**Clinical Orthopaedics
and Related Research®**
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

SURVEY

Network Meta-analysis: Users' Guide for Surgeons: Part I – Credibility

Clary J. Foote MD, Harman Chaudhry MD, Mohit Bhandari MD, PhD, Lehana Thabane PhD,
Toshi A. Furukawa MD, PhD, Brad Petrisor MD, Gordon Guyatt MD, MSc

Abstract Conventional meta-analyses quantify the relative effectiveness of two interventions based on direct (that is, head-to-head) evidence typically derived from randomized controlled trials (RCTs). For many medical conditions, however, multiple treatment options exist and not all have been compared directly. This issue limits the utility of traditional synthetic techniques such as meta-analyses, since these approaches can only pool and compare evidence across interventions that have been compared directly by source studies. Network meta-analyses (NMA) use direct and indirect comparisons to quantify the relative effectiveness of three or more treatment options. Interpreting the methodologic quality and results of NMAs may be challenging, as they use complex methods that may be unfamiliar to surgeons; yet for these surgeons to use these studies in their practices, they need to be able to determine whether they can trust the results of NMAs. The first judgment of trust requires an assessment of the credibility of the NMA methodology; the second judgment of trust requires a determination of certainty in effect sizes and directions. In this Users' Guide for Surgeons, Part I, we show the application of evaluation criteria for determining the credibility of a NMA through an example pertinent to clinical orthopaedics. In the subsequent article (Part II), we help readers evaluate the level of certainty NMAs can provide in terms of treatment effect sizes and directions.

C. J. Foote, H. Chaudhry (✉), M. Bhandari, L. Thabane,
B. Petrisor, G. Guyatt
Division of Orthopaedic Surgery, McMaster University, 293
Wellington Street N, Suite 110, Hamilton, ON L8L 2X2, Canada
e-mail: harman.chaudhry@medportal.ca

T. A. Furukawa
Department of Health Promotion and Human Behavior and
Department of Clinical Epidemiology, Kyoto University
Graduate School of Medicine/School of Public Health, Kyoto,
Japan

Box 1. Clinical Scenario

You are asked to see an active 36-year-old male patient who recently presented to the emergency department with a Gustilo Grade IIIA open mid-shaft tibial fracture. In your practice, you commonly perform unreamed intramedullary nailing for this fracture type. You present the case at morning rounds and several of your colleagues believe that given the amount of tissue destruction, an external fixator is the preferred option. Another colleague cites several studies that report high malunion rates with external fixation and he believes reamed intramedullary nailing leads to better bony stability and biologically enhanced union rates. You perform an uneventful, unreamed intramedullary nailing in the patient, who reports good function at 2-week followup. However, 9 months postoperatively, the patient continues to report residual pain, and radiographs show evidence of nonunion. You consider adjunctive revision procedures and wonder if either reaming or using an external fixator would have been better initial management options. You perform a literature search and come across a recently published network meta-analysis that evaluates outcomes for surgical treatment of open tibial shaft fractures [6]. What approaches can you use to evaluate the credibility of this work?

Network Meta-analysis: Background and Rationale

Systematic reviews use defined strategies for identification, inclusion, appraisal, and reporting of results to summarize

the literature addressing a specific clinical question [19, 21]. Meta-analysis refers to explicit statistical methods to generate a single pooled estimate summarizing the results of the included studies [21]. Traditionally, meta-analyses have evaluated the effectiveness of one intervention compared with another intervention using direct (head-to-head) comparisons; these studies sometimes are referred to as conventional or pairwise meta-analyses, and during the past 20 years have played an increasingly important role in orthopaedic science and practice [1, 3, 27].

The advantages of pairwise meta-analyses of randomized controlled trials (RCTs) compared with individual RCTs include: (1) greater precision in estimating treatment effects; (2) ability to assess variability (heterogeneity) in treatment effects among trials; (3) improved power to conduct subgroup analyses; (4) avoidance of unrepresentative results that arise from single studies owing to either random variation or bias; and (5) better informed determination of areas for future research.

Pairwise meta-analyses, however, are limited to addressing the effects of a single treatment versus a single alternative. For many conditions in orthopaedic surgery, several treatment options exist and the number of head-to-head clinical trials is limited. For instance, a systematic review and meta-analysis addressing management of periprosthetic femur fractures evaluated four treatment strategies: nonoperative treatment, nonlocked plating, locked plating, and retrograde intramedullary nailing [24]. Using pairwise meta-analysis to evaluate the relative effects of the four strategies would require six sets of head-to-head clinical trials. Further, the size of the trials would need to be large enough that, when pooled, they could provide sufficient precision for definitive decision-making [22]. Such robust clinical data are unlikely to be available for most clinical questions.

In response to the need to simultaneously evaluate all available treatments, new methods in meta-analysis—known as network meta-analysis (NMA)—have emerged (Box 1) [5, 14, 16, 17]. These are complex studies that should be conducted only with the support of an expert statistician. Also referred to as multiple-treatment−comparison meta-analyses, NMAs involve creating networks of treatments. Authors then apply statistical methods to these networks to estimate the effects of treatments shown through direct comparisons (head-to-head trials, A versus B) and indirect comparisons (making inferences about A versus C by looking at how 'A' compares with common comparator 'B' and how 'C' compares with common comparator 'B'). Investigators then combine direct and indirect comparisons to provide an overall pooled treatment effect [18].

To illustrate the concept of direct and indirect evidence, consider a hypothetical example of three surgical treatments for a tibial shaft fracture: reamed intramedullary

nailing, unreamed intramedullary nailing, and external fixation. If direct comparisons of reamed intramedullary nailing and external fixation are unavailable, deductions regarding their relative merit may require comparison with a third procedure, unreamed intramedullary nailing (Fig. 1). However, if head-to-head comparisons of all three treatments are available, direct and indirect evidence can be used to inform the comparison between reamed intramedullary nailing and external fixation (Fig. 2).

Indirect evidence, when combined with head-to-head comparisons, may enhance precision by increasing sample size and thus narrowing confidence intervals (CIs) [12]. A NMA also enables ranking available treatments and facilitates exploration of subgroup effects (eg, some
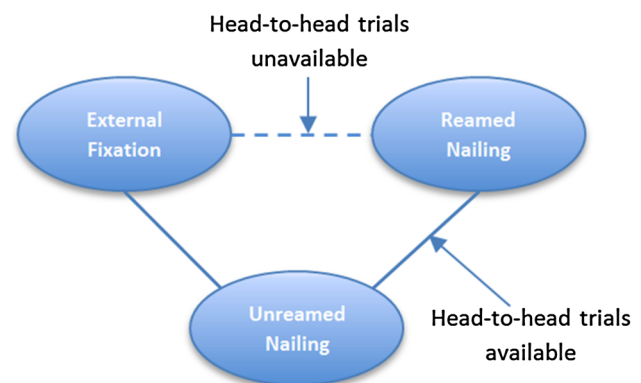

Fig. 1 The diagram shows the concept of an open network. In this hypothetical scenario, head-to-head (ie, direct) comparisons of unreamed nailing with reamed nailing and external fixation are available (as indicated by the solid lines). However, no available trials have compared reamed nailing with external fixation (as indicated by the dashed line). The loop enables estimates of effect between reamed nailing and external fixation to be calculated only indirectly through a common comparator (in this case, unreamed nailing).
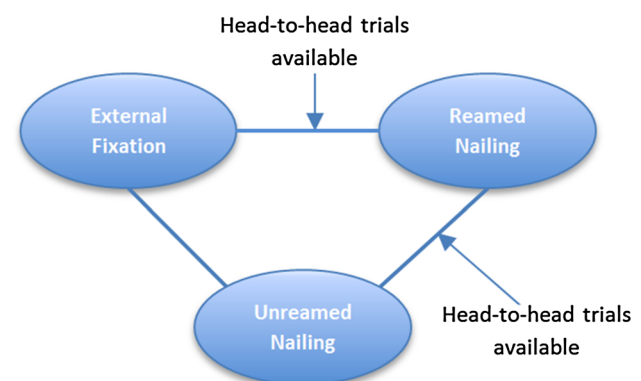

Fig. 2 The diagram shows the concept of a closed loop network. In this hypothetical scenario, head-to-head trials have compared external fixation with reamed and unreamed nailing (as indicated by the solid lines). Therefore, direct and indirect evidence (derived through the loop involving unreamed nailing) can inform the comparison between external fixation and reamed nailing.
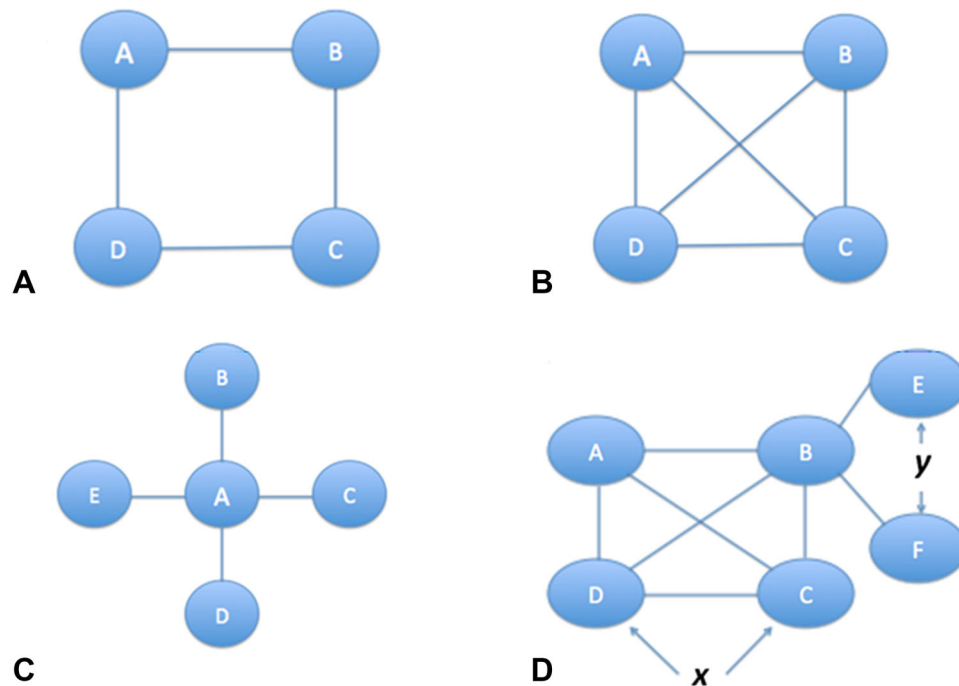
Fig. 3A–D Four network diagrams in increasing order of complexity are shown. The circles (nodes) represent treatments and the lines represent head-to-head trials (direct evidence). (**A**) This complex closed loop network shows that as more interventions are added without increasing linkages, the network indirectness increases. (**B**) This well-connected network shows that if an increase in interventions is complemented with an increase in linkages, then overall indirectness of the network may not increase. (**C**) A network with a single common comparator is called a "star" network. (**D**) Complex networks may have many linkages, although typically most studies have concentrated on only a few treatments comparisons. Parts of the network may be well-connected (as indicated by *x*), while other parts may be poorly connected (as indicated by *y*).

treatments may work better with less severe fractures and others with more severe fractures) through a process known as meta-regression [10].

The credibility of a NMA depends on consistency of results between direct and indirect comparisons. Most of the time direct and indirect comparisons yield similar results [26]. Dealing with inconsistency is an issue to which we will return.

The nature of the available direct and indirect comparisons will determine the architecture of the network of the NMA, typically displayed as a network diagram. The network diagram, which depicts available direct treatment comparisons, may, through the size, thickness, or color of the circles (called nodes) or connecting lines (called edges or links) signify the number of trials or participants. Depending on the number and arrangement of nodes and links, networks may be described as simple open loops (Fig. 1), simple closed loops (Fig. 2), or complex networks (Fig. 3).

Can You Trust the Results of a NMA?

Two fundamental problems can undermine trust in the results of a NMA: the search and screening process and the quality of the source material that the search identifies. An

Table 1. Guide for assessing credibility of the systematic review process

Did the review explicitly address a sensible clinical question?
Was the search for relevant studies exhaustive?
Are selection and assessments of studies reproducible?
Did the review present results that are ready for clinical application?
Did the review address confidence in effect estimates?

insufficiently rigorous search and screening process decreases the credibility of the review (that is, the extent to which the design and conduct of the NMA are likely to have protected against misleading results) (Table 1).

A second possible difficulty, even if the NMA is highly credible and has adhered to rigorous methodologic standards, is that there still may be low certainty in the estimates of effect sizes and directions that emerge from the NMA. This is because the underlying evidence (the studies that contribute to the review) may be limited by a high risk of selection, transfer, or assessment bias, imprecision (small sample size and wide CIs), inconsistent results from study to study, and publication bias. Underlying evidence causing decreased certainty in terms of the estimated effects in a NMA is discussed in detail in Part II of the Users' Guide for Surgeons.

Was the Systematic Review Process Credible?

Did the Review Explicitly Address a Sensible Clinical Question?

Formulating a sensible clinical question is the first step in any research endeavor, including systematic reviews and NMAs. Ideally, the question is framed in the Population, Intervention, Control (or Comparators), and Outcomes (PICO) format. If the patient population is too diverse, questions may be too broad to be useful. For instance, a hypothetical NMA that evaluates internal fixation for all hip fractures, regardless of patient age or severity of trauma, might be problematic. A displaced femoral neck fracture sustained by a 20-year-old man involved in a high-speed motor vehicle collision is likely to have a different response to available treatments than a low-energy fragility hip fracture sustained by an elderly woman with dementia [20]. In the latter case, a hemiarthroplasty may lead to good function and fewer complications, whereas the younger man might do better if his femoral neck were to be fixed and his native femoral head preserved. If the effects of available treatments differ in these two patients, pooling different patient group results will provide misleading inferences, perhaps by resulting in an intermediate effect size that applies to neither young nor old patients.

Selection of patients is not the only study characteristic that requires similarity across trials; interventions also must be similar. For example, if techniques have improved with time, combining older studies of a surgical approach to treatment with more recent studies of the same but improved approach may yield misleading results (an intermediate effect representing neither the old nor the new procedure) [2]. The same is true for differences in outcome measurement. For instance, effects may differ with time and including studies with short- and long-term followup again may yield intermediate results representing neither short- nor long-term effects.

Ensuring sufficient similarity in patients, interventions, and outcomes is important in any meta-analysis, but it is particularly important in NMAs. If there are differences in the groups, indirect comparisons can be particularly vulnerable. In other words, if we are making deductions regarding our comparison of interest (for example, A vs C) through a common comparator (B), important differences in patients, interventions, or outcomes between the A vs B trials and B vs C trials may bias the indirect comparison. The specific factors that differ, such as patients, interventions, or outcomes, are referred to as effect modifiers. For example, if treatment effects differ between young and old patients, then age is an effect modifier. When the A vs B and B vs C trials differ excessively in patients, interventions, or outcomes, we label the problem as a violation of transitivity [25].

One way to check whether transitivity has been violated is to ask the question: Could all patients and treatments in this NMA be included as independent arms in a single RCT? If the answer to the question is "no," then transitivity has been violated and the credibility of the NMA is compromised (Box 2).

If we expect treatment effects to be similar across age, technique, timing of outcomes, and other important factors, then broad eligibility criteria will result in more precise estimates of treatment effect sizes and directions by increasing the number of included studies and participants. Broad eligibility criteria also improve the generalizability of results; therefore, NMAs must strive to be as comprehensive as possible without risking serious intransitivity.

Box 2. Example of Transitivity

In the tibial shaft NMA [6], comparator groups were primary external fixation, reamed intramedullary nailing, unreamed intramedullary nailing, Ender nails, plate fixation, and primary Ilizarov fixation. Each of the procedures could reasonably make up one arm of a six-armed hypothetical clinical trial evaluating treatment options in the primary management of open tibial shaft fractures. Further, all trials enrolled predominantly adult patients (age older than 18 years) who had experienced open fractures of the tibial shaft, and all trials reported unplanned reoperation rates. Therefore the comparators are similar and the NMA meets the transitivity assumption.

Was the Search for Relevant Studies Exhaustive?

As for any systematic review, investigators should conduct a comprehensive literature search. Standard search strategies for randomized trials should include electronic medical databases, such as MEDLINE, EMBASE, and the Cochrane Library; conference proceedings and abstracts; clinical trial registries; and manual searches of reference lists. In general, searching multiple databases is an important strategy to avoid missing relevant articles.

Are Selection and Assessments of Studies Reproducible?

Authors should present evidence that the selection and assessment of studies was reproducible. Duplicate review eligibility and risk of bias assessment, including a measure of agreement, (chance-corrected agreement measured by κ often is used and is appropriate) enhances the credibility of a NMA [19].

Did the Review Present Results That Are Ready for Clinical Application?

Results should be presented in a way that they are useful to clinicians and can be applied to patient care. Clinicians

should be able to easily find best estimates of relative effect sizes and directions of each paired comparison, with 95% CIs or credible intervals. A credible interval conveys the same information as a CI, but is the term used when authors have used Bayesian rather than the standard frequentist statistical approaches [13]. Bayesian approaches are popular in NMAs because, unlike frequentist approaches, they allow for estimation of effects in terms of probabilities, which are more intuitively understood in the context of multiple treatments [11]. Therefore, readers often will find these articles reporting credible intervals.

Authors also should present estimates of absolute treatment effect sizes. A more costly procedure associated with a longer hospital stay may result in a 50% reduction in failure rates, which sounds impressive. The implications are very different, however, if it represents a reduction from 2% to 1% or from 20% to 10%. Thus, presentation of absolute differences is critical to informed decision-making.

In general, authors of NMAs should present results of each paired comparison in the network. Presentation of results of all paired comparisons, however, can become overwhelming when there are many competing alternative treatments. For instance, four treatments yield a manageable six comparisons, six treatments yield a challenging 15 comparisons, and 10 treatments yield an overwhelming 45 comparisons. One appealing format is a forest plot that presents the best estimates of treatment for all agents against one of the least effective. For instance, an extensive NMA addressing 21 surgical and nonsurgical treatment options for sciatica presented all results compared with an inactive control (conventional care) [14]. Even with only a small number of treatments, such a presentation can aid interpretation.

Did the Review Address Confidence in Effect Estimates?

In any NMA, it is almost certain that confidence in estimates will vary from comparison to comparison. Authors need to provide information that helps distinguish between comparisons that warrant strong inferences and those that do not. Issues that compromise the strength of inference and, therefore, the certainty of the evidence include: (1) high risk of bias in the included trials, which is reflected through issues such as lack of concealed randomization, lack of blinding, and high loss to followup. Readers may be familiar with common methods to describe this bias through scales such as the quality assessment scale described by Detsky et al. [4], the scale reported by Jadad et al. [9], or the Cochrane risk of bias tool [8]; (2) imprecision, which is reflected in very wide CIs or credible intervals; (3) inconsistency, which occurs when results differ from study to study, or between direct and indirect comparisons; (4) indirectness, which in a NMA we

refer to as intransitivity, typically resulting from differing populations, interventions, or outcomes; (5) publication bias, which occurs when negative trials remain unpublished; because systematic reviews are more efficient at finding published trials than unpublished ones, this can result in overestimates of treatment effect sizes and erroneous conclusions in support of newer treatments.

Credible NMAs ensure transparent reporting of necessary information to enable readers to make an informed assessment of the certainty of the evidence. Until recently, most NMAs dealt with issues of certainty of the evidence either in a cursory way or as inferences across the whole network [15]. Because certainty is likely to differ from comparison to comparison, such an approach is not very useful. Without certainty estimates for each paired comparison, clinicians are left to guess which results they can trust and which they cannot. Ideally, for each comparison, the authors present the estimate from the direct comparisons and its associated certainty, the estimate from the indirect comparison and its associated certainty, and the overall NMA estimate and its associated certainty. Methodology from the The Grades of Recommendation, Assessment, Development and Evaluation (GRADE) working group provides a system for making certainty ratings that considers risk of bias, precision, consistency, directness, and publication bias [7, 23].

Such a presentation tells readers where the evidence for each comparison comes from (ie, whether predominantly from direct or indirect evidence, or from both) and whether direct and indirect estimates are in agreement. Disagreement between direct and indirect estimates is referred to as incoherence and presents a serious problem when interpreting the results. (Incoherence is discussed in detail in Part II of the Users' Guide for Surgeons.)

Box 3. Revisiting the Clinical Scenario

The NMA addressing management of open tibial shaft fractures has posed a sensible question, patients are well defined and relevant to your clinical dilemma, comparator treatments are comprehensive yet similar, and outcomes are explicitly defined, patient-important, and measured during a consistent and logical time. The authors conducted an exhaustive search and performed duplicate review using predesigned, standardized forms. For the main outcomes, the authors used the 1-year critical unplanned reoperation rate; they also reported deep and superficial infection, nonunion, and malunion rates. They had planned to report postoperative functional outcomes and on health-related quality of life but data were sparse.

The authors present direct comparisons as forest plots, which include absolute event rates, odds ratios, and 95% CIs, and the indirect and combined direct-indirect comparisons in a concise table, along with odds ratios, 95% credible intervals, and GRADE assessments of confidence. The authors also provide an overall ranking of the four treatments with sufficient evidence. You conclude that the methodology is sufficiently robust to make it a credible NMA and therefore proceed to the results.

Conclusion

A NMA is a relatively new study methodology that provides simultaneous comparisons of multiple treatment options based on direct and indirect evidence. Assessing the credibility of the methodology is an important first step in critically appraising a NMA. As with conventional systematic reviews, assessing credibility involves evaluating the article for a sensible research question, an exhaustive search, reproducible selection and assessment of articles, presentation of clinically applicable results, and addressing certainty in effect estimates (Box 3).

In Part II of the Users' Guide for Surgeons we discuss the second important step in critical appraisal of a NMA: judging the certainty we can place in the results.

References

1. Brand RA. Editorial: CORR® criteria for reporting meta-analyses. *Clin Orthop Relat Res*. 2012;470:3261–3262.
2. Chaudhry H, Mundi R, Einhorn TA, Russell TA, Parvizi J, Bhandari M. Variability in the approach to total hip arthroplasty in patients with displaced femoral neck fractures. *J Arthroplasty*. 2012;27:569–574.
3. Chaudhry H, Mundi R, Singh I, Einhorn TA, Bhandari M. How good is the orthopaedic literature? *Indian J Orthop*. 2008;42:144–149.
4. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol*. 1992;45:255–265.
5. Ebrahim S, Mollon B, Bance S, Busse JW, Bhandari M. Low-intensity pulsed ultrasonography versus electrical stimulation for fracture healing: a systematic review and network meta-analysis. *Can J Surg*. 2014;57:E105–118.
6. Foote CJ, Guyatt GH, Vignesh KN, Mundi R, Chaudhry H, Heels-Ansdell D, Thabane L, Tornetta P 3rd, Bhandari M. Which surgical treatment for open tibial shaft fracture results in the fewest reoperations? A network meta-analysis. *Clin Orthop Relat Res*. 2015 Feb 28. [Epub ahead of print].
7. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64:383–394.
8. Higgins J, Altman DG, Sterne JA; Cochrane Statistical Methods Group and the Cochrane Bias Methods Group. Assessing risk of bias in included studies. In: Higgins JP, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions 5.1.0* [updated September 2011]. Available at: http://handbook.cochrane.org/. Accessed March 23, 2015.
9. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1–12.
10. Jansen JP, Cope S. Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*. 2012;12:152.
11. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. *Value Health*. 2008;11:956–964.
12. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, Lee K, Boersma C, Annemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health*. 2011;14:417–428.
13. Lee P. *Bayesian Statistics: An Introduction*. 4th ed.Chichester, United Kingdom: John Wiley & Sons, Ltd; 2012.
14. Lewis RA, Williams NH, Sutton AJ, Burton K, Din NU, Matar HE, Hendry M, Phillips CJ, Nafees S, Fitzsimmons D, Rickard I, Wilkinson C. Comparative clinical effectiveness of management strategies for sciatica: systematic review and network meta-analyses. *Spine J*. 2013 Oct 4. [Epub ahead of print] doi: 10.1016/j.spinee.2013.08.049.
15. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K; Ad Hoc Network Meta-analysis Methods Meeting Working Group. Network meta-analysis: highly attractive but more methodological research is needed. *BMC Med*. 2011;9:79.
16. Mills EJ, Bansback N, Ghement I, Thorlund K, Kelly S, Puhan MA, Wright J. Multiple treatment comparison meta-analyses: a step forward into complexity. *Clin Epidemiol*. 2011;3:193–202.
17. Mills EJ, Druyts E, Ghement I, Puhan MA. Pharmacotherapies for chronic obstructive pulmonary disease: a multiple treatment comparison meta-analysis. *Clin Epidemiol*. 2011;3:107–129.
18. Mills EJ, Wu P, Lockhart I, Thorlund K, Puhan M, Ebbert JO. Comparisons of high-dose and combination nicotine replacement therapy, varenicline, and bupropion for smoking cessation: a systematic review and multiple treatment meta-analysis. *Ann Med*. 2012;44:588–597.
19. Moher D Liberati A Tetzlaff J Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009 Jul 21. [Epub ahead of print].
20. Mundi S, Chaudhry H, Bhandari M. Systematic review on the inclusion of patients with cognitive impairment in hip fracture trials: a missed opportunity? *Can J Surg*. 2014;57:E141–145.
21. Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, Neumann I, Carrasco-Labra A, Agoritsas T, Hatala R, Meade MO, Wyer P, Cook DJ, Guyatt G. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA*. 2014;312:171–179.
22. Norman GR, Streiner DL. *Biostatistics: The Bare Essentials*. 3rd ed. Hamilton, Canada:BC Decker Inc; 2008.
23. Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Peterson R, Singh JA, Kessels AG, Guyatt GH; GRADE Working Group. A GRADE working group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*. 2014;349:g5630.
24. Ristevski B, Nauth A, Williams DS, Hall JA, Whelan DB, Bhandari M, Schemitsch EH. Systematic review of the treatment of periprosthetic distal femur fractures. *J Orthop Trauma*. 2014;28:307–312.
25. 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.
26. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, Holland R, Chen YF, Glenny AM, Deeks JJ, Altman DG. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ*. 2011;343:d4909.
27. Wright JG, Swiontkowski MF, Tolo VT. Meta-analyses and systematic reviews: new guidelines for JBJS. *J Bone Joint Surg Am*. 2012;94:1537.