

An overview of conducting systematic reviews with network meta-analysis

Deborah M Caldwell

Introduction

Systematic reviews with network meta-analysis (NMA) are published with increasing frequency in the health care literature. Prior to 2008, very few systematic reviews contained a NMA [1]; however, there has been a marked increase, to mid-2012 Lee recorded 201 published networks [2]. The statistical method has been available since 2002 [3,4] and owes its origins to much earlier work [5,6]. NMA has matured and models are available for all types of underlying data and summary effect measures [7-12] and can be readily implemented in both frequentist and Bayesian frameworks with pre-written programmes available in widely used softwares [8,13-15].

Recently, focus has shifted to making NMA more accessible [16,17]; however, the conduct of systematic reviews for NMA has received less attention [18]. In this special thematic series on network meta-analysis, the editors of *Systematic Reviews* are encouraging submissions of methodological papers concerning the conduct and reporting of meta-analyses and results papers (<http://www.systematicreviewsjournal.com/about/update/SysRevCFP>). As a preface to the series, this editorial provides an overview of the basic principles of NMA and summarises some of the key challenges for those conducting a systematic review.

The need for network meta-analysis in comparative effectiveness research

Why has NMA increased in popularity? To illustrate, consider the relative effectiveness of six psychotherapies vs. treatment as usual for treatment of moderate to severe depression [19]. In a pairwise meta-analysis, the systematic reviewer has three synthesis options: (1) “lump” all six psychotherapies together to form a single comparator, (2) conduct six separate pairwise meta-analyses in a single systematic review, or (3) conduct six separate systematic reviews. If the question of interest to the decision-maker is

“which psychotherapy should I recommend for depression?” the results of pairwise syntheses do not satisfactorily translate into practice. A clinician does not recommend an “average” psychotherapy to a patient but a specific one, such as cognitive behavioural therapy. To use results from options 2 and 3, the decision-maker must summarise across multiple analyses/reviews without formal assessment of whether the body evidence was coherent or similar enough to form a treatment recommendation. Such an approach makes effect estimates problematic to interpret and is not recommended [20].

NMA came to prominence within this decision-making context [21,22]. NMA is the simultaneous comparison of multiple competing treatments in a single statistical model [23]. In its simplest form, it is the combination of direct and indirect estimates of relative treatment effect, where indirect evidence refers to evidence on treatment C relative to B obtained from A vs. B and A vs. C studies. This is commonly depicted by the equation $\theta_{BC}^I = \theta_{AC}^D - \theta_{AB}^D$ where θ denotes the true underlying treatment effect estimate (e.g. log odds ratio, mean difference, etc.) and the superscript either *Direct* or *Indirect* evidence. If both direct and indirect estimates are available, they can be pooled to produce an internally coherent set of effect estimates of each treatment relative to every other whether or not they have been compared in head-to-head trials. It is also possible to calculate the probability of one treatment being the best for a specific outcome. Treatment options can then be ranked from the best to the worst for each outcome.

Systematic review process for a network meta-analysis

The rigorous conduct of a standard systematic review should apply equally to a NMA. For example, it is good practice to register a protocol for NMA on a repository such as PROSPERO [24] and report a thorough and reproducible literature search. Inclusion/exclusion criteria for a NMA should also be based on a well-defined population, intervention, comparator, outcome (PICO) research question, since it is the specification of the PICO which ensures the key assumption of transitivity is fulfilled.

Correspondence: d.m.caldwell@bristol.ac.uk
School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whately Road, Bristol BS8 2PS, UK

Transitivity suggests that intervention A is similar when it appears in the A vs. B and A vs. C studies [25]. Transitivity can be examined by comparing the distribution of potential effect modifiers across the different comparisons [26], since if there is an imbalance in the presence of effect modifiers across the A vs. B and A vs. C comparisons, the conclusions about B vs. C may be in doubt. Potential effect modifiers should be pre-specified in a protocol and are usually study level characteristics which are routinely extracted in systematic reviews, such as age, severity, dose, setting, etc. Identifying a lack of transitivity may be difficult and sufficient detail is not always available in published trials to allow a thorough assessment [27,28].

The statistical manifestation of the transitivity assumption is called consistency, which holds when the direct and indirect sources of evidence are in agreement, i.e. $\hat{\theta}_{BC}^I = \hat{\theta}_{AC}^D - \hat{\theta}_{AB}^D$ where $\hat{\theta}$ denotes observed estimates. Transitivity should *always* be examined in NMA; however, it is only possible to assess consistency when there are direct and indirect sources of evidence for a treatment contrast. Thus, inconsistency is a property of “loops” of evidence, here the loop A-B-C [29]. Empirical studies have reported the frequency of statistically significant inconsistency ranging from 2% to 14% of published “loops” of evidence [9,30]. It has been argued, however, that the detection of inconsistency in these studies may reveal less about the reliability of NMA and rather more about the problems associated with systematic review options 2 and 3 identified above [31]. Thus, the assessment of transitivity is of fundamental importance in the conduct of the systematic review.

Defining treatments and network size in NMA

Perhaps the biggest deviation from a pairwise systematic review is in the definition of treatments in the network. The identity of each distinct treatment can be preserved in NMA; there is no need to lump across doses or ignore co-treatments in order to conduct analysis. Indeed, the statistical inconsistency noted in empirical studies can often be explained by separating treatments into distinct doses or co-treatments [32].

Treatments included in the network can be divided into a *decision* and *supplementary* set. Treatments within the decision set are the focal treatments of interest to systematic review authors. However, a supplementary set of treatments may also be incorporated into the network to provide additional evidence on relative treatment effects of the decision set. For example, a placebo comparator is rarely of practical clinical interest but its inclusion might (i) connect an otherwise unconnected network of treatments, (ii) increase the precision of the treatment effect estimates of interest if the bulk of the evidence is on placebo comparisons, or (iii) improve estimates of between-

trial heterogeneity. Care must be taken to ensure that all treatments in the network are “jointly randomizable” [25]. That is, all treatments should be options for the population considered in the systematic review such that they could reasonably be compared in a single trial.

Sturtz and Bender [33] have referred to network size as an “unsolved issue” in NMA, and it is an area of developing interest [34,35]. The inclusion or exclusion of treatments from the network has the potential to modify treatment effect estimates and the treatment rankings [36]. A meticulous PICO and pre-specified strategy for extending the network [37] will mitigate but not eliminate the risk of *post hoc* inclusion/exclusion of treatments. Where unexpected interventions are identified by the literature search, a sensitivity analysis should be undertaken to examine the impact of its inclusion/exclusion. For the systematic reviewer, the most important consideration in determining network size is likely to be the resource implications of including additional treatments or searching for further evidence to connect existing networks. For example, although a search strategy for decision set treatments is also likely to return those studies also including a supplementary set comparator, the additional resource employed in title screening and eligibility checking is not inconsequential. The larger the network the more intensive the assessment of transitivity, data extraction, risk of bias assessment and tabulation of results is likely to be. Assuming the transitivity assumption holds, the systematic reviewer must balance this extra resource against the benefit of increasing network size.

Summarising and reporting network meta-analysis

An important source of guidance for systematic reviewers is the Cochrane Collaborations’ Comparing Multiple Interventions Methods Group. The group focuses on methodology for comparing multiple interventions in Cochrane Intervention Reviews; however, much of the work is generalizable. An example protocol for reviews containing a NMA is available, as is guidance on statistical methods and interpretation and presentation of results (see <http://cmimg.cochrane.org/comparing-multiple-interventions-cochrane-reviews>). Presenting the results from a systematic review with NMA can be challenging [38,39]. The number of treatments included in NMA can be large; Veroniki’s [9] findings are representative with a range of 4 to 17 treatments (median 6). The number of pairwise comparisons to report from 4 treatments is 6; from 17 treatments, it is 136.

It is commonplace in pairwise systematic reviews to consider the quality of the body of evidence and to summarise the confidence one can place in the conclusions. Attention is turning to how approaches, such as GRADE, can be extended to NMA [40,41]. There are no universally accepted standards for reporting either the methods or

results of a NMA, although there are a number of society and national technology assessment organisations who have produced in-house guidance [42,43]. Finally, journal editors and peer reviewers should be mindful that web appendices and supplementary files are a necessity in NMA and they can be large. International initiatives such as the forthcoming extension to PRISMA for reporting of NMA will provide systematic reviewers the much needed guidance here [44].

Competing interests

DMC is a co-convenor of the Cochrane Collaborations' Comparing Multiple Interventions Methods Group mentioned in this article.

Acknowledgements

I would like to thank Dr. Nicky J. Welton for the helpful feedback on the draft of this manuscript.

Funding

DMC is supported by a UK Medical Research Council Population Health Scientist fellowship award (G0902118).

Received: 1 September 2014 Accepted: 16 September 2014

Published: 29 September 2014

References

1. Song F, Loke YK, Walsh T, Glenny A-M, Eastwood AJ, Altman DG: **Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews.** *Br Med J* 2009, **338**:b1147.
2. Lee AW: **Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009.** *J Clin Epidemiol* 2014, **67**:138–143.
3. Lumley T: **Network meta-analysis for indirect treatment comparisons.** *Stat Med* 2002, **21**:2313–2324.
4. Lu G, Ades AE: **Combination of direct and indirect evidence in mixed treatment comparisons.** *Stat Med* 2004, **23**:3105–3124.
5. Hasselblad V: **Meta-analysis of multitreatment studies.** *Med Decis Making* 1998, **18**:3743.
6. Higgins J, Whitehead A: **Borrowing strength from external trials in a meta-analysis.** *Stat Med* 1996, **15**:2733–2749.
7. Caldwell DM, Welton NJ, Dias S, Ades AE: **Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis.** *Res Synth Methods* 2012, **3**:126–141.
8. Dias S, Sutton AJ, Ades AE, Welton NJ: **A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials.** *Med Decis Making* 2012, **33**:607–617.
9. Veroniki AA, Vasilidiadis HS, Higgins JP, Salanti G: **Evaluation of inconsistency in networks of interventions.** *Int J Epidemiol* 2013, **42**:332–345.
10. Eckermann S, Coory M, Willan AR: **Indirect comparison: relative risk fallacies and odds solution.** *J Clin Epidemiol* 2009, **62**:1031–1036.
11. Schmitz S, Adams R, Walsh C: **The use of continuous data versus binary data in MTC models: a case study in rheumatoid arthritis.** *BMC Med Res Methodol* 2012, **12**:167.
12. Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT: **Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials.** *Stat Med* 2013, **32**:914–930.
13. White IR: **Multivariate random-effects meta-analysis.** *Stata J* 2009, **9**:40–56.
14. Rucker G, Schwarzer G: **Package 'netmeta': network meta-analysis with R.** The R Project website: <http://cran.r-project.org/web/packages/netmeta/netmeta.pdf>.
15. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G: **Graphical tools for network meta-analysis in STATA.** *PLoS One* 2013, **8**:e76654.
16. Mills EJ, Ioannidis JA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH: **How to use an article reporting a multiple treatment comparison meta-analysis.** *JAMA* 2012, **308**:1246–1253.
17. Cipriani A, Higgins JPT, Geddes JR, Salanti G: **Conceptual and technical challenges in network meta-analysis.** *Ann Intern Med* 2013, **159**:130–137.
18. Bafeta A, Trinquart L, Seror R, Ravaud P: **Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review.** *Br Med J* 2013, **347**:f3675.
19. Churchill R, Moore TH, Furukawa TA, Caldwell DM, Jones H, Shinohara K, Imai H, Lewis G, Hunot V: **'Third wave' cognitive and behavioural therapies versus treatment as usual for depression.** *Cochrane Database Syst Rev* 2013, **10**:CD008705. doi: 10.1002/14651858.CD008705.pub2.
20. Caldwell DM, Welton NJ, Ades AE: **Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency.** *J Clin Epidemiol* 2010, **63**:875–882.
21. Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, Lu G: **Bayesian methods for evidence synthesis in cost-effectiveness analysis.** *Pharmacoeconomics* 2006, **24**:1–19.
22. Sutton A, Ades AE, Cooper N, Abrams K: **Use of indirect and mixed treatment comparisons for technology assessment.** *Pharmacoeconomics* 2008, **26**:753–767.
23. Caldwell DM, Ades AE, Higgins JPT: **Simultaneous comparison of multiple treatments: combining direct and indirect evidence.** *Br Med J* 2005, **331**:897–900.
24. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, Stewart L: **PROSPERO at one year: an evaluation of its utility.** *Syst Rev* 2013, **2**:4.
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. Jansen J, Naci H: **Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers.** *BMC Med* 2013, **11**:159.
27. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C: **Assessing key assumptions of network meta-analysis: a review of methods.** *Research Synthesis Methods* 2013, **4**:291–323.
28. Xiong T, Parekh-Bhurke S, Loke YK, Abdelhamid A, Sutton AJ, Eastwood AJ, Holland R, Chen YF, Walsh T, Glenny AM, Song F: **Overall similarity and consistency assessment scores are not sufficiently accurate for predicting discrepancy between direct and indirect comparison estimates.** *J Clin Epidemiol* 2013, **66**:184–191.
29. Lu G, Ades AE: **Assessing evidence inconsistency in mixed treatment comparisons.** *J Am Stat Assoc* 2006, **101**:447–459.
30. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, Holland R, Chen Y-F, Glenny A-M, Deeks JJ, Altman DG: **Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study.** *Br Med J* 2011, **343**:d4909.
31. Ades A, Dias S, Welton NJ: **Response: Song et al. have not demonstrated inconsistency between direct and indirect comparisons.** *Br Med J* 2011, **343**:d4909.
32. Caldwell DM, Gibb DM, Ades AE: **Validity of indirect comparisons in meta-analysis.** *Lancet* 2007, **369**:270.
33. Sturtz S, Bender R: **Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency.** *Research Synthesis Methods* 2012, **3**:300–311.
34. Dequen P, Sutton AJ, Scott DA, Abrams KR: **Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery.** *Value Health* 2014, **17**:416–423.
35. König J, Krahn U, Binder H: **Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons.** *Stat Med* 2013, **32**:5414–5429.
36. Mills EJ, Kanter S, Thorlund K, Chaimani A, Veroniki A-A, Ioannidis JPA: **The effects of excluding treatments from network meta-analyses: survey.** *Br Med J* 2013, **347**:f5195.
37. Hawkins N, Scott DA, Woods B: **How far do you go? Efficient searching for indirect evidence.** *Med Decis Making* 2009, **29**:273–281.
38. Salanti G, Ades AE, Ioannidis JPA: **Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial.** *J Clin Epidemiol* 2011, **64**:163–171.
39. Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ: **Novel presentational approaches were developed for reporting network meta-analysis.** *J Clin Epidemiol* 2014, **67**:672–680.
40. Dumville JC, Soares MO, O'Meara S, Cullum N: **Systematic review and mixed treatment comparison: dressings to heal diabetic foot ulcers.** *Diabetologia* 2012, **55**:1902–1910.

41. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT: **Evaluating the quality of evidence from a network meta-analysis.** *PLoS One* 2014, **9**:e99682.
42. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G: **Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force Report.** *Value Health* 2014, **17**:157-173.
43. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S: **Evidence synthesis for decision making 7: a reviewer's checklist.** *Med Decis Making* 2013, **33**:679-691.
44. Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, Mills E, Catala-Lopez F, Turner L, Altman DG, Moher D: **The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement.** *PLoS One* 2014, **9**:e92508.

doi:10.1186/2046-4053-3-109

Cite this article as: Caldwell: An overview of conducting systematic reviews with network meta-analysis. *Systematic Reviews* 2014 **3**:109.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

