REVIEW ARTICLE

CrossMark

Conducting and interpreting high-quality systematic reviews and meta-analyses

Rajat Kalra, MBChB,[a] Pankaj Arora, MD, FAHA,[b,d] Charity Morgan, PhD,[c]
Fadi G. Hage, MD, FASH, FACC, FASNC,[b,d] Ami E. Iskandrian, MD, MACC,[b] and
Navkaranbir S. Bajaj, MD[b,d]

[a] Department of Medicine, University of Alabama at Birmingham, Birmingham, AL
[b] Division of Cardiovascular Disease, University of Alabama at Birmingham, Birmingham, AL
[c] Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL
[d] Section of Cardiology, Birmingham Veterans Affairs Medical Center, Birmingham, AL

Systematic reviews and meta-analyses are powerful tools for summarizing existing literature and combining evidence from multiple studies. These methods employ complex searches, statistical techniques, and presentation techniques with which the clinical audience may not be very familiar. This review article aims to familiarize the clinical audience with the various techniques employed to conduct a high-quality systematic review and meta-analysis. (J Nucl Cardiol 2017;24:471–81.)

Key Words: Systematic review · meta-analysis · network meta-analysis

Abbreviations		NMA	Network meta-analysis (analyses)
MA	Meta-analysis (analyses)	SR	Systematic review(s)
MACE	Major adverse cardiac events		

INTRODUCTION

Systematic reviews (SR) and meta-analyses (MA) have a prominent role in clinical literature. Just like any other form of clinical research, both techniques have been subjected to criticisms about their validity and place in clinical evidence-based medicine.[1] Hence, we seek to outline what SR and MA can offer clinicians and discuss the modern-day methodological standards for conducting these studies.

DEFINITIONS

A SR is a method of summarizing an area of research literature done by aggregating critically appraised literature in an unbiased fashion, thus providing a thorough overview on a selected topic.[2] It provides insight into the clinical effectiveness of an intervention both quantitatively and qualitatively, and it may also be used to assess feasibility and cost-effectiveness. Like all clinical research, SR must be placed in an appropriate context in order to have clinical utility. SR can also be performed independently of a MA.[2]

MA is a quantitative method used to combine and analyze results from SR to derive a conclusion about the body of research in the prespecified area; among medical studies, it is considered to generate the highest level of evidence.[3,4]

MA can be performed to answer many different questions. In general, they are used to assess the strength

of evidence for questions related to a specific disease or treatment to determine whether an effect exists, whether it is positive or negative, and then obtain a summary estimate of the effect. When heterogeneity in results is identified, it can be used to generate a new hypothesis. MA can also be used to generate pooled estimates for rates of clinical outcomes. More recently, they have also been used to indirectly compare different treatments where direct comparisons are lacking.

The main benefit that MA offers over other types of research studies is its intrinsic capability to reduce bias by combining data from a broad range of studies. This advantage is dependent on having first performed a thorough SR, during which all relevant studies on a particular topic were identified. MA provides the opportunity to identify and estimate differences in subgroup populations that may not be otherwise detected in smaller individual studies. MA is also useful in fields where there is no clear consensus on the effectiveness of a treatment or intervention, as even studies with contradictory results may be combined and summarized.

HISTORICAL AND MODERN PERSPECTIVES

The works of Karl Pearson, among others, can be used to elucidate the origins of MA.[3,5] Pearson, a 19th century British statistician, was given the task of comparing the results of multiple studies evaluating mortality in British soldiers after inoculation. He arrived at the idea of creating a pooled estimate after combining the results from multiple smaller studies. Although his

work would not stand up to the rigorous methodology required in modern-day MA, the work of Pearson nevertheless introduced the idea of combining data from smaller studies to generate an estimate of overall effect.[3]

Since the publication of the first SR in 1904, MA have grown exponentially. Between 1991 and 2001, MA were the most frequently cited form of research.[6] At present, there are over 100,000 MA available on MEDLINE via PubMed (Figure 1). Multiple consensus societies, such as the American Heart Association and the American College of Cardiology, consider MA of high-quality randomized controlled trials as the highest level of evidence when determining the estimate of certainty, or precision, of treatment effect.[7]

CONDUCTING SR AND MA

Like all clinical research, every SR must first be commenced by the investigators identifying a question of interest. This may be a question addressing a controversy in the medical literature or consensus guidelines, or occasionally an area in clinical medicine in where data are lacking. Once the research question has been chosen, the investigators must determine which endpoints will be used to address this question. It is important to note that studies may frequently have differing definitions for endpoints or report the same endpoint in different ways (e.g., reporting risk ratios versus risk differences). Authors should therefore carefully define the endpoint or outcome of interest before beginning their literature search.
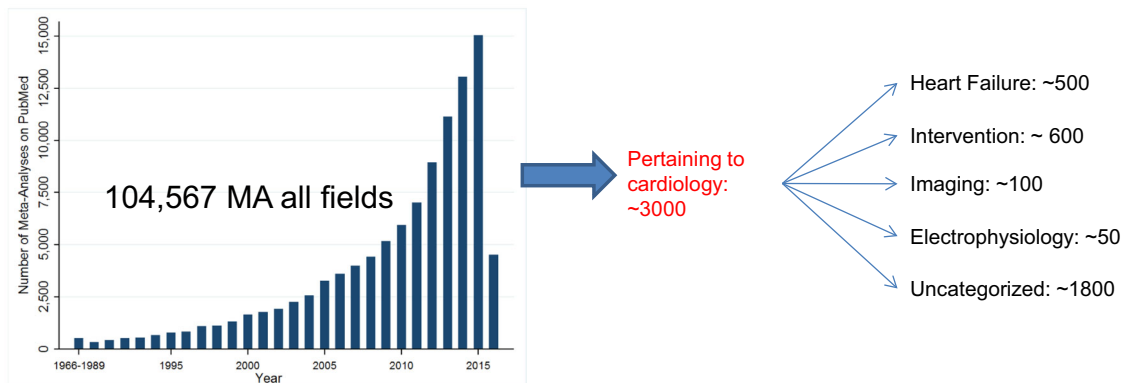
Search	Query	Items Found
6	Search (((meta-analysis) AND cardiology)) AND Intervention	565
5	Search (((meta-analysis) AND cardiology)) AND Imaging	128
4	Search (((meta-analysis) AND cardiology)) AND electrophysiology	49
3	Search (((meta-analysis) AND cardiology)) AND heart failure	478
2	Search (meta-analysis) AND cardiology	3004
1	Search (meta-analysis)	104567

Figure 1. Meta-analyses on PubMed by year. The *y* axis depicts the number of results in response to the search query 'meta-analysis' on PubMed.

Detailed PubMed search strategy:
(TITLE-ABS-KEY("STEMI") OR TITLE-ABS-KEY("ST elevation myocardial infarction") OR TITLE-ABS-KEY("Acute myocardial infarction") AND TITLE-ABS-KEY("PCI") OR TITLE-ABS-KEY("Percutaneous coronary intervention") OR TITLE-ABS-KEY("angioplasty") AND TITLE-ABS-KEY("Culprit") OR TITLE-ABS-KEY("non-culprit") OR TITLE-ABS-KEY("complete revascularization")) AND (EXCLUDE(DOCTYPE, "re") OR EXCLUDE(DOCTYPE, "no") OR EXCLUDE(DOCTYPE, "ed") OR EXCLUDE(DOCTYPE, "le") OR EXCLUDE(DOCTYPE, "Undefined"))

Figure 2. Example of an electronic search phrase. Adapted with permissions from: J Am Heart Assoc 2015 Dec 14;4(12).

Identification of Studies

Following the identification of the research question and selection of outcomes of interest, a search is undertaken to identify studies of relevance. The most critical aspects of the search are as follows: (1) to include all studies pertinent to the question being asked and (2) to make sure that the studies compare similar populations and similar outcomes. Authors can ensure similarity in the SR study population by including studies that have common inclusion and exclusion criteria for the hypothesis being tested. Based on the question being asked, the authors may choose to limit the search to observational studies or randomized controlled trials, or, in some cases, include both. For example, in the MA by Bajaj et al, the authors restricted their study population to randomized trials by including patients with multivessel coronary artery disease presenting with ST-elevation myocardial infarction and comparing complete versus culprit vessel revascularization.[8] They also used an a priori definition of MACE (major adverse cardiac events) to avoid bias as the

individual trials in this MA used varying definitions of MACE. Therefore, using the individual trial definitions may have led to comparison of different outcomes, thus invalidating the results of the MA. Hence, it is extremely important to identify a uniform definition to prevent misinterpretation of data.

Once the aforementioned criteria have been accounted for, a broad search strategy must be designed. This step consists of designing a search phrase that includes numerous smaller phrases, words, and MeSH keywords (Medical Subject Headings, as used by the National Library of Medicine to index articles for MEDLINE), which are targeted towards finding literature to answer the question at hand. Figure 2 provides an example of an electronic search phrase for SCOPUS from a recently published meta-analysis comparing revascularization approaches in patients with multivessel coronary disease presenting with ST-elevation myocardial infarction.[8] The primary outcome of interest in this meta-analysis is incidence of MACE. We strongly recommend that independent searches for each of the secondary endpoints be conducted to help reduce selection bias.

Once the search phrase has been chosen, the authors must then decide which databases will be queried with that phrase. An expert medical librarian may help the authors conduct this search and identify relevant abstracts or studies. Many different databases are available, including MEDLINE/PubMed (MEDLINE is the index of the National Library of Medicine and PubMed is the electronic index), SCOPUS (a database that includes MEDLINE and six other databases:
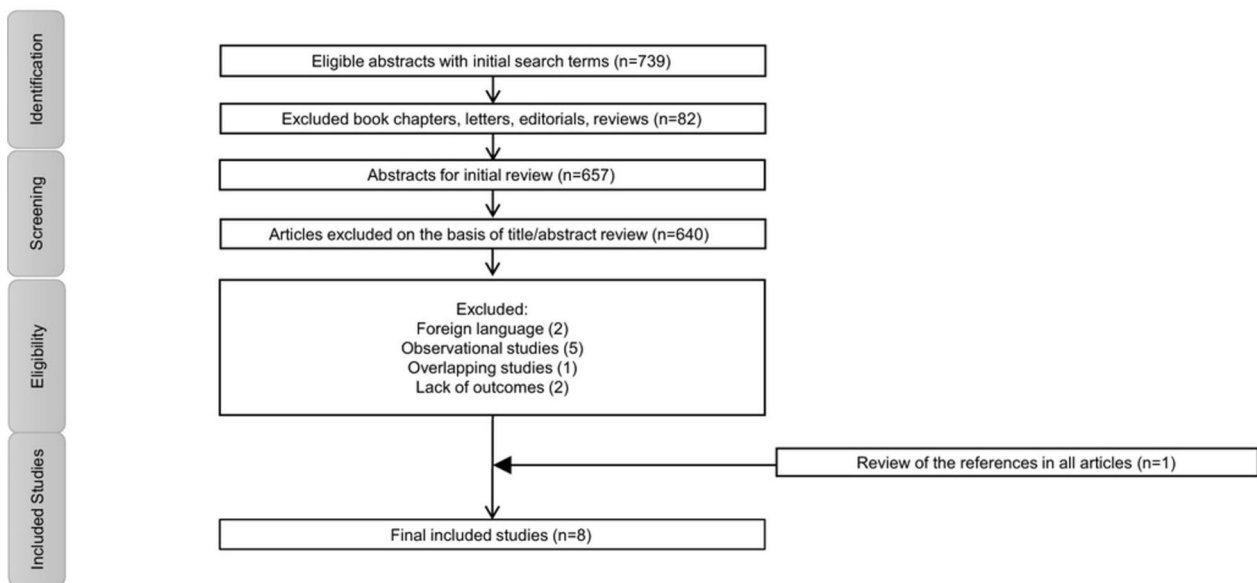

Figure 3. Example of literature search and study selection process. Adapted with permissions from: J Am Heart Assoc 2015 Dec 14;4(12).

Table 1. Hypothetical example of Jadad scoring of randomized controlled trials

Study name	Randomization	Double blind	Withdrawals/ dropouts	Total Jadad score
Study 1	Randomization used: 1 Randomization method described: 1 Randomization method appropriate: 0 *Total randomization score: 2*	Blinding used: 1 Blinding method described: 1 Blinding method appropriate: 0 *Total double blind score: 2*	Description of withdrawals/dropouts: 1	5
Study 2	Randomization used: 1 Randomization method described: 0 Randomization method appropriate: 0 *Total randomization score: 2*	Blinding used: 0 Blinding method described: 0 Blinding method appropriate: 0 *Total double blind score: 0*	Description of withdrawals/dropouts: 0	2
Study 3	Randomization used: 1 Randomization method described: 1 Randomization method inappropriate: -1 *Total randomization score: 1*	Blinding used: 1 Blinding method described: 1 Blinding method inappropriate: -1 *Total double blind score: 1*	Description of withdrawals/dropouts: 1	3

In this hypothetical example, two points are awarded for randomization if the paper described the method of randomization and it was appropriate. Two points are awarded for double blinding if it was both described and appropriate. If withdrawals/dropouts are described in the study, then the study is awarded a single point. If either randomization or double blinding was not described or it was not appropriate, the study loses one point in each category

Embase, Compendex, World Textile Index, Fluidex, Geobase, Biobase), Google Scholar, the COCHRANE database, and ClinicalTrials.gov as prominent examples. It is also requisite to search the citations of the included articles for additional references in the event that these articles include studies that are not otherwise listed on the above databases or are not identified using the search strategy. Other resources may include unpublished studies (for which data can be obtained by directly contacting the principal investigator or a designee), dissertations (usually found in national indexes of dissertations), and drug company and device studies (for which drug companies or the national regulatory board, such as the Food and Drug Administration in the USA, may need to be contacted directly). For studies published before MEDLINE indexing in 1966, Index Medicus may also be searched. It is a good practice to be over inclusive and retrieve full-text articles if there is any doubt about whether a study should be included in order to minimize study selection bias.

The initial search for studies is purposefully broad and often identifies many more studies than will be used in the final analysis. The next task in the literature search is to begin narrowing down the results of the search. For example, a commonly encountered problem is duplication of data from the same study population in sequential publications, particularly when the outcome of interest requires long-term follow-up. In such scenarios, the most complete dataset with the longest follow-up should be included, and datasets should not be duplicated. After retrieval of abstracts and full-text articles of the identified studies, there are a number of important steps to be taken to evaluate study eligibility for inclusion in the SR and MA. Studies must first be carefully evaluated for eligibility by comparing their inclusion, exclusion, and endpoint criteria against the authors' predefined criteria. Many reported MA include a flow chart detailing the results of the literature search and selection process. This figure not only reports how many studies were initially identified by the search but also describes how the authors arrived at the final set of included studies. Figure 3 provides an example of such a chart.[8]

Assessment of Study Quality

The quality of studies can be judged using any of a number of standardized scales available for grading both randomized and nonrandomized studies. Over 20 standardized quality assessment scales exist for randomized controlled trials.[9] The Jadad scale is among the most widely used in the literature.[9,10] This scale is used to judge whether randomization, blinding, withdrawals, and dropouts were appropriately handled and described within a randomized controlled trial. Table 1 depicts a

Table 2. Example of Newcastle-Ottawa scoring for cohort studies

Study name (total score)	Representativeness	Selection of nonexposed cohort	Ascertainment of exposure	Demonstration that outcome was not present at the start of study	Comparability of cohorts	Assessment of outcome	Length of follow-up	Adequacy of follow-up
Study 1 (6)	Truly representative (one point)	Drawn from different source as exposed cohort (zero points)	Secure record (one point)	Yes (one point)	Study controls for most important factor only (one point)	Independent blind assessment (one point)	Adequate follow-up (one point)	Complete follow-up recorded (one point)
Study 2 (5)	Not representative (zero points)	Drawn from the same source as exposed cohort (one point)	Structured interview (one point)	Yes (one point)	Study controls for most important factor and another factor (two points)	Self report (zero points)	Follow-up was too short (zero point)	Poor follow-up rate and no description of lost patients (zero points)
Study 3 (6)	Somewhat representative (one point)	No description (zero points)	No description (zero points)	Yes (one point)	Study does not control for any factor (one point)	Record linkage (one point)	Adequate follow-up (one point)	Some subjects lost to follow-up but acceptable (one point)

hypothetical example of study quality judged by the Jadad scale.

Nonrandomized studies can also be judged through a number of scales,[11] with the Newcastle-Ottawa scale being a commonly used set of criteria.[12] The Newcastle-Ottawa criteria examine study methodology by evaluating selection of exposed and nonexposed cohorts, comparability of cohorts, assessment of outcome, and length and adequacy of follow-up. Poor study quality should lead authors to at least consider exclusion of the study from analyses. Table 2 depicts a hypothetical example of study quality judged by the Newcastle-Ottawa scale.

Data Extraction

Data extraction for MA requires a meticulous approach. Data extraction is usually carried out by compiling information on baseline characteristics and study endpoints on structured extraction forms. Ideally, data extraction should be carried out by at least two authors to help reduce bias. Inconsistencies can be resolved by mutual consensus among all authors to reduce intraobserver variability.

Statistical Analysis

The goal of MA is to combine the results of the identified studies to produce a pooled estimate of the chosen endpoint. When studies are combined, they are weighted by their sample size and variability. As a result, larger and more consistent studies will have more influence on the final estimate than smaller and more variable studies.

In order to combine the data from the SR, two main MA data modeling techniques may be used to produce this pooled estimate.[13] The first method, *fixed effects modeling,* assumes that a common, true effect is shared by all of the included studies in a MA and provides an estimate of this effect.[14] In contrast, *random effects modeling* allows for heterogeneity in the effect sizes of the studies. That is, rather than assuming that the true effect sizes are equal, as in fixed effects modeling, random effects modeling assumes that the study effects are merely drawn from the same population. Random effects modeling aims, therefore, to estimate the mean of this population.[14] This heterogeneity may be due to small differences in the study populations or intervention. It is important to note that these differences in study populations and outcome measurement cannot be so significant so as to prevent comparison of the study populations. As random effects modeling assumes greater variability in effect sizes among the included studies, it provides a more conservative pooled estimate
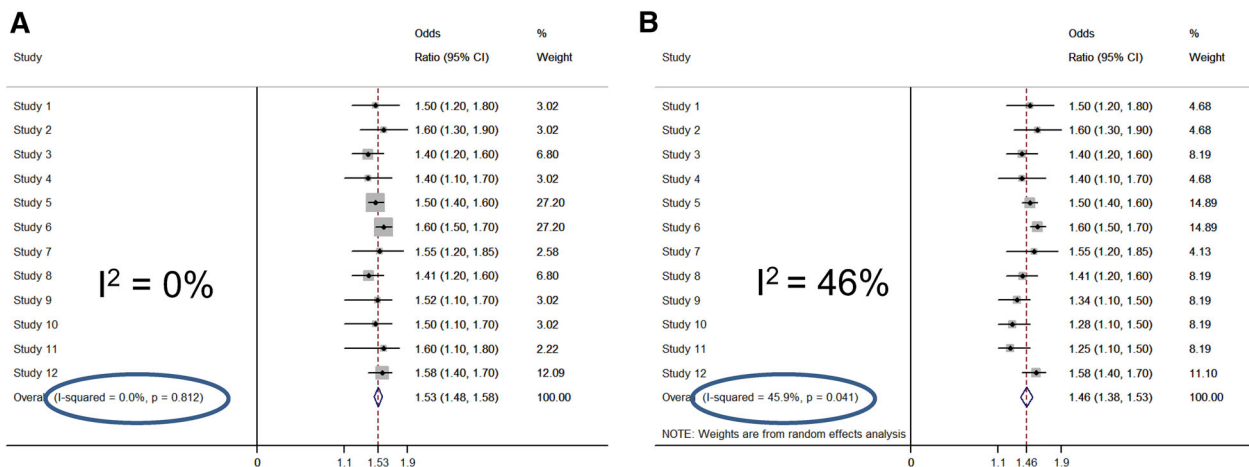
A

B

Figure 4. Hypothetical examples of modeling techniques. **A** Fixed effects modeling. **B** Random effects modeling.
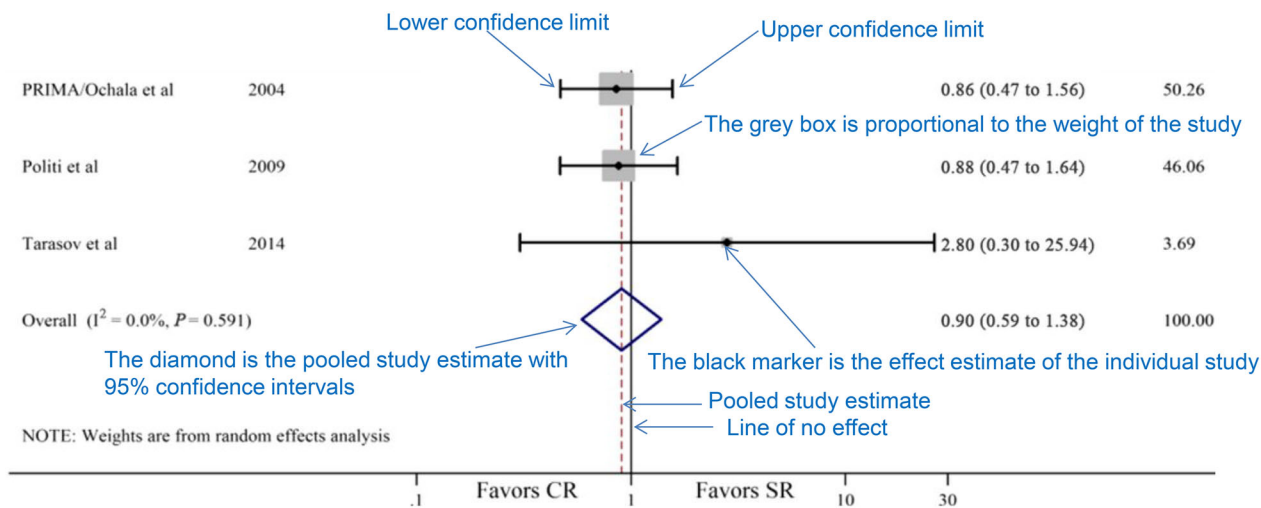

Figure 5. Example of a forest plot. Adapted with permissions from: J Am Heart Assoc 2015 Dec 14;4(12).

for the outcome in question. Consequently, random effects modeling has lesser power to detect a significant effect than fixed effects modeling.[4]

Determining whether a fixed or random effects model should be used requires assessing for heterogeneity in the estimated treatment effects from the included studies. The I^2 statistic may be used as a measure of heterogeneity. When using the I^2 statistic, for example, heterogeneity may be defined as potentially not being important between 0% and 20%; moderate in degree, if it is between 20% and 50%; substantial, if it is between 50% and 75%; and considerable, if greater than 75%.[15] Cochrane's Q statistic may also be used to formally test for the presence of heterogeneity.[16] Large I^2 values and/

or small P-values (usually $P < 0.05$) from testing the Q statistic indicate significant heterogeneity, in which case random effects models should be used. Figure 4A is a hypothetical example that highlights the use of the fixed effects model to provide a pooled estimate. The fixed effects model may be appropriate when the measured heterogeneity is low. Figure 4B is a hypothetical example that highlights the use of the random effects model to provide a pooled estimate.

Presentation of Results

Results of the statistical analysis are presented via a forest plot (Figure 5). The forest plot provides a visual
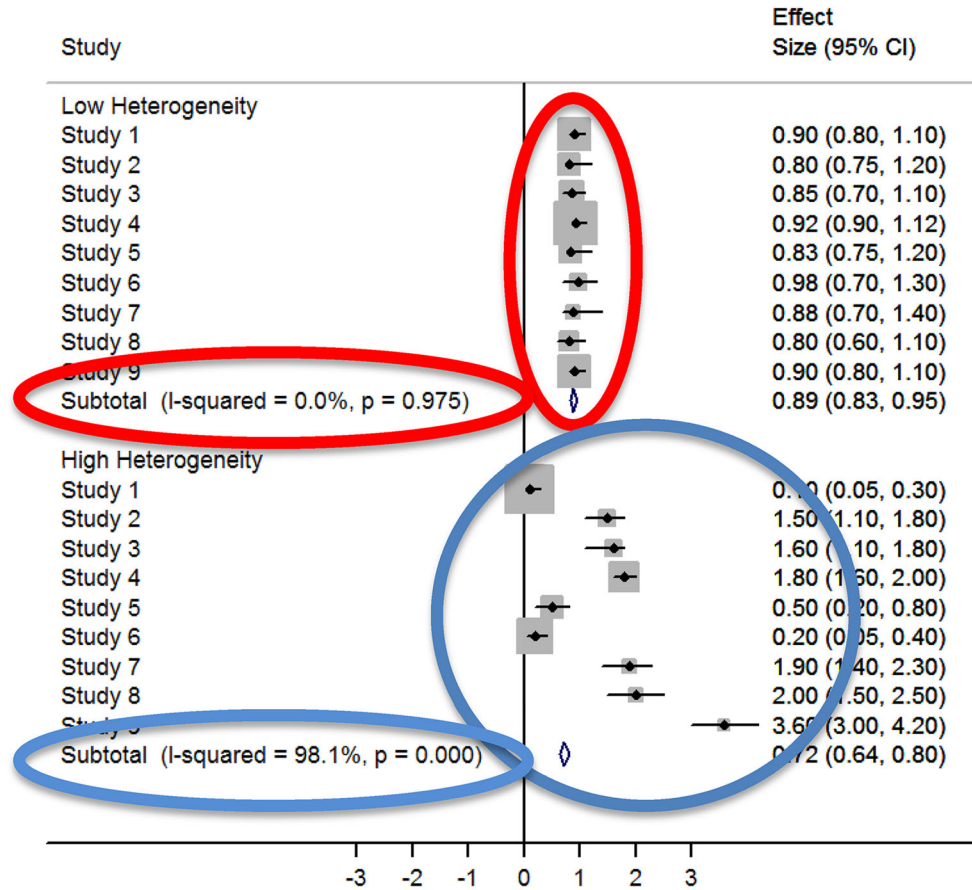
Figure 6. Visual assessment of a hypothetical forest plot for statistical heterogeneity. In the first study group, entitled 'Low Heterogeneity', the reader can see that the studies are closely clustered in a vertical fashion. The second study set, entitled 'High Heterogeneity', has a wide dispersion of study effect size estimates, indicating a greater degree of heterogeneity. The blue circles highlight the I^2 figures and their statistical significance.

representation of pooled estimate of the effect size. The estimated individual effect size (and its 95% confidence interval) for each study is also plotted; the size of the plotting symbol often represents the weight assigned to each study. Assessment of heterogeneity can be also informally be done by visual inspection of the forest plot.[17] This is demonstrated in the hypothetical example in Figure 6. Here, wider dispersion of the effect size estimates indicates greater heterogeneity in the second set of studies located near the bottom of the figure.

Interpretation of Results

Although heterogeneity in the study effect sizes may be addressed with a random effects model, it is important to investigate the source of this heterogeneity. A large degree of heterogeneity may suggest that populations that are unlike each other are combined (conceptual heterogeneity; 'combining apples with oranges') or that there is high

variability in treatment effects, but those effects are still measuring the same thing (statistical heterogeneity).[18] Investigation of heterogeneity may include conducting sensitivity analyses, meta-regression analyses, or subgroup analyses.

Meta-regression allows the authors to adjust for study-level factors when estimating the effect of interest. For example, if a treatment was thought to possibly be more effective in men than women, meta-regression could be used to determine whether studies with higher proportions of female subjects had smaller treatment effects. It is important to remember that meta-regression measures only associations between the effect size and study-level factors and cannot be used to explain causation.[19] Significant meta-regression results usually imply the need for further prospective investigative efforts to confirm these findings before they can be directly translated to clinical practice. An example of a meta-regression is shown in Figure 7, Panel A, where the hypothetical example
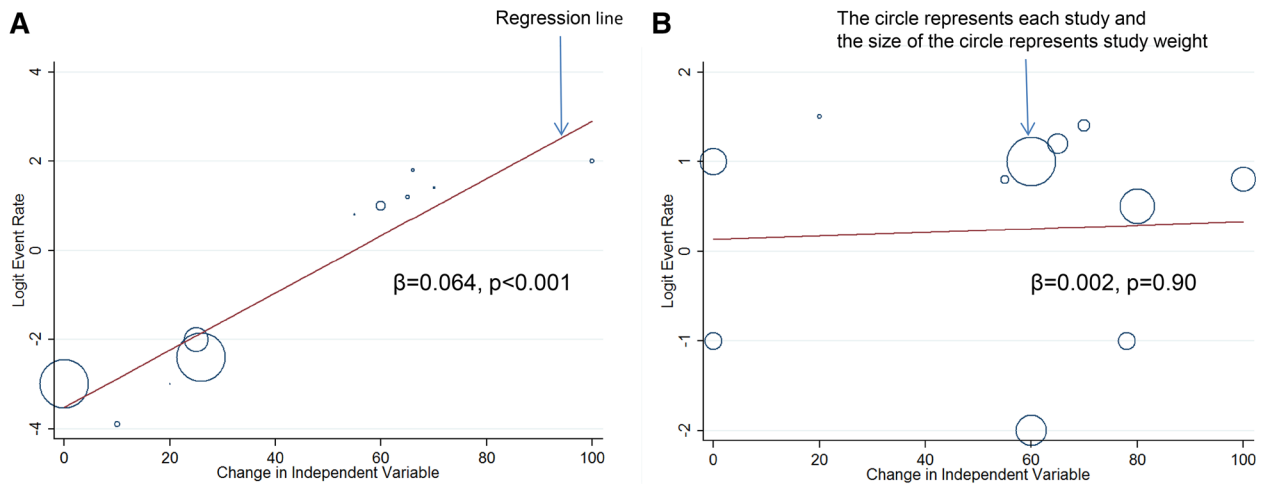
Figure 7. Example of meta-regression. The authors performed meta-regression with Logit event rate as the dependent variable. **A** demonstrates that there is an increase in Logit event rate as the proportion of the independent variable increases in individual studies increases. **B** there is no relationship between Logit event rate and the proportion of the independent variable in individual studies.

demonstrates that there is an increase in Logit event rate as the proportion of the independent variable in individual studies increases. In Figure 7B, there is no relationship between Logit event rate and the proportion of the independent variable in individual studies.

Subgroup analyses can also be employed when separate studies within a MA have patients that may be biologically or pathologically different from each other.[16]

In short, the presence of significant heterogeneity affects the generalizability of the results, and the MA should discuss potential reasons for the heterogeneity rather than simply focusing on the pooled estimate.

Network MA

Network MA (NMA) is an important concept that has emerged more recently. NMA, or mixed treatment comparison, is a statistical method used to combine direct and indirect comparisons from multiple studies examining a particular treatment effect.[20] For example, there may be two trials comparing mortality benefit for a given intervention—one comparing treatment A to treatment B and another comparing treatment B to treatment C. The purpose of the NMA would be to compare the benefit that treatments A and C confer over treatment B using direct comparisons, but then, using indirect comparisons, compare treatments A and C. This would be labeled an 'indirect' comparison if no such trial exists. The ability to produce indirect comparisons is the main benefit of NMA. Due to the advantage of making indirect comparisons, NMA also give authors the ability to concomitantly assess all treatments, thus allowing authors the ability to create a treatment

hierarchy for the intended outcome. One such example is the NMA conducted by Bajaj et al when comparing the major adverse cardiac event rate after complete revascularization, staged revascularization, and culprit vessel revascularization after ST-elevation myocardial infarction.[8] Figure 8 illustrates a network map and network forest plot of how individual studies contribute to direct and indirect comparisons of these various treatment strategies.[8]

TOOLS FOR MA

There are many statistical software packages available for meta-analyses.[21,22] There are also a number of freeware packages offered for computing platforms.[23]

MANUSCRIPT COMPILATION

There are many guidelines available to authors to encourage standardized conduction and reporting of systematic reviews, MA, and NMA. The Preferred Reporting Items for SR and MA (PRISMA) guidelines are available for both MA of randomized controlled trials[24] and NMA.[25] Additionally, the Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines are available to authors conducting a MA of observational studies.[26]

LIMITATIONS

MA have many well-described limitations.[27] In this section, we highlight some of the more prominent limitations of MA.
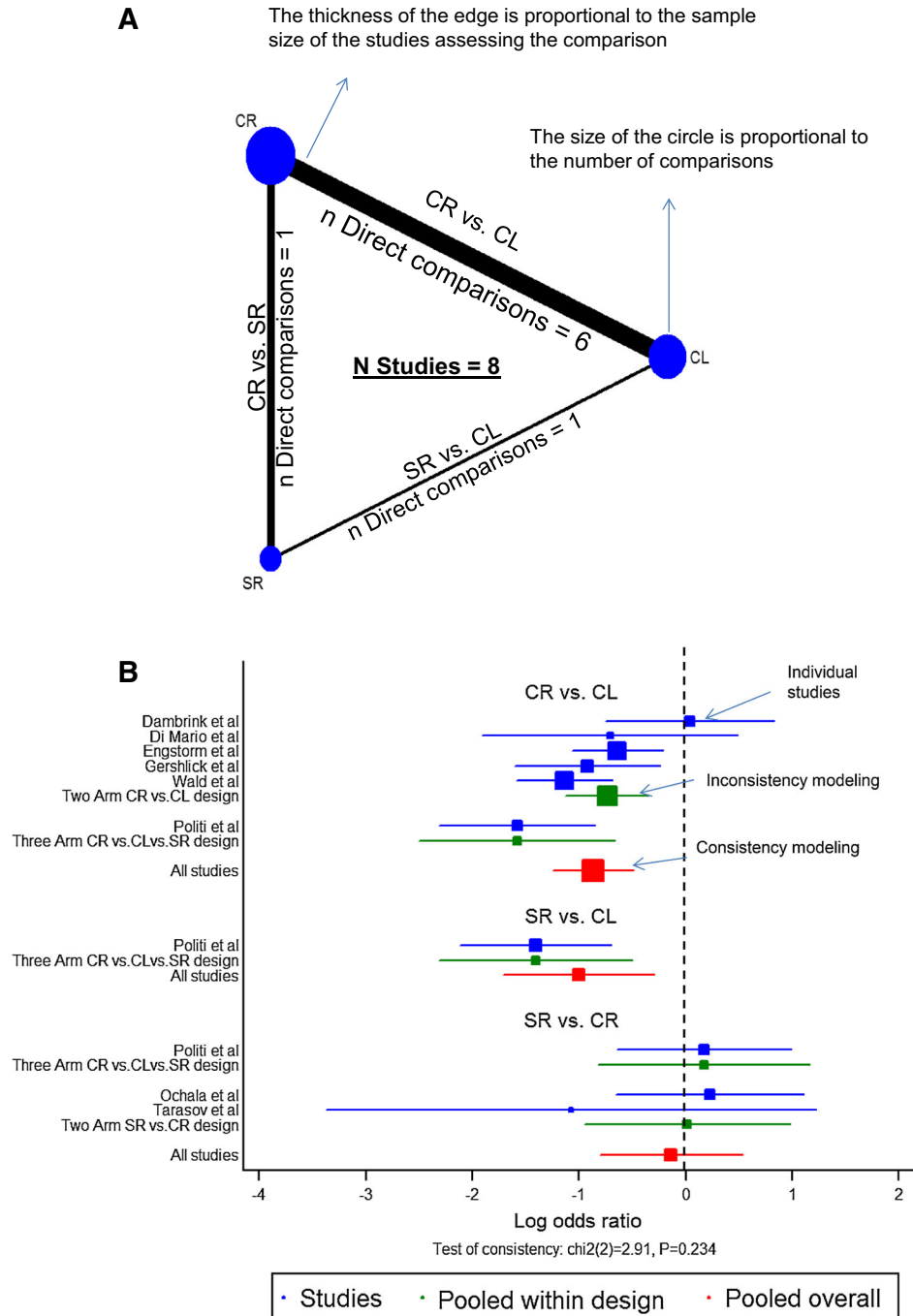
Figure 8. A Network for treatment comparison for primary outcome. The *solid blue circle* represents the treatment. The size of the *circle* corresponds to the total sample size of treatment from all included trials. The *solid black line* represents direct treatment comparisons. The thickness of line corresponds to total sample size assessing the comparison. *CR* complete revascularization at index angiogram, *SR* staged revascularization of nonculprit vessels after culprit lesion revascularization at index angiogram, and *CL* culprit lesion revascularization only at index angiogram. **B** Network forest plot on a logarithmic scale. The *solid blue lines* represent the confidence intervals for log odds ratios for each comparison in individual studies, and the *solid green lines* represent log odds ratio within study design. The *red solid line* represents respective overall log odds ratio using consistency modeling. The *dashed black line* is the line of no effect (odds ratio equal to 1). Adapted with permissions from: J Am Heart Assoc 2015 Dec 14;4(12).
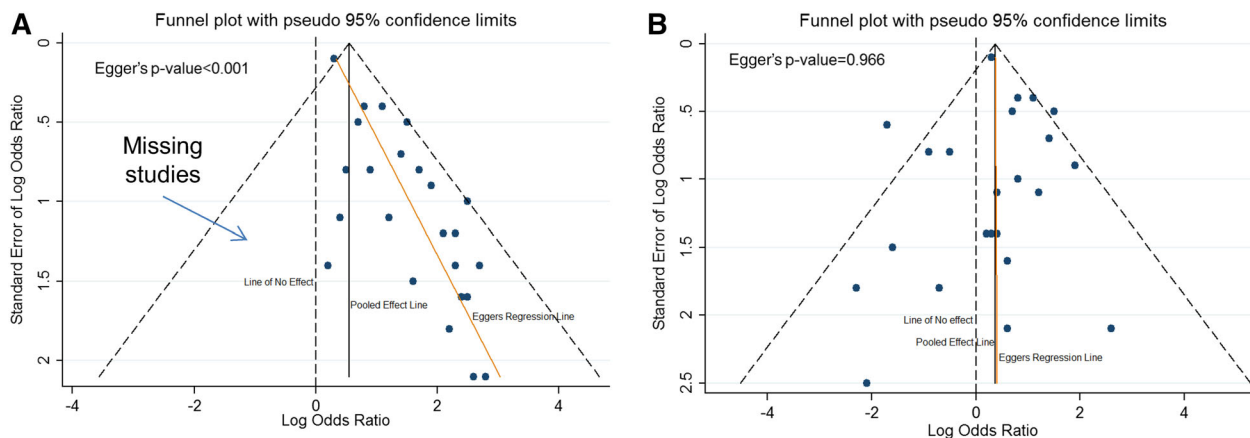
Figure 9. Hypothetical example of funnel plots. **A** Asymmetric funnel plot with publication bias. **B** Symmetric funnel plot with publication bias. *OR* odds ratio.

Inclusion of Poor Quality Studies

The strength of a MA is that it can provide a succinct summary of the literature that it encompasses. The generalizability of the summaries and estimates may be heavily limited when poor quality studies are included. At best, this may make the MA seem like an outlier and may cast judgment on the statistical techniques employed to conduct the MA. At worse, this may significantly alter treatment uptake and actually produce harm in patient care. Therefore, great care must be taken in quality assessment of studies for inclusion in MA. Even still, quality assessment is not devoid of pitfalls, and this is a known limitation of MA.[28]

Oversimplification of an Entire Field

It is difficult for a single summary effect produced by a MA to clearly encompass the clinical judgment required to make an important clinical decision. By producing a summary estimate, there is an inherent desire for authors and readers to do exactly that. However, both readers and authors alike should take caution in oversimplifying the results of MA. In most high-quality MA, the summary estimate is meant to reflect a treatment outcome in a well-defined clinical situation.

Disagreement with Individual Studies

The results of a MA may disagree with those of high-quality randomized controlled trials. This potential is exaggerated when the randomized controlled trials included in a MA have a wide dispersion in their individual treatment effects. In an effort to produce a summary effect, individual studies may be discarded or may disagree with the overall treatment effect. It should be understood that the goal of MA, however, is to provide the overall treatment summary estimate, even if the overall estimate disagrees with the individual studies used to calculate it. The latter concept is entitled 'Stein's Paradox'.[19]

While there is no clear remedy to the situation, this again highlights the need to employ careful inclusion/exclusion criteria to ensure comparison of like study populations and endpoints. Moreover, this also highlights that regardless of the rigor with which statistical analysis is sought, it is important to apply the research to the appropriate clinical context for it to have any generalizability. Finally, if there is a large amount of dispersion among randomized controlled trials, significant efforts must be made to explain why this dispersion is occurring in the first place.

Publication Bias

Identification of all studies relevant to the chosen topic is essential to produce a high-quality MA. However, studies with nonsignificant findings are less likely to be published, reducing the chance of them being identified and included in the MA. This biases MA in favor of finding significant effects. This problem can be partially mitigated by including sources for unpublished studies in the search strategy, as discussed above.

Nonetheless, the results of the search should be assessed for publication bias. A funnel plot provides a visual check for publication bias (Figure 9).[29] Funnel plots display estimated effect sizes for the included studies, plotted against their sample size or measure of sample size, such as standard error. Asymmetry in the funnel plot is an indicator that publication bias may be present (Figure 9A provides an example of funnel plot asymmetry).[29] In the absence of publication bias, the funnel plot will appear symmetric (Figure 9B). There are also methods available to statistically test for funnel plot asymmetry.[30]

NEW KNOWLEDGE GAINED

SR and MA are important methods of critically appraising, summarizing, and further analyzing large topics in an organized fashion. In turn, these tools can be used to provide summary estimates and generate new hypotheses to address important clinical controversies. We detail the basic rationale, methodology, and statistical techniques used in their composition.

CONCLUSION

SR and MA undoubtedly have a critical place in modern-day literature. The utility of these tools is evidenced by their importance in making pivotal clinical decisions. However, poor methodology and poor understanding of the appropriate methods may limit generalizability to the broader clinical setting. Therefore, sound methodology should be incorporated as part of the modern clinician's approach to evidence-based medical care in the appropriate clinical context.

Disclosures

None of the authors had any conflicts of interests or financial disclosures to declare.

References

1. Meta-analysis under scrutiny. Lancet 1997;350:675.
2. Lip GY, Lane DA. Stroke prevention in atrial fibrillation: A systematic review. JAMA 2015;313:1950-62.
3. O'Rourke K. An historical perspective on meta-analysis: Dealing quantitatively with varying study results. J R Soc Med 2007;100:579-82.
4. Haidich AB. Meta-analysis in medical research. Hippokratia 2010;14:29-37.
5. Egger M, Smith GD. Meta-analysis: Potentials and promise. BMJ 1997;315:1371-4.
6. Patsopoulos NA, Analatos AA, Ioannidis JA. Relative citation impact of various study designs in the health sciences. JAMA 2005;293:2362-6.
7. Jacobs AK, Anderson JL, Halperin JL. The Evolution and Future of ACC/AHA Clinical Practice Guidelines: A 30-Year Journey A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;64:1373-84.
8. Bajaj NS, Kalra R, Aggarwal H, Ather S, Gaba S, Arora G, et al. Comparison of approaches to revascularization in patients with multivessel coronary artery disease presenting with ST-segment elevation myocardial infarction: Meta-analyses of randomized control trials. J Am Heart Assoc. 2015;4:e002540.
9. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: A systematic review. Phys Ther 2008;88:156-75.
10. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 1996;17:1-12.
11. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:1-173.
12. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [cited 24 March 2016]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
13. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. Stat Med 1989;8:141-51.
14. Borenstein M, Hedges L, Rothstein H. Meta-analysis fixed effects vs. random effects [cited 25 March 2016]. Available from: https://www.meta-analysis.com/downloads/Meta-analysis%20fixed%20effect%20vs%20random%20effects.pdf.
15. Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration; 2011 [cited 23 March 2016], Version 5.1.0. Available from: www.cochrane-handbook.org.
16. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
17. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.
18. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. BMJ 2013;346:f3914.
19. Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? BMJ 1997;315:1610-4.
20. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105-24.
21. IBM. SPSS Software 2016 [cited 26 March 2016]. Available from: http://www-01.ibm.com/software/analytics/spss/.
22. Comprehensive Meta-Analysis (CMA) 2016 [cited 26 March 2016]. Available from: https://www.meta-analysis.com/.
23. Cochrane Collaboration. RevMan Cochrane Informatics and Knowledge Management Department: Cochrane Informatics and Knowledge Management Department; 2016 [cited 26 March 2016]. Available from: http://tech.cochrane.org/revman.
24. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. BMJ 2009;339:b2700.
25. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. Ann Intern Med 2015;162:777-84.
26. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008-12.
27. Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: Larger studies may be more reliable. Control Clin Trials 1997;18:568-79.
28. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: Pitfalls and hints. Heart Lung Vessels 2013;5:219-25.
29. Egger M, Smith GD. Bias in location and selection of studies. BMJ 1998;316:61-6.
30. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.