Systematic Review

Systematic review and meta-analysis: techniques and a guide for the academic surgeon

Kevin Phan[1], David H. Tian[1], Christopher Cao[1], Deborah Black[2], Tristan D. Yan[1]

[1]The Collaborative Research (CORE) Group, Macquarie University, Sydney, Australia; [2]Faculty of Health Sciences, University of Sydney, Sydney, Australia

Correspondence to: Tristan D. Yan. Professor of Cardiovascular and Thoracic Surgery, The Collaborative Research (CORE) Group, Macquarie University, Sydney, Australia. Email: tristanyan@annalscts.com.

With the rapidly growing literature across the surgical disciplines, there is a corresponding need to critically appraise and summarize the currently available evidence so they can be applied appropriately to patient care. The interpretation of systematic reviews is particularly challenging in cases where few robust clinical trials have been performed to address a particular question. However, risk of bias can be minimized and potentially useful conclusions can be drawn if strict review methodology is adhered to, including an exhaustive literature search, quality appraisal of primary studies, appropriate statistical methodology, assessment of confidence in estimates and risk of bias. Therefore, the following article aims to: (I) summarize to the important features of a thorough and rigorous systematic review or meta-analysis for the surgical literature; (II) highlight several underused statistical approaches which may yield further interesting insights compared to conventional pair-wise data synthesis techniques; and (III) propose a guide for thorough analysis and presentation of results.

Keywords: Systematic review; meta-analysis; surgery; outcomes; forest plot; network meta-analysis; meta-regression

Submitted Feb 13, 2015. Accepted for publication Feb 14, 2015.
doi: 10.3978/j.issn.2225-319X.2015.02.04
View this article at: http://dx.doi.org/10.3978/j.issn.2225-319X.2015.02.04

Introduction

With the rapidly growing literature across the surgical disciplines, there is a corresponding need to critically appraise and summarize the currently available evidence to date so they can be applied appropriately to patient care (1,2). Evidence-based medicine is the explicit, conscientious and judicious use of the currently best available evidence from research to guide health care decisions. Clinical decisions should be based on the totality of the available evidence, rather than based on the results of any individual study or trial (3). In the evidence-based approach to surgery of the modern era, formal and comprehensive reviews of the literature, with or without additional statistical analysis, remains critically important to the practicing surgeon as a source of updated information on diagnoses, prognoses, and the effectiveness of healthcare interventions (4).

While the popularity of systematic review and meta-analyses is increasing, there are also limitations, which surgeons must keep in mind prior to applying their conclusions directly to patient care. The quality of systematic reviews and meta-analyses is a function of the quality of the primary studies available as well as the degree of rigor to which the systematic reviews have been performed (5). A systematic review or meta-analysis that has been poorly performed may produce misleading results and conclusions, using methods or statistical approaches which may lack credibility (6,7).

The conduct and interpretation of systematic reviews is particularly challenging in the case where few robust clinical trials have been performed to address a particular question. However, risk of bias can be minimized and potentially useful conclusions can be drawn if strict review methodology is adhered to, including an exhaustive literature search, quality appraisal of primary studies, appropriate statistical methodology, assessment of confidence in estimates and risk of bias.

Therefore, the following article aims to: (I) summarize

© AME Publishing Company.

to the important features of a thorough and rigorous systematic review or meta-analysis for the surgical literature; (II) highlight several underused statistical approaches which may yield further interesting insights compared to conventional pair-wise data synthesis techniques; and (III) propose a guide for thorough analysis and presentation of results.

Framing the clinical question

Prior to the literature search, the review must ensure that the posed clinical question has a clear focus and is appropriate either for a systematic review or meta-analysis. It is important that the relevance and suitability of the question be carefully appraised so as to be better designed to improve current clinical knowledge, practice and guide policy and decisions.

A well-accepted methodology for this is the use of the PICO format, with a clearly defined study population (P), intervention studied (I), comparisons (C), and outcomes (O) (1). As an example, in a recent systematic review and meta-analysis on sutureless aortic valve replacement (8), the study population was defined as patients with aortic valve replacement requiring surgical intervention (P), the intervention was minimally invasive aortic valve replacement (I), the comparator was conventional aortic valve replacement (C), and outcomes include perioperative mortality and complications (O).

The challenge in developing a robust clinical question to set out to answer is defining the scope of patients and interventions. In order to define the scope of the question, the reviewers must have a comprehensive understanding of the existing literature, the potential gaps and uncertainties in the available evidence, and which gaps can be potentially addressed and answered via a systematic review or meta-analysis.

A scoping review (9) can be performed to explore the extent of the available evidence and to assist in determining the approach scope of the clinical question. Factors to take into consideration when developing the clinical question include the level of evidence and the design of available studies. For example, if there are an adequate number of randomized controlled trials (RCTs) available addressing the proposed clinical question, then the scope of the review should be limited to randomized studies to limit the effect of selection bias from non-randomized studies. In contrast, if there are few randomized trials available and that the evidence is mainly limited to observational studies, then the reviewers may consider broadening the scope of the review to include all observation and randomized trials. Subgroup analysis and sensitivity analysis can be performed as secondary analysis to elucidate any effect of their non-randomized design on the final effect sizes calculated (10,11). In the case of a meta-analysis, the question may be more narrow-focused, for example, with the inclusion of only trials that compare two particular interventions.

Similar logic also applies when considering the broadness or narrowness of scope for population, intervention, comparator and outcomes. For example, if the defined patient population is too broad, then any calculated estimate of effect that is generalized across a wide range of patients with varying risk factors could providing misleading trends and conclusions (12,13). For example, the durability of an implanted aortic valve would be different in young patients (<50 years) versus elderly patients (>70 years), based on their underlying comorbidities and operative risk, and thus pooling freedom from reoperation rates across all ages would providing an inaccurate indication of valve durability in different patient groups (14).

Without a clear question that is clinically relevant and has strictly defined population, intervention, comparator and outcome parameters, the systematic review performed risks being ambiguous, ill-structured, and heterogeneous with invalid interpretations of the results. We recommend the writing or publication of a research protocol prior to conducting the systematic review (15,16), with fully defined inclusion and exclusion criteria, participants, interventions, outcomes of interest and, strategy for statistical analysis.

Literature search

Next, the authors should decide on the appropriate inclusion and exclusion criteria *a priori*, based on the clinical question to be answered. These criteria should be explicitly stated in the final manuscript, and may include the design of studies to be included (RCTs *vs*. any study type; comparative studies only), the study population in terms of gender, age group, disease, the language of the published studies, or the time period of publication (e.g., inclusion of studies published after year 2000). Exclusion may include, but not limited to, low levels of evidence such as abstract-only articles, conference articles, editorials and expert opinions, removal of studies with repeat follow-up reports of the same population, removal of studies which report fewer than 10 patients per arm, etc. The potential bias generated by these inclusion and exclusion criteria should be considered and

© AME Publishing Company.

discussed along with the findings of the review.

To ensure a comprehensive and exhaustive literature search for appropriate primary studies, multiple databases should be systematically searched (17). Typical electronic databases include MEDLINE, EMBASE, PubMed, Cochrane Central Register of Controlled trials and ACP. Searching only a single electronic database is not recommended, as there is a high chance that relevant articles will be missed (18,19). It is important for the reviewers to use specific keywords and MeSH headings according to their clinical question posed, as well as Boolean operations such as "AND", "OR", and "NOT". As per recommended PRISMA guidelines (20), an example of at least one search strategy should be provided in the final manuscript, either as one of the main text tables or as a supplementary table or appendix data. The literature search should be conducted by at least two reviewers independently. Any discrepancies in the final list of articles to be included should be discussed and resolved by consensus. Furthermore, additional references should be identified via searches of trial registries, reference lists of included studies, foreign language literature and contacting experts in the field (21-23). The overall search strategy should be presented as a PRISMA flow-chart in the final manuscript.

Study quality appraisal

In order to assess intra-study risk of bias, which may undermine the validity of the final results obtained, quality appraisal should be performed according to Cochrane and MOOSE guidelines. There are different checklists and tools available to assess intra-study risk of bias (24,25). The Cochrane Collaboration Review Manager software has an inbuilt tool for the assessment of RCTs (26), based on the following domains: sequence generation, allocation concealment, blinding of participants and outcomes assessment, incomplete outcome data, selective outcome reporting, and other sources of bias. RCTs are often regarded as high-quality "gold-standard" studies (27) that systematic reviews and meta-analyses should ideally include. However, different RCTs can have variations in methodological and reporting quality, and thus it is important to assess all RCTs for quality appraisal. High-quality RCTs should follow the reporting standards set out by the Consolidated Standards of Reporting Trials (CONSORT) statement, which includes a 22-item checklist and flow diagrams (28).

There are also multiple checklists available for the assessment of observational studies (29-31). The Dutch Cochrane Collaboration Group has developed one checklist and another commonly used checklist is STROBE (29). The key domains assessed by the MOOSE tool includes (31): (I) clear definition of study population; (II) clear definition of outcomes and outcomes assessment; (III) independent assessment of outcome parameters; (IV) sufficient follow-up; (V) no selective loss during follow-up; and (VI) important confounders and prognostic factors identified. It is highly recommended that these checklists be used to rigorously assess the quality of included studies. Completed checklists should be included as a table in the final manuscript, or as a supplementary table. The risk of bias assessment should be performed by at least two reviewers, and any discrepancies should be resolved by consensus. An example of such a risk assessment checklist is demonstrated in *Table 1* from recent systematic reviews (8,32) in surgery. Studies which do not fulfil the a priori quality requirements should be carefully appraised to determine suitability of inclusion in further statistical analyses, such as via sensitivity analysis or cumulative meta-analysis (33-35) to detect heterogeneity or adjustment in effect size with time (34,36,37).

Data extraction

Data extraction should be performed according to a template with relevant demographic, operational parameters and outcomes that have been defined prior to review. Data extraction should be performed by at least two reviewers, and discrepancies resolved by consensus. This will reduce the risk of reviewer bias, error and subjectivity.

Statistical approaches

Summary statistics

The statistical techniques employed will depend on the type of review performed and available data. In the case of a systematic review of a single surgical intervention without comparison, it would be appropriate to report descriptive summary statistics in the form of mean, standard deviation, and range for parametric continuous measures. If the demographics of the study populations and inclusion/exclusion are relatively similar between the studies, this may warrant a meta-analysis of weighted proportions, using a random-effects model to pool one-arm cohort studies. An example of such analysis was performed to determine the weighted pooled paravalvular leak rate from sutureless aortic valve replacement (SU-AVR) (8) at 12-month follow-up (*Figure 1*).

© AME Publishing Company.

Table 1 Example of an intra-study risk of bias assessment table, according to Dutch Cochrane Group and MOOSE guidelines. Each cell can be filled in with either yes/no/unclear. Study 1 and 5 would be deemed of higher quality as all criteria was rated as "yes" for being free from intra-study bias

Questions	Study 1	Study 2	Study 3	Study 4	Study 5
Clear definition of study population?	Yes	Yes	Yes	Yes	Yes
Clear definition of outcomes and outcome assessment?	Yes	Yes	No	Yes	Yes
Independent assessment of outcome parameters?	Yes	No	No	Unclear	Yes
Sufficient duration of follow-up?	Yes	Yes	Yes	No	Yes
No selective loss during follow-up?	Yes	Yes	Yes	Yes	Yes
Important confounders and prognostic factors identified?	Yes	Yes	No	No	Yes

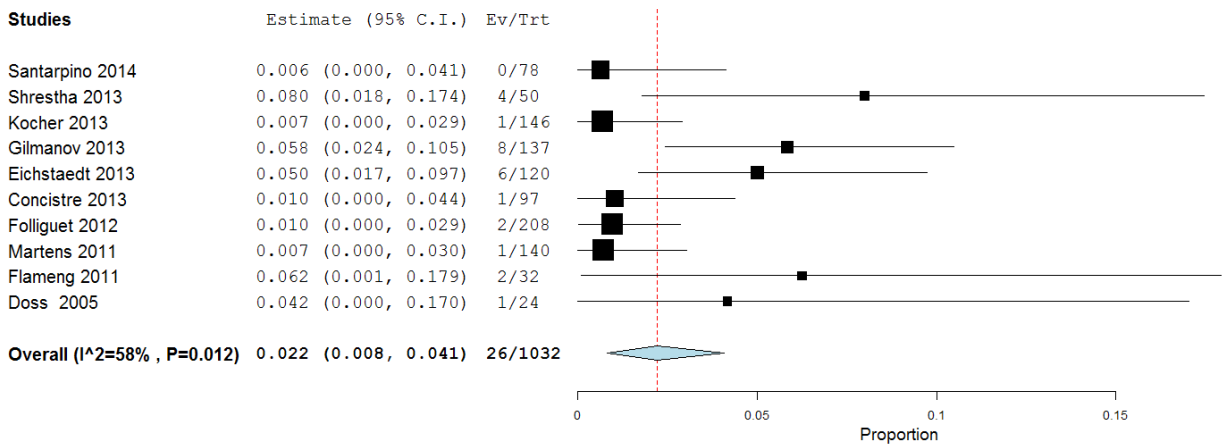


Figure 1 Meta-analysis of proportions of single-arm studies reporting paravalvular leak outcomes for SU-AVR, using a random-effects model. The weighted mean paravalvular leak rate here was 3.0%, with 95% confidence interval of 1.0-5.8%. Significant heterogeneity was demonstrated (I^2=72%, P<0.001). SU-AVR, sutureless aortic valve replacement.

For a systematic review and meta-analysis of comparative studies, presentation of summary statistics in the form of Forest plots is appropriate (38). Forest plots involve a weighted compilation of all the effect sizes reported by each study, and also provide an indication of heterogeneity between studies. An example is shown in *Figure 2*, from a recent meta-analysis comparing ministernotomy versus minithoracotomy approaches for minimally invasive aortic valve replacement (39). For each study, the effect size is represented by a square and horizontal line, representing the point estimate and 95% confidence interval, respectively. The size of the square is proportion to the weight assigned to that particular study for the meta-analysis. The pooled effect size following meta-analysis is represented by the black diamond, the width of which is indicated of the overall 95% confidence interval. If this diamond lies totally to one side of the solid vertical line in the center, then the pooled point estimate indicates a significant difference in effect size between the two interventions (38).

Typical summary statistics used for point estimates include relative risk (RR) or odds ratio (OR) for dichotomous parameters, and weighted mean difference (WMD) for continuous data. The analysis can be performed using fixed-effect or random-effects models. In the prior, the true effect size is assumed to be similar among the included studies, whilst in the latter, the included studies represent a random sample of effect sizes. The random-effects model is most appropriate in the case where there is risk of heterogeneity in the effect size reported.

In order to assess whether effect sizes were consistent across the included studies, heterogeneity should be quantified (40). There are two commonly used tests for the assessment of heterogeneity. The Cochran Q test provides a yes *vs.* no outcome for whether there is significant

© AME Publishing Company.

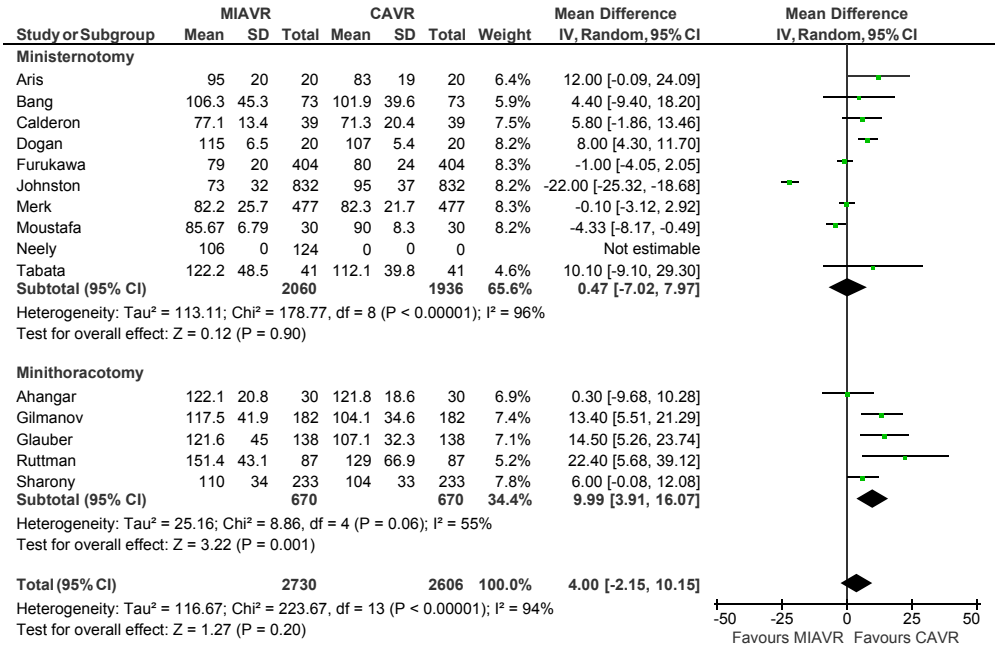


Figure 2 Example of forest plot comparing cardiopulmonary bypass duration between MIAVR versus CAVR, using weighted mean difference as summary statistic. Note that subgroup analysis was performed, comparing ministernotomy versus minithoracotomy approaches for MIAVR. Significant heterogeneity was observed for all subgroups tested here. MIAVR, minimally invasive aortic valve replacement; CAVR, conventional aortic valve replacement.

heterogeneity amongst the reported effect sizes (41). In comparison, the I^2 statistic provides a magnitude of variability, where 0% indicates that any variability is due to chance, whilst higher I^2 values indicate increasing levels of unexplained variability. Usually, I^2 value greater than 50% suggests significant heterogeneity in the reported effect sizes.

When significant heterogeneity arises, the source of this heterogeneity should be explained in the manuscript (40). There are several techniques that may be used to determine a potential source of variability. Firstly, subgroup analysis can be performed (42). The same analysis for the outcome of interest is performed for the subgroups, and a test of interaction can be performed to determine whether there are significant differences between the subgroups. If the calculated P value is significant usually set at 5%, there is a high chance that there is an association between the particular subgroup and outcome of interest. The other technique that can be used to assess heterogeneity is meta-regression analysis, described below.

Meta-regression analysis

Another technique used to assess heterogeneity is meta-regression analysis. This statistical approaches determines whether there is a significant association between an independent variable in form of study or intervention characteristics (for example age, study time point, operation duration) versus the dependent variable, the outcome of interest (43). A regression model is constructed, and the P value and regression coefficient (r) can be used to assess the strength of this association. A significant relationship may indicate that the study variable may be a source of variability observed (8,43-45). For example in *Figure 3*, meta-regression analysis was used to show a significant negative correlation between midpoint of study period with rate of paravalvular leak for the Perceval S sutureless prosthesis (8). This suggests that the learning curve of SU-AVR may be a source of heterogeneity in the pooled results for paravalvular leak.

Network meta-analysis

Network meta-analysis may be appropriate for assessment multiple interventions (>2) for the same disease or outcome. Also known as multiple-treatment comparison, network meta-analyses aim to pool all available direct and indirect

© AME Publishing Company.

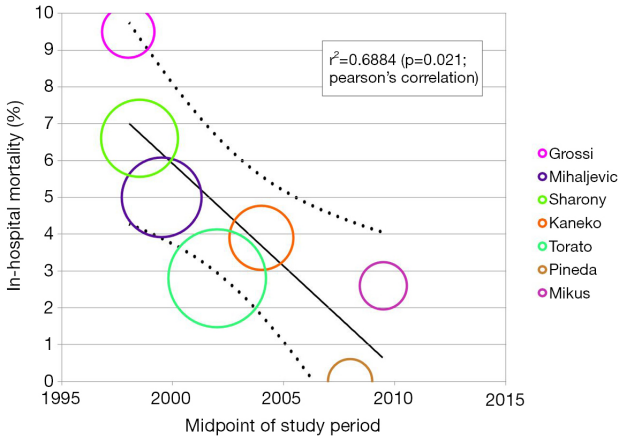


Figure 3 Meta-regression analysis of midpoint of study period versus paravalvular leak rate in Perceval S sutureless aortic valves. Significant negative correlation was found (r=–0.853, P=0.031), suggesting that there was a decrease in paravalvular leak rate with time. This is one potential source of heterogeneity in the meta-analysis of proportions in *Figure 1*, suggesting that paravalvular leak is a function of temporal learning curve.

comparisons for multiple interventions to provide an overall comparison (46-48). In contrast to traditional pair-wise meta-analysis, the advantage of network meta-analysis is that using indirect evidence, all available data can be synthesized to provide comparative effect estimates between interventions where there may not have been direct head-to-head trials performed to date.

While the technical statistical details of network meta-analyses is beyond the scope of the present article, it is suffice to say that available direct evidence (A versus B) and indirect evidence (A versus C, C versus B) can be applied to a "Bayesian" statistical model where Monte Carlo simulations are run (49). The model will converge to the likelihood effect estimate and provide a modeled comparison between A versus B versus C. The underlying assumption of this approach is that the comparator group for the interventions (i.e., C) is similar amongst the indirect comparison trials (50).

Typical software packages for network meta-analyses include WinBUGS and GeMTC packages. The general work process for network meta-analysis is similar to traditional meta-analysis, involving: (I) data extraction from direct and indirect comparative studies; (II) importing the data into the software package e.g., WingBUGS or GeMTC; (III) running the Bayesian model and Monte Carlo simulations.

In the first systematic review to compare median sternotomy versus ministernotomy versus minithoracotomy for minimally invasive aortic valve replacement, a Bayesian network meta-analysis was performed based on direct and indirect evidence (39). This is particularly pertinent, given so few studies were available providing a head-to-head comparison between ministernotomy and minithoracotomy. As such, network meta-analysis was performed to produce an integrated effect size for ministernotomy *vs.* minithoracotomy, based on all available direct and indirect evidence.

The caveats of multiple-intervention analysis are that it is more susceptible to the influence of heterogeneity compared to pair-wise analysis (51,52). There are several different models that can be run during a network meta-analysis to gauge the effect of heterogeneity. These statistical models include consistency, inconsistency and node-splitting models. If significant heterogeneity is detected, the inconsistency and node-splitting model results should be presented, and interpretations should be made with caution (53).

Time-to-event data synthesis

In systematic reviews and meta-analyses, time-to-event outcomes such as survival are ideally pooled using meta-analysis of hazard ratios (HR). However, in many studies, the HR is not reported and individual patient data (IPD) is not available. Some studies have simply estimated actuarial survival outcomes visually from Kaplan-Meier plots and reported these in systematic reviews and meta-analysis. However, this approach does not take into account the censoring and loss of patients to follow-up that occurs, and does not allow estimation of a HR. To address this, several statistical approaches have been proposed in the literature that allows estimation of HR based on other summary statistics published (54-57). The estimated HR can then be used for meta-analysis to allow data synthesis from all available evidence in the literature.

One common HR estimation technique was proposed and validated by Tierney *et al.* in 2007 (58). In this technique, available data from Kaplan Meier curves can be digitized using computer software such as Digitizelt, and the recorded number of patients at risk noted from the paper. Therefore the computerized actuarial survival rate and number of patients at risk for sequential follow-up periods can be accurately estimated. Tierney *et al.* also developed

© AME Publishing Company.

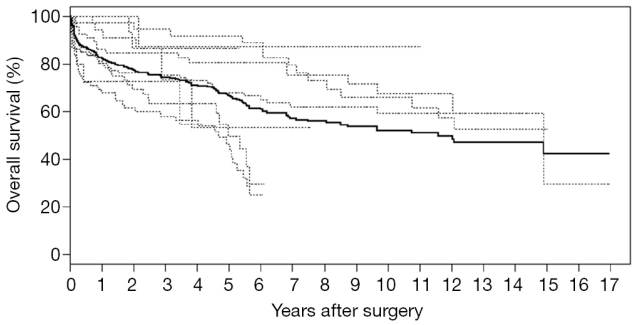


Figure 4 Overall survival based on reconstructed individual patient data. Data of 458 patients from seven studies on open surgical repair for chronic type B aortic dissection were reconstructed and presented. Dotted lines represent Kaplan-Meier curves of individual studies, while the solid line represents aggregate reconstructed survival data of the entire cohort.

an Excel spreadsheet (58), when these data are inputted and assuming constant censoring, an estimation of the original IPD for the particular study can be deduced. Performing a similar analysis amongst all included studies for a meta-analysis can allow data synthesis of reconstructed IPD, thus facilitating meta-analysis of time-to-event outcomes.

IPD can also be reconstructed using another recent approach proposed by Guyot and colleagues in 2012 (59). This group has produced an iterative algorithm that is able to solve Kaplan-Meier equations used to produce the graphs in the original publication. Similar to methods by Tierney *et al.*, a computerizing software such as Digitizelt can be utilized to digitize Kaplan-Meier curve data. This can then be entered into the iterative algorithm to determine optimal solutions to the Kaplan-Meier equations. Again, this algorithm assumes constant censoring and can be performed in the R statistical software. The reconstructed patient survival software can be aggregated to form combined survival curves. This approach was recently used to pool long-term time-to-event survival data for open surgical repair for chronic type B aortic dissection, producing aggregate Kaplan-Meier curves as shown in *Figure 4* (60). This approach has also been recently used other reviews in the cardiothoracic surgery literature (61), and its usage is expected to increase in the near future with the greater need for pooled time-to-event data.

Publication bias

Another inherent limitation of systematic reviews is publication bias (62,63). Often in the literature, studies that have produced negative results are more difficult to publish compared to studies which produce a positive result. As a result, there is often more "missing" negative result studies, which may skew the outcomes of meta-analysis can provide misleading results.

As such, it is important that a systematic review or meta-analysis evaluates the potential influence of publication bias. One commonly used approach for the assessment of publication bias is the use of funnel plots (64). This is a plot that graphs precision versus magnitude of effect treatment and has a shape of an inverted funnel. The horizontal axis represents the intervention effect, and the vertical axis represents the standard error. Ideally, in the case of minimal publication bias, the points of the funnel plot are symmetrically distributed around the mean effect size. Asymmetrical distribution indicates potential publication bias which may undermine the validity of conclusions. Begg and Egger's tests can be used in conjunction to statistically determine whether the asymmetry is significant or not (65).

To assess the effect of the "missing studies" due to publication bias on the calculated effect size, trim-and-fill analysis can be performed (66). This is an extension of the funnel plot technique, where "missing studies" are identified and estimated based on the symmetry of the funnel plot. The deduced missing studies can be imputed into the funnel plot, and the whether the resultant change in effect size is significant or not can assist in evaluation of publication bias. Recent meta-analyses in the cardiothoracic surgical literature have used this methodology to assess the effect of publication bias and "missing studies", and an example is presented in *Figure 5* (60,67,68). The caveat of this technique is that it is purely based on notion that no publication bias equates to a perfectly symmetry funnel plot, which may or may not be the case. Furthermore, the source or mechanism of publication bias is not addressed, and thus funnel plots and trim-and-fill analysis should be interpreted with caution.

Interpretation of results

There are several factors that must be considered when discussing results from the systematic review and meta-analysis. Firstly, the reviewers should assess the clinical significance of the results. For example, if there was a calculated significant difference in operation duration between two interventions as five minutes, would this significantly affect the outcomes of the patient? Prior

© AME Publishing Company.

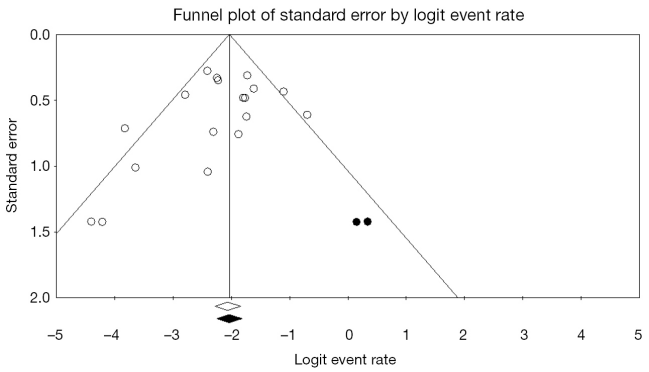


Figure 5 Example of a funnel plot with trim-and-fill analysis. Funnel plot for systematic review of open surgical repair for chronic type B dissection in all 19 included studies. The logit event rate for mortality (horizontal axis) is presented against the standard error (SE) of the log of logit event rate (vertical axis). The SE inversely corresponds to the study size. Asymmetry of the plot can indicate publication bias. Open circles indicate included studies, while the filed circles represent imputed studies identified through trim-and-fill analysis.

studies have suggested that the mean clinically important difference (MCID) of a therapy or intervention should be involved in the planning of clinical trials and interpretation of results (69). Secondly, the reviewer should discuss and explain potential sources of heterogeneity in the discussion. This may involve performing subgroup analysis and meta-regression analysis to determine which factors affect the outcome of interest. Thirdly, the review should discuss their strengths compared to prior reviews (if applicable) and limitations of the review, which may include but not limited to the inclusion of non-randomized studies (which may lead to risk of bias), small patient sample numbers, significant differences in baseline characteristics of comparative cohorts, short follow-up duration and heterogeneity in the surgical intervention techniques used amongst the included studies.

GRADE assessment of outcomes

The quality of scientific evidence and outcomes can further be assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (70,71). This is a transparent and structured approach for rating the evidence for a particular outcome. The GRADE methodology involves rating evidence for an outcome by upgrading or downgrading of evidence. Indications for upgrading evidence include having a large

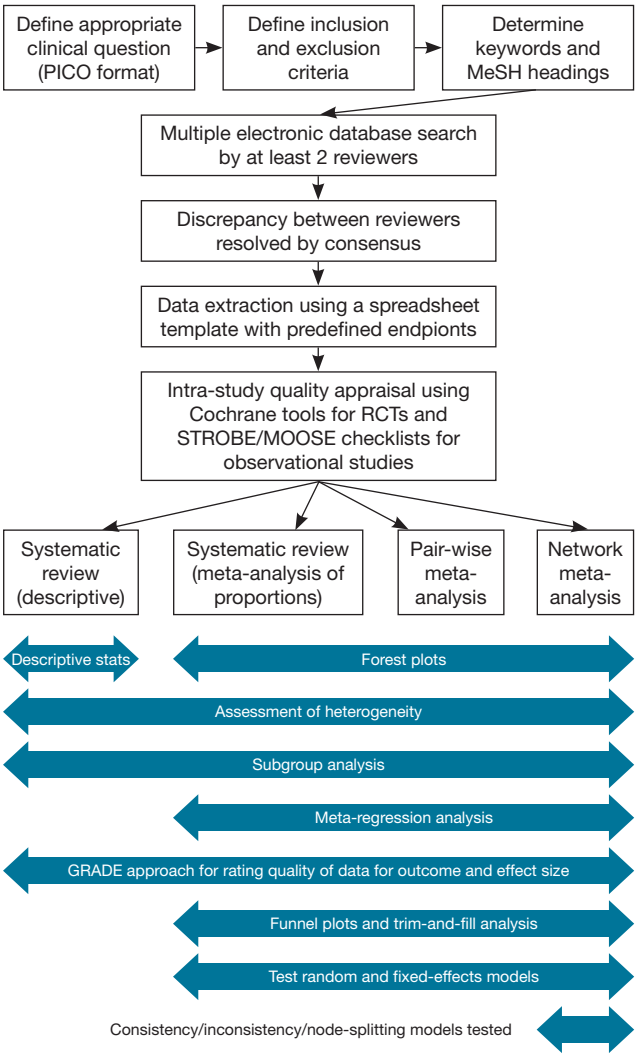


Figure 6 Overview of systematic review and meta-analysis workflow for the presentation of results.

effect size and dose-response gradient. Indications for downgrading quality of evidence include serious risk of bias, serious inconsistency between studies, serious indirectness, serious imprecision and likely publication bias. The GRADE approach may allow the reviewer to have increased or decreased confidence in the effect sizes presented, that is, a higher confidence in the true association (71,72).

Presentation of review results

An overview of the systematic review and meta-analysis process is demonstrated in *Figure 6*. In order to assist surgeons in their rigorous conduction of systematic reviews

© AME Publishing Company.

Table 2 The Collaborative Research (CORE) Group approach for presentation of results in a rigorously performed systematic review and meta-analysis

Item	Contents
Figure	PRISMA flow-chart for the search strategy process
	Forest plot for variable 1
	Forest plot for variable 2
	Forest plot for variable 3
	Risk of bias summary/graph, if applicable
	Meta-regression analysis graph, if applicable
Table	Study characteristics of included articles
	Baseline characteristics of included studies
	Table outlining details of surgical techniques, inclusion and exclusion criteria, follow-up protocols for each study
Supplementary Table	Search strategy for at least one database
	A completed PRISMA or AMSTAR checklist
Supplementary Figure	Any other relevant forest plots
	Funnel plot and trim-and-analysis plots. This can alternatively be included as a main figure

and meta-analyses, we recommend the following structure for presentation of results (*Table 2*).

Conclusions

Systematic reviews and meta-analyses are increasingly important in the surgical realm for data synthesis and quality appraisal of available evidence. However, surgeons should be wary of the quality of systematic reviews, which may seriously undermine the validity of presented results and conclusions. In order maintain high quality reviews with credible results, the review process should be standardized and strictly adhered. Here we have presented an overview of such a process to ensure optimal systematic review and meta-analysis outcomes and presentation.

Acknowledgements

Disclosure: The authors declare no conflict of interest.

References

1. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994;272:1367-71.
2. Swingler GH, Volmink J, Ioannidis JP. Number of published systematic reviews and global burden of disease: database analysis. BMJ 2003;327:1083-4.
3. Murad MH, Montori VM. Synthesizing evidence: shifting the focus from individual studies to the body of evidence. JAMA 2013;309:2217-8.
4. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376-80.
5. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.
6. Wilson P, Petticrew M; Medical Research Council's Population Health Sciences Research Network knowledge transfer project team, et al. Why promote the findings of single research studies? BMJ 2008;336:722.
7. Petticrew M. Why certain systematic reviews reach uncertain conclusions. BMJ 2003;326:756-8.
8. Phan K, Tsai YC, Niranjan N, et al. Sutureless aortic valve replacement: a systematic review and meta-analysis. Ann Cardiothorac Surg 2015;4:100-11.
9. Armstrong R, Hall BJ, Doyle J, et al. Cochrane Update. 'Scoping the scope' of a cochrane review. J Public Health (Oxf) 2011;33:147-50.
10. Wright CC, Sim J. Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. J Clin Epidemiol 2003;56:833-42.
11. Oxman AD, Guyatt GH. A consumer's guide to subgroup

© AME Publishing Company.

analyses. Ann Intern Med 1992;116:78-84.
12. Horwitz RI. "Large-scale randomized evidence: large, simple trials and overviews of trials": discussion. A clinician's perspective on meta-analyses. J Clin Epidemiol 1995;48:41-4.
13. Eysenck HJ. Meta-analysis and its problems. BMJ 1994;309:789-92.
14. Cao C, Ang SC, Indraratna P, et al. Systematic review and meta-analysis of transcatheter aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis. Ann Cardiothorac Surg 2013;2:10-23.
15. Van der Wees P, Qaseem A, Kaila M, et al. Prospective systematic review registration: perspective from the Guidelines International Network (G-I-N). Syst Rev 2012;1:3.
16. Booth A, Clarke M, Dooley G, et al. PROSPERO at one year: an evaluation of its utility. Syst Rev 2013;2:4.
17. Smith BJ, Darzins PJ, Quinn M, et al. Modern methods of searching the medical literature. Med J Aust 1992;157:603-11.
18. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994;309:1286-91.
19. Dickersin K, Hewitt P, Mutch L, et al. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. Control Clin Trials 1985;6:306-17.
20. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006-12.
21. Dickersin K, Chan S, Chalmers TC, et al. Publication bias and clinical trials. Control Clin Trials 1987;8:343-53.
22. McAuley L, Pham B, Tugwell P, et al. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000;356:1228-31.
23. Cook DJ, Guyatt GH, Ryan G, et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. JAMA 1993;269:2749-53.
24. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16:62-73.
25. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1-12.
26. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
27. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2.
28. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-9.
29. von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335:806-8.
30. von Elm E, Egger M. The scandal of poor epidemiological research. BMJ 2004;329:868-9.
31. 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;283:2008-12.
32. Phan K, Ha H, Phan S, et al. Early hemodynamic performance of the third generation St Jude Trifecta aortic prosthesis: systematic review and meta-analysis. J Thorac Cardiovasc Surg 2015. [Epub ahead of print].
33. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. J Clin Epidemiol 1995;48:45-57; discussion 59-60.
34. Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med 1992;327:248-54.
35. Villanueva EV, Zavarsek S. Evaluating heterogeneity in cumulative meta-analyses. BMC Med Res Methodol 2004;4:18.
36. Phan K, Xie A, Di Eusanio M, et al. A meta-analysis of minimally invasive versus conventional sternotomy for aortic valve replacement. Ann Thorac Surg 2014;98:1499-511.
37. Phan K, Xie A, La Meir M, et al. Surgical ablation for treatment of atrial fibrillation in cardiac surgery: a cumulative meta-analysis of randomised controlled trials. Heart 2014;100:722-30.
38. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322:1479-80.
39. Phan K, Xie A, Tsai YC, et al. Ministernotomy or minithoracotomy for minimally invasive aortic valve replacement: a Bayesian network meta-analysis. Ann Cardiothorac Surg 2015;4:3-14.
40. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351-5.
41. Rücker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 2008;8:79.
42. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323:42-6.
43. Baker WL, White CM, Cappelleri JC, et al. Understanding heterogeneity in meta-analysis: the role of meta-regression. Int J Clin Pract 2009;63:1426-34.
44. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med

© AME Publishing Company.

2002;21:1559-73.
45. Phan K, Zhou JJ, Niranjan N, et al. Minimally invasive reoperative aortic valve replacement: a systematic review and meta-analysis. Ann Cardiothorac Surg 2015;4:15-25.
46. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897-900.
47. Sutton A, Ades AE, Cooper N, et al. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008;26:753-67.
48. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol 2010;63:875-82.
49. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Research Synthesis Methods 2012;3:80-97.
50. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. BMC Med 2013;11:159.
51. Caldwell DM, Gibb DM, Ades AE. Validity of indirect comparisons in meta-analysis. Lancet 2007;369:270; author reply 271.
52. Sturtz S, Bender R. Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency. Research Synthesis Methods 2012;3:300-11.
53. Song F, Xiong T, Parekh-Bhurke S, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ 2011;343:d4909.
54. Dear KB. Iterative generalized least squares for meta-analysis of survival data at multiple times. Biometrics 1994;50:989-1002.
55. Arends LR, Hunink MG, Stijnen T. Meta-analysis of summary survival curve data. Stat Med 2008;27:4381-96.
56. Fiocco M, Putter H, van Houwelingen JC. Meta-analysis of pairs of survival curves under heterogeneity: a Poisson correlated gamma-frailty approach. Stat Med 2009;28:3782-97.
57. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815-34.
58. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8:16.
59. Guyot P, Ades AE, Ouwens MJ, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012;12:9.
60. Tian DH, De Silva RP, Wang T, et al. Open surgical repair for chronic type B aortic dissection: a systematic review. Ann Cardiothorac Surg 2014;3:340-50.
61. Phan K, Ha HS, Phan S, et al. New-onset atrial fibrillation following coronary bypass surgery predicts long-term mortality: a systematic review and meta-analysis. Eur J Cardiothorac Surg 2015. [Epub ahead of print].
62. Chalmers I. Underreporting research is scientific misconduct. JAMA 1990;263:1405-8.
63. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. Lancet 1991;337:867-72.
64. 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.
65. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.
66. 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;56:455-63.
67. Xie A, Phan K, Yan TD. Durability of continuous-flow left ventricular assist devices: a systematic review. Ann Cardiothorac Surg 2014;3:547-56.
68. Tsai YC, Phan K, Munkholm-Larsen S, et al. Surgical left atrial appendage occlusion during cardiac surgery for patients with atrial fibrillation: a meta-analysis. Eur J Cardiothorac Surg 2014. [Epub ahead of print].
69. Man-Son-Hing M, Laupacis A, O'Rourke K, et al. Determination of the clinical importance of study results. J Gen Intern Med 2002;17:469-76.
70. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6.
71. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-6.
72. Seco M, Cao C, Modi P, et al. Systematic review of robotic minimally invasive mitral valve surgery. Ann Cardiothorac Surg 2013;2:704-16.

Cite this article as: Phan K, Tian DH, Cao C, Black D, Yan TD. Systematic review and meta-analysis: techniques and a guide for the academic surgeon. Ann Cardiothorac Surg 2015;4(2):112-122. doi: 10.3978/j.issn.2225-319X.2015.02.04

© AME Publishing Company.