Draft Methods Research

Number xx—

Use of Network Meta-analysis in Systematic Reviews

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

This information is distributed solely for the purposes of predissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

Contract No.

Prepared by:

Investigators:

AHRQ Publication No. xx-EHCxxx
<Month Year>

This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [insert program e-mail address].

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation:

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Parivash Nourjah
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Peer Reviewers

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

Use of Network Meta-analysis in Systematic Reviews

Structured Abstract

Objectives: To summarize publically available guidance for, and current use of, meta-analytic methods for networks of studies with at least one closed loop [Bayesian mixed treatment comparison (MTC) or Lumley's network meta-analysis methods] in evidence synthesis; to identify analyses using these methods and summarize their characteristics; to gain insight regarding the rationale for selection, implementation, and reporting of such methods from content experts and/or researchers.

Methods: In Part One, we identified currently available guidance documents addressing the use of network meta-analysis of networks of studies with at least one closed loop in evidence synthesis by searching governmental agencies' and participating members' of the International Network of Agencies for Health Technology Assessment websites. Commonalities and disagreements amongst guidance documents were summarized qualitatively. Next, in Part Two, a systematic literature search for network meta-analyses using either a Bayesian MTC or Lumley's network meta-analysis method was undertaken. Characteristics of included analysis were summarized qualitatively. Last, in Part Three, we invited a random selection of nine investigators from the systematic literature search to participate in a focus group. Using a web-based series of questions, we queried respondents regarding their opinion of network meta-analysis and how elements of indirect statistical comparison methodology were chosen in their identified analysis. Responses were summarized qualitatively.

Results: Guidance documents were typically written in a fashion to be applicable to network meta-analysis in general and not to a specific methodology. Guidance documents stressed both approaches have strengths and limitations, while only one guidance document attempted to comprehensively address how to conduct a network meta-analysis and how to interpret and report results.

Our systematic review identified 53 network meta-analysis that used Bayesian MTC or Lumley's network meta-analysis methods, of which the majority used Bayesian MTC (88.7 percent). Bayesian analyses either used noninformative priors or did not report detail about priors used. Data regarding the evaluation of convergence, heterogeneity, and inconsistency were not consistently reported, and from those providing detail, it appears a broad range of methods were used.

Due to the infrequent use of Lumley's network meta-analysis and poor response rate to our focus group invitation, all respondents had conducted Bayesian MTC. Consequently, we were unable to compare/contrast the viewpoints of investigators who used these two different methods.

Conclusion: Additional guidance on how and when to conduct meta-analysis for networks of studies with at least one closed loop, as well as how to interpret and report results is needed. Published meta-analyses using these methods varied in how they conducted and reported results.

Contents

Introduction ... 1
Background ... 1
Objectives ... 1
Project Specific Terminology ... 2
Methods ... 3
Part One: Review of Existing Guidance Documents ... 3
Searching the Literature ... 3
Data Synthesis ... 3
Part Two: Systematic Review of Existing Network Meta-Analyses Using Bayesian MTC or Lumley's Network Meta-analysis Methods ... 4
Searching the Literature ... 4
Inclusion and Exclusion Criteria ... 4
Data Extraction ... 4
Data Synthesis ... 5
Part Three: Content Expert Focus Group ... 5
Composition of the Focus Group ... 5
Data Synthesis ... 6
Results ... 7
Part One: Review of Existing Guidance Documents ... 7
Key Points ... 7
Detailed Analysis ... 8
Part Two: Systematic Review of Existing Network Meta-Analyses Using Bayesian MTC or Lumley's Network Meta-analysis Methods ... 18
Results of the literature Search ... 18
Key Points ... 19
Detailed Analysis ... 20
Part Three: Content Expert Focus Group ... 28
Key points ... 28
Detailed Analysis ... 29
Discussion ... 35
References ... 37
Abbreviations ... 43

Tables

Table 1. Checklist of good research practices for conducting and reporting network meta-analyses. ... 10
Table 2. Strengths and weaknesses of approaches for synthesizing networks of evidence with at least one closed loop ... 12
Table 3. International Society for Pharmacoeconomics and Outcomes Guidance for the reporting of a network meta-analysis ... 16
Table 4. Journal level characteristics ... 20
Table 5. General characteristics of mixed treatment comparisons ... 21
Table 6. Methods characteristics in mixed treatment comparisons ... 23
Table 7. Outcomes and results reporting in mixed treatment comparisons ... 25

Table 8. General characteristics for analyses that used Lumley's network meta-analysis 26
Table 9. Methods characteristics analyses that used Lumley's network meta-analysis 27
Table 10. Outcomes and results reporting in analyses that used Lumley's network meta-analysis methods 28
Table 11. Training of respondents .....**Error! Bookmark not defined.**
Table 12. Role of respondents in their meta-analysis 30
Table 13. General network meta-analysis questions..... 30
Table 14. Question specific to the investigator's published Bayesian mixed treatment comparison 32
Table 15. Question specific to the investigator's published Bayesian mixed treatment comparison and how much the specific parameters influenced their decision to conduct a Bayesian mixed treatment comparison 33
Table 16. Information about how the code was derived 33
Table 17. Information about how and why the prior distributions were chosen 34

Figures

Figure 1. Network pattern types..... 11
Figure 2. Inclusion of identified citations for part two of this project 18

Appendixes

Appendix A. Verbatim quotes from guidance documents
Appendix B. Literature search for part two
Appendix C. Data extraction tool for part two
Appendix D. Focus group questions
Appendix E. Excluded studies
Appendix F. Evidence tables
Appendix G. Glossary

Introduction

Background

Clinicians and decision makers often have to select from multiple available interventions when determining the optimal treatment for disease. Ideally, high-quality randomized controlled trials (RCTs) that estimate the effectiveness of all possible interventions directly against one another would be available to guide decision making.[1,2] Of even higher regard in the hierarchy of evidence is a systematic review of such RCT data with pair-wise pooled estimates of effect. However, interventions are commonly compared with placebo or non-active control in RCTs rather than another active intervention and when direct comparative trials exist they are between two of a larger group of possible treatments. As such decision makers are faced with a lack of adequate direct comparative data to make their judgments.

In the absence of direct comparative data, indirect comparisons may provide valuable information. For example, if two different interventions have been evaluated against a common comparator, the relative effects of the two interventions compared with each other can be estimated indirectly.[1,2] Even in the presence of direct comparative data, indirect comparisons may add value to the interpretation of comparative effectiveness, as usually more than two treatments for a given disease are considered in clinical practice (even if all treatments are not directly compared).

According to the "Conducting Quantitative Synthesis When Comparing Medical Interventions" chapter of the Evidence-based Practice Center (EPC) methods guide, investigators may choose to implement a multi-treatment meta-analysis model in order to make indirect statistical comparisons between interventions.[3] Several methodologies exist to indirectly compare interventions, as do modes to implement such methodologies.[1,4-7] These include anchored indirect comparisons as described by Bucher et al.,[4] Lumley's Frequentist network meta-analysis (of networks with at least one closed loop)[5] and Bayesian network meta-analysis [commonly referred to as mixed treatment comparison (MTC)].[6,7] In the simplest form, interventions that are evaluated against a common comparator in separate trials can be evaluated to each other indirectly in the anchored indirect treatment comparison (not addressed in this report).[4] However, as a generalization of indirect comparisons, when more than two treatments are being compared indirectly, and at least one pair of treatments is being compared both directly and indirectly (a closed loop is present), both direct and indirect types of data can be used to estimate effects in a network meta-analysis.[1,2,4-7] Although these latter methodologies for synthesizing networks of studies with at least one closed loop are frequently employed, best practices for their use are unclear.[1,2,8]

Objectives

This report is divided into three parts, each with their own objective.

- Part One: Summarize publicly available guidance discussing when and how to conduct a meta-analysis of a networks of studies with at least one closed loop as well as how to interpret and report the results of such analysis. We will highlight guidance on methods to synthesize networks of evidence with at least one closed loop. However, we will also summarize guidance applicable to network meta-analysis in general when these methods also apply to networks of evidence with at least one closed loop.

- Part Two: Identify network meta-analyses that used either Bayesian MTC or Lumley's network meta-analysis methods and were published since 2006, and summarize their characteristics.
- Part Three: Gather insight from investigators who have conducted network meta-analyses with either Bayesian MTC or Lumley's network meta-analysis methods, as identified in part two of this project. More specifically, investigators will be queried about how elements of such methodology should be chosen and reported.

Project Specific Terminology

Throughout this report we will use the following specific definitions:

- **Network meta-analysis or indirect comparison** = Meant generically to define the simultaneous synthesis of evidence of all pairwise comparisons across more than two interventions[8,9]
- **Closed loop network of evidence** = A network of evidence where more than two interventions are being compared indirectly, and at least one pair of interventions is being compared both directly and indirectly
- **Mixed treatment comparison (MTC)** = An approach whereby both direct and indirect evidence for particular pair-wise comparisons are combined, and interventions that have not been compared directly are linked through common comparators[7,10,14]
- **Lumley's network meta-analysis** = The Frequentist approach originally described by Lumley whereby both direct and indirect evidence are combined when there is at least one closed loop of evidence connecting two interventions of interest [7,10,14]

Methods

Part One: Review of Existing Guidance Documents

Searching the Literature

We searched for publicly available guidance reports and manuals prepared by regulatory bodies or organizations engaged in evidence synthesis for guidance related to meta-analysis of networks of studies with at least one closed loop. More specifically, we searched the following websites: (1) Agency for Healthcare Research and Quality (AHRQ, http://www.ahrq.gov); (2) Centre for Reviews and Dissemination (CRD, http://www.crd.york.ac.uk/crdweb/); (3) Cochrane Collaboration (http://www.cochrane.org/); (4) National Institute for Health and Clinical Excellence (NICE, http://www.nice.org.uk/); (5) International Society of Pharmacoeconomics and Outcomes Researchers (ISPOR, http://www.ispor.org/); (6) Drug Effectiveness Review Program (DERP) of the Oregon Health & Science University (OHSU) Center for Evidence-based Policy (http://www.ohsu.edu/xd/research/centers-institutes/evidence-based-policy-center/derp/?WT_rank=1/); and (7) all other current members of the International Network of Agencies for Health Technology Assessment (INAHTA) (as listed on the INAHTA website http://www.inahta.org/on December 26, 2011). Finally, we supplemented the above with a Google search (http://google.com).

Data Synthesis

Each identified relevant document was read by a single researcher in detail, and key statements were summarized into the following three categories:

- Guidance on "when to conduct" meta-analysis of networks of studies with at least one closed loop, including definitions of network meta-analysis, justification for conducting a network meta-analysis and assumptions that should be met.
- Guidance on "how to conduct" meta-analysis of networks of studies with at least one closed loop, including planning and design, analysis framework, statistical modeling, detection and handling of potential bias/inconsistency, assessment of model fit and sensitivity analysis.
- Guidance on "how to report and interpret" meta-analysis of networks of studies with at least one closed loop, including requirements or suggestions for reporting and displaying results, types of permissible conclusions, limitations of analysis.

Data synthesis was not intended to be an exhaustive representation of the content of the source documents rather a summary. Verbatim quotes from the source documents can be found in Appendix A.

Part Two: Systematic Review of Existing Network Meta-Analyses Using Bayesian MTC or Lumley's Network Meta-analysis Methods

Searching the Literature

A systematic literature search was conducted in Medline (2006 to July 31, 2011), the Centre for Reviews and Dissemination Databases (July 31, 2011) (including the Database of Abstracts and Reviews of Effects, Health Technology Assessment and the National Institute for Health Research Economic Evaluation Database), The Cochrane Library (July 31, 2011), and the American College of Physicians Journal Club (2006 to July 31, 2011). The search strategy in Appendix B was used.

Inclusion and Exclusion Criteria

Two independent investigators assessed reviews for inclusion in a parallel manner based on a priori defined criteria. Systematic reviews that met the following criteria were included: (1) compared the clinical effectiveness or safety of two or more interventions (any treatment, dose, treatment regimen or clinical procedure) based on RCTs; (2) utilized either Bayesian MTC or Lumley's network meta-analysis methods to make a quantitative indirect comparison of two or more interventions; (3) was published in full-text; (4) was published in the English language; and (5) was published between January 1, 2006 and July 31, 2011. Of note, methodological publications that presented network meta-analyses for illustrative purposes and cost-effectiveness analyses were not considered, nor were individual patient data meta-analyses.

Data Extraction

Two reviewers used a standardized tool (Appendix C) to independently extract data; disagreements were resolved through discussion. For each included network meta-analysis, all published material including the manuscript, supplements, appendices, or external websites which the reader was referred to for additional data were used during data extraction. Therefore, the extraction of data in this project is predicated on the reporting of the information by the authors within these sources. When extracting data, we recorded what the authors reported without ourselves judging whether the methods were appropriate or not. If there was insufficient data from all available sources, we indicated "not reported" for that criterion on data extraction.

First, general characteristics were collected on the journals in which included network meta-analyses were published. Characteristics included journal name, impact factor, allowance of supplements or appendices, and limitations on word, table, and figure counts. Secondly, general characteristics of each network meta-analysis were extracted including: (1) the number of authors and if any authors were considered to be methodologists, (2) the number and type of intervention comparisons made; (3) country and year in which the review was conducted; (4) funding source and affiliations; (5) number of printed pages and use of supplement or appendix; (6) the number of trials and patients in the analyses; (7) clinical area (e.g., cardiology, endocrinology, etc.); and (8) the network pattern. For the purposes of this project, we defined a methodologist as an individual with affiliation to a department of statistics, biostatistics, epidemiology, clinical epidemiology, or public health services, as determined by author information and affiliations listed in the publication.[11] The country in which a review was

conducted was determined by the corresponding author's affiliation. The network pattern was determined by figures presented within the review. If a figure was not available investigators determined the pattern based on text descriptions of included trials.

We also extracted information regarding the indirect comparison method utilized including (1) methods/models applied (e.g., Bayesian MTC or Lumley's network meta-analysis methods); (2) whether a fixed-effect or random-effects model was applied; (3) description of model parameters (e.g., choices of prior distributions in Bayesian analysis and assumptions in Frequentist analysis); (4) method for assessment of model fit; (5) methods for handling of potential bias/inconsistency and heterogeneity (e.g., qualitative or quantitative); (6) use of covariate adjustment in models; (7) whether the model accommodated multi-arm trials; (8) software utilized (WinBUGS, OpenBUGS, wrappers, R, commercial software such as SAS/STATA/SPSS); and (9) availability of code. Finally we extracted data concerning the reporting of results. This included (1) type of endpoint (e.g., continuous versus binary); (2) effect size (e.g., odds ratio, relative risk, risk difference, weighted mean difference) assessed; (3) measure of variance (e.g., confidence or credible intervals); (4) use of other methods to report results (e.g., probability of treatment being best, claims of equivalence or non-inferiority); and (5) format/presentation of results (e.g., text, tables, figures, such as figure of network of studies, raw data tables).

Data Synthesis

The general characteristics of journals and network meta-analyses were summarized qualitatively. Categorical data is presented using frequencies and continuous data as means with standard deviations (SDs).

Part Three: Content Expert Focus Group

Composition of the Focus Group

Up to nine content experts were allowed to participate in this focus group. We randomly identified meta-analyses identified in Part Two of this project to invite, via email, either the first or corresponding author to participate in this group. If no response was obtained we sent a reminder email. If we still did not receive a response, we attempted to contact another author on the publication. After contacting two authors unsuccessfully, we selected another publication at random. Upon investigator-expressed interest, a link was sent to the investigator via email which redirected them to the web-based tool SurveyMonkey©. The investigators were asked to complete questions in regard to the unique network meta-analysis which we identified in Part Two of this project (Appendix D).

We intended on participation in this group being numerically similar between investigators who used Bayesian MTC and those who used Lumley's network meta-analysis methods. However, the number of Lumley's network meta-analyses identified in Part Two of this report was much fewer and author response was suboptimal. In an effort to identify additional models using Lumley's network meta-analysis, we re-ran the original literature search from Part Two without the publication year limit. No additional models were identified, and therefore we chose to continue to invite investigators of Bayesian MTCs until we met the target number of nine respondents.

Data Synthesis

Responses from members of the focus group were tallied, summarized, and reported in a de-identified format. Categorical data was summarized using frequencies and continuous data as means with SDs.

Results

Part One: Review of Existing Guidance Documents

Key Points

- Publicly available guidance discussing when and how to conduct a meta-analysis of a network of studies with at least one closed loop as well as how to interpret and report the results of such analysis is summarized.
  - The majority of guidance is applicable to network meta-analysis in general, regardless if the network includes at least one closed loop or follows an alternate pattern.
  - Guidance is provided from many organizations including: Health Information and Quality Authority (HIQA), ISPOR, AHRQ Effective Health Care (EHC) Program, DERP, CRD, Canadian Agency for Drugs and Technologies in Health (CADTH), Australian Department of Health and Ageing, NICE, and Cochrane Collaboration.
  - Guidance from these organizations is not comprehensive and many aspects are not fully commented on. This reflects the lack of definitive evidence in the literature on these approaches and the need for future research.
- The two main approaches to conduct a meta-analysis of a network with at least one closed loop include a Bayesian MTC approach or Lumley's network meta-analysis approach.
  - Both Bayesian MTC and Lumley's network meta-analysis approaches are limited in that they compromise the positive impact of randomization. Network meta-analysis of direct and indirect evidence with at least one closed loop is more akin to an observational approach than a traditional meta-analysis of direct comparative trials.
- There are three specific limitations of the Lumley's network meta-analysis approach: it is restricted to studies with at least one closed loop, it does not account for correlations that may exist between effect estimates when they are obtained from a single multi-arm trial, and there are weaknesses in situations where zero cells are common.
- There are likewise three weaknesses of the Bayesian MTC approach: it sometimes requires specification of arbitrary priors, it is more complex to understand, and more difficult to use the software.
- Regardless of the approach used, consistency of factors and event rates between direct and indirect trials is paramount if network meta-analysis is to be conducted.
  - Consistency should always be assessed for as an a priori component of the review process.
  - What is regarded as consistent enough is not well defined and is a subjective determination.
  - Some organizations recommend presenting direct and indirect evidence separately and if deemed consistent, performing network meta-analysis.
- Sensitivity analyses including alternative priors, inclusion or exclusion of covariates, and use of alternative models can be conducted to increase confidence in the network results.
- For the Bayesian MTC approach, assessment and reporting of model fit is recommended.
- ISPOR has the only comprehensive checklist for conducting and synthesizing networks with at least one closed loop.

- Reporting the study selection process, providing a description of included individual studies, and use of a graphical representation of the network results can help to improve transparency.

Detailed Analysis

Although our objective for Part One was to focus on guidance for conducting a meta-analysis of networks of studies with at least one closed loop, the majority of guidance available is applicable to network meta-analysis in general. When available, we also present guidance specific to networks of studies with at least one closed loop, using either Bayesian MTC or Lumley's network approach.

General Description of Guidance Documents

Searches identified 17 relevant documents from which we extracted information. These included documents from regulatory/government-affiliated groups and non-governmental organizations and collaborations involved in comparative effectiveness review and health technology assessment. Appendix A provides the verbatim statements from the 17 documents organized according to the categories listed in the Methods section. Most guidance is for network meta-analysis in general, regardless of the specific methodology used to conduct the analysis. The documents identified include:

- A guidance document from HIQA (2011)[7]
- A two part guidance document from the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices (2011)[8,9]
- A guidance document from AHRQ's EHC Program (2010)[3]
- A guidance document from the DERP (2011)[12]
- A guidance document from CRD (2009)[13]
- Two guidance documents from CADTH (2006 and 2009)[10,14]
- A published proceedings paper from the Ad Hoc Network Meta-analysis Methods Meeting Working Group (Li et al, 2011)[15]
- Two guidance documents from the Australian Government's Department of Health and Ageing (2008, 2008)[16,17]
- A guidance document from the NICE (2008)[18]
- Three guidance documents from NICE's Decision Support Unit (DSU) (2007, 2011, 2011)[19-21]
- Two guidance documents from the Cochrane Collaboration (2011)[22,23]

When to Conduct Meta-analysis of a Network of Studies with At Least One Closed Loop

The definition or meaning of the term network meta-analysis varies across the identified guidance documents.[8,9] Often these documents use terms such as "indirect treatment comparison", "multiple treatment comparison", "multiple treatment meta-analysis", "mixed treatment meta-analysis (MTM) and "mixed treatment comparison" as synonyms for network meta-analysis.[8,9,15,17,19,22] When used in this way, these terms are meant to represent the simultaneous synthesis of evidence of all pairwise comparisons across three or more interventions.[8,9,19,22] However, other documents use the above terms more definitely in order to differentiate the statistical analysis framework to be applied. In these cases, MTC is referred to

as “an extension”[22] or “special case”[8,9] of network meta-analysis for comparing multiple treatments of both direct and indirect evidence using a Bayesian approach.[7] Of note, MTC can be used for, but is not restricted to, synthesizing networks with at least one closed loop. Lumley’s network meta-analysis is used to describe the Frequentist method for “analyzing a combination of direct and indirect evidence where there is at least one closed loop of evidence connecting the two technologies of interest.”[7,10,14]

A key component of nearly all documents is a discussion around when conducting a network meta-analysis is justified. Here the documents are almost entirely in agreement that synthesizing direct evidence only (from sufficient head-to-head or randomized controlled trials) “should take precedence”[22] or is “preferred”[12,16] over analysis containing indirect evidence.[18] However, in the absence of sufficient direct evidence, network meta-analysis “can be considered as an additional analytic tool”,[3,18] although one document specifically states “pursuit of qualitative or quantitative indirect comparison is never required...”.[12] In cases where analysis of both direct and indirect comparisons is undertaken, two guidance documents suggest the approaches should be considered and reported separately.[18,22] Of note, a few documents appear to advocate for conducting network meta-analysis even in the presence of reasonable direct evidence, suggesting the combination of indirect and direct evidence may “add information that is not available from head-to-head comparison”,[18] “strengthen the assessment between treatments directly evaluated,”[8,9] and “yield a more refined and precise estimate of the interventions directly compared and broaden inference to the population sampled because it links and maximizes existing information within the network of treatment comparisons.”[8,9]

An additional key discussion theme of identified guidance documents revolves around the assumption of “consistency” (also referred to as “exchangeability” in some documents) that must be met in order to undertake network meta-analysis. Documents agreed that the validity of a network meta-analysis relies on the included studies or trials being similar in all factors (other than the intervention) that may affect outcomes, an assumption also important in standard pair-wise meta-analysis.

How to Conduct a Meta-analysis of a Network of Studies with At Least One Closed Loop

A number of the identified guidance documents reaffirmed that the same “good research practices” or “principles of good practice” used when conducting a traditional systematic review and meta-analyses should be carried over to conducting a network meta-analysis.[8,9,18] These documents often reminded readers, “to minimize error and ensure validity of findings from meta-analyses, the systematic review, whether it involves a standard, pair-wise meta-analysis or a network meta-analysis, must be designed rigorously and conducted carefully.”[15] This includes an a priori declaration of the intent to conduct a network meta-analysis and clearly stating in the protocol the methods and implementation methods to be utilized.

A particularly variable area of focus of these documents includes strategies for systematically searching for studies. While many documents suggest following “conventional guidance” when conducting systematic literature searches for a network meta-analysis, some documents also acknowledge the additional time and resources necessary to conduct a network meta-analysis search due to larger number of interventions to assess. While one document suggests an investigator might consider restricting a search to the minimum number of interventions of interest,[7] another document emphasizes that “different specification of eligibility criteria may result in differences in the structure or extent of a network, leading to discrepant findings for

network meta-analyses on the same topic."[23] Moreover, many documents acknowledged that as more interventions are included in a network meta-analysis, the greater that uncertainty is reduced,[21] precision is increased[23] and "the ability to establish whether various sources of evidence 'agree' with each other" is enhanced.[23] In doing so, the documents suggest that network meta-analyses may need to include comparisons not of direct interest (e.g., placebo controls and therapies no longer used in current practice) as they may provide valuable information for the primary comparison(s) through indirect means. To this end, documents propose various strategies to balance validity and efficiency. Some guidance suggests these strategies include restricting to direct evidence only and broadening the search only after demonstrating that no direct data exists,[16] using "iterative search methods" such as those proposed by Hawkins et al.,[8,9] and using previously published, good quality and up-to-date systematic reviews to augment a search.[15] While not uniformly done, some guidelines state [15] or imply[19] that evidence should be derived from RCTs only.[15,16]

Perhaps the most comprehensive guidance on the planning and design of a network meta-analysis is available in the ISPOR document, which provided "a checklist of good research practices".[9] Below is the checklist, which includes guidance in the areas of search strategies, data collection, statistical analysis planning, data analysis and reporting (Table 1). Of note, the checklist often refers researchers to conventional guidelines on conducting meta-analysis.

Table 1. Checklist of good research practices for conducting and reporting network meta-analyses.

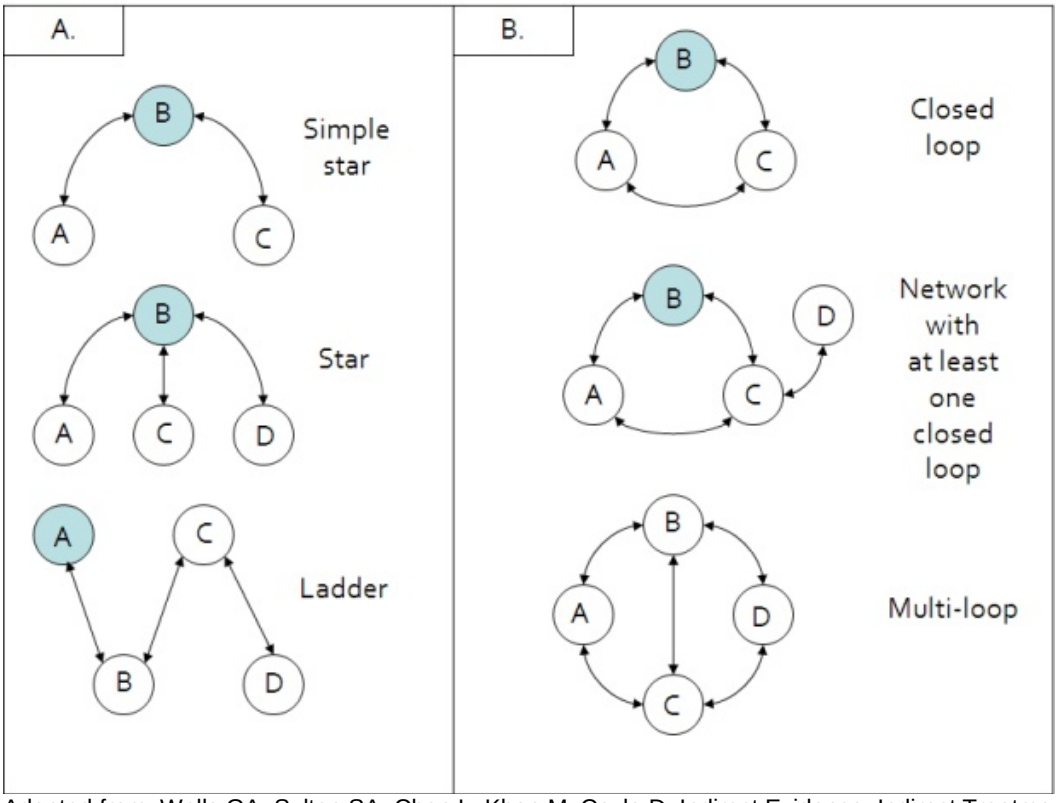
Item	Recommendation
Search strategies	• Follow conventional guidelines for systematic literature searches; be explicit about search terms, literature, and time frames, and avoid use of ad hoc data • Consider iterative search methods to identify higher-order indirect comparisons that do not come up in the initial search focusing on lower-order indirect comparisons
Data collection	• Set forth evidence network demonstrating direct and indirect linkages between treatments, based on identified study reports • Follow conventional guidelines for data collection; use a pre-specified protocol and data extraction form • Include sufficient study detail in data extraction to permit assessment of comparability and homogeneity (e.g., patient and study characteristics, comparators, and outcome measures)
Statistical analysis plan	• Prepare statistical analysis plan prior to data analysis, but permit modifications during data analysis, if necessary • Provide step-by-step descriptions of all analyses, including explicit statements of all assumptions and procedures for checking them • Describe analytic features specific to network meta-analysis, including comparability and homogeneity, synthesis, sensitivity analysis, subgroup analysis and meta-regression, and special types of outcomes
Data analysis	• Follow conventional guidelines for statistical model diagnostics • Evaluate violations of similarity or consistency assumption in evidence network • If similarity or consistency is a problem, consider use of meta-regression models with treatment x covariate interactions to reduce bias
Reporting	• Follow PRISMA statement for reporting of meta-analysis • Explicitly state the study research questions (e.g., in Introduction or Objectives section of report) • Provide graphical depiction of evidence network • Indicate software package used in the analysis and provide code (at least in an online appendix)

Adapted with permission from: Hoaglin DC, Hawkins N, Jansen JP et al. Conducting indirect-treatment comparisons and network meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices – part 2. Value Health 2011;14:429-437.

Abbreviations: PRISMA= preferred reporting in systematic review and meta-analysis

Many of the identified guidance documents provided advantages and disadvantages for the use of the different analysis frameworks (i.e., Frequentist and Bayesian approaches) to network meta-analysis. Documents highlight that the "pattern" of the network of included studies may often dictate the framework used.[7,10,14] Networks of studies that do not contain a "closed loop" such as a simple star, star or ladder pattern (Figure 1) cannot be analyzed using the Frequentist network meta-analysis method as described by Lumley, since a closed loop design is needed for calculating the estimate of incoherence, which is then used to construct 95% confidence intervals for the indirect estimate(s). However, those networks containing a closed loop (Figure 1) can be analyzed using the two of the more complex approaches, Lumley's network meta-analysis or Bayesian MTC. The Bayesian MTC method can be used to analyze any network pattern.

Figure 1. Network pattern types



Adapted from: Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009. Available at: http://www.cadth.ca (Last accessed on December 28, 2011).

Documents list a number of additional considerations when choosing between a Frequentist and Bayesian framework for analyzing these more complex closed loop networks of studies. These are summarized in Table 2 below and include considerations regarding the appropriateness, implementation and interpretation of such analyses. Perhaps the most frequent consideration noted is the potential advantage of Bayesian MTC in that "the method naturally leads to a decision framework that supports decision making"[8,9] by facilitating ranking of compared interventions.

Table 2. Strengths and weaknesses of approaches for synthesizing networks of evidence with at least one closed loop

Methodology	Strength	Weakness
Mixed Treatment Comparison (MTC)	• Allows for ranking of interventions according to the probability they are best • Handles multi-arm studies (trials with more than 2 treatment groups) • Allows for multi-/shared parameter models • Not restricted to networks of studies with at least one closed loop (can be applied to star- or ladder patterns) • "Elegant" • More substantially studies and used (experience) • Use of exact binomial and Poisson likelihoods (zero cell correction not required)	• Requires specification/selection of potentially "arbitrary" priors • Complexity of method may limit use • Difficult software implementation
Network Meta-Analysis Per Lumley	• The amount of agreement in results obtained from different paths for an indirect comparison can be quantified (incoherence statistic, ω^2)	• Restricted to studies with at least one closed loop • Does not account for correlations that may exist between effect estimates when they are obtained from a single multi-arm trial • Except for fixed effect model, only normal approximations and not binomial, multinomial or Poisson likelihoods are used, leading to problems as zero cells become common

With respect to statistical modeling, most guidance documents refer reviewers to the paper by Lumley (2002) for the statistical guidance in implementing Frequentist network meta-analysis, including the necessary code. For Bayesian MTC, the DSU of NICE has built a set of "core models" based upon the framework of generalized linear modeling. The guidance document provides for Normal, Binomial, Poisson and Multinomial likelihoods, with identity, logit, log, complementary log-log, and probit link functions. Moreover, these "core models" can accommodate the assumptions of fixed-effect and random-effects settings, as well as multi-arm trials and multi-/shared parameter models for trials reporting results in different formats (trial versus group level data).

Identified guidance documents also comment on additional statistical modeling issues related to Lumley's network meta-analysis and Bayesian MTC. The merits of using a fixed- or random-effects model are discussed in a number of documents. While fundamentally, either a fixed or random-effects model can be used,[8,9] at least one document[17] states a preference for using the random-effects approach "because the standard error obtained from a fixed effect analysis will be too small if there is heterogeneity between trials (beyond random variation)...", and due to the fact "that there may be additional heterogeneity in an indirect comparison compared to a direct comparison."[17] A few documents acknowledge the potential benefit of incorporating study-level covariates into the model (extending the network to include treatment-by-covariate interactions or meta-regression); however, they also note concerns in the implementation as too few studies are often included in such meta-analyses which increases the potential for ecological bias. Guidance from NICE[18] highlights that when a comparison of the results from single treatment arms from different RCTs is undertaken, the data must be treated as observational and

appropriate steps taken to adjust for possible bias and increased uncertainty (including extending network to include treatment-by-covariate interactions or meta-regression).

The implementation of Bayesian MTC was discussed in detail in many documents. While acknowledging the potential "arbitrary" nature of selection of priors in a Bayesian MTC, many of these documents suggested "vague priors" for such analyses, provided specific values (Appendix A) for different model parameters, and proposed alternative strategies for eliciting/determining priors (i.e., use of larger meta-analyses or expert clinicians in a field) when applicable. Documents also highlighted the need for checking convergence (i.e., running at least three chains, starting from widely different but sensible initial values, and examining posterior distributions visually for spikes and unwanted peculiarities) and running a "conservatively" large number of iterations for both the initial "burn-in" and the posterior sampling.

Additional statistical modeling discussion from the guidance documents included (1) the selection of the referent in Bayesian MTC (as this can affect the posterior estimates), (2) the inappropriateness of treating multi-arm trials as if they were separate trials in a network meta-analysis (the correlation among the effect estimates of pair-wise comparisons must be taken into account), (3) the potential need for multi-/shared parameter models to address situations where trials report results in different formats (i.e., binomial data versus summary log odds and variance), and (4) the summary effect measure to be chosen with documents often recommending relative versus absolute measures due to concerns regarding varying baseline risk; and odds ratios as the preferred relative measure as they are symmetrical around the line of unity.

Nearly all guidance documents addressed identification and handling of potential bias and inconsistency in network meta-analyses. Inconsistency was commonly defined by documents as a conflict between "direct" evidence and "indirect" evidence of a comparison. As noted by one of the NICE guidance documents, "like heterogeneity, inconsistency is caused by effect-modifiers, and specifically by an imbalance in the distribution of effect modifiers in the direct and indirect evidence."[20] Many documents reminded readers that network meta-analyses are akin to observational studies because the value of randomization does not hold across trials.[8,9,16,22] Consequently, they are prone to similar biases, particularly confounding bias. Other noted factors that might potentially influence effect estimates include the number of trials with two or more comparison arms and heterogeneity (as with traditional pair-wise meta-analysis).[15]

Documents unanimously agree that the "consistency" or "exchangeability" assumption must be assessed and should be an a priori component of the review protocol.[7-23] Both the CADTH and the ICWG documents provide guidance for determining whether the "consistency" or "exchangeability" assumption is met based upon a detailed review of included studies (Appendix A). Both frameworks include an assessment of comparability of the common or "linking" treatment and comparability of patients in trials for presence of clinical or methodological heterogeneity. The ICWG document more specifically suggests for the direct trials and indirect comparison, to assess whether the measure of comparative treatment effect is appropriate and assess the event rates of linking interventions. Another document further warned "with increased complexity and greater numbers of treatments, the prospect of inconsistency increases."[7]

Documents also suggest more quantitative methods for detecting inconsistency between direct and indirect evidence. As noted in the ISPOR document, many regulatory agencies require the direct estimates and indirect estimates be calculated separately and shown to be consistent before they are combined. Within a Bayesian MTC framework, a consistency model can be compared to an inconsistency model, with the residual deviance used as a test of "global

inconsistency". The same NICE DSU document that provided the core Bayesian MTC code also provides these models to assess inconsistency.[20] Other, less favored, statistical methods noted by documents for detecting inconsistency include node splitting and use of measures of inconsistency variance.

Guidance documents are clear in their cautions about conducting network meta-analysis if the "consistency" assumption is not met. Unfortunately, as pointed out by one document, even if inconsistency is detected, "no commonly accepted standard [defines] which studies are similar enough"[8,9] and that the determination is a "subjective" one. Moreover, some guidance documents stress that the validity of indirect comparisons may often be "unverifiable" because of limited detail in publications[3] and the underpowered nature of detecting heterogeneity,[20,17] and yet, another cautioned that inconsistency may affect different regions of a network of trials differently.[15] Therefore, many documents provide more unwavering recommendations against network meta-analysis in the presence of inconsistency, while others make more flexible statements such as: "large inconsistencies rule out meta-analysis, small inconsistencies should add uncertainty to the results"[10,14] and "…researchers must evaluate departures from consistency and determine how to interpret them".[8]

A number of documents discussed the importance of assessing model fit when conducting a Bayesian MTC, both to aid in fixed- versus random-effects model (or other competing model, i.e., with or without covariate interaction) selection, and to demonstrate that the overall model fit is adequate. Examination of residual deviance (the lower the residual deviance the better the fit) and deviance information criteria (DIC) statistics were most commonly recommended.

Some of the guidance documents emphasized researchers could conduct sensitivity analyses to check the robustness of their results to changes in model assumptions. Noted assumptions to be tested in sensitivity analysis included different priors, inclusion or exclusion of covariate/modifiers in the model, and use of a fixed- or random-effects model.

How to Report and Interpret a Meta-analysis of a Network of Studies with At Least One Closed Loop

The proper interpretation of network meta-analyses is of paramount importance given their propensity to inform both clinical decision making as well as coverage for third-party payers. A few guidance documents discussing the proper interpretation and reporting of network meta-analyses were identified in our literature search and are discussed here.[7-9,15,18]

When interpreting the results of a network meta-analysis, it is important to consider the internal validity of the analyses as this "maximizes transparency and avoid(s) errors in interpretation."[15] This can be achieved by assessing the appropriateness of inclusion criteria of the evidence network, the quality of the included studies, and the existence of confounding bias.[8,9] As mentioned previously, "good research practices" are necessary when conducting network meta-analyses, similar to traditional systematic reviews, and this includes use of "rigorous and extensive literature search methods"[8,9] to minimize the potential for publication bias. Moreover, the validity of the network meta-analysis also hinges on the internal validity of the studies included in the review. It is recommended that "each study included in the network meta-analysis should be critically evaluated for bias."[8,9] One of these determinants should be the similarity between the included trials. This involves evaluating the clinical and methodological characteristics of the included studies in order to identify potential sources of bias and includes (but is not limited to) assessing differences in patient populations, methods for outcomes measurement, protocol requirements, duration of follow-up, and the time-frame the study was

conducted.[8,9,15] Differences in these characteristics could affect the integrity of the network and potentially impact interpretation of its results if a treatment-by-covariate interaction exists. An example would be significant differences in "baseline risks and placebo responses across trials" which "can reflect additional important differences in study or patient characteristics across studies."[8,9]

In addition to assessing the internal validity of both the included studies as well as the network itself, decision makers should assess the external validity of the meta-analysis' findings and whether they apply to the population of interest.[8,9] This is important since many clinical trials are conducted using "selective homogenous populations, which compromises external validity." However, decision makers should embrace a certain level of dissimilarity between studies in a network meta-analysis, as this often times more closely reflects real-word clinical practice. It has been said that "some heterogeneity across trials in the network may arguably increase external validity."[8,9] This view should be interpreted with caution, as a high degree of heterogeneity within the direct comparisons may also significantly weaken the network and adversely affect its outputs.

As discussed above, probability statements regarding which intervention in a network meta-analysis is 'best' are commonplace. Is has been recommended that these "probability statements should be interpreted carefully since the difference between treatments might be small and not clinically meaningful."[15] Moreover, probabilities resulting from a Bayesian MTC —which themselves are estimates and contain inherent random variability—may (in certain situations) overstate or exaggerate the relative efficacy of an evaluated intervention leading to misinterpretation that can limit, rather than enhance, decision-making. For example, two interventions could demonstrate quite comparable safety and efficacy profiles, but may appear different based on their Bayesian probabilities. Additionally, this determination "cannot be made on the basis of efficacy endpoints alone."[8,9] This assessment should include evaluations of other available safety and effectiveness data not included in the network meta-analysis, including observational evidence. This will provide a more detailed picture of the totality of information for the intervention and allow the decision maker to more properly assess its place in medical practice.

Guidance documents have been published providing recommendations for the proper reporting of indirect comparisons and network meta-analyses.[8,9,11,15,24,25] A Task Force on Indirect Treatment Comparisons Good Research Practices by the ISPOR has proposed a simplified checklist to assist decision makers in the proper evaluation of a reported network meta-analysis.[8,9] The items included by this task force are included in Table 3. It should be noted that this list is not all-inclusive and does not include enough information to adequately assess either the internal or external validity of an indirect comparison or network meta-analysis.

Table 3. International Society for Pharmacoeconomics and Outcomes Guidance for the reporting of a network meta-analysis

Report section	Checklist item	What to look for in the paper
Introduction	Are the rationale for the study and the study objectives stated clearly?	A clear rationale for the review
Methods	Does the methods section include the following? Description of eligibility criteria Information sources Search strategy Study selection process Data extraction (validity/quality assessment of individual studies)	A systematic review of the literature in accordance with Centre for Reviews and Dissemination guidelines and PRISMA
	Are the outcome measures described?	Justification of outcome measures selected for analysis
	Is there a description of methods for analysis/synthesis of evidence? Do the methods described include the following? Description of analyses methods/models Handling of potential bias/inconsistency Analysis framework	Description and justification of statistical model(s) used: multiple meta-analysis of pairwise comparisons versus network meta-analysis models; fixed- versus random-effects models; models without or with covariate (interactions) Description of whether analyses were performed with a frequentist or Bayesian approach Description of how possible bias/inconsistency was evaluated (either qualitative or quantitative, e.g., comparison of direct evidence with the indirect evidence). If meta-regression models are used, rationale for selection of covariates in models Description of relative-effect estimates used for presentation of findings (e.g., odds ratio, relative risk, hazard ratio, difference in change from baseline) Description of whether relative-effect measures were transformed into expected (absolute) outcomes (e.g., proportion of responders)
	Are sensitivity analyses presented?	Rationale for and description of sensitivity analyses Studies included Prior distributions for model parameters in Bayesian framework
Results	Do the results include a summary of the studies included in the network of evidence? Individual study data? Network of studies?	Description of results of study identification and selection process Table/list of studies with information regarding study design and patient characteristics (that might act as effect modifiers); these are important to judge potential similarity/consistency issues Figure of network of studies Table with raw data by study and treatment as used for the analysis/model. (Optionally present relative effects of available direct comparisons of each study)
	Does the study describe an assessment of model fit? Are competing models being compared?	Justification of model results
	Are the results of the evidence synthesis (ITC/MTC) presented clearly?	Table/ figure with results for the pairwise comparisons as obtained with analyses; Point estimates and measure of uncertainty (95% CIs) In Bayesian framework, probability to reflect decision uncertainty (i.e., probability of which treatment is best if multiple treatments are being compared and probability that one treatment is better than the comparator)
	Sensitivity/scenario analyses	Description of (different) findings with sensitivity/scenario analysis

Report section	Checklist item	What to look for in the paper
Discussion	Does the discussion include the following? Description/summary of main findings Internal validity of analysis External validity Implications of results for target audience	Summary of findings Internal validity (individual trials, publication bias, differences across trials that might violate similarity and consistency assumptions) Discussion regarding generalizability of findings (given patient population within and across trials in network) Interpretation of results from a biological and clinical perspective

Adapted with permission from: Jansen JP, Fleurence R, Devine B et al. 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.

Abbreviations: CI=confidence interval; ITC=indirect treatment comparison; MTC=mixed treatment comparison; PRISMA=preferred reporting of systematic reviews and meta-analysis

This guidance document provides recommendations on items that should be included in the introduction, methods, results, and discussion sections of a network meta-analysis report as well as a detailed description of what to look for in each of these sections.[8,9] Many of the items discussed overlap with guidance on the proper reporting of traditional meta-analyses.[26] Aspects unique to conducting a network meta-analysis deserve special mention, much of which involves appropriate reporting of methods and results. If a Bayesian framework was used to perform the data analysis, it is recommended that "the choice of prior distributions for the model parameters should be defined."[8,9] If sensitivity analyses were conducted evaluating the prior distribution assumptions, these results should be also reported. In addition, the software package used to analyze the data as well as the written code from the program should be provided, "at least in an online appendix."[8,9]

When reporting the results of a network meta-analysis, the ISPOR Task Force suggests that a graphical representation of the network be provided to "improve transparency of the analyses."[8,9] In addition to discussing the study selection process and description of the individual studies, the report should provide results of both the pairwise comparisons as well as indirect treatment comparisons.[8,9,18] It has also been recommended that investigators "explain the difference between direct and indirect evidence based upon study characteristics."[3,18] Additionally recommended items for good reporting include goodness-of-fit of the data as well as calculations of residual deviance.[8,9]

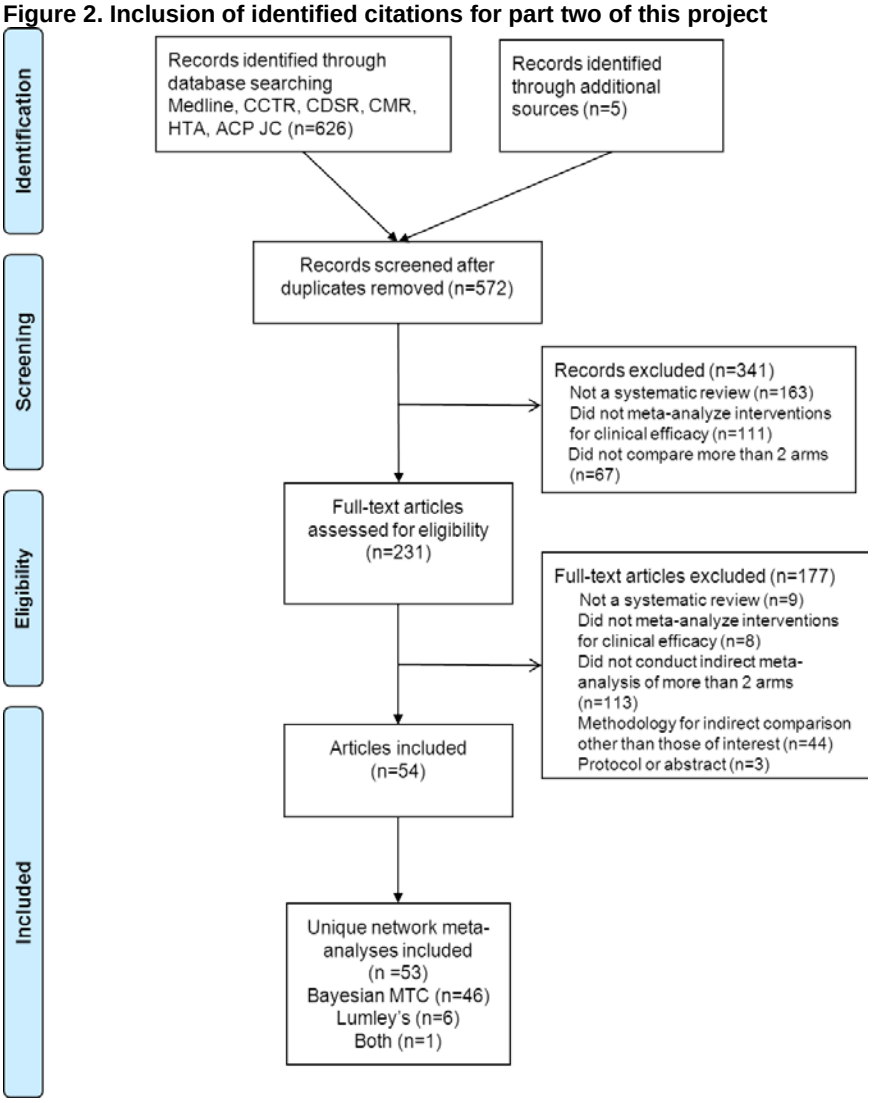
Additional guidance documents for reporting of studies using a Bayesian framework come from the Reporting Of Bayes Used in clinical Studies (ROBUST) criteria, BayesWatch (Bayesian analysis in biomedical research), and Bayesian Standards in Science (BaSiS).[11,24,25] Although these documents are intended for Bayesian analyses in general, they can also be applied to meta-analyses as well. The ROBUST criteria suggests that the following information should be included in any Bayesian study report: prior distributions used, including specified, justified, and sensitivity analysis, analyses run including the statistical model and analytical techniques, and results including central tendency, standard deviation or credible intervals.[11] The BayesWatch and BaSiS include more technical and computational items such as information about the model itself, including details about the software used, if Markov Chain Monte Carlo simulation was used, and if so the number and length of runs as well as convergence diagnostics, shape of the posterior densities, and use of appropriate Bayes factors, amongst others.[24,25] It has been questioned whether these more detailed requirements are important to include for a clinical journal and should be reserved for a more methodologically focused periodical.[11]

Part Two: Systematic Review of Existing Network Meta-Analyses Using Bayesian MTC or Lumley's Network Meta-analysis Methods

Results of the literature Search

A total of 626 citations were identified through the database search with an additional five citations identified manually (Figure 2). After duplicates were removed, 572 citations remained and were screened at the abstract level. Of the abstracts reviewed, 341 were excluded and 231 were considered at the full text level. After full text review, 54 articles representing 53 unique network meta-analyses that utilized either Bayesian MTC or Lumley's network meta-analysis methods were included. A list of excluded studies can be found in Appendix E.

Figure 2. Inclusion of identified citations for part two of this project



Abbreviations: ACP JC= American College of Physicians Journal Club; CCTR=Cochrane Central Register of Controlled Trials; CDSR=Cochrane Database of Systematic Reviews; CMR=Cochrane Methodology Register; HTA=Health technology Assessment; MTC=mixed treatment comparison

Key Points

- Of the included network meta-analyses, the majority use Bayesian MTC methods.
  - Forty-seven unique Bayesian MTC models were identified and conducted in 11 different countries. Fourteen disease categories were evaluated, with the most common being cardiovascular. Most analyses were funded by industry or government/foundation sources.
  - The most common network pattern was “at least one closed loop” and pharmacologic interventions were evaluated in the majority of networks.
  - The statistical code was rarely made available to the reader, although raw data was commonly published.
  - A similar percent of Bayesian MTC analyses either reported using vague priors or did not specify if the priors were intended to be vague or informative. Few models declared using informative priors. It was uncommon to find specific priors, and may be related to lack of code reporting. However, the majority of journals that published these Bayesian MTCs allowed supplement or appendix publication and several manuscripts did utilize this option.
  - Random effects models were used in the majority of Bayesian MTCs. A broad range of methods were used to evaluate convergence, heterogeneity, and inconsistency. Unfortunately, lack of reporting within manuscripts may or may not mean such evaluations were omitted.
  - It was common for authors to rank order interventions based on the probability of the intervention being best for a given outcome. Rarely did authors conclude equivalence or non-inferiority of interventions based on network meta-analysis results.
  - Most Bayesian MTCs evaluated binary outcomes and reported results as odds ratios or relative risks. However, most Bayesian MTCs did not specify whether these were mean or median values of the posterior distribution. All models reported 95 percent credible intervals. Of the models that reported continuous outcomes, the weighted mean difference was the effect measure used almost exclusively.
  - A mixture of tables, text, and figures was commonly used to report results of the MTCs.
  - Lumley’s network meta-analysis methods were infrequently used in the analyses identified by our systematic review of the literature. Seven unique analyses that used Lumley’s network meta-analysis methods were identified and conducted in four countries. Four disease categories were evaluated; cardiology, behavioral health, pain management and gastro-urology.
  - All but one analysis evaluated a network with at least one closed loop while the last pattern could not be determined with published data.
  - Most analyses evaluated pharmacologic interventions with on average seven interventions evaluated.
  - All analyses included a traditional meta-analysis as well. It was more common for heterogeneity to be evaluated in the traditional meta-analysis than in the network meta-analysis. Slightly over half of the analyses evaluated inconsistency.
  - None of the analyses made claims of equivalence, non-inferiority, or defined minimally important differences. All but one analysis reported binary outcomes with equal distribution of odds ratios or relative risks as the effect estimates. All analyses reported variance using 95 percent confidence intervals.

Detailed Analysis

The results are first presented for the journals in which identified network analyses were published followed by results according to the method used to conduct the Bayesian MTC or Lumley's network meta-analysis. Where applicable, mean values are accompanied by SDs (mean±SD). Not all data appear in both text and table format and some data are exclusively reported in within either format.

Journal Level Characteristics

Our systematic literature search identified 53 unique network meta-analyses that used either Bayesian MTC or Lumley's network meta-analysis methods. The majority of systematic reviews used Bayesian MTC methods (46 out of 53, 86.8 percent)[27-73] and few used Lumley's network meta-analysis method (6 out of 53, 11.3 percent).[75-80] One review (1.9 percent) used both methods.[74] Complete details of each journal in which at least one review was published and the journal's characteristics can be found in Appendix Table 2. The 53 analyses were published in 35 different journals, with a mean impact factor of 8.16±8.08 (Table 4). There were two journals which individually had the highest number of publications including British Medical Journal (7 of the 53 reviews, 13.2 percent) and Current Medical Research and Opinion (7 of 53 reviews, 13.2 percent). The majority of journals allowed online supplements or appendices, and also imposed word count limits (Table 4). However, the majority of these journals did not impose limitations on the number of tables or figures allowed.

Table 4. Journal level characteristics

Characteristic	Yes n/N (%) or Mean (SD)
Impact factor	8.16 (8.08)
Supplement or appendix allowed	27/35 (77.1)
Online	23/27 (85.2)
Not specified	4/27 (14.8)
Word count limit	22/35 (62.9)
Table count limit	14/35 (40.0)
Figure count limit	14/35 (40.0)

Bayesian MTC Network Meta-analyses

A summary of the results of Bayesian MTCs can be found in Table 5 to Table 7. Detailed characteristics of each analysis can be found in Appendix Tables 3 to 5. One analysis used both MTC and Lumley's network meta-analysis methods and is considered in both sections of the results.[74] The analysis by Orme et al.[41] included two individual networks and whether this analysis was considered once or twice for a given characteristic is defined within table legends.

General Characteristics

The majority of identified network meta-analyses identified in our literature search used Bayesian MTC to conduct the analysis (86.8 percent). On average, 6.5±5.4 authors were listed per publication and the majority of publications (52.2 percent) did not include a methodologist as an author. The most common country from which authors published reviews was the United Kingdom (36.1 percent), followed by the United States (14.9 percent) and Canada (10.6 percent).The remaining analyses were published in a variety of countries (Table 5). The majority of analyses were funded by industry (36.1 percent), followed by government/foundation sources (25.5 percent) and unfunded work (12.8 percent). Only three analyses (6.4 percent) identified an affiliation, two with the Health Technology Assessment Program and one with The Cochrane

Collaboration. The mean number of printed pages per publication was 15.0±30.8 and just over half (53.2 percent) published a supplement or appendix. However, from the publications that did not publish a supplement or appendix, four did not have the option given the journal specifications and two were affiliated reports that do not have word or page restrictions.

There were 14 different categories of disease states evaluated with a wide dispersion of categories. The most common category was cardiology (19.1 percent) with the remaining categories all representing 10.6 percent or less of included reviews (Table 5). The majority of analyses followed a pattern of at least one closed loop (70.2 percent), followed by star (17.0 percent), simple star (6.4 percent), ladder (4.3 percent), and closed loop (2.1 percent). The mean number of interventions included within the analyses was 7.4±4.0. The majority of analyses evaluated pharmacologic interventions (85.4 percent) with few evaluating devices (8.3 percent) or other interventions (4.2 percent), such as blood glucose monitoring. One analysis included both pharmacologic interventions and devices (2.1 percent). The mean number of trials included in the analyses was 32.1±28.6 and the mean number of patients included was 33,892±65,458.

Table 5. General characteristics of mixed treatment comparisons

Characteristic	n/N (%) or Mean (SD)
Number of authors	6.5 (5.4)
Was a methodologist an author on the manuscript?*	22/47 (46.8)
Country	
U.S.A.	7/47 (14.9)
United Kingdom	17/47 (36.1)
Canada	5/47 (10.6)
Brazil	1/47 (2.1)
China	2/47 (4.3)
Switzerland	4/47 (8.5)
Netherlands	2/47 (4.3)
Italy	3/47 (6.4)
Belgium	1/47 (2.1)
Greece	4/47 (8.5)
Multiple	1/47 (2.1)
Funding	
Industry	17/47 (36.1)
Government/Foundation	12/47 (25.5)
Unfunded	6/47 (12.8)
Other	2/47 (4.3)
Not reported	10/47 (21.3)
Declared affiliation	3/47 (6.4)
Health Technology Assessment Program	2/3 (66.6)
The Cochrane Collaboration	1/3 (33.3)
Number of printed pages	15.0 (30.8)
Supplement or appendix published	25/47 (53.2)
Disease state evaluated	
Behavioral health	4/47 (8.5)
Cardiology	9/47 (19.1)
Infectious disease	5/47 (10.6)
Endocrine	5/47 (10.6)
Pulmonary	2/47 (4.3)
Pain	3/47 (6.4)
Dermatology	2/47 (4.3)
Ophthalmology	2/47 (4.3)
Rheumatology	3/47 (6.4)
Gastroenterology	3/47 (6.4)
Dental	1/47 (2.1)
Oncology	5/47 (10.6)
Substance abuse	1/47 (2.1)
Neurology	2/47 (4.3)

Characteristic	n/N (%) or Mean (SD)
Network pattern	
≥1 closed loop	33/47 (70.2)
Closed loop	1/47 (2.1)
Star	8/47 (17.0)
Simple star	3/47 (6.4)
Ladder	2/47 (4.3)
Number of interventions compared*	7.4 (4.0)
Type of intervention*	
Pharmacologic	41/48 (85.4)
Devices	4/48 (8.3)
Other	2/48 (4.2)
Device and pharmacologic	1/48 (2.1)
Number of trials included in network*	32.1 (28.6)
Number of patients included in network*	33,892 (65,458)

*The trial by Orme et al. included two individual networks and they are considered separately for this characteristic

Methods Characteristics

The majority of analyses also included a traditional meta-analysis (68.1 percent) (Table 6). The most common model for Bayesian MTCs was a random-effects model (48.9 percent), followed by both a random and fixed effects model (23.4 percent), unspecified (17.0 percent), or a fixed-effects model (10.6 percent). The majority of analyses did not report information about whether there was adjustment for multiple arms (80.9 percent) or adjustment for covariates (80.9 percent). Approximately half of the analyses reported testing the model fit (51.1 percent), while the remaining did not comment on testing model fit. Of the 24 analyses that tested model fit, the most common method was use of residual deviance (45.8 percent) followed by using both residual deviance and the deviance information criterion (20.8 percent), solely the deviance information criterion (16.7 percent), median sum deviation (4.2 percent), Q-Q plots (4.2 percent), or unspecified methods (8.3 percent).

All but one analysis (97.9 percent) used WinBUGS software, the final analysis did not report the statistical software used. Two analyses also further specified additional software including BUGS XLA Wrapper (one network meta-analysis) and S-Plus (one network meta-analysis). The majority of analyses did not make their code available to the reader (83.0 percent), although of the eight analyses that did provide the code (17.0 percent) the most common presentation was within the online supplement (six network meta-analyses, 75.0 percent). Raw data was frequently available to the reader (61.7 percent of network meta-analyses) and of the 29 analyses that published raw data, the most common format was within the manuscript itself (25 network meta-analyses, 86.2 percent). Most analyses did not report evaluating convergence (66.0 percent). Of the 16 analyses (34.0 percent) that did evaluate convergence, the most common method was the Gelman Rubin statistic (50.0 percent), although several less frequent methods were used as well (Table 6). Totals of each individual method combined may not add up to the number of studies because one study may have used multiple methods.

Most analyses specifically described the prior distributions used as vague or non-informative (48.9 percent) although 44.7 percent of network meta-analyses did not specify whether the priors used were considered vague or informative. The remaining 6.4 percent of analyses used informative priors. It was uncommon for the actual prior distribution to be reported for d and sigma, as only 34.0 percent and 21.3 percent of network meta-analyses, respectively, reported the actual priors. Most analyses did not perform sensitivity analysis based on the priors used (89.4 percent).

Evaluation of heterogeneity within traditional meta-analyses was common (22 out of 32 network meta-analysis that included traditional meta-analysis, 68.8 percent). Some reported multiple means to test for heterogeneity and therefore the totals of each individual method combined may not add up to the number of studies. The most common method used was the I^2 statistic (81.8 percent) followed by the Cochrane Q-statistic (40.9 percent), among many less frequent methods (Table 6). Evaluation of heterogeneity within network meta-analyses was less common, reported in only 31.9 percent. Some analyses reported multiple means to test for heterogeneity and therefore totals of each individual method combined may not add up to the number of studies. Of these 15 analyses, the most common method used to assess heterogeneity was tau^2 (53.3 percent) followed by between study standard deviation (33.3 percent), among several other less frequent methods (Table 6).

Inconsistency was evaluated in 61.7 percent of analyses. One review reported being unable to evaluate inconsistency due to lack of direct data while the remaining reviews (17 reviews, 36.2 percent) did not report evaluating inconsistency. Totals of each individual method combined may not add up to the number of studies because one study may have used multiple methods. The majority of the 29 analyses that evaluated inconsistency did so through comparison of the results with either the results of their traditional meta-analysis or a previously conducted meta-analysis (48.3 percent) followed by unspecified methods (13.8 percent) or posterior mean residual deviance (13.8 percent), among several other less frequent methods (Table 6).

Table 6. Methods characteristics in mixed treatment comparisons

Characteristic	n/N (%)
Model	
Fixed effects	5/47 (10.6)
Random effects	23/47 (48.9)
Fixed and random effects	11/47 (23.4)
Not reported	8/47 (17.0)
Adjustment for covariates	9/47 (19.1)
Adjustment for multiple arms	9/47 (19.1)
Model fit tested	24/47 (51.1)
Residual deviance	11/24 (45.8)
Deviance information criterion	4/24 (16.7)
Residual deviance and deviance information criterion	5/24 (20.8)
Q-Q plots	1/24 (4.2)
Mean sum deviation	1/24 (4.2)
Method not reported	2/24 (8.3)
Code published	8/47 (17.0)
Online supplement	6/8 (75.0)
External website	2/8 (25.0)
Raw data published	29/47 (61.7)
Manuscript	25/29 (86.2)
Online supplement	3/29 (10.3)
External website	1/29 (3.4)
Evaluation of convergence*	16/47 (34.0)
Gelman Rubin statistic	8/16 (50.0)
Kernel density plot	1/16 (6.3)
Visual plot inspection	1/16 (6.3)
Observation of chain mix	2/16 (12.5)
Trace and time series plots	3/16 (18.8)
Varying burn-in	1/16 (6.3)
Method not reported	2/16 (12.5)
Priors	
Use of vague priors	23/47 (48.9)
Use of informative priors	3/47 (6.4)
Not specified	21/47 (44.7)

Characteristic	n/N (%)
Prior distribution of d reported	16/47 (34.0)
Prior distribution for sigma reported	10/47 (21.3)
Sensitivity analysis based on priors	5/47 (10.6)
Conducted traditional meta-analysis	32/47 (68.1)
Evaluation of heterogeneity in traditional meta-analysis*	22/32 (68.8)
I^2	18/22 (81.8)
Cochrane-Q statistic	9/22 (40.9)
Tau^2	3/22 (13.6)
Heterogeneity p-values	2/22 (9.1)
PICO statement	1/22 (4.5)
Plot visualization	2/22 (9.1)
L'Abbe plot	1/22 (4.5)
Evaluation of heterogeneity in network meta-analysis*	15/47 (31.9)
Tau^2	8/15 (53.3)
Between study SD	5/15 (33.3)
Heterogeneity p-values	1/15 (6.7)
Qualitative evaluation of confidence interval overlap	1/15 (6.7)
Trial population and outcomes	1/15 (6.7)
Evaluation of inconsistency*	29/47 (61.7)
Comparison to traditional or prior meta-analysis	14/29 (48.3)
Inconsistency factors	4/29 (10.3)
Posterior mean residual deviance	4/29 (13.8)
Method not reported	4/29 (13.8)
Trial sequential analysis	1/29 (3.4)
$\sigma^2 w$	1/29 (3.4)
Incoherence factors	2/29 (6.9)
Bayesian mixed p-values	1/29 (3.4)

*Studies that used multiple methods to test heterogeneity were counted multiple times, in the respective categories

Abbreviations: PICO=patient, intervention, comparator, outcome; SD=standard deviation

Outcome and Results Reporting

Few analyses presented graphical representation of the posterior distribution of outcomes (8.5 percent) (Table 7). The use of rank ordering of interventions based on the probability the given intervention was the best for a given outcome was reported in 59.6 percent of analyses. Few analyses made claims of equivalence (6.4 percent) or non-inferiority (12.8 percent). Of the seven that made claims of equivalence or non-inferiority, three (42.9 percent) did not define a minimally important difference while four network meta-analyses did. Four (8.5 percent) analyses defined minimally important differences although did not make specific claims of equivalence or non-inferiority.

Most analyses reported outcomes that were binary (72.0 percent of reviews) followed by both binary and continuous outcomes (17.0), solely continuous outcomes (10.6 percent), and one reported on a categorical non-binary outcome (2.1 percent). Of the 41 analyses that reported binary outcomes, three used multiple effect measures and are counted multiple times when calculating percentage. Odds ratios were the most commonly reported effect measure (58.5 percent), followed by relative risks (17.1 percent) and hazard ratios (14.6 percent), among other less frequent measures. Of the 13 analyses that reported continuous outcomes, two used multiple effect measures and are counted twice in calculating percentage. The weighted-mean difference was the most common effect measure (92.3 percent). Two network meta-analyses (15.4 percent) reported standardized mean difference and one (7.7 percent) reported a measure specific to the content (e.g., prevention fraction in a dental analysis). The one analysis that reported a categorical non-binary outcome used relative risk to measure effect. All analyses reported

variance with 95 percent credible intervals and one also reported standard errors. Most analyses (78.7 percent) did not report if the posterior distribution was the mean or median value. Presentation of results data varied although most analyses used multiple media (and were therefore counted multiple times) including tables, figures, and text. Of the 47 analyses, 39 used text (83.0 percent), 36 used tables (76.6 percent), and 28 used figures (59.6 percent) to present results.

Table 7. Outcomes and results reporting in mixed treatment comparisons

Characteristic	n/N (%) or Mean (SD)
Ranking of outcomes	29/47 (59.6)
Graphical representation of posterior distribution	4/47 (8.5)
Posterior distribution	
Mean	4/47 (8.5)
Median	5/47 (10.6)
Mean and median	1/47 (2.1)
Not reported	37/47 (78.7)
Claims of equivalence	3/47 (6.4)
Claims of non-inferiority	6/47 (12.8)
Minimally important difference	8/47 (17.0)
Type of outcome	
Binary	33/47 (72.0)
Continuous	5/47 (10.6)
Binary and continuous	8/47 (17.0)
Categorical non-binary	1/47 (2.1)
Binary effect measure*	41/47 (87.2)
Relative risk	7/41 (17.1)
Odds ratio	24/41 (58.5)
Hazard ratio	6/41 (14.6)
Multiple	3/41 (7.3)
Log odds ratio	1/41 (2.4)
Continuous effect measure*	13/47 (27.7)
Weighted mean difference	12/13 (92.3)
Standardized mean difference	2/13 (15.4)
Prevention fraction	1/13 (7.7)
Categorical non-binary effect measure	1/47 (2.1)
Relative risk	1/1 (100)
Presentation of Results*	
Table	36/47 (76.6)
Text	39/47 (83.0)
Figure	28/47 (59.6)

*Studies were counted multiple times when more than one method was used.

Network Meta-analyses that used Lumley's Network Meta-analysis Methods

A summary of the results of analyses that used Lumley's network meta-analysis methods can be found in Table 8 to Table 10. Detailed characteristics for each analysis can be found in Appendix Tables 6 to 8. One analysis used both Bayesian MTC and Lumley's network meta-analysis methods and is considered in both sections of the results.[74] Where applicable, mean values are accompanied by SDs (mean±SD).

General Characteristics

Very few of the analyses identified by our systematic review used Lumley's network meta-analysis methods (11.3 percent). On average, 5.3±2.0 authors were listed per publication and all publications were considered to have a methodologist as an author (Table 8). The most common

country from which authors published these network meta-analyses were from the United States (57.1 percent), followed by one each in the United Kingdom, Thailand, and France. The majority of analyses were funded by government/foundation sources (57.1 percent), followed by industry (28.6 percent) and academia (14.3 percent). One analyses identified an affiliation with the Health Technology Assessment Program. The mean number of printed pages per publication was 11.4±4.4 and most of the publications (85.7 percent) published supplements or appendices. The analysis affiliated with the Health Technology Assessment Program was the analysis without a supplement.

There were four different categories of disease states evaluated in the analyses. Three were based on cardiology topics (42.9 percent), two on behavioral health topics (28.6 percent), and one each on gastro-urology (14.3 percent) and pain management (14.3 percent). All but one analysis used a network with at least one closed loop. The last network pattern could not be clearly determined by the information provided in the text and author communication was unsuccessful. We therefore concluded that the pattern was unclear. The mean number of interventions included within the evaluated analyses was 7.0±1.9. Five analyses evaluated pharmacologic interventions (71.4 percent) while two evaluated multiple intervention types (28.6 percent). The mean number of trials included in the analyses was 41.6±30.0 and the mean number of patients included was 70,544.8±82,017.

Table 8. General characteristics for analyses that used Lumley's network meta-analysis methods

Characteristic	n/N (%) or Mean (SD)
Number of authors	5.3 (2.0)
Was a methodologist an author on the manuscript?	7/7 (100)
Country	
U.S.A.	4/7 (57.1)
United Kingdom	1/7 (14.3)
France	1/7 (14.3)
Thailand	1/7 (14.3)
Funding	
Industry	2/7 (28.6)
Government/Foundation	4/7 (57.1)
Academia	1/7 (14.3)
Declared affiliation	1/7 (14.3)
Health Technology Assessment Program	1/1 (100)
Number of printed pages	11.4 (4.4)
Supplement or appendix published	6/7 (85.7)
Disease state evaluated	
Behavioral Health	2/7 (28.6)
Cardiology	3/7 (42.9)
Gastro-urology	1/7 (14.3)
Pain	1/7 (14.3)
Network pattern	
≥1 closed loop	6/7 (85.7)
Unclear	1/7 (14.3)
Number of interventions compared	7 (1.9)
Type of intervention	
Pharmacologic	5/7 (71.4)
Pharmacologic and other	1/7 (14.3)
Procedure, device and pharmacologic	1/7 (14.3)
Number of trials included in network	41.6 (30.0)
Number of patients included in network	70544.8 (82017)

Methods Characteristics

All analyses also included a traditional meta-analysis. The language used to describe the model implemented in each analysis was heterogeneous and can be found in Appendix Table 7. Weighting of studies was not reported in most analyses (85.7 percent) while one (14.3 percent) weighted studies using inverse variance (Table 9). One analysis (14.3 percent) adjusted the model for covariates while the others did not report whether adjustments were made or not. Raw data was available in most analyses (85.7 percent) and of the six that published raw data, the format was equally distributed between online supplements and within the manuscript itself (50.0 percent for each). Four analyses (57.1 percent) used R as the software while one (14.3 percent) used SAS and the other (14.3 percent) used Stata.

Evaluation of heterogeneity within traditional meta-analyses was common (five of seven reviews, 71.4 percent). The most common method used in these five analyses was the I^2 statistic (60.0 percent) while one analysis used both the I^2 statistic and the Cochrane-Q statistic (14.3 percent) and one used the Riley Day test (14.3 percent). Evaluation of heterogeneity within network meta-analyses was less common, reported in only two of the seven analyses (28.6 percent). One used covariance statistics and standard error while the other review used chi^2 and tau^2. Inconsistency was evaluated in all analyses. The majority of analyses (57.1 percent) evaluated inconsistency by comparing results from the network meta-analysis to either the traditional meta-analysis or previously published literature. Other methods reported to evaluate inconsistency included evaluating incoherence values (28.6 percent) and t-tests based on odds ratios from the traditional and network meta-analyses (14.3 percent).

Table 9. Methods characteristics analyses that used Lumley's network meta-analysis methods

Characteristic	n/N (%)
Weighting of studies	
Inverse variance	1/7 (14.3)
Not reported	6/7
Adjustment for covariates	1/7 (14.3)
Raw data published	6/7 (85.7)
Manuscript	3/6 (50.0)
Online supplement	3/6 (50.0)
Heterogeneity assessed in traditional meta-analysis	5/7 (71.4)
I^2	3/5 (60.0)
Cochrane-Q statistic and I^2	1/5 (20.0)
Riley Day test	1/5 (20.0)
Heterogeneity assessed in network meta-analysis	2/7 (28.6)
Covariance and SE	1/2 (50.0)
Tau^2 and chi^2	1/2 (50.0)
Inconsistency evaluation	7/7 (100)
Comparison to traditional or prior meta-analysis	4/7 (57.1)
Incoherence statistic	2/7 (28.6)
T-test	1/7 (14.3)

Outcome and Results Reporting

None of the analyses made claims of equivalence, non-inferiority, or defined a minimally important difference (Table 10). All but one analysis reported outcomes that were binary (85.7 percent). The remaining analysis reported a continuous outcome (14.3 percent). Of the six analyses that reported binary outcomes, half used relative risks and half used odds ratios as effect measures. All analyses reported variance with 95 percent confidence intervals. Presentation of

results data varied although most reviews used multiple media including tables, figures, and text. Of the seven analyses, six used text (85.7 percent), 3 used tables (42.9 percent), and four used figures (57.1 percent) to present results.

Table 10. Outcomes and results reporting in analyses that used Lumley's network meta-analysis methods

Characteristic	n/N (%) or Mean (SD)
Claim of equivalence	0/7 (0)
Claim of non-inferiority	0/7 (0)
Minimally important difference	0/7 (0)
Type of outcome	
Binary	6/7 (85.7)
Continuous	1/7 (14.3)
Binary effect measure	
Relative risk	3/6 (50.0)
Odds ratio	3/6 (50.0)
Continuous effect measure	
Weighted mean difference	1/1 (100.0)
Presentation of Results*	
Table	3/7 (42.9)
Text	6/7 (85.7)
Figure	4/7 (57.1)

*Studies can be counted multiple times based on format used to present results

Part Three: Content Expert Focus Group

Key points

- Nine individuals participated in our focus group, all of whom were authors of a Bayesian MTC identified in Part Two of this report. Unfortunately despite all efforts, none of the limited number of investigators who conducted Lumley's network meta-analysis replied to our invitation or participated in the group.
- The majority of respondents were from academic settings, have been trained in network meta-analysis methods and have conducted at least two such analyses. The respondents seemed to be involved in a variety of the steps in conducting the identified network meta-analysis.
- Respondents seem to feel the term "network meta-analysis" is used ambiguously and inconsistently in the medical literature, although they do not feel the same about the terms "mixed treatment comparison" or "frequentist network meta-analysis".
- Of the questions asking general opinion of network meta-analysis, most responses to questions were on average a neutral response on a 5-point scale. Of the comments which had clear majority opinions were:
    - Disagreement that investigators should consider restricting their search to the minimum number of interventions of interest when conducting a network meta-analysis
    - Agreement that the combination of indirect and direct evidence adds valuable information that is not available from head-to-head comparisons.
    - Agreement that network meta-analysis should provide a graphical depiction of the evidence network.
- When asked specifically about Bayesian MTC, respondents provided a variety of strengths and limitations. Although many were unique, the limitation mentioned most

commonly was in regards to the software while there was no commonly mentioned strength.

- When asked specifically about their Bayesian MTC, most respondents built the code from scratch or adapted the code from a previously published code. Unfortunately we did not gain insight as to how or why prior distributions were chosen but rather what the priors chosen were.
- Additionally respondents were asked to rate 11 criteria on how influential each was in their decision to use Bayesian MTC for their network meta-analysis. The most influential criteria, on average, were the method's ability to handle multi-arm studies and collaborator's or respondent's prior expertise and/or experience. The least influential criterion was the requirement to specify priors which are often arbitrary.

Detailed Analysis

Tables are used throughout this section to present results for each individual focus group question or to present free text responses. Not all data appear in both text and table format and some data are exclusively reported in within either format.

Composition of the Focus Group

The focus group was comprised of nine individuals (hereafter respondents), who authored a unique Bayesian MTC identified in Part Two of this project. Despite all efforts to contact the authors of the analyses using Lumley's network meta-analysis methods, no authors successfully replied or participated in the group. Therefore, the presented results represent the views of investigators who have used Bayesian MTC to conduct their meta-analysis. Most respondents work in academic settings (66.7 percent) and consider themselves to have the expertise needed to implement a network meta-analysis themselves (77.8 percent). Most respondents (88.9 percent) have received either formal or informal training in network meta-analysis methods (Table 11).

Table 11. Training of respondents

"I read many published analysis I initiated myself working on Anne Whitehead book I followed Bayesian courses on evidence synthesis"
"Systematic Review & Meta-Analysis of Direct, Indirect and mixed Treatment Evidence, University of Glasgow Indirect and Mixed Treatment Comparisons Course Leicester"
"3-day course on indirect comparison and MTC by Leicester & Bristol university staff"
"Took a course in Bayesian analysis, including meta-analysis"
"Dedicated course (actually after I did my first analyses)"
"2005 Bristol course. In addition involved in development of methods"
"…research fellowship, AHRQ workshop"

Three respondents are affiliated with an organization involved in conducting synthesis, systematic review, or meta-analysis, including AHRQ (n=2) and Cochrane (n=1). The referenced meta-analysis was not the first in which any of the respondents used such methods. All respondents have conducted at least two network meta-analyses and three of the nine respondents (33.3 percent) have conducted five or more of these analyses. When asked to select

which activities described their involvement in the given analysis, it appears that the respondents were involved in multiple steps of the process (Table 12).

Table 12. Role of respondents in their meta-analysis

Clinical advice, clinical interpretation, policy development (n=6, 66.7%)
Protocol development (n=8, 88.9%)
Developed search strategy (n=8, 88.9%)
Data extraction (n=7, 77.8%)
Statistical advice/methodology (n=7, 77.8%)
Writing or critical revision of manuscript/report (n=8, 88.9%)
Obtaining funding (n=3, 33.3%)
Other (please specify) (n=0)

General Questions Regarding Network Meta-analysis

Respondents were asked a series of 14 questions, using a 5-point Likert scale, regarding general principles and views of network meta-analysis. The results for each question are presented in Table 13. In summary, mixed results were obtained when asking the respondents their opinion as to the ambiguity and consistency in which certain terms were used in the literature. Respondents felt that the term “network meta-analysis” is used ambiguously and inconsistently in the medical literature, whereas the term “mixed treatment comparison” was consistently and unambiguously used. Last, most respondents were neutral to how the term “frequentist network meta-analysis” is used in the literature. All respondents agreed that the combination of indirect and direct evidence adds valuable information that is not available from head-to-head comparisons as well as the necessity for a network meta-analysis to provide a graphical depiction of the evidence network. The majority of respondents disagreed that “when conducting a network meta-analysis, an investigator should consider restricting a search to the minimum number of interventions of interest”. All respondents agreed or were neutral with the statement “the combination of direct and indirect evidence yields a more refined and precise estimate of the interventions directly compared” and the statement “the combination of direct and indirect evidence broadens the external validity of the analysis”. The remaining questions had a mixture of responses that did not have a majority representation.

Table 13. General network meta-analysis questions

Question	Strongly disagree n (%)	Disagree n(%)	Neutral n (%)	Agree n (%)	Strongly agree n (%)	Mean (SD)
The term “network meta-analysis” is used unambiguously and consistently in the medical literature.	2 (22.2%)	3 (33.3%)	4 (44.4%)	0	0	2.2 (0.8)
The term “mixed treatment comparison” is used unambiguously and consistently in the medical literature.	0	1 (11.1%)	4 (44.4%)	3 (33.3%)	1 (11.1%)	3.4 (0.9)
The term “frequentist network meta-analysis” is used unambiguously and consistently in the medical literature.	0	1 (11.1%)	6 (66.7%)	2 (22.2%)	0	3.1 (0.6)

Question	Strongly disagree n (%)	Disagree n(%)	Neutral n (%)	Agree n (%)	Strongly agree n (%)	Mean (SD)
Synthesizing direct evidence only form sufficient head-to-head or randomized controlled trials takes precedence over analysis containing indirect evidence.	0	3 (33.3%)	3 (33.3%)	3 (33.3%)	0	3 (0.9)
The combination of indirect and direct evidence adds valuable information that is not available from head-to-head comparisons.	0	0	0	5 (55.6%)	4 (44.4%)	4.4 (0.5)
The combination of indirect and direct evidence yields a more refined and precise estimate of the interventions directly compared.	0	0	3 (33.3%)	6 (66.7%)	0	3.7 (1.1)
The combination of indirect and direct evidence broadens the external validity of the analysis.	0	0	5 (55.6%)	4 (44.4%)	0	3.4 (0.5)
When analysis of both direct and indirect comparisons is undertaken, each approach should be considered and reported separately.	0	2 (22.2%)	1 (11.1%)	4 (44.4%)	2 (22.2%)	3.7 (1.1)
When conducting a network meta-analysis, an investigator should consider restricting a search to the minimum number of interventions of interest	1 (11.1%)	5 (55.6%)	2 (22.2%)	1 (11.1%)	0	2.3 (0.9)
When conducting a network meta-analysis, an investigator should consider including comparisons not of direct interest (e.g. placebo controls and therapies no longer used in practice)	0	3 (33.3%)	2 (22.2%)	2 (22.2%)	2 (22.2%)	3.3 (1.2)
The more interventions that are included in a network meta-analysis, the greater uncertainty is reduced, precision is increased, and the ability to establish whether various sources of evidence agree with each other is enhanced.	0	2 (22.2%)	3 (33.3%)	3 (33.3%)	1 (11.1%)	3.3 (1.0)
Network meta-analysis should provide a graphical depiction of the evidence network.	0	0	0	2 (22.2%)	7 (77.8%)	4.8 (0.4)
The specific statistical code used should be available either as part of the manuscript, appendix/supplement material, or available on an external website for the reader to freely access	0	1 (11.1%)	2 (22.2%)	4 (44.4%)	2 (22.2%)	3.8 (1.0)
Current guidance on how to conduct and report a network meta-analysis is sufficient.	0	2 (22.2%)	2 (22.2%)	5 (55.6%)	0	3.3 (0.9)

Questions Specific to Bayesian MTC

The respondents were asked a series of open-ended questions. First, they were asked to list the three most significant barriers of Bayesian MTC, results of which are found in Table 14 . All respondents listed at least one barrier, seven listed two barriers, and five listed three barriers. Respondents were also asked to list the three most significant strengths of Bayesian MTC, results of which are listed in Table 14. All respondents listed at least one barrier, eight listed two barriers, and six listed three barriers.

Table 14. Question specific to the investigator's published Bayesian mixed treatment comparison

Listed as the three most significant barriers to conducting a Bayesian mixed treatment comparison	Listed as the three most significant strengths to conducting a Bayesian mixed treatment comparison
Data quality-heterogeneity	Treatment efficacy ranking
Knowledge of the method	Ability to use both direct and indirect evidence
To construct the code	To analyze comparisons that were not conducted directly
Availability of researchers with the necessary expertise	High quality
Ease of use of the software	Practical factors listed on the previous page
WinBUGS	Allows for indirect comparisons when direct evidence is lacking
I cannot see any	Because we do "informal" MTC everyday
Implementing the code	Model estimation
Ability of reader to interpret the method	Multi-arm trial adjustment
Analyze and interpret data	Handling uncertainty
Time	Ability to report ranking of interventions
Lack of user friendliness of WinBUGS	Impact
Proper understanding of the results	Intuitive interpretation
The amount of data to digest	Is growing in acceptance
Adequately reporting on methods in manuscript	Comprehensive picture of the evidence
Availability of evidence to form adequate network	Multi arm trials
Resources	No adjustment for zero cells
Acceptance of the method at time of publishing	Thorough check of the available evidence
How to present results	Novel
	Answers questions that are otherwise unanswerable
	Presentation of results (Bayesian inference in general)
	Survival endpoints

Questions Specific to Respondent's Bayesian MTC

Respondents were asked to rate 11 criteria based on how influential each criterion was in their decision to conduct a Bayesian MTC. A 5-point scale was used ranging from "not at all" to "extremely". The responses to each question can be found in Table 15. On average, the criteria with the most influence were the method's ability to handle multi-arm trials and the collaborator's or respondent's prior experience and/or expertise. The next most influential criterion was the amount of methodological research supporting this method followed by the method's ability to allow rank ordering of interventions according to the probability they are best. The remaining criteria were less influential in the respondents' decision-making to use Bayesian MTC (Table 15).

Table 15. Question specific to the investigator's published Bayesian mixed treatment comparison and how much the specific parameters influenced their decision to conduct a Bayesian mixed treatment comparison

Question	Not at all n (%)	A little n (%)	Moderately n (%)	Quite a bit n (%)	Extremely n (%)	Mean (SD)
The method allows for the ranking of interventions according to the probability they are best.	1 (11.1%)	2 (22.2%)	3 (33.3%)	2 (22.2%)	1 (11.1%)	3 (1.2)
The method allows investigators to check and compare the fit of a model	1 (11.1%)	3 (33.3%)	4 (44.4%)	0	1 (11.1%)	2.7 (1.1)
The method's ability to handle multi-arm studies (those with more than 2 treatment groups)	0	1 (11.1%)	4 (44.4%)	3 (33.3%)	1 (11.1%)	3.4 (0.9)
Frequency of use in previously published network meta-analyses	3 (33.3%)	3 (33.3%)	1 (11.1%)	2 (22.2%)	0	2.2 (1.2)
Ease of software implementation	3	2	2	1	1	2.4 (1.4)
The amount of methodological research supporting this method	2	1	1	4	1	2.7 (1.4)
The methods ability to combine trials reporting results in different formats, for example binomial data and summary log odds with variance (multi- or shared parameter models)	4	0	3	0	2	2.6 (1.7)
Access to pre-built models	3	1	1	4	0	2.7 (1.4)
Requirement to specify priors which are often arbitrary	3	4	2	0	0	1.9 (0.8)
Collaborator(s) or your prior experiences and/or expertise	1	1	2	3	2	3.4 (1.3)
The method's ability to handle studies with "zero cells"	3	1	1	2	2	2.9 (1.7)

In response to a true/false question, five of nine respondents involved a researcher/collaborator solely due to their methodological expertise in Bayesian MTC. Eight of the nine respondents did not use formal guidance to guide how the Bayesian MTC was conducted. One respondent replied that there was no guidance available at the time of their analysis. Respondents were asked how the code used in the analysis was derived. Three codes were adapted from a previously published code, three codes were built from scratch, one code was built from scratch with the help of WinBUGS examples, one code was adapted from a publically available code, and the last instance the respondent was unsure how the code was derived (Table 16). The last open-ended question asked how prior distributions were chosen for the meta-analysis and why they were chosen over others. Unfortunately, the responses collected do not seem to provide insight as to how or why, but rather what the prior distributions were (Table 17).

Table 16. Information about how the code was derived

How was the code used in your analysis derived (e.g. built from scratch, used/adapted previously published/publically available code, used a wrapper such as BUGSXLA or other to generate code, or other source)
Built from scratch plus help from WinBUGS examples (n=1)
Built from scratch (n=3)

Don’t know (n=1)
Adapted from previously published code (n=3)
Adapted from publically available code (n=1)

Table 17. Information about how and why the prior distributions were chosen

How were your prior distributions chosen and why were these distributions chosen over others? (only 8 replies)
Only noninformative distribution used
I don’t know
Noninformative
I don’t know
Used Gaussian noninformative distribution priors for treatment effects as recommended. Model didn’t converge well with flat prior for between study SD, so used empirical informative half-normal prior
We chose flat, noninformative priors
We tried three different priors
Noninformative and sensitivity analyses on heterogeneity priors default in code

Discussion

This report provides the results of a three part methods project that aimed to first review existing guidance on network meta-analysis, secondly to identify previously published network meta-analyses using either Bayesian MTC or Lumley's network meta-analysis approach and summarize their characteristics, and finally, to gather insight from investigators who have conducted network meta-analyses using these methods.

Our review of publicly available guidance documents from various governmental and evidence synthesis groups found that the majority of these documents were typically written in a fashion applicable to network meta-analysis in general, and not specific to any one methodology type. In regards to methods used to conduct meta-analyses of networks of trials containing at least one closed loop, the two approaches typically discussed by guidance included the Bayesian MTC and Lumley's network meta-analysis approaches. Guidance documents stressed that both these approaches have decreased internal validity because they compromise the positive impact of individual study randomization. Common limitations of the Lumley's network meta-analysis approach discussed by guidance documents included the approaches' inability to synthesize networks of studies lacking at least one closed loop, the fact that the method does not account for correlations that may exist between effect estimates when they are obtained from a single multi-arm study, and a weaknesses in situations where zero cells are common. The Bayesian MTC approach was often criticized for requiring specification of arbitrary priors, its complexity to understand, and the need to use non-user-friendly software to implement. Guidance noted some similarities between the methods as well. Regardless of the approach discussed, guidance documents stressed the need for consistency of factors and event rates between direct and indirect trials and the importance of assessing for consistency/inconsistency and heterogeneity. The International Society of Pharmacoeconomics and Outcomes Researchers was the only group that attempted to comprehensively address how to conduct, interpret and report a network meta-analysis. Additional guidance on how to conduct, interpret and report a network meta-analysis is needed.

Our systematic review identified 53 unique network meta-analyses that used either Bayesian MTC or Lumley's network meta-analysis methods. Of these analyses the vast majority used Bayesian MTC methods. Although the presence of a closed loop is a prerequisite to conducting a network meta-analysis using Lumley's methods, the majority of Bayesian models followed a pattern of at least one closed loop. Therefore, despite the option of conducting either a Bayesian MTC or Lumley's network meta-analysis, most investigators chose a Bayesian MTC. Of the analyses that utilized Bayesian MTC, there was a wide distribution of disease states evaluated although cardiology was the most common area. Most analyses evaluated pharmacologic interventions and were funded by industry or a government/foundation source. The statistical code used in the analysis was rarely made available to the reader, despite the majority of journals allowing publication of a supplement or appendix, although raw outcomes data were more commonly published. A similar number of analyses used vague priors or did not specify whether priors were intended to be vague and few analyses used informative priors. However, it was uncommon for authors to report specific priors used. Most models used a random effects model. Unfortunately, data regarding the evaluation of convergence, heterogeneity, and inconsistency were inconsistently reported and often times not mentioned throughout the publication. From the analyses that reported evaluating these three parameters, it appears that a broad range of methods are being utilized. We cannot say with certainty though that a lack of reporting means these parameters were not evaluated. Perhaps with more clear guidance in the future, as to how to

conduct and report these types of network meta-analyses, a more consistent approach may be taken. When investigators reported results of their findings, it was common that interventions were rank ordered based on the probability of the intervention being best for a given outcome. Rarely did authors conclude equivalence or non-inferiority of interventions based on network meta-analysis results. The most common types of outcomes evaluated were binary outcomes, measured with relative risks or odds ratios and 95 percent credible intervals.

As there were very few network meta-analyses identified by our systematic review that used Lumley's network meta-analysis methods, summarizing similarities and differences across the analyses is difficult. Only seven analyses used these Frequentist type methods, despite the option of doing so amongst the majority of Bayesian MTC models. Unfortunately, we did not gain any insight as to the decision-making and opinions of the investigators of the Frequentist models because of a lack of response to our focus group invitation. All of the respondents had conducted a Bayesian MTC and therefore we could not compare and contrast the viewpoints between investigators who used Bayesian methods versus Frequentist methods.

The group of respondents did not appear to be new to Bayesian MTC methods as all had conducted at least two such analysis and appeared to be involved in a variety of steps in the process. On average, the group felt the term "network meta-analysis" is used ambiguously and inconsistently in the medical literature, although did not feel the same about the terms "mixed treatment comparison" or "frequentist network meta-analysis". In general, there were neutral opinions on average regarding network meta-analyses principles. However, clear majority was seen for the following: disagreement that investigators should consider restricting their search to the minimum number of interventions of interest when conducting a network meta-analysis; agreement that the combination of indirect and direct evidence adds valuable information that is not available from head-to-head comparisons; and agreement that network meta-analysis should provide a graphical depiction of the evidence network. The respondents identified several strengths and limitations of Bayesian MTC. Although most were unique statements, there was a common limitations suggested regarding the software used to run the analyses.

Respondents were asked specifically about their Bayesian MTC which we had identified in part two of this project. The most influential criteria in deciding to use Bayesian MTC, on average, were the method's ability to handle multi-arm studies and collaborator's or respondent's prior expertise and/or experience. The least influential criterion was the requirement to specify priors which are often arbitrary. Most respondents built the code from scratch or adapted the code from a previously published code. Unfortunately we did not gain insight as to how or why prior distributions were chosen rather what the priors chosen were.

References

1. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct andindirect treatment comparisons in meta-analysis of randomized controlled trials. J ClinEpidemiol. 1997;50:683-91. PMID:9250266

2. Lumley T. Network meta-analysis for indirect treatment comparisons. Stats Med 2002;21:2313-2324. PMID:12210616

3. Fu R, Gartlehner G, Grant M, et al. Conducting Quantitative Synthesis When Comparing Medical Interventions: AHRQ and the Effective Health Care Program. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews [posted October 2010]. Rockville, MD. Available at: http://effectivehealthcare.ahrq.gov/ (Last accessed on December 28, 2011).

4. Sutton A, Ades AE, Cooper N et al. Use of indirect and mixed treatment comparisons for technology assessments. Pharmacoeconomics 2008;26:753-767. PMID:18767896

5. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stats Med 2004;23:3105-3124. PMID:15449338

6. Caldwell DM, Ades AE, Higgins PT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897-900. PMID:16223826

7. Health Information and Quality Authority. Guidelines for Evaluating the Clinical Effectiveness of Health Technologies in Ireland. Dublin: Health Information and Quality Authority; 2011. Available at: http://www.hiqa.ie (Last accessed on December 28, 2011)

8. Jansen JP, Fleurence R, Devine B, et al. 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;14:417-28. PMID:21669366

9. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health. 2011;14:429-37. PMID:21669367

10. Wells GA, Sultan SA, Chen L, et al. Indirect Evidence: Indirect TreatmentComparisons in Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009. Available at: http://www.cadth.ca (Last accessed on December 28, 2011).

11. Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. J Clin Epidemiol 2005;58:261-268.

12. Drug Effectiveness Review Project. Systematic Review Methods and Procedures. Portland, OR, 2011. Available at: http://www.ohsu.edu/xd/research/centers-institutes/evidence-based-policy-center/derp/documents/upload/DERP_METHODS_WEB_Final_January-2011-2.pdf (Last accessed on December 28, 2011).

13. Centre for Reviews and Dissemination. CRD's Guidance for Undertaking Reviews in Health Care. University of York, 2009. Available at: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf (Last accessed on December 28, 2011).

14. Guidelines for the economic evaluation of health technologies: Canada [3rd Edition]. Ottawa:Canadian Agency for Drugs and Technologies in Health; 2006. Available at: http://www.cadth.ca (Last accessed on December 28, 2011).

15. Li T, Puhan MA, Vedula SS, et al. 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. PMID:21707969

16. Pharmaceutical Benefits Advisory Committee. Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (version 4.3); Australian Government, Department of health and Ageing. 2008. Available at: http://www.health.gov.au/internet/main/publishing.nsf/content/AECB791C29482920CA25724400188EDB/$File/PBAC4.3.2.pdf (Last accessed on December 28, 2011).

17. Report of the Indirect Comparisons Working Group (ICWG) to the Pharmaceutical Benefits Advisory Committee: Assessing Indirect Comparisons. Australian Government, Department of health and Ageing. Available at: http://www.health.gov.au/internet/main/publishing.nsf/Content/B11E8EF19B358E39CA25754B000A9C07/$File/ICWG%20Report%20FINAL2.pdf (Last accessed on December 28, 2011).

18. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, 2008. Available at: http://www.nice.org.uk (Last accessed on December 28, 2011).

19. Dias, S., Welton, N.J., Sutton, A.J. et al. A.E. NICE DSU Technical Support Document 2:A Generalised Linear Modeling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials, 2011; last updated August 2011. Available at: http://www.nicedsu.org.uk (Last accessed on December 28, 2011).

20. Dias, S., Welton, N.J., Sutton, A.J. et al. A.E. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. 2011. Available at: http://www.nicedsu.org.uk (Last accessed on December 28, 2011).

21. National Institute for Health and Clinical Excellence. Briefing paper for methods review workshop on evidence synthesis(indirect and mixed treatment comparisons), 2007. Available at: http://www.nice.org.uk/media/4A6/2F/EvidenceSynthesisBriefingPaper.pdf (Last accessed on December 28, 2011).

22. Higgins JPT, Green S (eds). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (Last accessed on December 28, 2011).

23. Evolution of Cochrane Intervention Reviews and Overviews of Reviews to better accommodate comparisons among multiple interventions. Report from a meeting of the Cochrane Comparing Multiple Interventions Methods Groups, Milan, March 2011. Available at:http://cmimg.cochrane.org/sites/cmimg.cochrane.org/files/uploads/CMIMG%20summary%20of%20meeting%20Milan%20March%202011.pdf (Last accessed on December 28, 2011).

24. The BaSiS Group. Bayesian standards in science (BaSiS). Available at: http://lib.stat.cmu.edu/bayesworkshop/2001/BaSiS.html (Last accessed on September 21, 2011).

25. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. Health Technol Assess 2000;4:1-130.

26. Preferred reporting items for systematic reviews and meta-analyses. Available at: http://www.prisma-statement.org/statement.htm (Last accessed April 4, 2012)

27. Bangalore S, Kumar S, Kjeldsen E et al. Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analysis of 324168 participants from randomized trials. Lancet Oncol 2011;12:65-82. PMID:21123111

28. Mills EJ, Wu P, Chong G, et al. Efficacy and safety of statin treatment for cardiovascular disease: a network meta-analysis of 170,255 patients from 76 randomized trials. QJM 2011;104(2):109-124. PMID:20934984

29. Freemantle N, Tharmanathan P, Herbrecht R. Systematic review and mixed treatment comparison of randomized evidence for empirical, pre-emptive and directed treatment strategies for invasive mould disease. J Antimicrob Chemother 2011 66(Suppl 1):25-35. PMID:21177401

30. Gross JL. Kramer CK. Leitao CB. Et al. Effect of antihyperglycemics agents added to metformin and a sulfonylurea on glycemic control and weight gain in type 2 diabetes: a network meta-analysis. Ann Intern Med 2011;154(10):672-679. PMID:21576535

31. Hartling L, Fernandes RM, Bialy L, et al Steroids and bronchodilators for acute bronchiolitis in the first two years of life: systematic review and meta-analysis. BMJ 2011;342:1714. PMID:21471175

32. Jansen JP, Bergman GJ, Huels J, Olson M. The efficacy of bisphosphonates in the prevention of vertebral, hip, and nonvertebral-nonhip fractures in osteoporosis: a network meta-analysis. Semin Arthritis Rheum 2011;40(4):275-84.e1-2. PMID:20828791

33. Maund E, McDaid C, Rice S, Wright K, Jenkins B, Woolacott N. Paracetamol and selective and non-selective non-steroidal anti-inflammatory drugs for the reduction in morphine-related side-effects after major surgery: a systematic review. Br J Anaesth 2011;106(3):292-297. PMID:21285082

34. McDaid C, Maund E, Rice S, Wright K, Jenkins B, Woolacott N. Paracetamol and selective and non-selective non-steroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery: a systematic review. Health Technol Assess 2010;14(17):1-153. PMID:20346263

35. Sciarretta S, Palano F, Tocci G, Baldini R, Volpe M. Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. Arch Intern Med 2011;171(5):384-394. PMID:21059964

36. Trelle S, Reichenback S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ 2011;342:c7086. PMID:21224324

37. van de Kerkhof P, de Peuter R, Ryttov J, Jansen JP. Mixed treatment comparison of a two-compound formulation (TCF) product containing calcipotriol and betamethasone dipropionate with other topical treatments in psoriasis vulgaris. Curr Med Res Opin 2011;27(1):225-238. PMID:21142833

38. Van den Bruel A, Gailly J, Devriese S, Welton NJ, Shortt AJ, Vrijens F. The protective effect of ophthalmic viscoelastic devices on endothelial cell loss during cataract surgery: a meta-analysis using mixed treatment comparisons. Br J Ophthalmol 2011;95(1):5-10. PMID:19965836

39. Bergman GJ, Hochberg MC, Boers M, Wintfeld N, Kielhorn A, Jansen JP. Indirect comparison of tocilizumab and other biologic agents in patients with rheumatoid arthritis and inadequate response to disease-modifying antirheumatic drugs. Semin Arthritis Rheum 2010;39(6):425-441. PMID:20223500

40. Dakin H, Fidler C, Harper C. Mixed treatment comparison meta-analysis evaluating the relative efficacy of nucleos(t)ides for treatment of nucleos(t)ide-naive patients with chronic hepatitis B. Value Health 2010;13(8):934-945. PMID:20825624

41. Orme M, Collins S, Dakin H, Kelly S, Loftus J. Mixed treatment comparison and meta-regression of the efficacy and safety of prostaglandin analogues and comparators for primary open-angle glaucoma and ocular hypertension. Curr Med Res Opin 2010;26(3):511-528. PMID:20014995

42. Phung OJ, Scholle JM, Talwar M, Coleman CI. Effect of noninsulin antidiabetic drugs added to metformin therapy on glycemic control, weight gain, and hypoglycemia in type 2 diabetes. JAMA 2010;303(14):1410-1418. PMID:20388897

43. Uthman OA, Abdulmalik J. Comparative efficacy and acceptability of pharmacotherapeutic agents for anxiety disorders in children and adolescents: a mixed treatment comparison meta-analysis. Curr Med Res Opin 2010;26(1):53-59. PMID:19905879

44. Vissers D, Stam W, Nolte T, Lenre M, Jansen J. Efficacy of intranasal fentanyl spray versus other opioids for breakthrough pain in cancer. Curr Med Res Opin 2010;26(5):1037-1045. PMID:20199140

45. Walsh T, Worthington HV, Glenny A, Appelbe P, Marinho CCV, Shi X. Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents. Cochrane Database of Systematic Reviews 2011;1 Art No CD007868. PMID:20091655

46. Wandel S, Juni P, Tendal B, Nuesch E, Villiger PM, Welton NJ, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. BMJ 2010;341:4675. PMID:20847017

47. Wang H, Huang T, Jing J, Jin J, Wang P, Yang M, et al. Effectiveness of different central venous catheters for catheter-related infections: a network meta-analysis. J Hosp Infect 2010;76(1):1-11. PMID:20638155

48. Woo G, Tomlinson G, Nishikawa Y, Kowgier M, Sherman M, Wong DK, et al. Tenofovir and entecavir are the most effective antiviral agents for chronic hepatitis B: a systematic review and Bayesian meta-analyses. Gastroenterology 2010;139(4):1218-1229. PMID:20600036

49. Baker WL, Baker EL, Coleman CI. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. Pharmacotherapy 2009;29(8):891-905. PMID:19637942

50. Bansback N, Sizto S, Sun H, Feldman S, Willian MK, Anis A. Efficacy of systemic treatments for moderate to severe plaque psoriasis: systematic review and meta-analysis. Dermatology 2009;219(3):209-218. PMID:19657180

51. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009;373(9665):746-758. PMID:19185342

52. Thompson Coon JS, Liu Z, Hoyle M, Rogers G, Green C, Moxham T, et al. Sunitinib and bevacizumab for first-line treatment of metastatic renal cell carcinoma: a systematic review and indirect comparison of clinical effectiveness. Br J Cancer 2009;101(2):238-243. PMID:19568242

53. Edwards SJ, Lind T, Lundell L, Das R. Systematic review: standard- and double-dose proton pump inhibitors for the healing of severe erosive oesophagitis – a mixed treatment comparison of randomized controlled trials. Aliment Pharmacol Ther 2009;30(6):547-556. PMID:19558609

54. Edwards SJ, Smith CJ. Tolerability of atypical antipsychotics in the treatment of adults with schizophrenia or bipolar disorder: a mixed treatment comparison of randomized controlled trials. Clin Ther 2009;31(Pt 1):1345-1359. PMID:19698898

55. Edwards SJ, Clarke MJ, Wordsworth S, Welton NJ. Carbapenems versus other beta-lactams in the treatment of 40eneralized40 patients with infection: a mixed treatment comparison. Curr Med Res Opin 2009;25(1):251-261. PMID:19210156

56. Golfinopoulos V, Pentheroudakis G, Salanti G, Nearchou AD, Ioannidis JP, Pavlidis N. Comparative survival with diverse chemotherapy regimens for cancer of unknown primary site:

multiple-treatments meta-analysis. Cancer Treat Rev 2009;35(7):570-573. PMID:19539430

57. Jansen JP, Bergman GJ, Huels J, Olson M. Prevention of vertebral fractures in osteoporosis: mixed treatment comparison of bisphosphonate therapies. Curr Med Res Opin 2009;25(8):1861-1868. PMID:19530978

58. ManzoliL, Georgia S, De Vito C, Boccia A, Ionnidis JPA, Villari P. Immunogenicity and adverse events of avian influenza A H5N1 vaccine in healthy adults: multiple-treatments meta-analysis. Lancet Infect Dis 2009;9:481-92. PMID:19628173

59. Meader N. A comparison of methadone, buprenorphine and alpha2 adrenergic agonists for opiod detoxification: a mixed treatment comparison meta-analysis. Drug and Alcohol Dependence 2010;108:110-114. PMID:20074867

60. Mills EJ, Perri D, Cooper C, Nachega JB, Wu P, Tleyjeh I, et al. Antifungal treatment for invasive Candida infections: a mixed treatment comparison meta-analysis. Ann Clin Microbiol Antimicrob 2009;8:23. PMID:19558681

61. Quilici S, Chancellor J, Lothgren M, et al. Meta-analysis of duloxetine vs. pregabalin and gabapentin in the treatment of diabetic peripheral neuropathic pain. BMC Neurol 2009;9:6. PMID:19208243

62. Coleman CI, Baker WL, Kluger J, White CM. Antihypertensive medication and their impact on cancer incidence: a mixed treatment comparison meta-analysis of randomized controlled trials. J Hypertens 2008;26(4):622-629. PMID:18327066

63. Mills EJ, Rachlis B, Wu P, Devereaux PJ, Arora P, Perri D. Primary prevention of cardiovascular mortality and events with statin treatments: a network meta-analysis involving more than 65,000 patients. J Am Coll Cardiol 2008;52(22):1769-1781. PMID:19022156

64. Mauri D, Polyzos NP, Salanti G, Pavlidis N, Ionnidis JPA. Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. J Natl Cancer Inst 2008;100:1780-1791. PMID:19066278

65. Quilici S, Abrams KR, Nicolas A, et al. Meta-analysis of the efficacy and tolerability of pramipexole versus ropinirole in the treatment of restless legs syndrome. Sleep Med 2008;9(7):715-726. PMID:18226947

66. Stettler C, Allemann S, Wandel S, et al. Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis. BMJ 2008;337:a1331. PMID:18757996

67. Golfinopoulos V, Salanti G, Pavlidis N, Ioannidis JP. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: a meta-analysis. Lancet Oncol 2007;8(10):898-911. PMID:17888735

68. Lam SK, Owen A. Combined 41eneralized41ation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of 41eneralize controlled trials. BMJ 2007;335(7626):925. PMID:17932160

69. Nixon R, Bansback N, Brennan A. The efficacy of inhibiting tumour necrosis factor alpha and interleukin 1 in patients with rheumatoid arthritis: a meta-analysis and adjusted indirect comparisons. Rheumatology 2007;46(7):1140-1147. PMID:17478472

70. Stettler C, Wandel S, Allemann S, et al. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. Lancet 2007;370(9591):937-948. PMID:17869634

71. Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. Arch Intern Med 2006;166(12):1269-1275. PMID:16801509

72. Jansen JP. Self-monitoring of glucose in type 2 diabetes mellitus: a Bayesian meta-analysis of direct and indirect comparisons. Curr Med Res Opin 2006;22(4):671-681. PMID:16684428

73. Kyrgiou M, Salanti G, Pavlidis N, Paraskevaidis E, Ioannidis JP. Survival benefits with diverse chemotherapy regimens for ovarian cancer: meta-analysis of multiple treatments. J Natl Cancer Inst 2006;98(22):1655-1663. PMID:17105988

74. Baldwin D, Woods R, Lawson R, Taylor D. Efficacy of drug treatments for 42eneralized anxiety disorder: systematic review and meta-analysis. BMJ 2011;342:1199. PMID:21398351

75. Freemantle N, Lafuente-Lafuente C, Mitchell S, Eckert L, Reynolds M. Mixed treatment comparison of dronedarone, amiodarone, sotalol, flecainide, and propafenone, for the management of atrial fibrillation. Europace 2011;13(3):329-345. PMID:21227948

76. Anothaisintawee T, Attia J, Nickel JC, et al. Management of chronic prostatitis/chronic pelvic pain syndrome: a systematic review and network meta-analysis. JAMA 2011;305(1):78-86. PMID:21205969

77. Hansen RA, Gaynes BN, Gartlehner G, Moore CG, Tiwari R, Lohr KN. Efficacy and tolerability of second-generation antidepressants in social anxiety disorder. Int Clin Psychopharmacol 2008;23(3):170-179. PMID:18408531

78. Jalota L, Kalira V, George E, et al. Prevention of pain on injection of propofol: systematic review and meta-analysis. BMJ 2011;342:1110. PMID:21406529

79. Trikalinos TA, Alsheikh-Ali AA, Tatsioni A, Nallamothu BK, Kent DM. Percutaneous coronary interventions for non-acute coronary artery disease: a quantitative 20-year synopsis and a network meta-analysis. Lancet 2009;373(9667):911-918. PMID:19286090

80. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007;369(9557):201-207. PMID:17240286

Abbreviations

AHRQ	Agency for Healthcare Research and Quality
BaSiS	Bayesian Standards in Science
CADTH	Canadian Agency for Drugs and Technologies in Health
CRD	Center for Reviews and Dissemination
DERP	Drug Effectiveness Review Program
DIC	deviance information criterion
DSU	Decision Support Unit
EHC	Effective Health Care
EPC	Evidence-based Practice Center
HIQA	Health Information and Quality Authority
ICWG	Indirect Comparisons Working Group
INAHTA	International Network of Agencies for Health Technology Assessment
ISPOR	International Society of Pharmacoeconomics and Health Outcomes Researchers
MTC	mixed treatment comparison
MTM	mixed treatment meta-analysis
NICE	National Institute for Health and Clinical Excellence
OHSU	Oregon Health & Science University
RCT	randomized-controlled trial
ROBUST	Reporting of Bayes Used in Clinical Studies
SD	standard deviation