

Meta-analysis: Everything you wanted to know but were afraid to ask

Dr R Rodseth MBChB, FCA(Anaes), Cert Crit Care, MMed(Anaes), MSc(HRM), PhD

Department of Anaesthetics, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Dr LC Marais MBChB, FC(Orth)(SA), MMed(Orth), CIME, PhD

Department of Orthopaedics, University of KwaZulu-Natal, Pietermaritzburg, South Africa

Corresponding author:

Dr LC Marais

Department of Orthopaedic Surgery

Grey's Hospital

Private Bag X9001

3201 Pietermaritzburg

Email: maraisl@ukzn.ac.za

Tel: +27 033 897 3424

Fax: +27 33 897 3409

Abstract

The perceived ability of systematic reviews and meta-analyses to concisely and definitively summarise existing literature regarding a specific medical issue has made them the most highly read and cited form of academic literature. However, the evidence they provide is only as good as the data evaluated and the methodology followed when doing them. In order to conduct a rigorous systematic review, all the elements of the research question should be clearly stated, all relevant studies should be identified, the inclusion criteria should be appropriate and the included studies valid, heterogeneity should be identified, and finally the results of the meta-analyses should be appropriately presented. This paper aims to provide an overview of systematic reviews and meta-analyses. Furthermore, we examine the rationale for conducting such a study, overview the methodology, and highlight pitfalls and weaknesses inherent in the process.

Key words: systematic review, meta-analysis, review

<http://dx.doi.org/10.17159/2309-8309/2016/v15n4a4>

Introduction

Systematic reviews and meta-analyses have exploded into prominence in the medical literature. While traditional narrative reviews remain informative, they are prone to bias as the author is free to select only those articles they believe are important. Systematic reviews are specifically aimed at reducing bias by identifying, appraising and synthesising all relevant studies. A systematic review becomes the basis for the next step where all data is synthesised into a single estimate or summary of effect. This process is called a meta-analysis. The perceived ability of systematic reviews and meta-analyses to concisely and definitively summarise existing literature regarding a specific medical question or issue has made them the most highly read and cited form of academic literature.¹ However, systematic review and meta-analyses are not without their critics.^{2,3}

This paper will provide an overview of systematic review and meta-analysis and will: 1) examine the rationale for conducting such a study, 2) overview the methodology, and 3) highlight pitfalls and weaknesses inherent in the process.

Why conduct a systematic review and meta-analysis?

We have all experienced the medical equivalent of the Mexican standoff. One consultant quotes a study from a prominent orthopaedic journal, showing that bipolar hip arthroplasty is superior to unipolar hemiarthroplasty in femoral neck fractures, only to be rebutted by another consultant who cites more recent research from another journal showing that there is no difference in outcome. In the aftermath the trainees, patients and even the consultants themselves remain unsure about which implant they should use and even more confused about how to make clinical decisions when faced with disparate trial results. The field of systematic review and meta-analysis had been developed as a tool to deal with this problem. This research tool is also invaluable in understanding the current state of the literature with regard to a specific question. A good systematic review is able to determine whether the literature has already answered the question you are asking, or whether more research is needed. For this reason,

conducting a meta-analysis before embarking on a big research trial is essential.

Clinical research generally aims to identify the best treatment for a specific group of patients and the randomised controlled trial is the tool most commonly used to do this. Two groups of patients randomly receive two different treatment options (e.g. unipolar vs bipolar hip arthroplasty). The researcher then compares outcomes in the two groups to determine which option is best. The basic assumption in all randomised controlled trials is that the two patient groups are identical to each other and that all the outcome differences are due only to the treatment being compared. But this is clearly not always the case. Clinical trials occur in the real world with real people who are not identical. One group may have more smokers, or exercise less, or may have a higher incidence of chronic diseases – all factors that will affect the trial outcome. Small trials, which are cheaper and easier to do, are very vulnerable to imbalances between study groups and can give wildly disparate results driven solely by imbalance between patient groups. Increasing the number of patients in a trial decreases the impact that individual patients have on the outcome and so large trials provide more robust and reliable findings.

Meta-analysis tries to pool all available trials, irrespective of size or date of publication, into one big patient pool. This dramatically increases the patient population of this combined 'trial' and hopefully determines the 'true' effect of the intervention. At its heart meta-analysis is an attempt to increase the statistical power of clinical trials studying an intervention.

How it's done

As the name suggests there are two parts to a systematic review and meta-analysis. At the outset it is important to realise that both of these parts have become highly standardised. Where in the past you could do your literature review in any way that seemed good to you, now a highly structured and transparent process must be followed. Similarly, the processes by which studies are identified and evaluated, and data extracted and finally pooled has become very rigid. These rules, encompassed in the PRISMA guidelines, have been developed to ensure that published meta-analyses, much like clinical trials, conform to best practice and are of a high standard.⁴ The PRISMA guidelines provide a checklist that can be used by both authors and journals to ensure that a meta-analysis meets the required level of rigour. More and more journals are insisting that authors comply with these guidelines.

So how do you begin?

The clinical question

The first step is to begin with a clinical question that you want answered. For example – Should we be using unipolar or bipolar implants in femur neck fractures? This question should then be carefully honed using the PICOT framework (Table 1).⁵

Table 1: The PICOT framework for the development of the research question for systematic literature review

	Factors	Example
P -	Population in question	Femoral neck fractures
I -	Intervention of interest	Bipolar hemiarthroplasty
C -	Comparator	Unipolar hemiarthroplasty
O -	Outcome measure	Functional outcome
T -	Timeframe	Short-term (at 2 years)

Typically a reworked question will read something like this: In patients ≥ 55 years of age requiring hip replacement due to a displaced femoral neck fracture, does a cemented unipolar hip replacement provide better functional outcomes at 2 years after replacement, as measured by activities of daily living, as compared to a cemented bipolar hip replacement? This question clearly defines the study population, the intervention and comparison arms, as well as the time frame at which outcomes will be assessed. Rather than focusing only on one outcome as is done in a clinical trial, a meta-analysis should try to identify all clinically important outcomes that have been reported in individual studies. This includes common events such as readmissions and infections, but also rare events such as cardiovascular collapse, anaphylaxis and death. A second important aspect, unique to meta-analysis, is to define the type of research that will be analysed. Will you only be looking at randomised controlled trials, or are you going to be looking at observational studies? Your question should now read, 'In randomised controlled trials of patients ≥ 55 years of age requiring ...'

Conduct a quick literature search to find major clinical trials dealing with this topic and then review the question to make sure it is correctly focused. Also take this time to search for meta-analyses that may already have been conducted on the topic. Finding a recent meta-analysis on the topic does not automatically mean that you need to terminate yours. The existing analysis might have a different focus, have used an inferior methodology, or be outdated.

The protocol

At this point a formal protocol, using the structure provided by the PRISMA guidelines, should be written to guide the meta-analysis. The protocol has the same function as that of a protocol written for a clinical trial. It ensures that investigators adhere to a predetermined process and ensures transparency. This step is critically important. A succinct well-structured protocol with a carefully considered description of the study methodology is invaluable as a guide during a long and complex review and analysis. Failure to write a protocol will detract from the rigour of the study and may also leave the researcher confused 6 months down the line when the primary aims

and methods of the study become blurred. Once written it is strongly recommended that the review be registered with PROSPERO (<http://www.crd.york.ac.uk/prospero/search.asp>) and the protocol published. This again ensures transparency.

The literature search

The foundation of a high quality systematic review is the literature search. As already mentioned this search must be structured and logical, and each step must be recorded so that the process can be duplicated. Work in conjunction with someone experienced in searching medical databases or with your librarian. Take key words from your research question and then combine these to form search terms but take care to cast your net as wide as possible. Using search terms that are too restrictive at the start of the search runs the risk of excluding eligible trials. Using our example – the key words could be hip replacement / hip arthroplasty and randomised controlled trial. It is not sufficient to only search a single database such as Pubmed or Medline. Pubmed only indexes a fraction of the world's medical research and it is essential to find all trials that have examined this topic. At a minimum databases such as EMBASE or SCOPUS should be included, but the Cochrane Database, Web of Science and ProQuest should also be considered. An attempt should then be made to search the grey literature. There are a significant number of trials that never make it to formal publication in the medical literature and many can be identified by searching the abstracts of conference proceedings.

It is important not to include any outcome terms in your initial search as this will exclude trials that might have recorded the outcome of interest but did not report it as their main finding. Your search terms must be recorded and an example reported in the study findings.

Screening for trial eligibility

A wide search of this nature will often identify thousands of abstracts. These must now be screened in duplicate by two or more investigators. First round screening is done by reading only the study title and the abstract. Abstracts that do not address the research question are then excluded. In our example this would exclude all letters to the editor, review articles, trials conducted in animals, trials conducted in patients < 55 years of age, traumatic hip replacements, and all observational studies. All abstracts thought to be eligible by any investigator after first round screening are then included in the second round.

Second round screening is again done in duplicate but this time using the full paper rather than the abstract only. Papers are only excluded once both investigators agree that they are not eligible. At the end of this sifting process you should be left with all available research pertinent to your research question. During this whole process careful record must be kept of the number of excluded abstracts

and trials as well as the reasons for their exclusion. Transparency and repeatability are key components of a well performed meta-analysis.⁶ A flow diagram, illustrating the study selection process, must be included in the report (*Figure 1*).

Data extraction

From these trials pertinent data must now be extracted. There are three parts to this process. The first is extracting demographic data. What did each study population look like in terms of age, gender and co-morbidities? Also record the number of cases in each treatment arm. The second phase is to determine the quality of the trials. Each trial contributing data to your meta-analysis must be critically evaluated to determine how well the trial was conducted and what its risk of bias was (*Table II*).⁷

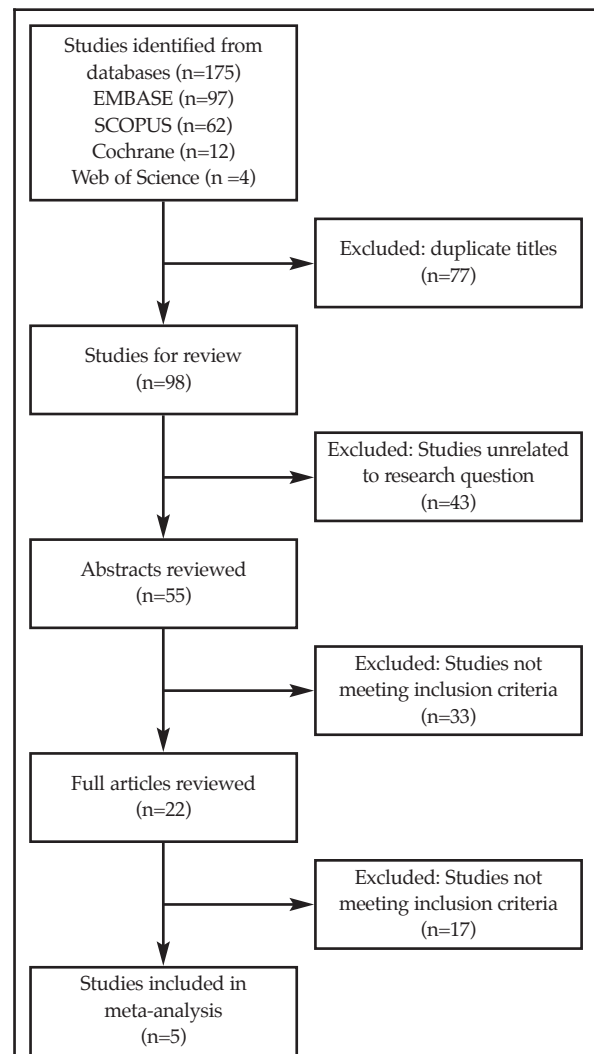


Figure 1. An example of a trial inclusion flow diagram, which illustrates the study selection process followed during the systematic review

Table II: Examples of bias that need to be considered during the systematic review, as described by the Cochrane Collaboration's tool for assessing risk of bias in randomised trials⁸

Type of bias	Source of bias
Selection	Enrolment of patients into different cohorts inherently different
Performance	Improper blinding of participants or personnel in terms of intervention
Detection	Improper blinding of outcome assessment in relation to the intervention
Attrition	Incomplete outcome data due to exclusions or loss to follow-up
Reporting	Selective reporting of results
Other	Study design bias, chronological bias, recall bias, citation bias

Was the trial randomised, were the participants blinded, were there excessive dropouts and were the results reported consistently. The Cochrane data base and the review manager (RevMan) provide a risk of bias tool that allows these aspects to be recorded and visually presented.^{8,9} As described by the Cochrane Collaboration, the emphasis should be on the assessment of internal validity of the trial results. This should then be recorded as low, unclear or high risk of bias for each trial. Finally, the outcome data you are interested in must be extracted for analysis purposes.

Analysis

The primary outcome for this research question is function at 2 years, as measured by activities of daily living. But what about other important outcome measures such as quality adjusted life years, or pain scores. One of the strengths of the systematic review is that because of the wide net it casts many adverse outcomes, not often identified in single trials or in narrative reviews, can be captured. Examples of adverse outcomes that may be considered are need for re-operation, infection rates or mortality rate.

The unit, or type of data, that is used to compare the outcome of interest is important to understand. Outcomes can be expressed as dichotomous data – such as mortality or number of infections. This would then be analysed as the number of deaths or infections in the one arm as compared to the number of deaths in the other. The risk comparison is then reported either as an odds ratio or a risk ratio. Outcomes can also be presented as continuous data. In this analysis the outcome is activities of daily living and is expressed with a score ranging from 4 (totally independent) to 18 (totally dependent). We now need to compare the average score, across all eligible studies, between patients who have received cemented unipolar as compared to cemented bipolar hip replacements. This is done by comparing the mean difference between the two arms. If different instruments have been used to measure the same type of patients then these can be compared using a standardised mean difference, but this can be difficult to interpret.

The actual pooling and analysis of the data is done using a statistical software package such as RevMan (Cochrane Collaboration), Comprehensive Meta-Analysis (NIH), Stata or R. Data is not just simply averaged but rather each trial is assigned a weighting based on its size and the precision of the study findings (measured by the confidence interval). Two statistical methods can be used: a fixed effects or a random effects method. Fixed effects models assume that the intervention has the same size of effect across all the studies, and results using this model tend to have narrower confidence intervals. Random effects models assume that this effect size may vary between trials and produce results with wider confidence intervals. In most cases though the summed effects differ very little.

Reporting results

Meta-analysis reporting follows a standardised pattern. First, the number of citations screened and the results of this screening process are reported in the text as well as in a flow diagram. Then, for each of the included trials, patient demographics and details of what was done in each trial must be reported – this data should also be reported in a table. After this, details regarding the quality of the trials must be reported in a second table.

The results of a meta-analysis are commonly reported in a forest plot. An example of a forest plot is shown in Figure 2. The forest plot provides the reader with a visual representation of the trials included in the analysis and the results for each of the trials. For each trial the forest plot shows the treatment effect (the mid-point of the square box), the size of the trial (represented by the size of the box) and the trial precision as measured by the confidence intervals (represented by the horizontal lines). In the middle of the forest plot you will find a vertical line that represent a relative risk of one. This point represents no difference between the two arms of the trial. At the bottom of the plot you will find a diamond that represents the summary of the pooled treatment effect. The size of the diamond represents the total size of the pooled study population, the mid-point of the diamond is the point estimate of the treatment effect, and the horizontal corners of the diamond represent the width of the confidence intervals. Pooling many precise large trials will result in a large long thin diamond (precise estimate of effect) while a meta-analysis that contains few small and imprecise trials will result in a small wide diamond (imprecise estimate of effect).

But how reliable are these results? Can they be trusted? There are two tools that are used to answer this question – heterogeneity tests and funnel plots.

Tests of heterogeneity include the the χ^2 and the I^2 tests. Both these tests ask the question whether the difference seen between the individual trial results are due to chance or if the difference is due to a true effect. In low heterogeneity meta-analyses there is not much difference in trial results between the different studies. In high heterogeneity

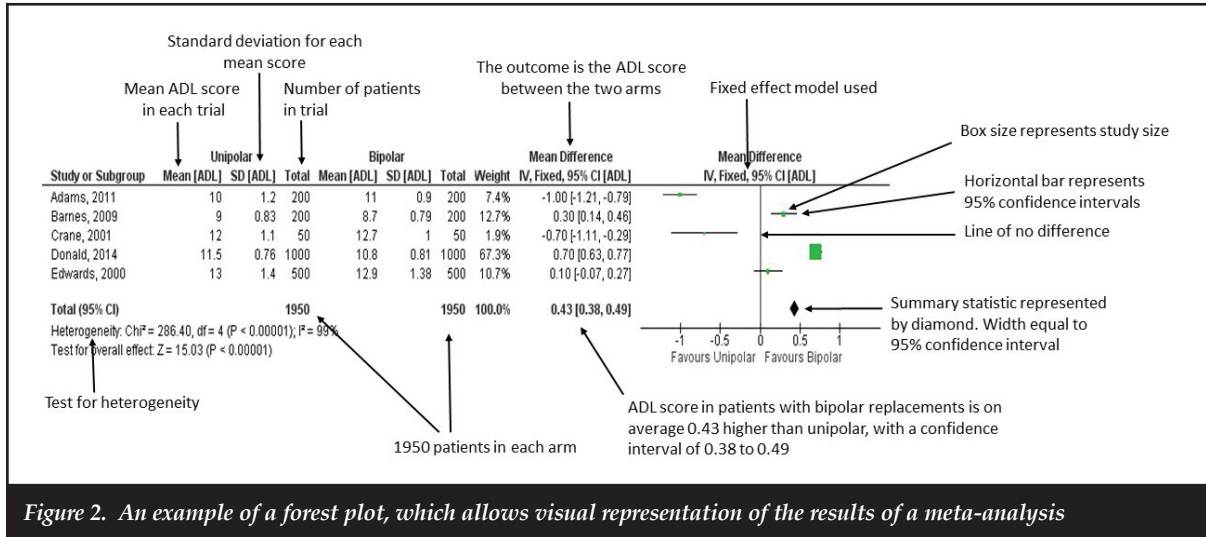


Figure 2. An example of a forest plot, which allows visual representation of the results of a meta-analysis

meta-analyses there is a large difference and this is cause for concern. In this example (Figure 1) both tests show that there is high heterogeneity (individual trial results are very different) with the chi² test showing a probability of p<0.000001 that these result are due to chance only. Thus, the chance that the difference in results occurred by accident is less than 1 out of a million. An I² of 30% to 60% suggests moderate heterogeneity while values above 70% suggest considerable heterogeneity. In this example the I² is 99% – very high heterogeneity.

The second tool commonly used is the **funnel plot** that tries to capture the effectiveness and reliability of the intervention. The plot looks like an inverted funnel with precise trials (small confidence interval) plotted at the top of the funnel and less precise trials (larger confidence interval) plotted at the bottom. The x-axis of the plot represents the treatment effect as determined by the trial (Figure 3). Journals generally tend to publish articles that show a difference and ignore articles that don't. A funnel plot that is asymmetrical

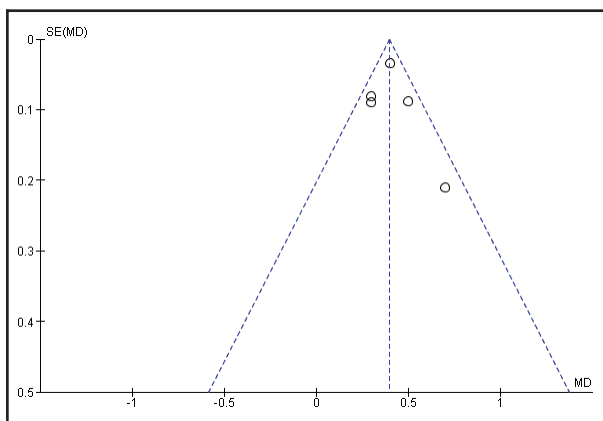


Figure 3. An example of a funnel plot; a scatter plot of the intervention effect (on the x-axis) against some measure of the study's size or precision (on the y-axis)¹⁰

may be an indicator that the meta-analysis is at risk of publication bias. The more pronounced the asymmetry, with gaps in the bottom right-hand corner of the plot, the more likely that the meta-analysis will be over-estimating the effect of the intervention.¹⁰

Weaknesses and pitfalls

Oxford University's Centre for Evidence-Based Medicine (CEBM) provides useful guidelines for the critical appraisal of systematic reviews.¹¹ To determine if the results of a meta-analysis are valid, six aspects have to be assessed (Table III). Bhandari *et al.* found that, prior to the year 2000, 88% of meta-analyses in orthopaedics had methodological flaws that could limit their validity.¹² In order to maximise the potential value of a meta-analysis several factors need to be considered.

The first point to be aware of is that a meta-analysis is only as good as the underlying trials that have been conducted in the field. Poorly designed trials will result in unreliable results. It is therefore important to evaluate the underlying methodology and bias in the individual trials. In this same vein it is important to keep in mind that a meta-analysis can only reflect the research that has been done in the field. This means that if insufficient trials have been conducted the meta-analysis, instead of providing a definitive result, will demonstrate that more needs to be done in the field.

The second point to keep in mind is that a systematic review and meta-analysis is a tool for analysing data and deriving summary results. And in the same way that tools can be abused so too meta-analysis can be abused. Common mistakes are searching only in English, using incorrect or incomplete search terms, searching in only one database and choosing outcomes that are not clinically important. When it comes to synthesising the data and doing a meta-analysis some judgment is called for.

Table III: The aspects that need to be assessed during the critical appraisal of a meta-analysis, according to the Centre for Evidence-Based Medicine (CEBM)¹¹

Questions to ask in order to determine if the results of a systematic review are valid

1.	Has the main research question (PICOT framework) been clearly stated?
2.	Is it unlikely that important relevant studies were missed?
3.	Were the inclusion and exclusion criteria appropriate?
4.	Were the included studies valid in terms of the research question being asked?
5.	Was the heterogeneity of studies identified and adequately explored?
6.	Were the results adequately presented?

If there are fewer than three trials and the outcomes used in the trials are very different, does it make clinical sense to merge them into a single analysis? There are undoubtedly cases where it is wiser not to conduct a meta-analysis and to leave it at a systematic review.

Finally, keep in mind that in many cases a single large trial is the primary contributor to the signal seen in the meta-analysis result. In cases like this it may be better to evaluate the dominant trial rather than trying to do a meta-analysis. Abdullah *et al.* recently illustrated that 28% of orthopaedic randomised controlled trials (RCT) reporting a negative finding were underpowered.¹³ As stated by these authors, if an RCT is statistically underpowered to show a clinically relevant absence of a difference in outcome between the two groups, it would be inappropriate to declare the two procedures equivalent.

Conclusion

While systematic reviews and meta-analyses are traditionally considered higher level evidence, the evidence they provide is only as good as the data it evaluates and the methodology that was followed. In order to satisfy the criteria for a well performed systematic review all the elements of the research question should be clearly stated; all relevant studies should be identified; the inclusion criteria should be appropriate; included studies must be valid; heterogeneity should be identified; and finally, the results of the meta-analyses should be appropriately presented.

Compliance with ethics guidelines

• Conflict of interest statement

The authors declare that they have no conflict of interests and that no financial support was received for this study.

• Authors' contributions

All authors made contributions toward the conception and design of the research, acquisition of data and drafting of the manuscript. The final manuscript was read and approved by all the authors.

• Funding

The study was funded by departmental resources.

• Disclosure

Dr RR is supported by an incentive grant from the South African National Research Foundation.

References

1. Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA* 2005;**293**:2362-66.
2. Meta-analysis under scrutiny [editorial] *Lancet* 1997;**350**:675.
3. Bailar JC. The promise and problems of meta-analysis [editorial] *N Engl J Med* 1997;**337**:559-61.
4. Moher D, Shamseer L, Clarke M, *et al.* Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 2015;**4**:1.
5. Thabane L, Thomas T, Ye C, Paul J. Posing the research question: not so simple. *Can J Anaesth* 2009;**56**:71-79.
6. Lefavre KA, Slobogean GP. Understanding Systematic Reviews and Meta-analyses in Orthopaedics. *J Am Acad Orthop Surg* 2013;**21**:245-55.
7. Pannucci CJ, Wilkins EG. Identifying and Avoiding Bias in Research. *Plast Reconstr Surg* 2010;**126**(2):619-25.
8. Higgins JPT, Altman DG, Gotzsche PC, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928 doi:1136/bmj.d5928.
9. Available from: <http://tech.cochrane.org/revman/download>. Date last accessed: 27 May 2016.
10. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
11. Available from: <http://www.cebm.net/critical-appraisal/>. Date last accessed: 30 May 2016.
12. Bhandari M, Morrow F, Kulkarni AV, Tornetta P. Meta-analyses on orthopaedic surgery. A systematic review of their methodologies. *J Bone Joint Surg Am* 2001;**83**(1):15-24.
13. Abdullah L, Davis DE, Fabricant PD, *et al.* Is There Truly 'No Significant Difference'? Underpowered Randomized Controlled Trials in the Orthopaedic Literature. *J Bone Joint Surg Am* 2015;**97**(24):2068-73.

This article is also available online on the SAOA website (www.saoa.org.za) and the SciELO website (www.scielo.org.za). Follow the directions on the Contents page of this journal to access it.