re-analyse the published statistical results from a large number of independent studies on a specific topic for the purpose of integrating the findings in the context of educational research. He felt that when faced with an

abundance of information, the goal should be to extract the quantifiable information, group them in an orderly fashion according to categories, pooling the data and summarizing these results. Meta-analysis is described as 'a statistical analysis that combines or pools the results of several independent clinical trials considered by the analyst to be "combinable" in which the primary aim is attaining an estimate of average effect size attributable to a certain intervention presented in the same metric' (Egger and Smith, 1997; Smith and Egger, 1998; Huedo-Medina et al, 2006).

Appraisal of meta-analysis

Quality of studies included

Quality assessment of a study is necessary in order to prevent misleading results based on invalid or poor quality studies. Quality (or validity) involves some measure of the methodological strength of the relevant study, or how able it is, through its design and its conduct, to prevent systematic errors, or bias. Pooling results from low levels of evidence, e.g. retrospective trials, with those with a high level of evidence, e.g. randomized controlled trials, reduces the quality of the synthesized results and may lead to invalid conclusions.

Several methods have been described to evaluate trial quality. The Jadad scale or Oxford Quality Scoring System (Jadad et al, 1996), based on reporting of randomization, blinding and withdrawals, is the most widely used assessment tool because of its simplicity. Another tool is the Cochrane Risk of Bias Tool, which is based on a number of domains such as selection bias, performance bias, detection bias, attrition bias, reporting bias and other bias (*Figure 1*) (Higgins et al, 2011). The Newcastle–Ottawa scale (Stang, 2010) (*Figure 2*) evaluates the quality of observational studies based on three broad categories: the selection of the study groups, the comparability of the groups, and the ascertainment of either exposure or outcome of interest for case control or cohort studies respectively.

Other tools for non-randomized study assessment include the Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I) assessment tool (Sterne et al, 2016). It outlines seven domains where biases might occur: two in the 'pre-intervention' phase, one in the 'at intervention' phase and four in the 'post-intervention' phase. It is beyond the scope of this article to provide a detailed account of these tools.

Reporting variability of summary statistics and how to combine the data

The summary statistics or aggregate data of the outcome variables are reported in the individual studies using different units of measurement. Depending on the type of outcome variable, the summary statistics could be mean and standard deviation, or correlation coefficient (measuring the relationship of two quantitative outcome variables) for continuous outcome variables, and odds ratio, risk ratio and risk difference for categorical outcome variables, along with the sample size.

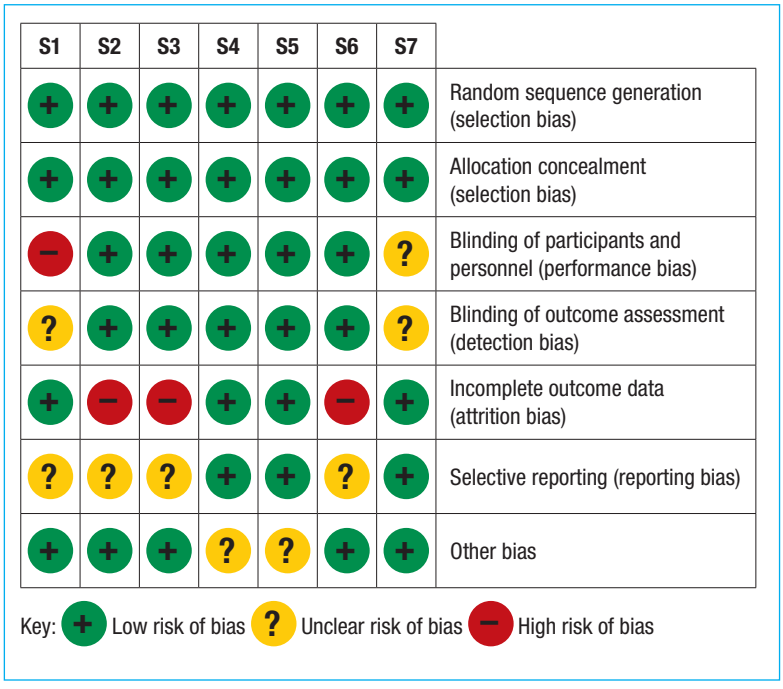

Figure 1. The Cochrane Risk of Bias Tool representing various domains. S1–S7 represent various randomized controlled trials.

For the computation of any confidence interval for an unknown population effect size, the point estimate of effect size for each of the individual studies, along with their standard deviation and sample size, is essential. Also, the sampling distribution of the estimator of the population effect size must be identifiable in order to be able to determine the critical value of the underlying statistic at a

Author/ Year	Selection			Comparability		Outcome assessment		Total quality
	1	2	3	4	5	6	7	
A1/1990	*	*	*	*	*	*	*	*******
B2/1995	*	*	*	*		*		*****
C3/2000	*	*	*	*	*	*	*	*******
D4/2005	*	*	*		*	*	*	******
E5/2010	*	*	*	*	*	*		******
F6/2015	*	*	*	*	*	*		******

Selection: *1) Is the case definition described? 2) Was the sample truly representative of the total population? 3) How was the ascertainment of exposure done? (Each affirmative answer gets one star)*
Comparability: *4) Did the study have no difference between study 1 and study 2 groups? The main factors taken into account while calculating this were various outcomes, e.g. prior treatment, pain score, complications, mentioned. 5) One more star was given if the following factors were comparable in the two groups; age, gender of patient, operative time recorded. (If both were affirmative then two stars, even if one or more of the above mentioned criteria were absent. However, no stars were given if the groups differed entirely.)*
Outcome assessment: *6) Clear assessment of outcomes via record linkage. 7) Adequacy of the cohort follow ups – whether the follow ups were completed or less than 20% of patients were lost to follow up.*

Figure 2. Newcastle–Ottawa scale for assessing the quality of non-randomized studies. A1–F6 represent authors of these studies.

> Meta-regression is an extension to subgroup analyses that allows the effect of continuous as well as categorical characteristics to be investigated.

predetermined confidence level. Such a critical value and the standard error of the estimator are used to calculate the margin of error for the confidence interval.

If the raw (individual patient) data from all selected studies are available, then one could analyse the data by using mega-analysis methods. Unfortunately, the sample effect size (in individual studies) is a random variable as it differs from study to study. The study-specific values of the estimated effect size are not only different but may have opposing results, producing contradictory evidence. In the face of conflicting evidence from different primary studies, the challenge is to reconcile the results to come up with a valid estimated common effect size.

In many cases, different studies use different measurements such as the median, the minimum and maximum values and/or the interquartile range instead of reporting mean and standard deviation of the quantitative outcome variables. Therefore, the effect size of all selected studies must be converted to the same unit of measurement before pooling them. Wan et al (2014), using Hozo's methodology and improving upon it, provided details on estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range.

Inverse variance method and redistribution of weights

The main objective of any meta-analysis is to pool statistics from independent studies with a view to synthesizing them to calculate an estimate of the common effect size. There are different ways of pooling statistics, depending on the type of weight used for individual studies in computing the pooled estimate.

Conventionally, the inverse variance weights are commonly used in meta-analysis (Borenstein et al, 2009). This way the weights are redistributed among the studies depending on the extent of spread of the individual studies. It is well known that the sample variance (a measure of the spread or the differences between the observed value of the relevant outcome variable and their arithmetic mean) is inversely related to the sample size (n), and hence larger studies (with higher sample size) should receive higher weights under the principle of inverse variance weight than the smaller studies. But in many cases this principle may lead to allocating higher than appropriate weight to the lower quality smaller studies and vice versa. The redistribution of weights varies significantly under various commonly used statistical models especially if there is significant heterogeneity among the studies.

Heterogeneity

Since meta-analyses are based on the summary statistics of individual studies, the between-studies variation often plays a significant role in determining the estimate of the effect size. It is therefore essential to check if heterogeneity is present in the data (Sauerland and Seiler, 2005; Ng et al, 2006). If the between-studies variation is not significant, the meta-analysis becomes simple and straightforward. However, in many cases the studies are heterogeneous and therefore meta-analysis must address this fact in computing the pooled effect size and the confidence interval. Some of the commonly used methods to overcome the heterogeneity are discussed below. Unfortunately, not all of them are equally effective or provide a real remedy for the problem.

Assessing heterogeneity

The presence of heterogeneity among the effect size measure is assessed by performing the Cochrane's Q-statistic or I^2 statistic (Huedo-Medina et al, 2006). The main problem with the Q-statistic is that its value increases as the number of studies in a meta-analysis grows larger. While it is useful to detect heterogeneity and inform on the degree of its statistical significance, it is unable to describe the extent of the presence of true heterogeneity (Huedo-Medina et al, 2006).

An alternative method for assessing heterogeneity is the I^2 statistic used in conjunction with its confidence intervals (Higgins and Thompson, 2002). The I^2 statistic presents as a percentage of the total variability that can be attributed to true heterogeneity within a set of effect sizes. Therefore, an I^2 statistic equalling 0% suggests that there is no between-study variability and that all variation observed is a result of sampling error. Conversely, when I^2 approaches 100%, it suggests that the observed variation is the result of between-study variability rather than sampling error. A rough guide for the interpretation of I^2 statistic is: 0–40% might not be important; 30–60% may represent moderate heterogeneity; 50–90% may represent substantial heterogeneity and 75–100% is considerable heterogeneity (Huedo-Medina et al, 2006; Higgins and Green, 2011).

Subgroup analysis

One way to minimize the impact of heterogeneity in a meta-analysis is to group the studies based on the value of the individual study variance (*Figures 3a* and *b*) (Higgins and Green, 2011). Studies with similar values of the variance are sub-grouped and separate meta-analyses are conducted on each of the subgroups along with the overall meta-analysis. However, the problem remains with the estimation of the common effect size of all studies based on combining the results from all the subgroups because of significant heterogeneity among the subgroups.

Meta-regression

Meta-regression is a technique for performing a regression analysis to assess the relationship between the treatment effects and the study characteristics of interest (e.g. suture *vs* prosthesis) or factors concerning the execution of the study (e.g. allocation sequence concealment) (*Figures 3a* and *b*) (Thompson and Higgins, 2002; van Houwelingen et al,

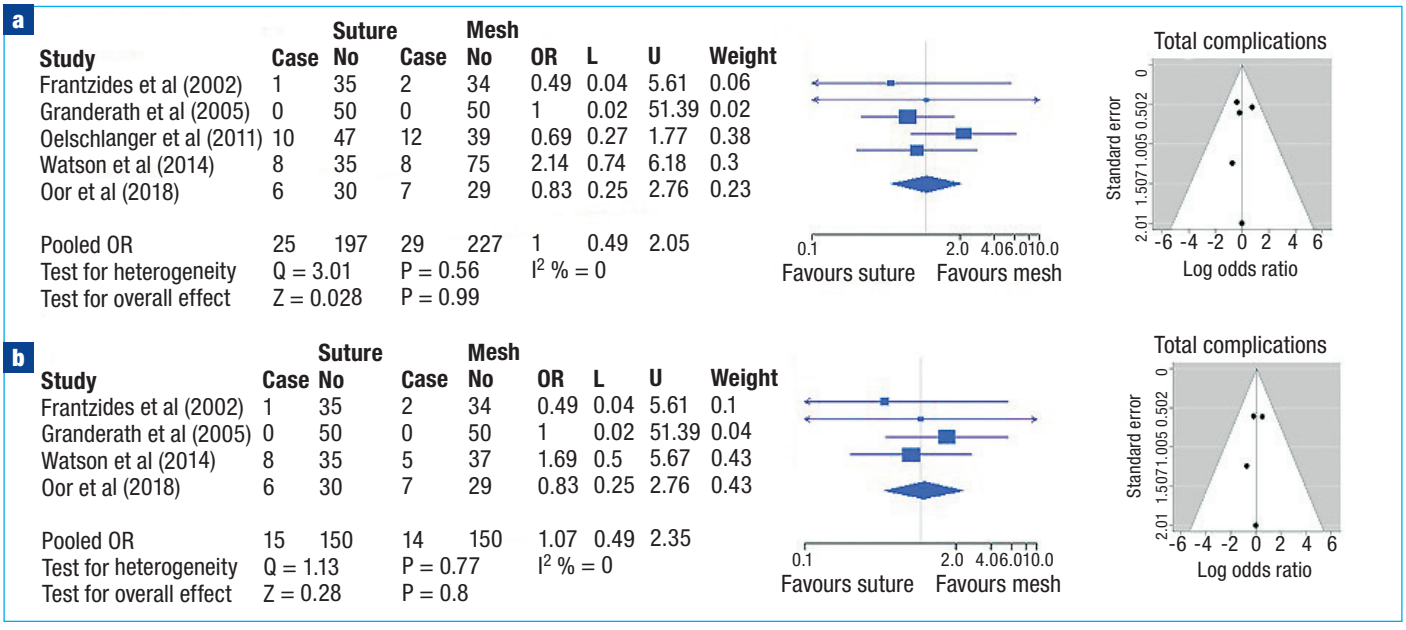

Figure 3. Forest and funnel plots. In these graphs which were created using random effects model, squares indicate point estimates of treatment effect with the size of the squares representing the weight attribute to each study. The horizontal lines represent 95% confidence interval for odds ration. The pooled estimate for complication rate (is the pooled odds ratio obtained by combining all odds ratios of the five studies using the inverse variance weighted method) and is represented by the diamond and the size of the diamond depicts the 95% confidence interval. Funnel plot showing no outliers (dots within the triangle) and therefore, no publication bias. a. Analysis based on a total sample of meshes (i.e. both absorbable and non-absorbable meshes). b. Subgroup analysis based only on non-absorbable meshes. No. = total number of cases; Case = number of patients with complication; OR = odd ratio; L = lower limit of confidence interval; U = upper limit of confidence interval.

2002; Greenland and O'Rourke, 2008). Meta-regression is an extension to subgroup analyses that allows the effect of continuous as well as categorical characteristics to be investigated, and in principle allows the effects of multiple factors to be investigated simultaneously (although this is rarely possible because there are inadequate numbers of studies) (Greenland and O'Rourke, 2008). Meta-regression should generally not be considered when there are fewer than ten studies in a meta-analysis.

Sensitivity analysis

Sensitivity analysis is often used to check the impact on the result of the estimate of the common effect size as a result of the inclusion of one particular study or a group of similar studies in a meta-analysis (Stroup et al, 2000; Sauerland and Seiler, 2005). Studies with significantly higher variance than others could be used in the sensitivity analysis to see how the results are impacted by these studies. This is not a solution to the heterogeneity problem, but it provides some useful insight into the problem and how individual study effect size impacts on the meta-analysis.

Statistical models for meta-analyses

Different statistical models are used for meta-analysis under different conditions. The main difference among the models is the way they allocate the inverse variance weights to the individual studies. The objective of redistribution of weights among the studies, under different statistical models, is to find a more precise estimator of the common effect size to achieve a shorter confidence interval.

Fixed effects model

The fixed effects model assumes that there is a common unknown true effect size for all the studies under investigation (Borenstein et al, 2009).

Random effects model

This model is widely used to handle heterogeneity. It assumes that the true effect size is not identical for every study. However, independent studies have enough in common to justify the synthesis of results to produce an 'average effect size'. The random effects model redistributes the weights to the individual studies in computing the pooled estimate of the common effect size using two sources of variation, i.e. within-study or random error and between-study or variation of the true effect size (Borenstein et al, 2009). Under the random effects model, larger studies normally receive smaller weights. Thus, the random effects model method of synthesizing the common effect size takes away weights from the larger studies and redistributes them to smaller studies.

The DerSimonian and Laird approach is commonly used for a random effects meta-analysis. However, it may lead to too many statistically significant results when the number of studies is small and there is moderate or substantial heterogeneity. An alternative method described by Hartung and Knapp and by Sidik and Jonkman (IntHout et al, 2014) is claimed to be simple and robust especially when there is heterogeneity and the number of studies in the meta-analysis is small. It is beyond the scope of this article to provide a detailed account of these methods.

Inverse variance heterogeneity model

Doi et al (2015a) introduced the inverse variance heterogeneity model. It emphasizes that the fixed effects model based estimator variance can be made closer to the observed variance by modelling over-dispersion through a quasi-likelihood approach. This implies that the meta-analysis is performed under a fixed effects assumption and the variance of the estimator of the common effect size is inflated to account for the heterogeneity. This has the advantage of being based purely on the variance-to-mean relationship, rather than on distributional assumptions, with variance appropriately inflated using a scale parameter (Kulinskaya and Olkin, 2014; Doi et al, 2015b).

Publication and reporting bias in meta-analysis

Publication bias is a serious problem in meta-analysis. It arises because studies with negative or non-significant effects are not normally submitted or accepted for publication in professional journals (Duval and Tweedie, 2000; Sauerland and Seiler, 2005; Higgins and Green, 2011). This phenomenon impacts on the ultimate results of the meta-analysis, sometimes without realizing the extent of the exclusion and their potential impact.

The funnel plot (*Figure 3*) is used to assess the publication bias in a meta-analysis. It is used primarily as a visual aid for detecting bias or systematic heterogeneity. A funnel plot is a scatterplot of treatment effect against a measure of study size such as the estimated standard error or sample size of each of the studies. Asymmetry in funnel plots indicate publication bias in meta-analysis, but the shape of the plot in the absence of bias depends on the choice of axes (Sterne and Egger, 2001). Lastly, most articles are published in English language journals and therefore publications in other languages are excluded from systematic review or meta-analysis. This not only introduces language bias but may also exclude some of the evidence, leading to erroneous conclusions (Morrison et al, 2012).

Presentation and reporting of meta-analyses

The usual way to present the results of meta-analysis is to show the confidence intervals of individual studies and the combined meta-analysis on the same graph in the form of a forest plot (Higgins and Thompson, 2002; Petrie and Sabin, 2009). The middle of the confidence interval of the individual studies is marked by dark squares and the size or area of the associated squares represents the level of weight of the study. For the meta-analysis of the common effect size, the confidence interval is represented by a diamond. The horizontal edges of the diamond represent the limits of the confidence interval. The relative location of the diamond with respect to the no-effect vertical line indicates which intervention is favoured by the data. If appropriate, subgroup analyses are also included in the forest plot along with the combined meta-analysis.

Benefits and drawbacks of meta-analyses

Benefits

Meta-analysis based on well-conducted randomized controlled trials provides the highest level of objective evidence by controlling extraneous variation and biases. The selection and implementation of the correct statistical model produces highest quality of evidence. The results are accurate if the underlying model assumptions are met and there is no selection bias and no error in the extraction of data.

Meta-analysis can combine quantitative summary statistics of individual studies to estimate the common effect size even if the results of the individual studies are inconclusive, and conflicting. Meta-analyses provide more statistical power because of their increased sample size, leading to precise and reliable results. Meta-analysis can be performed to estimate the common effect size for a subset of the selected studies (subgroup analysis).

Drawbacks

The validity of the results of meta-analysis depend on the quality and the design of the trials included in the synthesis, presence of publication bias and presence of heterogeneity.

Meta-analysis cannot be used if the measurement of outcome variables is not similar for all studies.

The validity of forest plots representing confidence intervals depends on the correct identification of the sampling distribution of the effect size estimator which enables choosing the correct critical value in calculating the margin of error.

Selection of the wrong model for heterogeneity will result in misleading results.

In the presence of significant publication bias or reporting anomalies, the results of meta-analysis will be inaccurate, unreliable and invalid.

Conclusions

As the practice of evidence-based decision-making continues to grow, it is important that everyone involved in conducting meta-analysis follows Khan et al's (2003) five steps to ensure its objectivity. Meta-analysis permits the detection of statistically significant differences among study groups that may not have been possible in individual reports because of their small sample size or underpowered trials. However, the quality of the results produced by a meta-analysis will never be superior to the quality of the statistics reported in the individual studies, which again is directly dependent on the design of the study. Meta-analysis can provide much needed high-quality quantitative evidence for making appropriate decisions if the underlying processes, protocols and methods are properly and strictly observed. Moreover, every step in any meta-analysis must be scrutinized for potential bias, from the formulation of the research question to the interpretation and discussion of the results, to ensure the quality and value of the final product (Bernard, 2014). **BJHM**

Conflict of interest: none.

Bernard RM. Things I have learned about meta-analysis since 1990. Reducing bias in search of 'the big picture'. Can J Learn Technol. 2014;40:1–17.
Borenstein M, Hedges LV, Higgins JPT, Rothstein RH. 2009. Introduction of Meta-analysis. Chichester: John Wiley & Sons Ltd:421.
Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance heterogeneity model. Contemp Clin Trials. 2015a Nov;45(Pt A):130–138. https://doi.org/10.1016/j.cct.2015.05.009
Doi SAR, Barendregt JJ, Khan S, Thalib L, Williams GM. Simulation comparison of the quality effects and random effects methods of meta-analysis. Epidemiology. 2015b Jul;26(4):e42–e44. https://doi.org/10.1097/EDE.0000000000000289
Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000 Jun;56(2):455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x
Egger M, Smith GD. Meta-analysis: potentials and promise. BMJ. 1997 Nov 22;315(7119):1371–1374. https://doi.org/10.1136/bmj.315.7119.1371
Eysenck HJ. Systematic Reviews: meta-analysis and its problems. BMJ. 1994 Sep 24;309(6957):789–792. https://doi.org/10.1136/bmj.309.6957.789
Glass G. Primary, secondary, and meta-analysis of research. Educ Res. 1976 Nov;5(10):3–8. https://doi.org/10.3102/0013189X005010003
Greenland S, O'Rourke K. 2008. Meta-analysis. In: Modern Epidemiology. Rothman KJ, Greenland S, Lash TL. 3rd edn. Wolters Kluwer/Lippincott Williams & Wilkins: Philadelphia: 652–682.
Higgins JPT, Altman DG, Gøtzsche PC et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928. https://doi.org/10.1136/bmj.d5928
Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002 Jun 15;21(11):1539–1558. https://doi.org/10.1002/sim.1186
Higgins JP, Green S. 2011. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. (accessed 6 September 2019) https://handbook-5-1.cochrane.org/
Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? Psychol Methods. 2006;11(2):193–206. https://doi.org/10.1037/1082-989X.11.2.193
IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014 Dec;14(1):25. https://doi.org/10.1186/1471-2288-14-25
Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996 Feb;17(1):1–12. https://doi.org/10.1016/0197-2456(95)00134-4
Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. J R Soc Med. 2003 Mar;96(3):118–121. https://doi.org/10.1177/014107680309600304
Khan S, Doi SAR, Memon MA. Evidence based decision and meta-analysis with applications in cancer research studies. Appl Math Inf Sci. 2016 May 1;10(3):815–822. https://doi.org/10.18576/amis/100301
Kulinskaya E, Olkin I. An overdispersion model in meta-analysis. Statistical Modelling: An International Journal. 2014 Feb;14(1):49–76. https://doi.org/10.1177/1471082X13494616
Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Lancet. 1999 Nov;354(9193):1896–1900. https://doi.org/10.1016/S0140-6736(99)04149-5
Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009 Jul 21;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097
Moher D, Hopewell S, Schulz KF et al; Consolidated Standards of Reporting Trials Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol. 2010 Aug;63(8):e1–e37. https://doi.org/10.1016/j.jclinepi.2010.03.004
Moher D, Shamseer L, Clarke M et al; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015 Dec;4(1):1. https://doi.org/10.1186/2046-4053-4-1
Morrison A, Polisena J, Husereau D et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012 Apr;28(2):138–144. https://doi.org/10.1017/S0266462312000086
Ng TT, McGory ML, Ko CY, Maggard MA. Meta-analysis in surgery. Arch Surg. 2006 Nov 01;141(11):1125–1130, discussion 1131. https://doi.org/10.1001/archsurg.141.11.1125
Petrie A, Sabin C. 2009. Medical Statistics at a Glance. Oxford: Wiley-Blackwell: 180.
Sauerland S, Seiler CM. Role of systematic reviews and meta-analysis in evidence-based medicine. World J Surg. 2005 May;29(5):582–587. https://doi.org/10.1007/s00268-005-7917-7
Shamseer L, Moher D, Clarke M et al; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 02;349 jan02 1:g7647. https://doi.org/10.1136/bmj.g7647
Smith GD, Egger M. Meta-analysis: unresolved issues and future developments. BMJ. 1998 Jan 17;316(7126):221–225. https://doi.org/10.1136/bmj.316.7126.221
Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010 Sep;25(9):603–605. https://doi.org/10.1007/s10654-010-9491-z
Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis. J Clin Epidemiol. 2001 Oct;54(10):1046–1055. https://doi.org/10.1016/S0895-4356(01)00377-8
Sterne JAC, Hernán MA, Reeves BC et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919. https://doi.org/10.1136/bmj.i4919
Stroup DF, Berlin JA, Morton SC et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283(15):2008–2012. https://doi.org/10.1001/jama.283.15.2008
Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002 Jun 15;21(11):1559–1573. https://doi.org/10.1002/sim.1187
van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002 Feb 28;21(4):589–624. https://doi.org/10.1002/sim.1040
Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014 Dec;14(1):135. https://doi.org/10.1186/1471-2288-14-135

KEY POINTS

- Meta-analysis is a systematic review that uses quantitative methods to summarize the results. It allows a re-examination of treatment effect of several studies and provides a single, overall measure of the treatment effect.
- Meta-analysis is the aggregation of information (from several studies) leading to a higher statistical power and more robust point estimate when compared to any individual study.
- Meta-analysis represents the highest level (level 1) of evidence among the hierarchy of evidence for evidence-based research.
- Meta-analysis results can be generalized to a large population (in the vast majority of cases).
- The validity of evidence produced by meta-analyses, and hence the decisions based on them, are reliant on the quality of the data, the methodology used to extract them, the robustness and totality of the evidence and relevant methods to analyse them.