Systematic reviews and meta-analyses in spine surgery, neurosurgery and orthopedics: guidelines for the surgeon scientist

Kevin Phan, Ralph J. Mobbs

The NeuroSpine Surgery Research Group (NSURG), Sydney, Australia
Correspondence to: Kevin Phan. The NeuroSpine Surgery Research Group (NSURG), Neuro Spine Clinic, Suite 7a, Level 7 Barker Street, Randwick, New South Wales 2031, Australia. Email: kphan.vc@gmail.com.

The research evidence in the realm of surgery is expanding at a rapid pace, and thus corresponds with an increasing need to critically appraise and synthesize the available literature. Particularly in fields such as spine surgery, neurosurgery and orthopedics which traditionally have little Class I randomized clinical data, reviews are important to pool the available evidence on clinical questions which are otherwise difficult to answer. Whilst systematic reviews and meta-analyses have the potential to provide critical and updated surgical evidence to guide clinical decisions, poorly performed analyses and misinterpretation of such reviews may have a detrimental effect on patient care and outcomes. We present a summary of the critical steps in performing a systematic review and meta-analysis, allowing the surgeon scientist to better interpret and perform their own systematic reviews and meta-analyses.

Keywords: Systematic review; meta-analysis; evidence; clinical outcomes; academic surgery; forest plot; heterogeneity; surgery; spine

Introduction

The research evidence and literature in the realm of surgery is expanding at a rapid pace. In the face of a multitude of resources and conflicting evidence, clinical decision making and applying the most appropriate care for patients may become difficult and challenging (1). As such, there is an increasing need to critically appraise and synthesize the available evidence in order to guide health care policies, interventions and decisions (2,3).

Systematic reviews and meta-analyses aim to critically appraise and pool all available literature to produce a set of recommendations or directions for future studies (3). Particularly in fields such as spine surgery, neurosurgery and orthopedics which traditionally have little Class I randomized clinical data, reviews are important to pool the available data and provide evidence on clinical questions which are otherwise difficult to answer. Furthermore, the rapid evolution of new surgical techniques and technologies

in these fields also lends itself to continually updating the evidence base. For example, in the case of spine surgery, the evolution of posterior (4), transforaminal (5), anterior (4), lateral (6) and most recently oblique techniques for lumbar spinal fusion has been propelled forward with judicious and continual appraisal required and updating of the surgical evidence necessary (7).

The validity of review articles is highly dependent on its methodological quality. In contrast to a narrative review, which are often biased, systematic reviews use informal and subjective methodology to find, extract and appraise all available evidence to the clinical question being studied (8). Whilst systematic reviews and meta-analyses have the potential to provide critical and updated surgical evidence to guide clinical decisions, poorly performed analyses and misinterpretation of such reviews may have a detrimental effect on patient care and outcomes. This is particularly pertinent for clinical questions where few high-

quality randomized controlled trials have been conducted. Inadequate literature searches, lack of rigorous quality assessment, inappropriate methodology or statistical analysis may also lead to misleading results (9-11).

As systematic reviews and meta-analyses become more commonly understood and accepted in fields such as spine surgery, neurosurgery and orthopedics, we believe they will ultimately be the format expected for reviews in our journal. The aim of this paper is to summarize the critical steps in performing a systematic review and meta-analysis, allowing readers to better interpret and perform their own systematic reviews and meta-analyses.

Getting started: systematic review versus meta-analysis

Prior to conducting a review, it is important to understand the differences between a systematic review and meta-analysis. These terms are often used interchangeably, albeit erroneously. A systematic review is a qualitative, high-level evidence synthesis of primary research. All studies are chosen according to pre-determined selection criteria, and systematic methodology is employed to minimize bias.

Compared to a systematic review, a meta-analysis employs additional statistical techniques and analyses to provide a quantitative synthesis of evidence from pooled data (12). The conclusions of meta-analyses are often reported in terms of a pooled effect size such as relative risk (RR) or weighted mean difference (WMD), with a 95% confidence interval (CI). Results are often displayed graphically in the form of forest plots and additional analyses are performed to determine the heterogeneity among included studies. A systematic review may or may not include a meta-analysis component, whilst some but not all meta-analyses are systematic.

Research question

A defining feature of a systematic review and meta-analysis in clinical medicine is that it sets out to answer specific healthcare questions, rather than providing general summaries of available literature on a topic. As such, identifying and framing the clinical question to answer is arguably one of the most important and significant checkpoints of conducting a rigorous systematic review. Typically, a scoping search of the literature (13) is performed to allow the reviewer to garner a better understanding of the clinical problem, its boundaries, the current knowledge and

what is not known. Furthermore, it is important to ensure that the research question composed can be answered or addressed in the form of a systematic review or meta-analysis.

From the scoping search, the main aim of the systematic review should be identified: to highlight the strengths or limitations of the available literature, to address a controversy or conflicting advice on a particular topic, to improve the precision of an effect size known about an intervention or therapy, or to avoid future unnecessary trials.

The type of research question that is addressed can fall into several categories. A review focusing on etiology will aim to determine the potential causes of a certain pathology or disease. A question focusing on diagnosis will assess whether a method or strategy is good for detecting a particular condition. A review focusing on prognosis will address the question of what is the probability of developing a particular outcome. However, the most common form of review question in the surgical arena is one of intervention, to assess the benefits and risks of a particular treatment or surgical procedure compared with another intervention or therapy.

A scaffold which can be used to ensure a structured and reliable approach for defining the research question is "PICOTS" (14), which stands for Population, Intervention, Comparison, Outcome, Time, and Study Design. This scaffold ensures that the clinical question addressed is not too broad, which would limit the applicability of results, or too narrow, which would limit the generalizability of the outcomes. A recent meta-analysis (4) on anterior lumbar interbody fusion (ALIF) versus transforaminal lumbar interbody fusion (TLIF) will be used to demonstrate the use of the PICOTS scaffold. In this review, the population defined was all adults requiring a fusion procedure. The intervention was ALIF whereas the comparator was TLIF. Outcomes examined included fusion rates, operative duration, blood loss, hospital stay, changes in radiological disc height, segmental lordosis, lumbar lordosis, and functional outcomes [oswestry disability index (ODI), visual analogue scale (VAS) leg/back pain scores]. For this particular study, the Time was defined as outcomes reported postoperatively upon follow-up beyond perioperative 30 days. Finally, the study types included all comparative studies of ALIF versus TLIF, including randomized controlled studies, prospective and retrospective observational studies.

Eligibility of studies and study outcomes

The eligibility criteria for study selection in a systematic

review should be established prior to the process of identifying, locating and retrieving the articles. The eligibility criteria specifically define the types of articles to be included and excluded. The inclusion and exclusion criteria can be derived from PICOTS, however, they are typically applied liberally initially to ensure that all relevant articles are included and that articles are not excluded without thorough assessment. The eligibility criteria typically states the study population, nature of intervention studied, outcome variables, time period (e.g., time frame beyond year 2000), and linguistic range (studies reported in English only?). Additional criteria may include methodological quality and level of evidence (e.g., randomized controlled trials (RCTs) only, or inclusion of observational studies). Exclusion criteria may include low-level evidence such as abstracts, conference articles, editorials, comments and expert opinions. For multiple studies published from the same patient cohort but with different follow-up durations, earlier publications may be excluded. Exclusion criteria may also include studies with low number of patients studied, for example, fewer than 10 patients per arm of a comparative analysis.

In recent years, there has been an increased uptake in "minimally invasive" technologies and techniques in the arena of spine surgery. However, there is still a lack of clear consensus and guidelines on the precise definition of minimally invasive spinal surgery. As such, the selection criteria of such reviews focusing on minimally invasive spinal approaches must define what types of interventions will be included as "minimally invasive surgery". A meta-analysis (5) comparing outcomes of minimally invasive TLIF (MI-TLIF) and open-TLIF (O-TLIF) defined minimally invasive surgery as surgery conducted through a "*tube, cylindrical retractor blades or sleeves via a muscle-dilating or muscle-splitting approach*". The conventional or O-TLIF approach was defined as an approach which included "*elevating or stripping the paraspinal muscles to gain access to the spine, even via a limited midline incision*". It is important to define these terms when setting out the eligibility criteria, particularly because of the heterogeneous use of "minimally invasive", "mini-open", and "open" in the spine literature.

The review authors should also clearly define their study outcomes, specifically the primary and secondary outcomes. The primary outcomes should be similar to the equivalent in a randomized controlled study or clinical trial. There should be greater statistical power for the primary outcome in a review, compared to secondary outcomes and post-hoc comparative analysis.

Research protocol

A research protocol summarizing the clinical question formulated, and techniques for database searching, article screening, data extraction, statistical analysis should be clearly stated. The systematic review protocol should clearly state: (I) the objective(s) for the systematic review or meta-analysis; (II) definition of the population to be studied; (III) intervention(s) or therapy investigated; (IV) outcome measures including primary and secondary outcomes; (V) follow-up duration of outcomes included; (VI) search strategy used for the review, including the databases to be searched, and inclusion and exclusion criteria; and (VII) the statistical analyses that will be employed for the review, if appropriate. Ideally, the review should follow recommended guidelines from Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). To ensure transparency, review protocols can be published in known protocol databases such as PROSPERO (15), or submitted for publication in our journal.

Search methodology

The search methodology should be based on the PICOTS format. Prior to beginning the literature review, the authors should be familiar with the search strategy flow-chart created by PRISMA. The flowchart includes four main phases of the search strategy: (I) identification of articles from database searching or additional sources; (II) screening and removal of duplicate articles; (III) application of inclusion and exclusion criteria to assess articles for eligibility; and (IV) deciding on a final set of articles to be included for qualitative and quantitative synthesis.

To ensure a complete and exhaustive literature search, multiple electronic databases (16) should be searched using pre-determined key terms and medical subject headings (MeSH) headings as defined by PICOTS. Standard databases used for systematic reviews and meta-analyses include MEDLINE, PubMed (Central), Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), American College of Physicians (ACP) database, Database of Abstracts of Reviews of Effects (DARE) and EMBASE. Some reviews may also seek to use additional databases such as Google Scholar and Web of Science. According to the original eligibility criteria, the author should decide whether it is appropriate to limit the search by time, and the level of evidence the studies may be restricted to.

The search strategy should be sensitive, specific and systematic to ensure repeatability by readers. The search strategy should employ keywords and MeSH terms, keywords developed by the National Library of Medicine. These terms should be combined with Boolean operators such as AND, OR and NOT. Ideally, the exact search strategy should be published in the final manuscript to ensure transparency and reproducibility.

After identifying relevant studies, the reference lists of these studies should also be searched to identify further appropriate studies. Hand searching of hardcopies of journals for the most recent six months is also recommended, particularly since some articles may not have had an electronic version stored online until after print or press (17). The overall search should be performed by at least two reviewers, each one applying the inclusion and exclusion criteria to decide on which articles to include. Any discrepancies should be discussed and resolved by consensus, to come to a final set(s) of articles for inclusion in the systematic review or meta-analysis. It is important to summarize the final search in the manuscript in the form of a PRISMA flow-chart.

Quality appraisal of included studies

The included studies for evidence synthesis should be assessed for risk of bias. There are a multitude of tools and checklists available to assess intra-study risk of bias. For systematic reviews and meta-analyses which are restricted to only randomized controlled studies, Cochrane Collaboration's RevMan software has an inbuilt tool for assessing risk of bias (18). The main domains of assessment include "random sequence generation", "allocation concealment", "blinding of participants and personnel", "blinding of outcome assessment", "incomplete outcome data", "selective reporting" and "other bias". Each domain can be rated as yes, no, or unclear bias present.

Other checklists are also available for use by reviewers, although the majority will have similar checklist items as contained in the consort statement: recommendations for improving the quality of reports of parallel-group randomized trials (19,20). Some common checklists used in systematic reviews and meta-analyses in the spine field include the Furlan Checklist (21,22) for randomized studies, and the Cowley checklist for non-randomized studies (23). The items on the checklist are scored with "yes", "no", or "unsure", and a total score is calculated for each study. Typically, a Furlan score of 6 or more out of a possible 12,

or a Cowley score of 9 or more out of a possible 17 reflects "high methodological quality". Similar to the systematic literature search strategy, the studies should be scored independently by two reviewers, and any discrepancies resolved by discussion and consensus.

Extraction of study data

Following identification of relevant studies and quality appraisal, a data extraction form is created. This is often an electronic form, which allows simultaneous data extraction and entry. The data form commonly includes study characteristics (authors, study year, study period, institution, country, number of patients, follow-up duration, and intervention details), baseline demographic characteristics, operation parameters (operation duration, specific intervention or surgery details, spinal levels etc.), perioperative outcomes and complications, follow-up outcomes and complications, functional outcomes (e.g., VAS and ODI scores), and radiographic outcomes if appropriate.

Statistical analysis

The appropriate statistical methods will be determined according to the type of data available and the overall aim of the investigation. For a systematic review without meta-analysis, it is appropriate to report a descriptive summary of the available data, often in the form of mean and standard deviation or range, in a tabular format. Tables should also include information such as study details, surgery or therapy specifics and outcomes.

In systematic reviews of single-arm studies, a meta-analysis of weighted proportions may be conducted, using fixed- or random-effects models. This was the approached used to synthesize nine studies to determine a pooled 36.2% prevalence of bacteria in patients with symptomatic degenerative disc disease (24). For meta-analyses of comparative studies, the summary statistics are commonly presented in the form of a forest plot. A forest plot is a graphical representation of effect sizes from all studies, and finally a pooled effect size once all evidence is synthesized (25). In the example forest plot (4) in *Figure 1*, the middle square represents the effect size of each individual study, and the horizontal line represents the 95% CI. Towards the bottom of the figure, an overall black diamond represents the pooled summary effect size. The midpoint of this diamond represents the effect size while the width of the diamond represents the 95% CI of
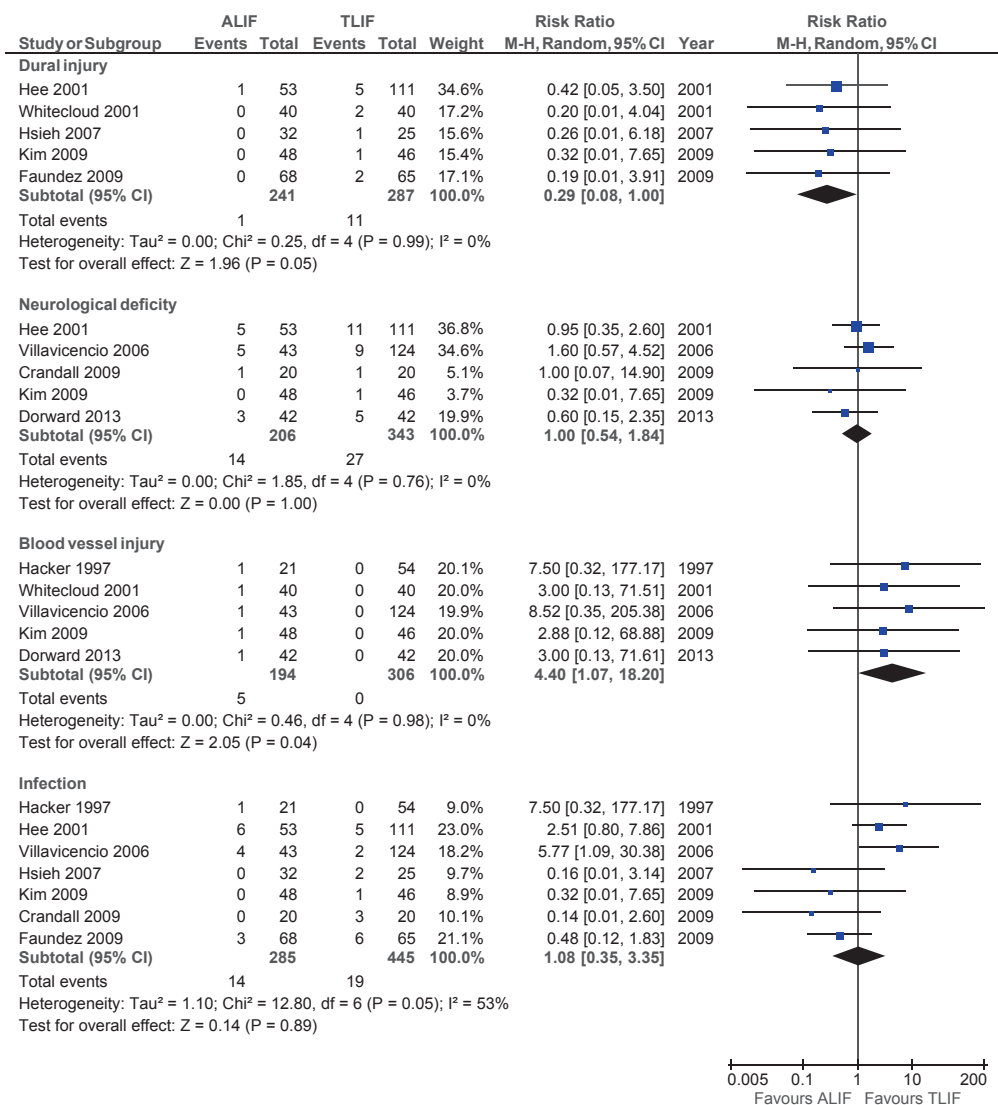
Figure 1 Example of forest plot, comparing anterior lumbar interbody fusion (ALIF) and transforaminal lumbar interbody fusion (TLIF) in terms of dural injury, neurological deficit, blood vessel injury, and infection rates. The blue square represents the effect size of the adjacent study, whilst the horizontal line represents the 95% confidence interval (CI). These effect sizes were pooled to produce an overall effect size, represented by the black diamond in the lowest row. When the black diamond is completely on either side of the midline, this indicates a statistically significant difference (e.g., dural injury pooled effect size). The width of the diamond represents the 95% CI of the pooled data. Heterogeneity analysis is determined as per I^2 test, where heterogeneity $I^2 > 50\%$ is considered statistically significant.

the pooled data. For dichotomous variables, typical effect sizes used include RR and odds ratio (OR). For continuous variables, WMD is commonly used.

In addition to the above analyses, heterogeneity should also be quantified (26). The Cochran Q test provides a yes *vs*. no outcome for whether there is significant heterogeneity amongst the reported effect sizes (27). 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. I^2 value greater than 50% is commonly used as a cut-off for significant heterogeneity in the reported effect sizes. If significant heterogeneity is detected, the source of heterogeneity should be explained. Subgroup analysis (28,29) and meta-regression analysis (30) are further statistical techniques which may be used to
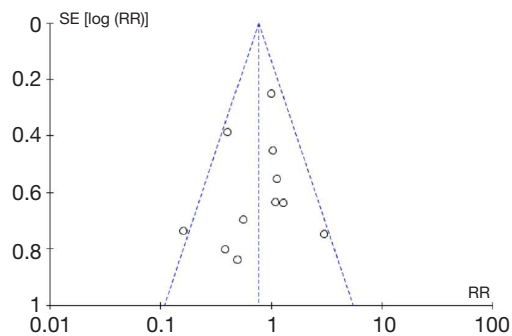
Figure 2 Example of funnel plot of total complications, in a meta-analysis comparing minimally invasive transforaminal lumbar interbody fusion (MI-TLIF) versus open-TLIF (O-TLIF). No significant asymmetry or publication bias was detected using Begg's and Egger's tests. RR, relative risk.

assess the source of heterogeneity.

As mentioned above, the summary effect can be estimated using a fixed-effect model or random-effects model. The fixed-effect model makes the assumption that the true effect size is identical across the studies, but the reason for variation seen in effect size is attributed to sampling error (31). As such, the analysis weights smaller studies less since a more reliable estimate of the true effect size can be derived from large sample size studies. In contrast, a random-effects analysis aims to estimate the mean of a distribution of effect sizes. In essence, this approach does not discount smaller sample size studies, nor does it overweigh large size studies (31). If a review is at risk of significant statistical heterogeneity, as calculated using Cochrane Q or I^2 tests, the random-effects model is often chosen to minimize the impact of outlier studies (32). A systematic review may choose to present results of both fixed-effect and random-effects model analyses.

Interpretation of results

There are several factors to consider when interpreting the results from a systematic review and meta-analysis.

Firstly, the review results may be significantly influenced by publication bias. In essence, there is a tendency for positive results to be published in the literature, whilst negative or inconclusive results are less likely to be published (33,34). As such, the results of a meta-analysis of studies identified from the literature may be misleading since it is "missing" the unpublished negative results and data. In order to assess the influence of publication bias

on the meta-analysis results, a funnel plot analysis can be performed, as demonstrated in example (5) in *Figure 2*. This graph plots treatment effect on the horizontal axis and standard error on the vertical axis. In an ideal pooled analysis with no publication bias, the points are equally and symmetrically distributed around the mean effect size. If the points are distributed asymmetrically around the mean effect size, then this suggests that there may be publication bias. To demonstrate whether the publication bias is statistically significant, Begg's and Egger's tests can be performed (35). Trim-and-fill analysis is an additional test for the funnel plot analysis (36), which determines the number of "missing studies" due to publication bias, and whether the effect size, after adjusting for these missing studies, would be significantly different.

Secondly, the quality of evidence also needs to be considered. The Cochrane Collaboration has recommended the use of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (37) for rating the evidence for a particular outcome. The quality of evidence is downgraded based on (I) limitations of design; (II) inconsistency; (III) indirectness; (IV) imprecision of results; and (V) publication bias. Evidence is upgraded based on (I) significant effect size of at least 2-fold reduction or increased risk; (II) another upgrade if 5-fold difference, and (III) upgrade for dose-response gradient shown. The GRADE approach may allow the reviewer to have increased or decreased confidence in the effect sizes presented.

Thirdly, the clinical relevance of any differences between interventions detected during meta-analysis must also be considered. For example, is a difference in VAS leg pain score of 10% considered clinically significant? Although there are no universally accepted guidelines on precisely what differences are considered clinical significant, some prior reviews have used a threshold of 30% improvement in pain scores and in function outcomes from baseline as clinically important (38).

There are several potential caveats with performing systematic reviews and meta-analyses. Firstly, the quality of the systematic review is dictated by the level of evidence of primary research studies included (39). Particularly in the realm of spine surgery, which does not lend itself to high level of evidence studies, the levels of conclusions from reviews cannot exceed the level of studies reviewed. Secondly, a poorly performed systematic review with missing studies, an ill-constructed search strategy and an unsuitable clinical question may also lead to bias in results presented and misleading conclusions (39). Other common

Item	Content
Table 1 Systematic review presentation checklist, adapted from PRISMA checklist	
Title	• Should identify as systematic review and/or meta-analysis
Introduction	• Clearly states the objective and rationale for the review, according to PICOTS layout • Clear clinical question to address stated
Methods	• Defines important key terms e.g., "*minimally invasive*" • Primary and secondary outcomes described • States which electronic databases were used • States year range for literature search • States/summarizes the search strategy used, based on PICOTS • States inclusion and exclusion criteria • States how to deal with abstracts, conference reports, editorials, duplicates studies etc. • States the reviewers (should be 2 or more) performing the literature search • Statistical methodology described • Publication bias assessment and heterogeneity analysis described • Defining "clinical relevance" of outcome, e.g., 30% improvement in pain score as clinically relevant
Results	• Provides information on PRISMA search strategy workflow e.g. number of studies screened, and numbers of studies finally included • Study characteristics presented (usually in tabular form), including year, study enrolment, level of evidence, design, country, demographics, operative parameters, complications, functional outcomes and radiographic outcomes • Intra-study risk of bias assessment • Inter-study risk of bias assessment • Quality appraisal of included studies • Summary table for outcomes of individual studies • Descriptions of meta-analysis and heterogeneity analysis results in words • Additional analysis should also be described
Discussion	• Main findings summarized • Key limitations and strengths discussed • Overall general interpretation and future directions of research discussed • Conclusions should be appropriate according to the data presented
Funding	• Funding, acknowledgements, and conflicts of interest mentioned
Figures	• PRISMA flow-chart for search strategy • Forest plots of primary and secondary outcomes • Funnel plot for publication bias • Trim-and-fill analysis plots • Subgroup analysis forest plots/meta-regression graphs for heterogeneity assessment if appropriate
Supplementary Tables	• Search strategy for at least one database • GRADE approach for assessment of outcomes • Checklist for quality appraisal, e.g., Furlan, Cowley, Newcastle-Ottawa assessment scale • PRISMA checklist, abbreviation not previously touched on checklist

pitfalls of systematic reviews include non-exclusion of duplicate study populations, failure to recognize and report bias from included studies, and making claims in conclusions that are beyond the facts and results presented in the review (40,41). Systematic review and meta-analysis guidelines should be followed strictly in order to ensure a critical evaluation and synthesis of all available evidence. An adapted checklist with items important for a well-performed systematic review is presented in *Table 1*.

Conclusions

In summary, systematic reviews and meta-analyses represents effective methods of synthesizing all relevant evidence to

26

Phan and Mobbs. Systematic review and meta-analysis guidelines

address a well-constructed clinical question. These types of studies are particularly important especially in areas such as spine surgery, where there is a lack of high-quality Class I evidence. We present a summary of the critical steps in performing a systematic review and meta-analysis, allowing the surgeon scientist to better interpret and perform their own systematic reviews and meta-analyses.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

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. Phan K, Thayaparan GK, Mobbs RJ. Anterior lumbar interbody fusion versus transforaminal lumbar interbody fusion - systematic review and meta-analysis. Br J Neurosurg 2015;29:705-11.

5. Phan K, Rao PJ, Kam AC, et al. Minimally invasive versus open transforaminal lumbar interbody fusion for treatment of degenerative lumbar disease: systematic review and meta-analysis. Eur Spine J 2015;24:1017-30.

6. Phan K, Rao PJ, Scherman DB, et al. Lateral lumbar interbody fusion for sagittal balance correction and spinal deformity. J Clin Neurosci 2015;22:1714-21.

7. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med 2008;358:794-810.

8. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997;126:376-80.

9. Petticrew M. Why certain systematic reviews reach uncertain conclusions. BMJ 2003;326:756-8.

10. 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.

11. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. BMJ 1998;316:140-4.

12. Garg AX, Hackam D, Tonelli M. Systematic review and meta-analysis: when one study is just not enough. Clin J Am Soc Nephrol 2008;3:253-60.

13. 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.

14. Farrugia P, Petrisor BA, Farrokhyar F, et al. Practical tips for surgical research: Research questions, hypotheses and objectives. Can J Surg 2010;53:278-81.

15. Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012;1:2.

16. Suarez-Almazor ME, Belseck E, Homik J, et al. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. Control Clin Trials 2000;21:476-87.

17. Hopewell S, Clarke M, Lefebvre C, et al. Handsearching versus electronic searching to identify reports of randomized trials. Cochrane Database Syst Rev 2007;(2):MR000001.

18. 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.

19. 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.

20. 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.

21. Furlan AD, Pennick V, Bombardier C, et al. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine (Phila Pa 1976) 2009;34:1929-41.

22. van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine (Phila Pa 1976) 2003;28:1290-9.

23. Cowley DE. Prostheses for primary total hip replacement. A critical appraisal of the literature. Int J Technol Assess Health Care 1995;11:770-8.

24. Ganko R, Rao PJ, Phan K, et al. Can bacterial infection by low virulent organisms be a plausible cause for symptomatic disc degeneration? A systematic review. Spine

(Phila Pa 1976) 2015;40:E587-92.

25. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001;322:1479-80.

26. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351-5.

27. 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.

28. Wright CC, Sim J. Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. J Clin Epidemiol 2003;56:833-42.

29. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.

30. 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.

31. Schmidt FL, Oh IS, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. Br J Math Stat Psychol 2009;62:97-128

32. Phan K, Tian DH, Cao C, et al. Systematic review and meta-analysis: techniques and a guide for the academic surgeon. Ann Cardiothorac Surg 2015;4:112-22

33. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. Lancet 1991;337:867-72.

34. Chalmers I. Underreporting research is scientific misconduct. JAMA 1990;263:1405-8.

35. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.

36. 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.

37. 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.

38. Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976) 2008;33:90-4.

39. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896-900.

40. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. Lancet 1998;351:123-7.

41. Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? BMJ 1997;315:1610-4.