

Cochrane Handbook for Systematic Reviews of Interventions	3
Chapter 1: Introduction	3
Key points	3
1.1 Cochrane	3
1.1.1 What is Cochrane?	4
1.1.2 A brief history of Cochrane	4
1.1.3 Cochrane organization and structure	5
1.2 Systematic reviews	5
1.2.1 The need for systematic reviews	5
1.2.2 What is a systematic review?	5
1.3 Cochrane Reviews	6
1.3.1 Reviews of the effects of interventions	6
1.3.2 Reviews of diagnostic test accuracy	6
1.3.3 Overviews of Reviews	6
1.3.4 Reviews of methodology	7
1.4 Publication of Cochrane Reviews	7
1.4.1 The Cochrane Library	7
1.5 Handbook structure	8
1.6 Chapter information	8
1.7 References	8
Chapter 3: Reporting the review	9
3.1 Introduction	10
3.2 Reporting of protocols of new Cochrane Intervention Reviews	10
3.3 Reporting of new Cochrane Interventions Reviews	11
3.3.1 Abstracts	11
3.3.2 Background and Objectives	12
3.3.3 Methods	13
3.3.3.1 Criteria for considering studies for this review	13
3.3.3.2 Search methods for identification of studies	14
3.3.3.3 Data collection and analysis	14
3.3.4 Results	15
3.3.5 Discussion	17
3.3.6 Conclusions	18
3.3.7 Administrative information	19
3.4 Reporting of plain language summaries in new Cochrane Intervention Reviews	19
3.5 Chapter information	19
3.6 References	20
Chapter 11: Undertaking network meta-analyses	21
11.1 Introduction	21
11.2 Important concepts	22
11.2.1 Indirect comparisons	22
11.2.2 Transitivity	24
11.2.3 From indirect comparisons to network meta-analysis	27
11.3 Planning a Cochrane Review to compare multiple interventions	29
11.3.1 Expertise required in the review team	29
11.3.2 The importance of a well-defined research question	29
11.3.3 Selecting outcomes to examine	31
11.3.4 Study designs to include	31
11.4 Synthesis of results	31
11.4.1 What does a network meta-analysis estimate?	31
11.4.2 Synthesizing direct and indirect evidence using meta-regression	32
11.4.3 Performing network meta-analysis	33
11.4.4 Disagreement between evidence sources (incoherence)	35
11.5 Evaluating confidence in the results of a network meta-analysis	37
11.5.1 Available approaches for evaluating confidence in the evidence	37
11.6 Presenting network meta-analyses	41
11.6.1 Presenting the evidence base of a network meta-analysis	41
11.6.2 Tabular presentation of the network structure	42
11.6.3 Presenting the flow of evidence in a network	43
11.6.4 Presentation of results	44

11.7 Concluding remarks	46
11.8 Chapter information	47
11.9 References	47
Handbook information	50

Cochrane Handbook for Systematic Reviews of Interventions



Handbook for Systematic Reviews of Interventions

Version x.x

[Updated]

Editors:

Chapter 1: Introduction

UNDER CONSTRUCTION - PLEASE DO NOT USE

Authors: Jackie Chandler, Julian PT Higgins, Jonathan J Deeks, Clare Davenport, Mike J Clarke.

This chapter should be cited as: Chandler J, Higgins JPT, Deeks JJ, Davenport C, Clarke MJ. Chapter 1: Introduction. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.2.0 (updated February 2017), Cochrane, 2017. Available from [Cochrane Community](#).

Key points

- Systematic reviews seek to collate all evidence that fits pre-specified eligibility criteria in order to address a specific research question.
- Systematic reviews aim to minimize bias by using explicit, systematic methods documented in advance with a protocol.
- Cochrane prepares, maintains and promotes systematic reviews to inform decisions about health and social care (Cochrane Reviews).
- Cochrane Reviews are published in the Cochrane Database of Systematic Reviews in the Cochrane Library.
- The Cochrane Handbook for Systematic Reviews of Interventions contains methodological guidance for the preparation and maintenance of Cochrane Intervention Reviews, Overviews of Reviews and Methodology Reviews.
- Methodological advice on Cochrane Diagnostic Test Accuracy Reviews can be found in the separate Cochrane Handbook for Diagnostic Test Accuracy Reviews.
- Cochrane has developed conduct and reporting standards.

1.1 Cochrane

1.1.1 What is Cochrane?

Trusted evidence. Informed decisions. Better health.

Cochrane is a global independent network of health practitioners, researchers, patient advocates and others, responding to the challenge of making the vast amounts of evidence generated through research useful for informing decisions about health (www.cochrane.org). Previously known as The Cochrane Collaboration, it is a not-for-profit organization where collaborators aim to produce credible, accessible health information that is free from commercial sponsorship and other conflicts of interest.

Cochrane's mission is to promote evidence-informed health decision-making by producing high quality, relevant, accessible systematic reviews and other synthesized research evidence. The work of Cochrane is underpinned by a set of 10 key principles, listed in Box 1.1.a.

Box 1.1.a: The 10 principles of Cochrane

1. Collaboration	by fostering global co-operation, teamwork, and open and transparent communication and decision-making.
2. Building on the enthusiasm of individuals	by involving, supporting and training people of different skills and backgrounds.
3. Avoiding duplication of effort	by good management, co-ordination and effective internal communications to maximize economy of effort.
4. Minimizing bias	through a variety of approaches such as scientific rigour, ensuring broad participation, and avoiding conflicts of interest.
5. Keeping up-to-date	by a commitment to ensure that Cochrane Systematic Reviews are maintained through identification and incorporation of new evidence.
6. Striving for relevance	by promoting the assessment of health questions using outcomes that matter to people making choices in health and health care.
7. Promoting access	by wide dissemination of our outputs, taking advantage of strategic alliances, and by promoting appropriate access models and delivery solutions to meet the needs of users worldwide.
8. Ensuring quality	by applying advances in methodology, developing systems for quality improvement, and being open and responsive to criticism.
9. Continuity	by ensuring that responsibility for reviews, editorial processes and key functions is maintained and renewed.
10. Enabling wide participation	in our work by reducing barriers to contributing and by encouraging diversity.

1.1.2 A brief history of Cochrane

The Cochrane Collaboration was founded in 1993, a year after the establishment of the UK Cochrane Centre in Oxford, UK. The UK Cochrane Centre arose from a vision to extend a ground-breaking programme of work by Iain Chalmers and colleagues in the area of pregnancy and childbirth to the rest of health care. Inspired by Archie Cochrane's claim that "It is surely a great criticism of

our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials” (Cochrane 1979), Chalmers and colleagues developed the *Oxford Database of Perinatal Trials* and a series of systematic reviews published in *Effective Care in Pregnancy and Childbirth* (Chalmers 1989). The database became a regularly updated electronic publication in 1989, developed into *Cochrane Pregnancy and Childbirth Database* in early 1993, and formed the basis of the broader *Cochrane Database of Systematic Reviews (CDSR)*, launched in 1995. Work on a handbook to support authors of Cochrane Reviews had begun in 1993, and the first version was published in May 1994. Over its first 20 years, Cochrane has grown from an initial group of 77 people from nine countries who met at the first Cochrane Colloquium in Oxford in 1993 to over 31,000 contributors from more than 120 countries in 2015, making it the largest organization involved in this kind of work (Allen 2006; Allen 2007; Allen 2011). Cochrane is now an internationally renowned initiative (Clarke 2005; Green 2005).

1.1.3 Cochrane organization and structure

Cochrane currently involves over fifty Cochrane Review Groups (CRGs), responsible for supporting the production and publication of reviews within specific areas of health. The review authors working with these groups include researchers, health professionals and people using healthcare services (consumers), all of whom share a common enthusiasm for generating reliable, up-to-date evidence relevant to the prevention and treatment of specific health problems or groups of problems.

CRGs are supported in this work by Methods Groups, Centres, Fields and by the Cochrane Editorial Unit (CEU). Cochrane Methods Groups provide a forum for methodologists to discuss development, evaluation and application of methods used to conduct Cochrane Reviews. They play a major role in the production of the Cochrane Handbook for Systematic Reviews of Interventions and, where appropriate, chapters in this volume contain information about relevant Methods Groups. Members of these Methods Groups have made major contributions to systematic review methodology (Chandler 2013). Cochrane Centres are located in different countries. Collectively, they represent all regions of the world and provide training and support for review authors and CRGs in addition to advocacy and promotion of access to Cochrane Reviews. Cochrane Fields focus on broad dimensions of health, such as the setting of care (e.g. primary care), the type of consumer (e.g. children), or the type of intervention (e.g. vaccines). People associated with Fields help to ensure that priorities and perspectives in their sphere of interest reflect the work of CRGs. The CEU provides strategic support and direction, and leads initiatives to improve and assure the quality of review activity across Cochrane.

1.2 Systematic reviews

1.2.1 The need for systematic reviews

Healthcare providers, consumers, researchers, and policy makers are inundated with unmanageable amounts of information, including evidence from health research. It is unlikely that they will have the time, skills and resources to find, appraise and interpret all this evidence and to incorporate it into healthcare decisions. Cochrane Reviews respond to this challenge by identifying, appraising and synthesizing research-based evidence and presenting it in an accessible format (Mulrow 1994). The requirement for systematic reviews to appraise the ever-growing proliferation of individual research studies has, if anything, become more important in recent years (Mallett 2003; Bastian 2010).

1.2.2 What is a systematic review?

A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made (Antman 1992; Oxman 1993). The key characteristics of a systematic review are:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that meet the eligibility criteria;
- an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies.

Many systematic reviews contain meta-analyses. Meta-analysis is the use of statistical methods to summarize the results of independent studies (Glass 1976). By combining information from all relevant studies, meta-analyses can provide more precise

estimates of the effects of health care than those derived from the individual studies included within a review (see Chapter 9, Section 9.1.3). Meta-analyses facilitate investigations of the consistency of evidence across studies, and the exploration of differences across studies.

1.3 Cochrane Reviews

Cochrane has developed a rigorous approach to the preparation of systematic reviews, with a structured review model. Cochrane publishes four main types of systematic reviews, summarized in Sections 1.3.1 to 1.3.4 and has a programme to explore development of review methods for other types of research question.

1.3.1 Reviews of the effects of interventions

Most Cochrane Reviews consider evidence on the effects of health or healthcare interventions. These reviews focus primarily on randomized studies as the most robust research design for assessment of the effects of interventions. Where evidence is unlikely to be found in randomized studies, for example for many adverse effects of interventions, or for large-scale interventions such as in public health or organizational change, reviews include non-randomized studies. Intervention reviews may additionally address broader issues such as economic issues or patient experiences of the intervention.

Cochrane has recently developed quality standards for the conduct and reporting of reviews. These standards summarize attributes for the conduct, and reporting, of reviews of interventions as set out in this *Handbook* (see Chapter 2, Section 2.4).

1.3.2 Reviews of diagnostic test accuracy

Cochrane has published systematic reviews of diagnostic test accuracy (DTA) in CDSR since 2008 (Leeflang 2013). These reviews evaluate how correctly a test detects the presence or absence of a target condition. Cochrane DTA reviews cover target conditions across health, including both pathologically defined diseases and more loosely defined indications for which treatments may be available. All types of tests are eligible, including: signs and symptoms from the patient history and examination; questionnaire-based tools, scores and decision rules; laboratory tests including biochemical, immunological, genetic, genomic and other 'pan-omic' technologies; imaging tests; and physiological measurements. Evaluation of the accuracy of a test is one component of the assessment of whether test use could lead to improvement in patient outcomes. Direct evaluation of how a test (and consequent decision-making and interventions) actually affect patient outcomes is best assessed by randomized studies that incorporate the effects of interventions that follow the test result. Such studies fit within the structure of Cochrane Intervention Reviews. However, randomized studies of test use are rare (especially outside the context of screening; Ferrante di Ruffano 2012), whereas accuracy studies are relatively common and provide most of the available evidence to guide test use, which makes them worthy of detailed systematic review. Although the stages in a DTA review are the same as for reviews of interventions, specific methodological challenges are encountered at each step: from formulation of review questions, through searching for and locating studies, assessing study quality, meta-analysis and interpretation of findings. Full methodological details are described in a separate Cochrane Handbook for Diagnostic Test Accuracy Reviews (<http://srdta.cochrane.org/handbook-dta-reviews>).

1.3.3 Overviews of Reviews

Cochrane Overviews of Reviews (Overviews) compile evidence from multiple systematic reviews into a single accessible and usable document. They are intended primarily to synthesize multiple Cochrane Reviews addressing a set of related interventions, populations, outcomes, or conditions, although other published non-Cochrane reviews may also be included. Cochrane Overviews provide the reader with a quick and comprehensive guide to reviews relevant to a specific decision. Overviews are aimed at decision makers, such as clinicians, policy makers, or informed consumers, who are accessing the CDSR for evidence on a specific problem. Overviews of Reviews on the effects of interventions are addressed in detail in Chapter 22 (see Section 22.1).

An overview of systematic reviews of diagnostic test accuracy (DTA Overview) can be used to synthesize and compare findings from a related set of test accuracy reviews. For example, an overview might bring together and compare the findings of separate reviews of alternative tests used to diagnose the same condition at the same point in the patient pathway. DTA Overviews also have a role in evaluating the accuracy of tests for the detection of closely related target conditions (particularly when they form part of a set of differential diagnoses), and in evaluating the performance of the same test across different settings. DTA Overviews are best planned when commencing on a portfolio of related individual systematic reviews, and plans for incorporation in a DTA Overview should be mentioned in the protocols of the individual reviews.

1.3.4 Reviews of methodology

Cochrane Methodology Reviews seek to answer questions about various aspects of the methods for systematic reviews, randomized studies and other evaluations of health and social care. They provide an evidence base for the methods of these evaluations, as well as providing descriptive accounts of other relevant issues, for example, to show the scale of problems faced by researchers working on systematic reviews or making decisions about health and social care. Cochrane Methodology Reviews use the widest range of study designs of Cochrane Reviews, including:

- experimental studies such as randomized studies to compare different strategies to increase response rates to surveys;
- comparative observational studies to examine the relationship between the use of reporting guidelines and the quality of research reports; and
- descriptive observational studies of the proportion of studies presented at conferences that are also published in full.

Cochrane Methodology Reviews have a particular structure, based on the structure of Cochrane Intervention Reviews but with changes to some of the headings and sub-headings. The Cochrane Methodology Review Group has editorial responsibility for all Methodology Reviews. Appendix A provides a guide to the contents of a Cochrane Methodology protocol and review.

1.4 Publication of Cochrane Reviews

1.4.1 The Cochrane Library

Cochrane Reviews are published in full online in the *CDSR*, which is a core component of the Cochrane Library (www.thecochranelibrary.com). The Cochrane Library was first published in 1996, and is now an online collection of six databases (listed in 1.4.a) published by Wiley-Blackwell. In addition to the *CDSR*, the Cochrane Library includes additional resources that are provided by the Centre for Reviews and Dissemination (CRD) in York, UK. It is available free at the point of use in some countries, thanks to national licences and free one-click access provided by Wiley-Blackwell and Cochrane in most low- and middle-income countries, in association with Evidence Aid. Elsewhere it is subscription based, or pay-per-view. Since February 2013, reviews that have been published in full, or updated in full for the first time, now become freely available to all 12 months after their initial publication under an open access model.

Box 1.4.a: Databases published in the Cochrane Library

Active databases

The CDSR contains the full text (including methods, results and conclusions) for Cochrane Reviews and protocols.

The Cochrane Central Register of Controlled Trials (CENTRAL) is a highly concentrated source of reports of randomized and quasi-randomized studies. The majority of CENTRAL records are taken from bibliographic databases (mainly MEDLINE and Embase), but records are also derived from other published and unpublished sources.

The Health Technology Assessment database contains details of completed and ongoing health technology assessments (studies of the medical, social, ethical, and economic implications of healthcare interventions). It is produced by CRD, using information obtained from members of International Network of Agencies for Health Technology Assessment (INAHTA) and other health technology assessment organizations.

Archived databases

The Database of Abstracts of Reviews of Effects (DARE), assembled and previously maintained by CRD, contains critical assessments and structured abstracts of other systematic reviews, conforming to explicit quality criteria. This database was archived in March 2015.

The Cochrane Methodology Register (CMR) contains bibliographic information on articles and books on the science of reviewing research, and a prospective register of methodological studies. This database was archived in July 2012.

NHS Economic Evaluation Database (EED) contains appraised economic evaluations highlighting their relative strengths and weaknesses. It was produced by CRD. This database was archived in March 2015.

1.5 Handbook structure

There are three parts to the Handbook. Part 1 provides general information on Cochrane, its principles and the specific structure of Cochrane Reviews, their preparation, reporting, publication and maintenance. Part 2 provides the requisite methods to conduct a review with the required minimum standards. Part 3 covers a range of special topics for consideration when undertaking a Cochrane Review.

1.6 Chapter information

Acknowledgements: We thank previous chapter authors Sally Green, Philip Alderson, Cynthia Mulrow and Andrew Oxman on whose text this version is based. We also thank Ruth Foxlee for her contribution to 1.4.a.

1.7 References

Allen 2006

Allen C, Clarke M. International activity in Cochrane Review Groups with particular reference to China. *Chinese Journal of Evidence-based Medicine* 2006; 6: 541-545.

Allen 2007

Allen C, Clarke M, Tharyan P. International activity in Cochrane Review Groups with particular reference to India. *National Medical Journal of India* 2007; 20: 250-255.

Allen 2011

Allen C, Richmond K. The Cochrane Collaboration: international activity within Cochrane Review Groups in the first decade of the twenty-first century. *Journal of Evidence Based Medicine* 2011; 4: 2-7.

Antman 1992

Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. *JAMA* 1992; 268: 240-248.

Bastian 2010

Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: How will we keep up? *PLoS Medicine* 2010; 7: e1000326.

Chalmers 1989

Chalmers I, Enkin M, Keirse MJNC. *Effective care in pregnancy and childbirth*. Oxford (UK): Oxford University Press.

Chandler 2013

Chandler J, Hopewell S. Cochrane methods - twenty years experience in developing systematic review methods. *Systematic Reviews* 2013; 2: 76.

Clarke 2005

Clarke M. Cochrane Collaboration. In: Armitage P, Colton T, editor(s). *Encyclopedia of Biostatistics*. 2nd edition. Chichester (UK): John Wiley & Sons, 259-262.

Cochrane 1979

Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. In: Teeling-Smith G, Wells N, editor(s). *Medicines for the year 2000*. London (UK): Office of Health Economics, 1-11.

Ferrante di Ruffano 2012

Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *Journal of Clinical Epidemiology* 2012; 65: 282-287.

Glass 1976

Glass GV. Primary, secondary and meta-analysis of research. *Educational Researcher* 1976; 5: 3-8.

Green 2005

Green S, McDonald S. The Cochrane Collaboration: More than systematic reviews? *Internal Medicine Journal* 2005; 35: 4-5.

Leeflang 2013

Leeflang MMG, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. *Systematic Reviews* 2013; 2: 82.

Mallett 2003

Mallett S, Clarke M. How many Cochrane Reviews are needed to cover existing evidence on the effects of healthcare interventions? *Evidence Based Medicine* 2003; 8: 100-101.

Mulrow 1994

Mulrow CD. Rationale for systematic reviews. *BMJ* 1994; 309: 597-599.

Oxman 1993

Oxman AD, Guyatt GH. The science of reviewing research. *Annals of the New York Academy of Sciences* 1993; 703: 125-133.

Chapter 3: Reporting the review

To cite this draft chapter, please use:

Page MJ, Cumpston M, Chandler J, Lasserson T. Chapter III: Reporting the review. Draft version (2 November 2018) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston MS, Li T, Page MJ, Welch V (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane.

Copyright © 2018 The Cochrane Collaboration.

Key points

- Clear reporting of a systematic review allows readers to evaluate the rigour of the methods applied, and to interpret the findings appropriately. Transparency can facilitate attempts to verify or reproduce the results, and make the review more usable for health care decision makers.
- The target audience for Cochrane Reviews is people making decisions about health care, including healthcare professionals, consumers and policy makers. Cochrane Reviews should be written so that they are easy to read and understand by someone with a basic sense of the topic who may not necessarily be an expert in the area.
- Methodological Expectations of Cochrane Intervention Reviews (MECIR) are recommendations available to guide the conduct and reporting of new review protocols, new reviews, and updates of reviews of interventions.
- Guidance on the composition of plain language summaries of Cochrane Intervention Reviews is also available to help review authors specify the key messages in terms that are accessible to consumers and non-expert readers.

- Review authors should ensure that reporting of objectives, important outcomes, results, caveats and conclusions is consistent across the main text, the abstract, and any other summary versions of the review (e.g. plain language summary).

3.1 Introduction

The effort of undertaking a systematic review is wasted if review authors do not report clearly what they did, and what they found (Glasziou et al 2014). Clear reporting enables others to replicate the methods used in the review, which can facilitate attempts to verify or reproduce the results (Page et al 2018). Transparency can also make the review more usable for health care decision makers. For example, clearly describing the interventions assigned in the included studies can help users determine how best to deliver effective interventions in practice (Hoffmann et al 2017). Also, comprehensively describing the eligibility criteria applied, sources consulted, analyses conducted, and post-hoc decisions made, can reduce uncertainties in assessments of risk of bias in the review findings (Whiting et al 2016). For these reasons, transparent reporting is an essential component of all systematic reviews.

Surveys of the transparency of published systematic reviews suggest that many elements of systematic reviews could be reported better. For example, Page and colleagues evaluated a random sample of 300 systematic reviews of biomedical research indexed in MEDLINE in February 2014 (Page et al 2016). They found that in at least a third of the reviews there was no information on eligible publication types, the years of coverage of the search, the methods used to collect data and appraise studies, or the funding source of the review. However, Cochrane Reviews, which accounted for 15% of the sample, had more complete reporting than all other types of systematic reviews (Page et al 2016).

Possible reasons why more complete reporting of Cochrane Reviews has been observed include the use of software (RevMan, <https://revman.cochrane.org/>) and strategies in the editorial process that promote good reporting. RevMan includes many standard headings and subheadings which are designed to prompt Cochrane Review authors to document their methods and results clearly. In addition, the Methodological Expectations of Cochrane Intervention Reviews (MECIR) are recommendations available to guide the conduct and reporting of these reviews (Higgins JPT et al 2018).

The MECIR guidelines were developed in consultation with review authors, editors and methodologists from the Cochrane community, and form the basis of quality assurance work undertaken by the Cochrane Editorial and Methods Department. They cover both conduct and reporting for new review protocols, new reviews, and updates of reviews of interventions. The guidelines distinguish between conduct and reporting for good reason: good conduct does not necessarily lead to good reporting, good reporting cannot improve poor conduct, and poor reporting can obscure good or poor conduct of a review.

The MECIR reporting guidance is consistent with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2009 statement (Liberati et al 2009, Moher et al 2009). Review authors and Cochrane Review Groups are expected to follow the guidance developed specifically for Cochrane Reviews, in conjunction with PRISMA should that prove to be helpful. Reporting guidance is currently being developed for the methods and reporting of syntheses of quantitative data where meta-analysis was not possible or appropriate (the ICONS-Quant reporting guideline) (Campbell et al 2018). Review authors are advised to consult such guidance alongside MECIR once it is available.

Guidance on the composition of plain language summaries of Cochrane Reviews of interventions is also available. The guidance outlines the key messages from Cochrane Reviews that should be included in a plain language summary, in terms that are accessible to consumers and non-expert readers.

The structure of this chapter is built around the MECIR reporting guidance for new Cochrane Review protocols (Section III.2) and new Cochrane Reviews (Section III.3) of interventions, and guidance for reporting plain language summaries (Section III.4). The MECIR expectations of conduct are embedded in the relevant Handbook chapters. MECIR conduct and reporting guidance for updates of Cochrane Reviews of interventions are presented in Chapter IV. For the latest version of all MECIR conduct and reporting guidance, readers should consult the MECIR web pages, available at <https://methods.cochrane.org/mecir>.

Many of the standard headings recommended for use in Cochrane Reviews are referred to in this chapter, although the precise headings available in RevMan may be amended as new versions are released. New headings can be added and some standard headings can be deactivated; if the latter is done, review authors should ensure that all information expected (as outlined in the MECIR reporting guidelines) is still reported somewhere in the review.

3.2 Reporting of protocols of new Cochrane Intervention Reviews

Preparing a well-written review protocol is important for many reasons (see Chapter 1). The protocol is a public record of the

question of interest and the intended methods before results of the studies are fully known. This helps readers to judge how the eligibility criteria of the review, stated outcomes and planned methods will address the intended question of interest. It also helps anyone who evaluates the completed review to judge how far it fulfilled its original objectives (Lasserson T et al 2016). Investing effort in the development of the review question and planning of methods also stimulates review authors to anticipate methodological challenges that may arise, and helps minimise potential for non-reporting biases by encouraging review authors to publish their review and report results for all pre-specified outcomes (Shamseer et al 2015).

See the online [MECIR Manual](#) for the 44 MECIR reporting items for protocols of new Cochrane Intervention Reviews. They include guidance for reporting of the:

- Background;
- Objectives;
- Criteria for considering studies for inclusion in the review;
- Search methods for identification of studies (e.g. a list of all sources that will be searched, a complete search strategy to be implemented for at least one database);
- Data collection and analysis (e.g. types of information that will be sought from reports of included studies and methods for obtaining such information, how risk of bias in included studies will be assessed, and any intended statistical methods for combining results across studies);
- Other information (e.g. acknowledgments, contributions of authors, declarations of interest, and sources of support).

These sections correspond to the same sections in a completed review, and further details are outlined in Sections III.3.2, III.3.3 and III.3.7.

One key difference between a review protocol and completed review is that the 'Methods' section in a protocol should be written in the future tense. Because Cochrane Reviews are updated as new evidence accumulates, methods outlined in the protocol should generally be written as if a suitably large number of studies will be identified to allow the objectives to be met (even if this is assumed likely not to be the case at the time of writing).

It is important to recognize that the MECIR guidelines reflect the minimum expectations for good reporting of a review protocol. Further guidance on the level of planning required for each aspect of the review methods and the detailed information recommended for inclusion in the protocol is given throughout this Handbook in the relevant chapters.

3.3 Reporting of new Cochrane Interventions Reviews

The main text of a Cochrane Review should be succinct and readable. Although there is no formal word limit for Cochrane Reviews, review authors should consider 10,000 words an absolute maximum for the main text of the review unless there is a special reason to write a longer review, such as when the question is unusually broad or complex. Most reviews should be substantially shorter.

People making decisions about health care are the target audience for Cochrane Reviews. This includes healthcare professionals, consumers and policy makers with a basic understanding of the underlying disease or problem, and reviews should be accessible to these audiences. Cochrane Reviews should be written so that they are easy to read and understand by someone with a basic sense of the topic who is not necessarily an expert in the area. Some explanation of terms and concepts is likely to be helpful, and perhaps even essential. However, too much explanation can detract from the readability of a review. Simplicity and clarity are also vital to readability. The readability of Cochrane Reviews should compare to that of a well written article in a general medical journal.

Review authors should ensure that reporting of objectives, important outcomes, results, caveats and conclusions is consistent across the main text, the tables and figures, the abstract, and any other summary versions of the review (e.g. 'Summary of findings' table and plain language summary). Although this sounds simple, it can be challenging in practice; authors should review their text carefully to ensure that readers of a summary version are likely to come away with the same overall understanding of the conclusions of the review as readers accessing the full text.

Plagiarism is not acceptable and all sources of information need to be cited. Also, the unattributed reproduction of text from other sources must be avoided. Quotes from other published or unpublished sources need to be indicated and attributed clearly, and permission may be required before reproducing any published figures.

See the online [MECIR Manual](#) for all MECIR reporting items for new Cochrane Intervention Reviews. In the remainder of this section we summarize the reporting guidance relating to different sections of a Cochrane Review.

3.3.1 Abstracts

All reviews must include an abstract of not more than 1,000 words, although in the interests of brevity, authors should aim to include no more than 700 words, without sacrificing important content. Abstracts should be targeted primarily at healthcare decision makers (clinicians, consumers and policy makers) rather than just to researchers. Terminology should be reasonably comprehensible to a general rather than a specialist healthcare audience. Abbreviations should be avoided, except where they are widely understood (e.g. HIV). Where essential, other abbreviations should be spelt out (with the abbreviations in brackets) on first use. Names of drugs and interventions that can be understood internationally should be used wherever possible. Trade or brand names should not be used and generic names are preferred.

Abstracts of Cochrane Reviews are made freely available on the internet and published in bibliographic databases that index the Cochrane Database of Systematic Reviews (e.g. MEDLINE, Embase). However, some readers may be unable to access the full review, or the full text may not have been translated into their language, so abstracts may be the only source they have to understand the review results (Beller et al 2013). It is important therefore that they can be read as stand-alone documents. The abstract should summarize the key methods, results and conclusions of the review and should not contain any information that is not in the review.

The content of a Cochrane Review abstract should include:

- Background (a summary of the rationale and context of the review);
- Objectives of the review;
- Search methods (including an indication of databases searched, and the date of the last search from which records were evaluated);
- Selection criteria (including a summary of eligibility criteria for study designs, participants, interventions and comparators);
- Data collection and analysis (including a summary of any noteworthy methods for selecting studies, collecting data, evaluating risk of bias and synthesizing results; for many reviews it may be sufficient to state “We used standard methodological procedures expected by Cochrane”);
- Main results (including the findings of all important benefit and harm outcomes, irrespective of the statistical significance, magnitude or direction of the result, along with GRADE ratings);
- Author’s conclusions (including both implications for practice and research).

See the online [MECIR Manual](#) for reporting guidance for the abstract of a Cochrane Intervention Review.

3.3.2 Background and Objectives

Well-formulated review questions occur in the context of an already-formed body of knowledge. The ‘Background’ section should address this context, including a description of the condition or problem of interest. It should help clarify the rationale for the review, and explain why the questions being addressed are important. It should be concise (generally around one page when typeset printed) and be understandable to the users of the intervention(s) under investigation.

It is important that the eligibility criteria and other aspects of the methods build on ideas that have been developed in the ‘Background’ section. For example, if there are uncertainties in how variation in setting, dose of intervention or timing of outcome assessment influence the intervention effect, then it would be important to acknowledge them as a reason for doing the review and consider how the relevant aspects of the methods have been designed to identify relevant evidence and explore these uncertainties.

There are four standard subheadings in the ‘Background’ section of a Cochrane Review that are intended to facilitate a structured approach to the context and overall rationale for the review:

- **Description of the condition:** A brief description of the condition being addressed and its significance is a useful way to begin the review. It may include information about the biology, diagnosis, prognosis, prevalence, incidence and burden of the condition.
- **Description of the intervention:** A description of the experimental intervention(s) should place it in the context of any standard or alternative interventions, remembering that standard practice may vary widely according to context. The role of the comparator intervention(s) in standard practice should also be made clear. For drugs, basic information on clinical pharmacology should be presented where available, such as dose range, metabolism, selective effects, half-life, duration and any known interactions with other drugs. For more complex interventions, such as behavioural or service-level interventions, a description of the main components should be provided (see Chapter 17).
- **How the intervention might work:** This section should provide theoretical reasoning as to why the interventions under

review may have an impact on potential recipients, for example, by relating a drug intervention to the biology of the condition. Authors may refer to a body of empirical evidence such as similar interventions having an impact on the target recipients or identical interventions having an impact on other populations. Authors may also refer to a body of literature that justifies the possibility of effectiveness. For reviews of complex interventions, a logic model (Kneale et al 2015) or conceptual framework may be useful to illustrate the proposed mechanism of action of the intervention and its components. This will also provide review authors with a framework for the methods and analyses undertaken throughout the review to ensure that the review question is clearly and appropriately addressed. More guidance on the conduct of reviews of complex interventions is presented in Chapter 17.

- **Why it is important to do this review:** Review authors should explain clearly why the questions being asked are important. Rather than justifying the review on the grounds that there are known eligible studies, it is more helpful to emphasize what aspects of or uncertainties in the accumulating evidence base now justify a systematic review. For example, it might be the case that studies have reached conflicting conclusions, that there is debate about the evidence to date, or that there are competing approaches to implementing the intervention.

Immediately following the background section of the review, review authors should declare the review objectives. They should begin with a precise statement of the primary objective of the review, ideally in a single sentence. Where possible the style should be of the form “To assess the effects of [intervention or comparison] for [health problem] for/in [types of people, disease or problem and setting if specified]”. This might be followed by a series of secondary objectives relating to different participant groups, different comparisons of interventions or different outcome measures. If relevant, any objectives relating to the evaluation of economic or qualitative evidence should be stated. It is not necessary to state specific hypotheses.

See the online [MECIR Manual](#) for reporting guidance relevant to the ‘Background’ and ‘Objectives’ sections of a Cochrane Review.

3.3.3 Methods

The ‘Methods’ section in a completed review should be written in the past tense, and should describe what was done to obtain the results and conclusions of the current review. Authors should describe the use of methods that are in accordance with the MECIR conduct guidance.

Review authors are expected to cite their protocol to make it clear that there was one. Often a review is unable to implement all of the methods outlined in the protocol. For example, planned investigations of heterogeneity (e.g. subgroup analyses) and small-study effects may not have been conducted because of an insufficient number of studies. In such circumstances, we recommend that the methods that were not implemented be removed from the main ‘Methods’ section and outlined in the section headed ‘Differences between protocol and review’ or in an Appendix. A description of the methods not implemented can serve as a protocol for future updates of the review.

The ‘Methods’ section of a Cochrane Intervention Review includes three main subsections, within which are a series of standard headings to guide authors in reporting all the relevant information. See Sections III.3.3.1, III.3.3.2 and III.3.3.3 for a summary of content recommended for inclusion under each subheading.

3.3.3.1 Criteria for considering studies for this review

Review authors should declare all criteria used to help decide whether or not to include a study in the review. Doing so will help readers understand the scope of the review and recognize why particular studies they are aware of were not included. Eligible study designs should be described, with a focus on specific features of a study’s design rather than design labels (e.g. how groups were formed, whether the intervention was assigned to individuals or clusters of individuals) (Reeves et al 2017). Review authors should describe eligibility criteria for participants, including any restrictions based on age, diagnostic criteria, location and setting. If relevant, it is useful to describe how studies including a subset of relevant participants were addressed (e.g. when children up to the age of 16 years only were eligible but a study included children up to the age of 18 years). Eligibility criteria for interventions and comparators should be stated also, including any criteria around delivery, dose, duration, intensity, co-interventions and characteristics of complex interventions.

Review authors should specify the primary and secondary outcomes of interest to the review. The review’s primary outcomes should normally reflect at least one potential benefit and at least one potential harm. Additional information about outcomes of interest is helpful to include, including a description of how multiple variants of outcome measures (e.g. definitions, assessors, scales, time points) were addressed. Typically, studies should not be excluded from a review solely because no outcomes of interest were reported, because failure to report an outcome does not mean it was not assessed (Dwan et al 2017). However, on occasion it will be appropriate to include only studies that measured particular outcomes. For example, a review of a multi-

component public health intervention promoting healthy lifestyle choices, focussing on reduction in smoking prevalence, might legitimately exclude studies that do not measure smoking rates. Review authors should specify if measurement of a particular outcome was used as an eligibility criterion for the review, and justify why this was done.

See the online [MECIR Manual](#) for reporting guidance relevant to the eligibility criteria for the review. Further guidance on planning eligibility criteria is presented in Chapter 3.

3.3.3.2 Search methods for identification of studies

It is essential that users of systematic reviews are given an opportunity to evaluate the methods used to identify studies for inclusion. Such an evaluation is possible when review authors report their search methods comprehensively. This involves specifying all sources consulted, including databases, trials registers, websites, and a list of individuals or organizations contacted. If particular journals were hand searched, this should be noted, but it is not necessary to describe hand searching done routinely to populate a Cochrane Specialized Register. Specifying the dates of coverage of all databases searched and the date of the last search can help users determine how up to date the review is. Review authors should also declare any limits placed on the search (e.g. by language, publication date or publication format).

To facilitate replication of a search, review authors should include in an Appendix the exact search strategy (or strategies) used for each database, including any limits and filters used. Search strategies can be exported from bibliographic databases, and these should be copied and pasted in lieu of re-typing each line, which can introduce errors.

See the online [MECIR Manual](#) for reporting guidance relevant to the search methods used to identify studies, and refer to Chapter 4 for conduct guidance on search methods.

3.3.3.3 Data collection and analysis

Cochrane Intervention Reviews include several standard subheadings to enable a structured, detailed description of the methods used for data collection and analysis. Additional headings should be included where appropriate to describe additional methods implemented in the review, e.g. those specific to the analysis of qualitative or economic evidence. See the online [MECIR Manual](#) for guidance relevant to the reporting of data collection and analysis methods.

Selection of studies: There should be a description of how the eligibility criteria were applied, from screening of search results through to the final selection of studies for inclusion in the review. The number of people involved at each stage of the process should be stated, such as two authors working independently, along with an indication of how any disagreements were resolved. See Chapter 4 for conduct guidance on the study selection process.

Data extraction and management: Review authors should specify how data were collected from reports of included studies. This includes describing the number of people involved, whether they worked independently, how any disagreements were resolved, and whether standardized data collection forms were used (and if so, whether they were piloted in advance). A brief description of the data items (e.g. participant characteristics, intervention details) extracted from each report is recommended. If study authors or sponsors were contacted to obtain missing information or to clarify the information available, this should be stated. If methods for transforming or processing data in preparation for analysis were necessary (e.g. converting standard errors to standard deviations, extracting numeric data from graphs), these methods should be described. See Chapter 5 for conduct guidance on data collection.

Assessment of risk of bias in included studies: There should be a description of the approach used to assess risk of bias in the included studies. This involves specifying the risk of bias tool(s) used, how many authors were involved in the assessment, and how the assessments were incorporated into the analysis or interpretation of the results. Cochrane Review authors are expected to use the tools recommended by Cochrane (described in Chapter 8 and Chapter 25). If an existing risk of bias tool was modified for the purposes for the review, the modifications should be described and justified. See Chapter 7 for conduct guidance on study risk of bias assessment.

Measures of treatment effect: The effect measures used by the review authors to describe results in any included studies or meta-analyses (or both) should be stated. Examples of effect measures include the odds ratio (OR), risk ratio (RR) or risk difference (RD) for dichotomous data, mean difference (MD) or standardized mean difference (SMD) for continuous data, and appropriate measures for other anticipated outcome types (e.g. hazard ratio for time-to-event data). See Chapter 6 for more guidance on effect measures.

Unit of analysis issues: If the review includes study designs that can give rise to a unit-of-analysis error (when the number of

observations in an analysis does not match the number of 'units' randomized), the approaches taken to address these issues should be described. Studies that can give rise to unit-of-analysis errors include cross-over trials, cluster-randomized trials, and studies where interventions are assigned to multiple parts of the body of the same participant. See Chapter 23 for guidance on handling unit of analysis issues.

Dealing with missing data: Review authors may encounter various types of missing data in their review. For example, there may be missing information about the methods of the included studies (e.g. when the method of randomization is not reported), or missing statistics (e.g. when standard deviations of mean scores are not reported). Strategies to deal with such missing data should be reported. This may include attempts to obtain the missing data, and approaches to the analysis and interpretation of results in light of missing data (e.g. imputing missing standard deviations). See Chapter 5 for conduct guidance on dealing with missing data.

Assessment of heterogeneity: Review authors should describe their approach to identifying statistical heterogeneity (e.g. non-quantitative assessment, I^2 , τ^2 , or statistical test). See Chapter 10 for conduct guidance on assessment of heterogeneity.

Assessment of non-reporting biases: Any methods used to assess the risk of non-reporting biases in a synthesis should be described. Such methods may include consideration of the number of studies missing from a synthesis due to selective non-reporting of results, or investigations to assess small-study effects (e.g. funnel plots), which can arise from the suppression of small studies with 'negative' results (amongst other reasons). See Chapter 13 for a description of methods for assessing risk of non-reporting biases in a synthesis.

Data synthesis: Review authors should describe any methods used for combining results across studies (e.g. meta-analysis, network meta-analysis). Where data have been combined in statistical software external to RevMan, authors should reference the software, commands and settings used to run the analysis. If relevant, other synthesis methods used when meta-analysis was not possible or appropriate should be described. See Chapter 10 for guidance on undertaking meta-analysis, Chapter 11 for guidance on undertaking network meta-analysis, and Chapter 12 for a description of other synthesis methods.

Subgroup analysis and investigations of heterogeneity: If subgroup analyses (or meta-regression) were performed, review authors should specify the potential effect modifiers explored, the rationale for each, whether they were identified before or after the results were known, and how they were compared (e.g. using a statistical test for interaction). See Chapter 10 for more information on investigating heterogeneity.

Sensitivity analyses: If any sensitivity analyses were performed to explore the robustness of meta-analysis results, review authors should specify the basis of each analysis (e.g. removal of studies at high risk of bias, imputing alternative estimates of missing standard deviations). See Chapter 10 for more information on sensitivity analyses.

Summarizing findings and assessing certainty of the evidence: Review authors should describe any methods for summarizing the findings of the review, and assessing the certainty of the body of evidence for each main outcome (e.g. using the GRADE approach). If review authors used an alternative to the GRADE approach to assess certainty of the body of evidence, or deviated from standard GRADE methods, they should say so and provide a rationale. Review authors should also indicate which populations, interventions, comparisons and outcomes are addressed in 'Summary of findings' tables. For more details on completing 'Summary of findings' tables and using the GRADE approach, see Chapter 14.

3.3.4 Results

A narrative summary of the results of a Cochrane Intervention Review should be provided under the three standard subheadings in the 'Results' section (see Sections [III.3.4.1](#), [III.3.4.2](#) and [III.3.4.3](#) for a summary of content recommended for inclusion under each subheading). Details about the effects of interventions (including summary statistics and effect estimates for each included study and for meta-analyses) can be presented in various tables and figures (see Section [III.3.4.4](#)).

III.3.4.1 Description of studies

The results section should start with a summary of the results of the search (for example, how many references were retrieved by the electronic searches, how many were considered as potentially eligible after screening, and how many were included). Review authors are encouraged to include a PRISMA-type flow diagram demonstrating the flow of studies throughout the selection process (Moher et al 2009). Such flow diagrams can be created within RevMan.

To help readers determine the applicability of the review findings, review authors should describe the characteristics of the studies included in the review. In the 'Results' section, a brief narrative summary of the included studies should be provided (by specifying the number of participants and summarizing characteristics of the study populations and settings, interventions, comparators, outcomes and funding sources). More details about each included study should be presented in the 'Characteristics of included

studies' table. This table should include (at a minimum) the following information about each included study:

- basic study design or design features;
- baseline demographics of the study sample (e.g. age, sex/gender);
- sample size;
- details of all interventions (including what was delivered, by whom, in which setting, and how often; for more guidance see (Hoffmann et al 2017));
- outcomes measured (with details on how and when they were measured);
- funding source;
- declarations of interest among the primary researchers.

Certain studies that may appear to some readers to meet the eligibility criteria, but which were excluded, should be listed in the 'Characteristics of excluded studies' table, and the reason for exclusion should be provided (one reason is usually sufficient). It is not necessary to include every study excluded at the full text screening stage in the table; rather, authors should use their judgement to identify those studies most likely to be considered eligible by readers, and hence most useful to include here. A succinct summary of the reasons why studies were excluded from the review should be provided in the 'Results' section.

It is helpful to make readers aware of any completed studies that have been identified as potentially eligible but have not been incorporated into the review. This may occur when there is insufficient information to determine whether the study meets the inclusion criteria of the review, or when an updated search is run immediately prior to publication and the review authors consider it unlikely that inclusion of the study would change the review conclusions substantially. A description of such studies can be provided in the 'Characteristics of studies awaiting classification' table. Readers should also be made aware of any studies that meet the eligibility criteria for the review, but which are still in progress and hence have no results available. This serves several purposes. It will help readers assess the stability of the review findings, alert research funders about ongoing research activity, and can serve as a useful basis for deciding when an update of the review may be needed. A description of such studies can be provided in the 'Characteristics of ongoing studies' table.

See the online [MECIR Manual](#) for reporting guidance relevant to the description of studies.

III.3.4.2 Risk of bias in included studies

To help readers determine the trustworthiness of the results of included studies, review authors should present and summarize their risk of bias assessments (see the online MECIR Manual for relevant reporting guidance). A 'Risk of bias' table for each included study should be presented, indicating judgements about risk of bias for each result assessed, along with explicit support for these judgements. Forest plots created in RevMan can present the risk of bias judgements relating to each included study. The current subheadings in this section of the review are by domain, but review authors should also provide in the Results section a narrative summary of the risks of bias among results contributing to key outcomes of the review.

III.3.4.3 Effects of interventions

There are 24 MECIR items relevant to the reporting of effects of interventions (see the online [MECIR Manual](#)). We provide a summary of them in this and the following section.

Review authors should summarize in text form the results for all pre-specified review outcomes, regardless of the statistical significance, magnitude or direction of the effects, or whether evidence was found for those outcomes. The text should present the results in a logical and systematic way. This can be done by organizing results by population or comparison (e.g. by first describing results for the comparison of drug versus placebo, then describing results for the comparison of drug A versus drug B), and retaining a distinction between primary and secondary outcomes.

If meta-analysis was possible, synthesized results should always be accompanied by a measure of statistical uncertainty, such as a 95% confidence interval. It is helpful to also indicate the amount of information (numbers of studies and participants) contributing to each meta-analysis. If no data were available for particular review outcomes of interest, review authors should say so, so that all pre-specified outcomes are accounted for. Guidance on summarizing results from meta-analysis is provided in Chapter 10, for results when meta-analysis was not possible or appropriate is provided in Chapter 12.

It is important that the results of the review are presented in a manner that ensures the reader can interpret the findings accurately. The direction of effect (increase or decrease, benefit or harm), should always be clear to the reader, and the minimal important difference in the outcome (if known) should be specified. Review authors should consider presenting results in formats that are easy to interpret. For example, standardized mean differences are difficult to interpret because they are in units of standard deviation, but can be re-expressed in more accessible formats (see Chapter 15). A common mistake to avoid is the confusion of 'no evidence of an effect' with 'evidence of no effect'. When there is inconclusive evidence, it is wrong to claim that it shows that an intervention has

'no effect' or is 'no different' from the control intervention. In this situation, it is better to report the data, with a confidence interval, as being uncertain, for example when the confidence interval is compatible with either a reduction or an increase in the outcome, or with a negligible difference.

In addition to summarizing the effects of interventions, review authors should also summarize the results of any subgroup analyses (or meta-regression), sensitivity analyses, and assessments of the risk of non-reporting bias (if performed) that are relevant to each synthesis. A common issue in reporting the results of subgroup analyses that should be avoided is the misleading emphasis placed on the intervention effects within subgroups without reference to the between-subgroup difference.

A 'Summary of findings' table is a useful means of presenting findings for the most important outcomes, whether or not evidence is available for them. A 'Summary of findings' table typically:

- includes results for one clearly defined population group;
- indicates the intervention and the comparator;
- includes seven or fewer patient-important outcomes;
- describes the characteristics of the outcomes (e.g. scale, scores, follow-up);
- indicates the number of participants and studies for each outcome;
- presents at least one baseline risk for each dichotomous outcome (e.g. study population or median/medium risk) and baseline scores for continuous outcomes (if appropriate);
- summarizes the intervention effect (if appropriate), and;
- includes an assessment of the certainty of the body of evidence for each outcome.

The assessment of the certainty of the body of evidence should follow the GRADE approach, which includes considerations of risk of bias, indirectness, inconsistency, imprecision and publication bias (see Chapter 14).

III.3.4.4 Presenting results of studies and syntheses in tables and figures

Simple summary data for each intervention group (such as means and standard deviations), as well as estimates of effect (such as mean differences), should be presented for each study for each outcome of interest to the review. This is achieved primarily by using the 'Data and analyses' section of the review. The 'Data and analyses' section has a hierarchical structure, presenting results in forest plots or other table formats, grouped first by comparison, and then for each outcome assessed within the comparison. Authors can also record in each table the source of all results presented, in particular, whether it was obtained from published literature, by correspondence, from a trials register, or from another source (e.g. clinical study report). Presenting such information facilitates attempts by others to verify or reproduce the results (Page et al 2018).

Forest plots display effect estimates and confidence intervals for each individual study and the meta-analysis (Lewis and Clarke 2001). Forest plots created in RevMan typically illustrate:

1. the summary statistics (e.g. number of events and sample size of each group for dichotomous outcomes) for each study;
2. point estimates and confidence intervals for each study, both in numeric and graphic format;
3. a point estimate and confidence interval for the meta-analytic effect, both in numeric and graphic format;
4. the total numbers of participants in the experimental and control groups;
5. labels indicating the interventions being compared and the direction of effect;
6. percentage weights assigned to each study;
7. estimates of heterogeneity (e.g. Tau²) and inconsistency (I²);
8. a statistical test for the meta-analytic effect.

Review authors should present the results of single studies in the review with the expectation that results of additional studies will be added in future when they become available.

If meta-analysis was not possible or appropriate, review authors might consider presenting alternative figures to present the results of included studies. These may include a harvest plot (Ogilvie et al 2008), effect direction plot (Thomson and Thomas 2013) or albatross plot (Harrison et al 2017) (see Chapter 12 for more details). Such plots may be produced in software other than RevMan and included as an 'Additional figure' in a Cochrane Review.

Authors must ensure that all statistical results presented in the main review text are consistent between the text and tables or figures.

3.3.5 Discussion

A structured discussion can help readers consider the implications of the review findings. The 'Discussion' subheadings in RevMan provide the structure for this section of the review.

Summary of main results: It is useful to provide a concise description of results for the main outcomes of the review, but this should not simply repeat text provided elsewhere. If the review has a number of comparisons this section should focus on those that are most prominent in the review, and that address the main review objectives.

Overall completeness and applicability: This section should present an assessment of how well the evidence identified in the review addressed the review question. It should indicate whether the studies identified were sufficient to address all of the objectives of the review, and whether all relevant types of participants, interventions and outcomes have been investigated.

Certainty of the evidence: Review authors should summarize the considerations that led to downgrading or upgrading the certainty of the evidence in their implementation of GRADE. This information can be based on explanations for downgrading decisions alongside the 'Summary of findings' tables in the review. Note that in the current version of RevMan this subheading defaults to 'Quality of the evidence'.

Potential biases in the review process: It is important for review authors to reflect on and report any decisions they made that might have introduced bias into the review findings. For example, rather than emphasizing the comprehensiveness of the search for studies, review authors should consider whether any aspects of the design or execution of the search could have led to studies being missed. This might occur because of the complexity and low specificity of the search, because the indexing of studies in the area is poor, or because searches beyond bibliographic databases did not occur. If attempts to obtain relevant data were not successful, this should be stated. Additional limitations to consider include contestable decisions relating to the inclusion or exclusion of studies, or synthesis of study results. For example, review authors may have decided to exclude particular studies from a synthesis because of uncertainty about the precise details of the interventions delivered, or measurement instrument used. If data were imputed and alternative approaches to achieve this could have been undertaken, this might also be acknowledged. It may be helpful to consider tools that have been designed to assess the risk of bias in systematic reviews (such as the ROBIS tool (Whiting et al 2016)) when writing this section.

Agreements and disagreements with other studies or reviews: Review authors should also discuss the extent to which the findings of the current review agree or disagree with those of other reviews. Authors could briefly summarize the conclusions of previous reviews addressing the same question, and if the conclusions contrast with their own, discuss why this may have occurred (e.g. because of differences in eligibility criteria, search methods or synthesis approach).

See the online [MECIR Manual](#) for all reporting guidance relevant to the 'Discussion' section. Further guidance on issues for consideration in the 'Discussion' section is presented in Chapter 14 and Chapter 15.

3.3.6 Conclusions

There are two standard sections in Cochrane Intervention Reviews devoted to the authors' conclusions:

Implications for practice: In this section, review authors should provide a general interpretation of the evidence so that it can inform healthcare or policy decisions. The implications for practice should be as practical and unambiguous as possible, should be supported by the data presented in the review and should not be based on additional data that were not systematically compiled and evaluated as part of the review. Recommendations for how interventions should be implemented and used in practice must not be given in Cochrane Reviews, as they may be inappropriate depending on the different settings and individual circumstances of readers. Authors may be helpful to readers by identifying factors that are likely to be relevant to their decision making, such as the relative value of the likely benefits and harms of the intervention, participants at different levels of risk, or resource issues.

Implications for research: This section of a Cochrane Review is often used by people making decisions about future research, and review authors should try to write something that will be useful for this purpose. Implications for how research might be done and reported (e.g. the need for randomized trials rather than other types of study, for better descriptions of interventions, or for the routine collection of patient-important outcomes) should be distinguished from what future research should be done (e.g. research in particular subgroups of people, on an as yet untested experimental intervention). Any factors that led to downgrading the evidence as part of a GRADE assessment may provide suggestions to be addressed by future research. This section could also usefully draw on what is known about any ongoing studies identified from trials register searches, and use any information about ongoing or recently completed studies to guide recommendations on whether new studies need to be initiated. It is important that this section is as clear and explicit as possible. General statements that contain little or no specific information, such as "Future research should be better conducted" or "More research is needed" are of little use to people making decisions, and should be avoided.

See the online [MECIR Manual](#) for reporting guidance relevant to the conclusions of a review.

3.3.7 Administrative information

A Cochrane Intervention Review should include several pieces of administrative information, many of which are standard in other journals. These include acknowledgements, contributions of authors, declarations of interest, differences between the protocol and review, and sources of support (see the online MECIR Manual for relevant reporting guidance).

Acknowledgements: Review authors should acknowledge the contribution of people not listed as authors of the review, including any assistance from the Cochrane Review Group responsible for handling the review, and any contributions to searching, data collection, study appraisal or statistical analysis performed by people not listed as authors.

Contributions of authors: The contributions of each author to the review should be described. It is helpful to specify which authors were involved in each of the following tasks: conception of the review; design of the review; coordination of the review; search and selection of studies for inclusion in the review; collection of data for the review; analysis of data; interpretation of data, and; writing of the review.

Declarations of interest: All review authors must report any present or recent (three years prior to declaration) affiliations or other involvement in any organization or entity with an interest in the review's findings that might lead to a real or perceived conflict of interest. They must also include the dates of the involvement. If there are no known conflicts of interest, this should be stated explicitly, for example, by writing "None known". Authors should make themselves aware of the restrictions in place on authorship of Cochrane Reviews where conflicts of interest arise. The full policy on conflicts of interest is available in the [Cochrane Editorial and Publishing Policy Resource](#).

Differences between protocol and review: Review authors may sometimes use different or additional methods from those described in the review protocol (e.g. making post hoc changes to eligibility criteria, or adding subgroup analyses). This could occur because methods for dealing with a particular issue had not been specified in the protocol, pre-specified methods could not be applied due to insufficient data, or methods were changed because a preferable alternative arose or more recent guidance was identified. All changes of methods from protocol to review must be fully described and justified in this section of the review.

Sources of support: Authors should acknowledge grants that supported the review, and other forms of support, such as support from their university or institution in the form of a salary. Sources of support are divided into 'internal' (provided by the institutions at which the review was produced) and 'external' (provided by other institutions or funding agencies). Each source, its country of origin and what it supported should be provided. Authors should make themselves aware of the restrictions in place on funding of Cochrane Review by commercial sources where conflicts of interest may arise. The full policy on conflicts of interest is available in the [Cochrane Editorial and Publishing Policy Resource](#).

3.4 Reporting of plain language summaries in new Cochrane Intervention Reviews

The plain language summary is a stand-alone summary of the systematic review. It should convey succinctly and clearly the key question and key findings of the review, in plain English that can be understood by consumers and non-expert readers. Authors writing a plain language summary should consider the target audience, which may include people with a health condition, carers, health care workers or policy makers. Some topics may need more explanation than others based on the likely familiarity the target audience has with the topic, and the same term may mean different things to different people.

Writing in plain language is a skill, different to writing for a scientific audience, and review authors are encouraged to seek assistance to ensure that the summary is readily understood by a non-expert audience.

A complementary initiative to MECIR, the Plain Language Expectations for Authors of Cochrane Summaries (PLEACS), produced a set of specific reporting guidelines for plain language summaries (available at [Editorial and Publishing Policy Resource](#)). These guidelines were developed collaboratively between consumers, representatives of Cochrane Review Groups, and methodologists. Subsequent to their release, Glenton and colleagues produced further guidance on writing a Cochrane Review plain language summary and provides examples (see <https://www.cochrane.no/plain-language-summary-format>).

3.5 Chapter information

Contributing authors: Matthew J Page, Miranda Cumpston, Jacqueline Chandler, Toby Lasserson

Acknowledgments: This chapter builds on an earlier version of the Handbook (Chapter 4: Guide to the contents of a Cochrane protocol and review), edited by Julian Higgins and Sally Green. We thank them for their contributions to the earlier chapter. We thank Rachel Churchill for her feedback on an earlier draft of this chapter.

Declarations of interest: Toby Lasserson and Jacqueline Chandler are members of the core group who developed MECIR guidance, and were members of the PLEACS development group.

3.6 References

Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, Gøtzsche PC, Lasserson T, Tovey D, for the PfAG. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS Med* 2013;10: e1001419.

Campbell M, Katikireddi SV, Sowden A, McKenzie JE, Thomson H. Improving Conduct and Reporting of Narrative Synthesis of Quantitative Data (ICONS-Quant): protocol for a mixed methods study to develop a reporting guideline. *BMJ Open* 2018;8: e020064.

Dwan KM, Williamson PR, Kirkham JJ. Do systematic reviews still exclude studies with "no relevant outcome data"? *BMJ* 2017;358: j3919.

Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383: 267-276.

Harrison S, Jones HE, Martin RM, Lewis SJ, Higgins JPT. The albatross plot: A novel graphical tool for presenting results of diversely reported studies in a systematic review. *Res Synth Methods* 2017;8: 281-289.

Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. *Methodological Expectations of Cochrane Intervention Reviews*, Version 1.06. London: Cochrane; 2018.

Hoffmann TC, Oxman AD, Ioannidis JP, Moher D, Lasserson TJ, Tovey DI, Stein K, Sutcliffe K, Ravaut P, Altman DG, Perera R, Glasziou P. Enhancing the usability of systematic reviews by improving the consideration and description of interventions. *BMJ* 2017;358: j2998.

Kneale D, Thomas J, Harris K. Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews. *PLoS One* 2015;10: e0142187.

Lasserson T, Churchill R, Chandler J, Tovey D, Higgins JPT. Standards for the reporting of protocols of new Cochrane Intervention Reviews. In: Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R, editors. *Methodological Expectations of Cochrane Intervention Reviews*. London: Cochrane; 2016.

Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001;322: 1479-1480.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine / Public Library of Science* 2009;6: e1000100.

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine / Public Library of Science* 2009;6: e1000097.

Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, Worthy G. The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Med Res Methodol* 2008;8: 8.

Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catala-Lopez F, Li L, Reid EK, Sarkis-Onofre R, Moher D. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLoS Med* 2016;13: e1002028.

Page MJ, Altman DG, Shamseer L, McKenzie JE, Ahmadzai N, Wolfe D, Yazdi F, Catala-Lopez F, Tricco AC, Moher D. Reproducible research practices are underused in systematic reviews of biomedical interventions. *Journal of Clinical Epidemiology* 2018;94: 8-18.

Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. *Journal of Clinical Epidemiology* 2017;89: 30-42.

Shamseer L, Moher D, Clarke M, Gherzi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, Group P-P. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349: g7647.

Thomson HJ, Thomas S. The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Res Synth Methods* 2013;4: 95-101.

Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology* 2016;69: 225-234.

Chapter 11: Undertaking network meta-analyses

Anna Chaimani, Deborah M Caldwell, Tianjing Li, Julian PT Higgins, Georgia Salanti

This is a draft version of this chapter and is subject to change before finalization. It is made available for personal use of Cochrane members only, and is not for general distribution. All content remains the copyright of Cochrane.

To cite this draft chapter, please use:

Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. Draft version (16 September 2018) for inclusion in: Higgins JPT, Thomas J, Chandler J, Cumpston MS, Li T, Page MJ, Welch V (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane.

Copyright © 2018 The Cochrane Collaboration.

Key Points

- Network meta-analysis is a technique for comparing three or more interventions simultaneously in a single analysis by combining both direct and indirect evidence across a network of studies.
- Network meta-analysis allows for estimating the relative effects between any pairs of interventions in the network and usually yields more precise estimates than a single direct or indirect estimate. It also allows for the estimation of the ranking and hierarchy of interventions.
- A valid network meta-analysis relies on the assumption that the different sets of studies included in the analysis are similar, on average, in all important factors that may affect the relative effects.
- Incoherence (also called inconsistency) occurs when different sources of information (e.g. direct and indirect) for a relative effect disagree.
- Grading the confidence in the evidence in a network meta-analysis begins by evaluating each risk of bias domain for each direct comparison. Then the domain-specified assessments are combined to determine the overall confidence in the evidence.

11.1 Introduction

Most Cochrane Reviews present comparisons between pairs of interventions (an experimental intervention and a comparator intervention) for a specific condition and in a specific population or setting. However, it is usually the case that several, perhaps even numerous, competing interventions are available for any given condition. In these cases, people who need to decide between these interventions would benefit from a single review that includes all relevant interventions, and presents their comparative effectiveness and potential for harm. Network meta-analysis provides an analysis option for such a review.

This chapter provides an overview of the concepts, assumptions and methods that relate to network meta-analyses and to the indirect intervention comparisons on which they are built. Section 11.2 first describes what an indirect comparison is and how it can be conducted. It then introduces the notion of **transitivity** as the core assumption underlying the validity of an indirect comparison. Examples are provided where this assumption is likely to be held or violated. An introduction to the ideas of network meta-analysis and the assumption of **coherence** follows. Section 11.3 provides guidance on the design of a Cochrane Review with multiple interventions and the appropriate definition of the research question with respect to selecting studies, outcomes and interventions. Section 11.4 briefly describes the available statistical methods for synthesizing the data, estimating the relative ranking and

assessing incoherence in a network of interventions. Finally, Sections 11.5 and 11.6 provide approaches for evaluating the confidence in the evidence and presenting the evidence base and the results form a network meta-analysis. Note that the present chapter only provides an introduction to the statistical aspects of network meta-analysis; authors will need a knowledgeable statistician to plan and execute these methods.

In addition to Cochrane Reviews on the effects of health interventions (the subject of this Handbook), Cochrane produces Overviews of Reviews. Cochrane Overviews may also address multiple interventions for the same condition, by compiling evidence from multiple systematic reviews on a set of closely related interventions into one accessible and usable document. A central aim is to serve as a 'friendly front end' to the evidence in the Cochrane Database of Systematic Reviews. Cochrane Overviews present a synthesis of existing **systematic reviews**, while network meta-analysis within intervention reviews provide a synthesis of data from **individual randomized trials**. This leads to important distinctions in the methods employed, particularly for the search strategy and analysis plan, and also in the types of conclusions that can be drawn. These differences are summarized in Table 11.1.a. We recommend against undertaking network meta-analyses in the context of an Overview of Reviews unless the included reviews were undertaken as a series of reviews following a common protocol.

Table 11.1.a. Differences between a Cochrane Overview of reviews and a standard Cochrane Review when addressing multiple interventions for the same condition		
Review type	Cochrane Overview	Cochrane Review on interventions (with network meta-analysis)
Aim	Collate multiple systematic reviews about the effectiveness of interventions for the same condition to extract and analyse their results across important outcomes. [Overviews may also be used for other purposes. These are described in specific guidance for Cochrane Reviews]	Re-analyse data from randomized trials of multiple interventions for the same condition to make inferences about their comparative effectiveness or safety
Focus of search strategy	Reviews of primary studies	Primary studies
Focus of statistical synthesis	Review data	Study data
Focus of data collection	Summary estimates based on existing meta-analyses from the included reviews	Estimates from individual randomized trials

11.2 Important concepts

At the heart of network meta-analysis methodology is the concept of an indirect comparison. Indirect comparisons are necessary to estimate the relative effect of two interventions when no studies have compared them directly.

11.2.1 Indirect comparisons

Indirect comparisons allow us to estimate the relative effects of two interventions that have not been compared directly within a trial. For example, suppose there are randomized trials directly comparing provision of dietary advice by a dietitian (which we refer to as intervention A) with advice given by a doctor (intervention B). Suppose there are also randomized trials comparing dietary advice given by a dietitian (intervention A) with advice given by a nurse (intervention C). Suppose further that these randomized trials have been combined in standard, pairwise meta-analyses separately to derive direct estimates of intervention effects for A versus B (sometimes depicted 'AB') and A versus C ('AC'), measured as mean difference (MD) in weight reduction (see Chapter 6, Sections 6.4.1.1). The situation is illustrated in Figure 11.2.a, where the solid straight lines depict available evidence. We wish to learn about the relative effect of advice by a doctor versus a nurse (B versus C); and the dashed line depicts this comparison, for which there is no direct evidence.

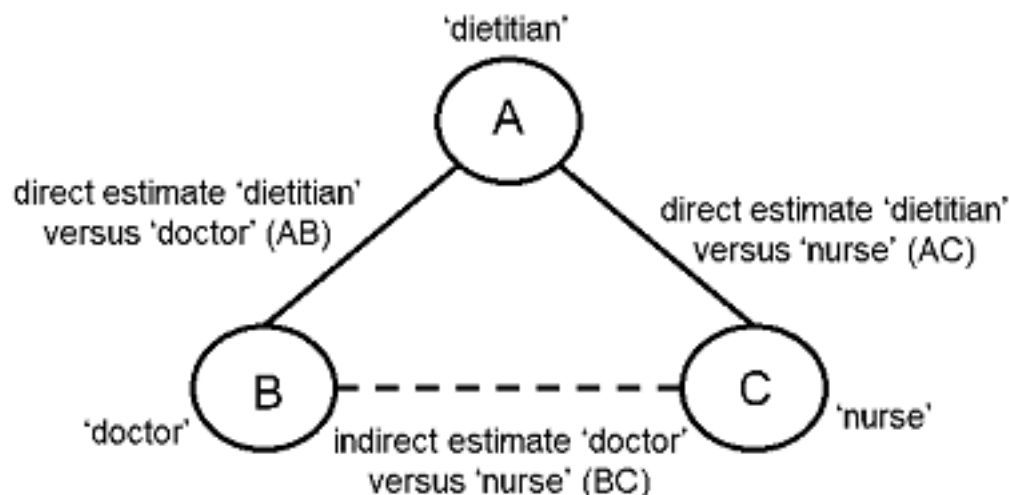


Figure 11.2.a. Illustration of an indirect estimate that compares the effectiveness of 'doctor' (B) and 'nurse' (C) in providing dietary advice through a common comparator 'dietitian' (A).

One way to understand an indirect comparison is to think of the BC comparison (of B versus C) as representing the benefit of B over C. All else being equal, the benefit of B over C is equivalent to the benefit of B over A plus the benefit of A over C. Thus, for example, the indirect comparison describing benefit of 'doctor' over 'nurse' may be thought of as the benefit of 'doctor' over 'dietitian' plus the benefit of 'dietitian' over 'nurse' (these 'benefits' may be positive or negative; we do not intend to imply any particular superiority among these three types of people offering dietary advice). This is represented graphically in Figure 11.2.b. Mathematically, the sum can be written:

$$\text{indirect MD (BvsC)} = \text{direct MD (BvsA)} + \text{direct MD(AvsC)}$$

We usually write this in the form of subtraction:

$$\text{indirect MD(BvsC)} = \text{direct MD(AvsC)} - \text{direct MD(AvsB)}$$

such that the difference between the summary statistics of the intervention effect in the direct A versus B and A versus C meta-analyses provides an indirect estimate of the B versus C intervention effect.

For this simple case where we have two direct comparisons (three interventions) the analysis can be conducted by performing subgroup analyses using standard meta-analysis routines (including RevMan): studies addressing the two direct comparisons (i.e. A versus B and A versus C) can be treated as two subgroups in the meta-analysis. Subtracting the summary effect from each subgroup gives an estimate for the indirect comparison.

Most software will provide a P value for the statistical significance of the difference between the subgroups based on the estimated variance of the indirect effect estimate (Bucher et al 1997):

$$\text{Variance [indirect MD(BvsC)]} = \text{Variance [direct MD(AvsC)]} + \text{Variance [direct MD(AvsB)]}$$

where $\text{variance}[\text{direct MD(AvsC)}]$ and $\text{variance}[\text{direct MD(AvsB)}]$ are the variances of the respective direct estimates (from the two subgroup analyses).

A 95% confidence interval for the indirect summary effect is constructed by the formula:

$$[\text{indirect MD (BvsC)} \pm 1.96 \times \sqrt{\text{Variance [indirect MD (BvsC)]}}]$$

This method uses the intervention effects from each group of randomized trials and therefore preserves within-trial randomization. If we had instead pooled single arms across the studies (e.g. all B arms and all C arms, ignoring the A arms) and then performed a direct comparison between the pooled B and C arms (i.e. treating the data as if they came from a single large randomized trial), then our analysis would violate within-trial randomization (Li and Dickersin 2013). This approach should not be used.

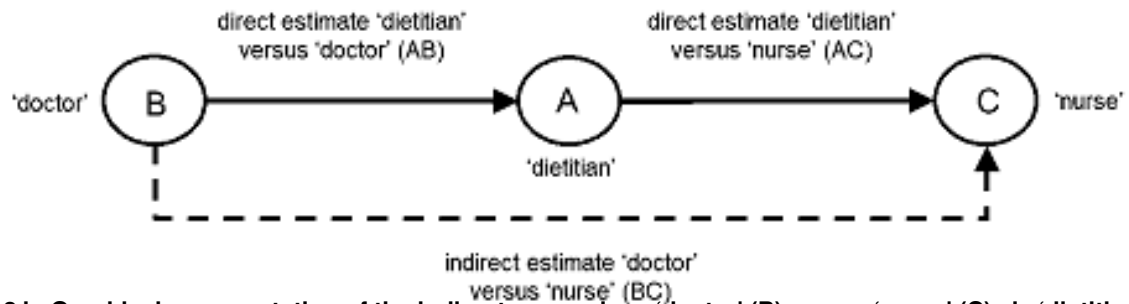


Figure 11.2.b. Graphical representation of the indirect comparison 'doctor' (B) versus 'nurse' (C) via 'dietitian' (A).

When four or more competing interventions are available, indirect estimates can be derived via multiple routes (Hughes 2010). The only requirement is that two interventions are 'connected' and not necessarily via a single common comparator. An example of this situation is provided in Figure 11.2.c. Here 'doctor' (B) and 'pharmacist' (D) do not have a common comparator, but we can compare them indirectly via the route 'doctor' (B) – 'dietitian' (A) – 'nurse' (C) – 'pharmacist' (D) by an extension of the arguments above.

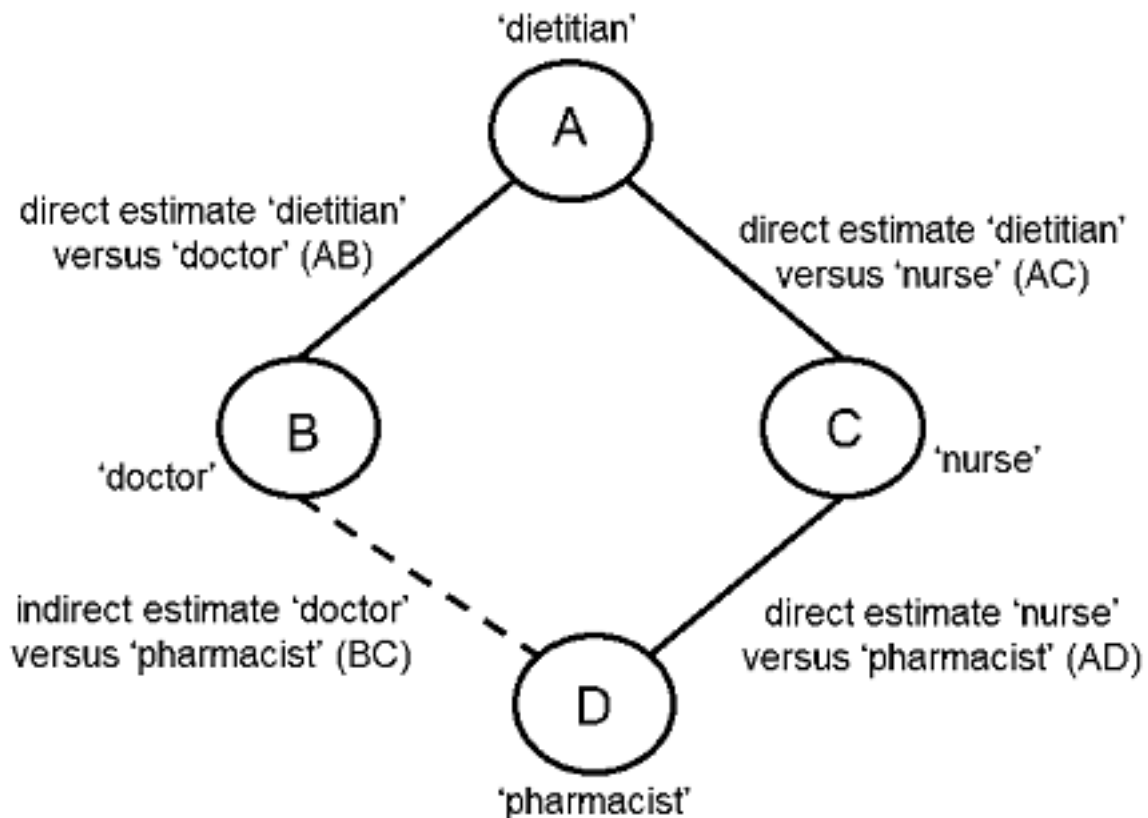


Figure 11.2.c. Example of deriving indirect estimate that compares the effectiveness of 'doctor' (B) and 'pharmacist' (D) in providing dietary advice through a connected loop.

11.2.2 Transitivity

11.2.2.1 Validity of an indirect comparison

Indirect comparisons provide observational evidence across randomized trials and may suffer the biases of observational studies, such as confounding (see Chapter 10, Section 10.11.5). The validity of an indirect comparison relies on the assumption that the different sets of randomized trials are similar, on average, in all important factors other than the intervention comparison being made (Song et al 2003, Glenny et al 2005, Donegan et al 2010, Salanti 2012). Studies that compare different interventions may differ in a wide range of characteristics. Sometimes these characteristics are associated with the effect of an intervention. These are often known as effect modifiers, and are sources of heterogeneity in pairwise meta-analyses. If the A versus B and A versus C randomized trials differ with respect to such effect modifiers, then it would not be appropriate to make an indirect comparison.

11.2.2.2 What is transitivity?

The underlying assumption of indirect comparisons is that the common comparator intervention A allows a **transitive** relationship between the A versus B and A versus C effects. This transitive relationship can be written mathematically as:

effect of B versus C = (effect of A versus C) – (effect of A versus B)

In words, this means that we can compare interventions B and C via intervention A (Figure 11.2.a).

Transitivity requires that intervention A is similar when it appears in A versus B studies and A versus C studies with respect to characteristics (effect modifiers) that may affect the two relative effects (Salanti et al 2009). For example, in the dietary advice network the common comparator 'dietitian' might differ with respect to the frequency of sessions with the participants between randomized trials that compare dietitian with doctor (A versus B) and trials that compare dietitian with nurse (A versus C). If the participants visit the dietitian once a week in AB studies and once a month in AC studies, transitivity may be violated. Similarly, any other effect modifiers should not differ between AB and AC studies.

Transitivity requires all competing interventions of a systematic review to be **jointly randomizable**. That is, we can imagine all interventions being compared simultaneously in a single multi-arm randomized trial. Another way of viewing this is that the 'missing' interventions (those not included in the identified studies) may be considered to be missing for reasons unrelated to their effects (Caldwell et al 2005, Salanti 2012).

11.2.2.3 Assessing transitivity

Clinical and methodological differences are inevitable between studies in a systematic review. Researchers undertaking indirect comparisons should assess whether such differences are sufficiently large to cause **intransitivity**. In principle, transitivity can be evaluated by comparing the distribution of effect modifiers across the different comparisons (Salanti 2012, Cipriani et al 2013, Jansen and Naci 2013). Imbalanced distributions would threaten the plausibility of the transitivity assumption and thus the validity of indirect comparison. In practice, however, this requires that the effect modifiers are known and have been measured.

Extended guidance on considerations of potential effect modifiers is provided in discussions of heterogeneity in Chapter 10 (Section 10.11). For example, we may believe that age is a potential effect modifier so that the effect of an intervention differs between younger and older populations. If the average age in A versus B randomized trials is substantially older or younger than in A versus C randomized trials, transitivity may be implausible, and an indirect comparison B versus C may be invalid.

Figure 11.2.d shows hypothetical examples of valid and invalid indirect comparisons for the dietary advice example. Suppose a single effect modifier is severity of disease (e.g. obesity measured by the BMI score). The top row depicts a situation in which all patients in all trials have moderate severity. There are AB studies and AC studies in this population. Estimation of BC is valid here because there is no difference in the effect modifier. The second row depicts a similar situation in a second subgroup of patients who all have severe disease. Analysing the two subgroups of randomized trials separately also gives a valid indirect estimate of B versus C for this population. In the third row we depict a situation in which all AB trials are conducted only in moderately obese populations and all AC trials are conducted only in severely obese populations. In this situation, the distribution of effect modifiers is different in the two direct comparisons, so the indirect effect based on the entire population is invalid (due to intransitivity). In practice, differences in effect modifiers are usually less extreme than this hypothetical scenario; for example, AB randomized trials may have 80% moderately obese population and 20% severely obese, and AC randomized trials may have 20% moderately obese and 80% severely obese population. Intransitivity still would invalidate the indirect estimate B versus C if severity is an important effect modifier.

Population with moderate disease

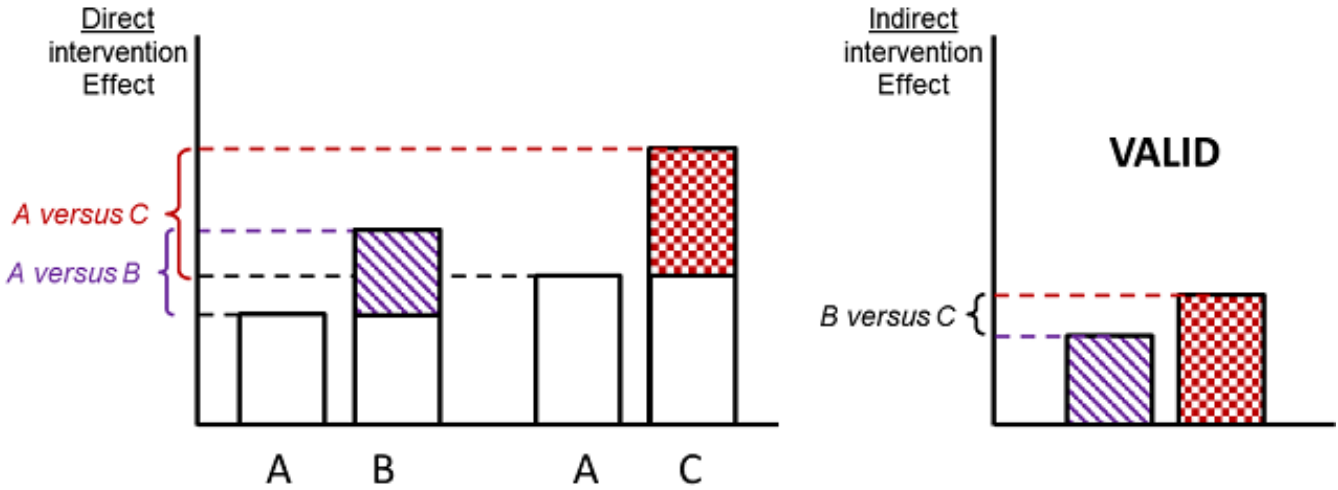


Figure 11.2.d. Example of valid and invalid indirect comparisons when the severity of disease acts as effect modifier and its distribution differs between the two direct comparisons. The shaded boxes represent the treatment effect estimates from each source of evidence (striped box for A versus B and checked box for A versus C). In the first row, randomized trials of A versus B and of A versus C are all conducted in moderately obese populations; in the second row randomized trials are all conducted in severely obese populations. In both of these the indirect comparisons of the treatment effect estimates would be valid. In the last row, the A versus B and A versus C randomized trials are conducted in different populations. As severity is an effect modifier, the indirect comparison based on these would not be valid (Jansen et al 2014).

11.2.3 From indirect comparisons to network meta-analysis

11.2.3.1 Combining direct and indirect evidence

Often there is direct evidence for a specific comparison of interventions as well as a possibility of making an indirect comparison of the interventions via one or more common comparators. If the key assumption of transitivity is considered reasonable, direct and indirect estimates should be considered jointly. When both direct and indirect intervention effects are available for a particular comparison, these can be synthesized into a single effect estimate. This summary effect is sometimes called a **combined** or **mixed** estimate of the intervention effect. We will use the former term in this chapter. A combined estimate can be computed as an inverse variance weighted average (see Chapter 10, Section 10.3) of the direct and indirect summary estimates.

Since combined estimates incorporate indirect comparisons, they rely on the transitivity assumption. Violation of transitivity threatens the validity of both indirect and combined estimates. Of course, biased direct intervention effects for any of the comparisons also challenge the validity of a combined effect (Madan et al 2011).

11.2.3.2 Coherence (or consistency)

The key assumption of transitivity refers to potential clinical and methodological variation across the different comparisons. These differences may be reflected in the data in the form of disagreement in estimates between different sources of evidence. This is the statistical manifestation of transitivity and is typically called either coherence or consistency. We will use the former to distinguish the notion from inconsistency (or heterogeneity) within standard meta-analyses (e.g. as is measured using the I^2 statistic; see Chapter 10, Section 10.10.2). Coherence implies that the different sources of evidence (direct and indirect) agree with each other.

The coherence assumption is expressed mathematically by the coherence equations, which state that the true direct and indirect intervention effects for a specific comparison are identical:

$$\text{'true' MD (BvsC)} = \text{'true' MD (AvsC)} - \text{'true' MD (AvsB)}$$

Some methods for testing this assumption are presented in Section 11.4.4.

11.2.3.3 Network meta-analysis

Any group of studies that links three or more interventions via direct comparisons forms a **network of interventions**. In a network of interventions there can be multiple indirect intervention effects for each comparison. Then, the combined estimates for any pairwise comparisons in a network may incorporate direct or several indirect estimates, or both. **Network meta-analysis** combines direct and indirect estimates across a network of interventions. Synonymous terms, less often used, are **mixed treatment comparison** and **multiple treatments meta-analysis**.

11.2.3.4 Network diagrams

A **network diagram** is a graphical depiction of the structure of the network (Chaimani et al 2013). It consists of nodes representing the interventions in the network and lines showing the available direct comparisons between pairs of interventions. Distinct lines forming closed loops can be added in a network diagram to illustrate the presence of multi-arm studies. For example, a triangular loop would represent a three-arm study (see Figure 11.2.e). For large and complex networks this presentation of multi-arm studies may give complicated and unhelpful network diagrams. In this case it might be preferable to show multi-arm studies in a tabular format (see Section 11.6.1). Further discussion on network diagrams is available in Section 11.6.1.

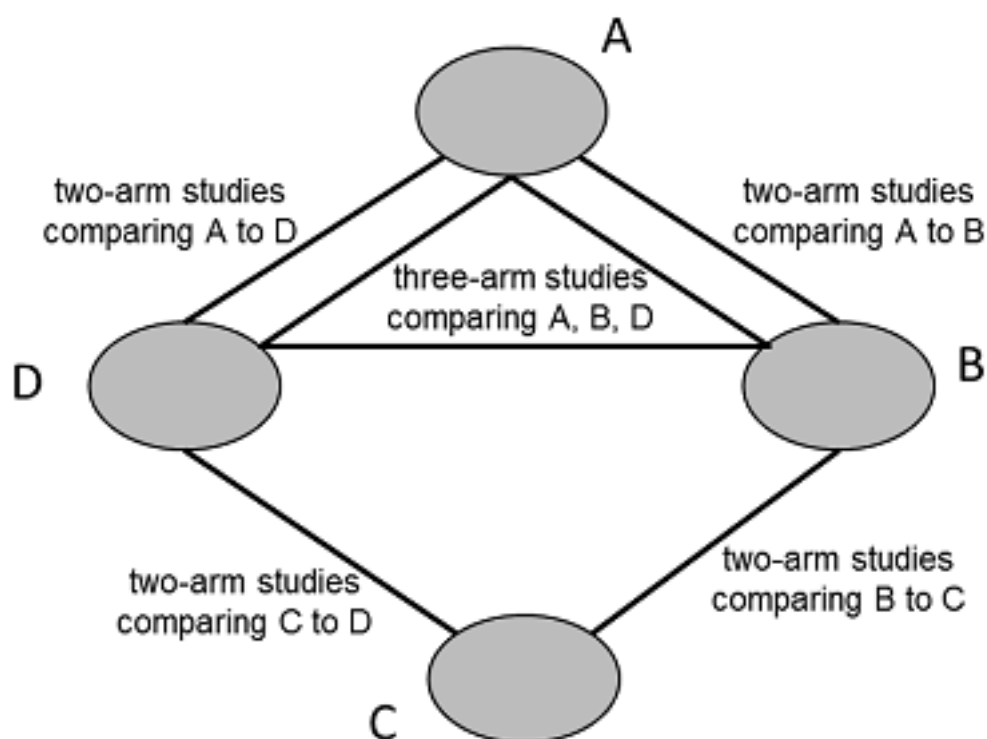


Figure 11.2.e. Example of network diagram with four competing interventions and information on the presence of multi-arm randomized trials.

11.2.3.5 Advantages of network meta-analysis

A network meta-analysis exploits all available direct and indirect evidence. Empirical studies have suggested it yields more precise estimates of the intervention effects in comparison with a single direct or indirect estimate (Cooper et al 2011, Caldwell et al 2015). In addition, network meta-analysis can provide information for comparisons between pairs of interventions that have never been evaluated within individual randomized trials. The simultaneous comparison of all interventions of interest in the same analysis enables the estimation of their relative ranking for a given outcome. More extensive discussion on the relative ranking of

interventions is provided in Section 11.4.3.3.

11.2.3.6 Validity of network meta-analysis

The validity of network meta-analysis relies on the fulfilment of underlying assumptions. Transitivity should hold for every possible indirect comparison, and coherence should hold in every loop of evidence within the network (see Section 11.4.4). Considerations about heterogeneity within each direct comparison in the network should follow the existing recommendations for standard pairwise meta-analysis (see Chapter 10, Section 10.10).

11.3 Planning a Cochrane Review to compare multiple interventions

11.3.1 Expertise required in the review team

Because of the complexity of network meta-analysis, it is important to establish a multidisciplinary review team that includes a statistician skilled in network meta-analysis methodology early and throughout. Close collaboration between the statistician and the content area expert is essential to ensure that the studies selected for a network meta-analysis are similar except for the interventions being compared (see Section 11.2.2). Because basic meta-analysis software such as RevMan does not support network meta-analysis, the statistician will have to rely on statistical software packages such as Stata, R, WinBUGS or OpenBUGS for analysis.

11.3.2 The importance of a well-defined research question

Defining the research question of a systematic review that intends to compare multiple interventions should follow the general guidelines described in Chapter 2 and should be stated in the objectives of the review. In this section, we summarize and highlight key issues that are pertinent to systematic review with a network meta-analysis.

Because network meta-analysis could be used to estimate the relative ranking of the included interventions (Salanti et al 2011, Chaimani et al 2013), reviews that aim to rank the competing interventions should specify this in their objectives (Chaimani et al 2017). Review authors should consider obtaining an estimate of relative ranking as a secondary objective to supplement the relative effects. An extended discussion on the relative ranking of interventions is provided in Section 11.4.3.3.

11.3.2.1 Defining the population and choosing the interventions

Populations and interventions often need to be considered together given the potential for intransitivity (see Section 11.2.2). A driving principle is that any eligible participant should be eligible for randomization to any included intervention (Salanti 2012, Jansen and Naci 2013). Review authors should select their target population with this consideration in mind. Particular care is needed in the definition of the eligible interventions, as discussed in Chaimani et al (Chaimani et al 2017). For example, suppose a systematic review aims to compare four chemotherapy regimens for a specific cancer. Regimen (D) is appropriate for stage II patients exclusively and regimen (A) is appropriate for both stage I and stage II patients. The remaining two regimens (B) and (C) are appropriate for stage I patients exclusively. Now suppose A and D were compared in stage II patients, and A, B and C were compared in stage I patients (see Figure 11.3.a). The four interventions forming the network are unlikely to satisfy the transitivity assumption because regimen D cannot be given to the same patient population as regimens B and C. Thus, a four-arm randomized trial comparing all interventions (A, B, C and D) simultaneously is not a reasonable study to conduct.

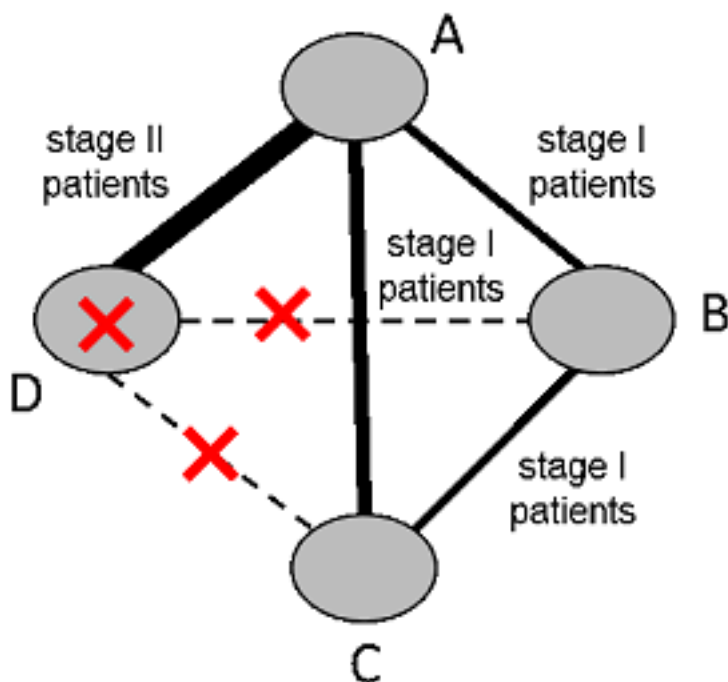


Figure 11.3.a. Example of a network comparing four chemotherapy regimens, where transitivity is violated due to incomparability between the interventions.

11.3.2.2 Decision sets and supplementary sets of interventions

Usually there is a specific set of interventions of direct interest when planning a network meta-analysis, and these are sometimes referred to as the **decision set**. These are the options among which patients and health professionals would be choosing in practice with respect to the outcomes under investigation. In selecting which competing interventions to include in the decision set, review authors should ensure that the transitivity assumption is likely to hold (see also Section 11.2.2) (Salanti 2012).

The ability of network meta-analysis to incorporate indirect evidence means that inclusion of interventions that are not of direct interest to the review authors might provide additional information in the network. For example, placebo is often included in network meta-analysis even though it is not a reasonable treatment option, because many studies have compared active interventions against placebo. In such cases, excluding placebo would result in ignoring a considerable amount of indirect evidence. Similar considerations apply to historical or legacy interventions.

We use the term **supplementary set** to refer to interventions, such as placebo, that are included in the network meta-analysis for the purpose of improving inference among interventions in the decision set. The full set of interventions, the decision set plus the supplementary set, has been called in the literature the **synthesis comparator set** (Ades et al 2013, Caldwell et al 2015).

When review authors decide to include a supplementary set of interventions in a network they need to be cautious regarding the plausibility of the transitivity assumption. In general, broadening the network challenges the transitivity assumption. Thus, supplementary interventions should be added when their value outweighs the risk of violating the transitivity assumption. The addition of supplementary interventions in the analysis might be considered more valuable for sparse networks that include only a few randomized trials per comparison. In these networks the benefit of improving the precision of estimates by incorporating supplementary indirect evidence may be quite important. There is limited empirical evidence to inform the decision of how far one should go in constructing the network evidence base (König et al 2013, Caldwell et al 2015). Inevitably it will require some judgment, and the robustness of decisions can be evaluated in sensitivity analyses and discussed in the review.

11.3.2.3 Grouping variants of an intervention (defining nodes in the network diagram)

The definition of nodes needs careful consideration in situations where variants of one or more interventions are expected to appear in the eligible randomized trials (James et al 2018). The appropriateness of merging, for example, different doses of the same drug or different drugs within a class depends to a large extent on the research question. Lumping and splitting the variants of the competing interventions might be interesting to both review authors and evidence users; in such a case this should be stated clearly in the objectives of the review and the potential for intransitivity should be evaluated in every network. A decision on how the nodes

of an expanded network could be merged is not always straightforward and researchers should act based on pre-defined criteria where possible. These criteria should be formed in such a way that maximizes similarity of the interventions within a node and minimizes similarity across nodes.

The following example refers to a network that used two criteria to classify electronic interventions for smoking cessation into five categories:

“To be able to draw generalizable conclusions on the different types of electronic interventions, we developed a categorization system that brought similar interventions together in a limited number of categories. We sought advice from experts in smoking cessation on the key dimensions that would influence the effectiveness of smoking cessation programmes. Through this process, two dimensions for evaluating interventions were identified. The first dimension was related to whether the intervention offered generic advice or tailored its feedback to information provided by the user in some way. The second dimension related to whether the intervention used a single channel or multiple channels. From these dimensions, we developed a system with five categories... , ranging from interventions that provide generic information through a single channel, e.g. a static Web site or mass e-mail (category e1) to complex interventions with multiple channels delivering tailored information, e.g. an interactive Web site plus an interactive forum (category e5)” (Madan et al 2014).

Empirical evidence is currently lacking on whether more or less expanded networks are more prone to present important intransitivity or incoherence. Extended discussions of how different dosages can be modelled in network meta-analysis are available (Giovane et al 2013, Owen et al 2015, Mawdsley et al 2016).

11.3.2.4 Defining eligible comparisons of interventions (defining lines in the network diagram)

Once the nodes of the network have been specified, every study that meets the eligibility criteria and compares any pair of the eligible interventions should be included in the review. The exclusion of specific direct comparisons without a rationale may introduce bias in the analysis and should be avoided.

11.3.3 Selecting outcomes to examine

In the context of a network meta-analysis, outcomes should be specified a priori regardless of the number of interventions the review intends to compare or the number of studies the review is able to include. Review authors should be aware that some characteristics may be effect modifiers for some outcomes but not for other outcomes. This implies that sometimes the potential for intransitivity should be examined separately for each outcome before undertaking the analyses.

11.3.4 Study designs to include

Randomized designs are generally preferable to non-randomized designs to ensure an increased level of validity of the summary estimates (see Chapter 3). Sometimes observational data from non-randomized studies may form a useful source of evidence (Chapter 24). In general, combining randomized with observational studies in a network meta-analysis is not recommended. In the case of sparse networks (i.e. networks with a few studies but many interventions), observational data might be used to supplement the analysis; for example, to form prior knowledge or provide information for baseline characteristics (Schmitz et al 2013, Soares et al 2013).

11.4 Synthesis of results

11.4.1 What does a network meta-analysis estimate?

In a connected network, the coherence equations provide mathematical links between the intervention effects, so that some effects can be computed from others using transitivity assumptions. This means that not all pairwise comparisons are independently estimated. In fact, the number of comparisons that need to be estimated in a network meta-analysis equals the number of interventions minus one. In practice, we select a particular set of comparisons of this size, and we often label these the **basic comparisons** for the analysis (Lu and Ades 2006). For example, in the network of four interventions for heavy menstrual bleeding illustrated in Figure 11.4.a we might choose the following basic comparisons: ‘Hysterectomy versus first generation hysteroscopic techniques’, ‘Mirena versus first generation hysteroscopic techniques’ and ‘second generation non-hysteroscopic techniques’.

versus first generation hysteroscopic techniques'. All other (non-basic) comparisons in the network (e.g. 'Mirena versus hysterectomy', 'Mirena versus second generation non-hysteroscopic techniques', etc.) can be computed from the basic comparisons and are called **functional comparisons**.

The main result of each is a set of **network estimates** of the intervention effects for all basic comparisons. We obtain estimates for the functional comparisons after the analysis using the coherence equations (see Section 11.2.3.2). It does not matter which set of comparisons we select as the basic comparisons. Often we would identify one intervention as a reference, and define the basic comparisons as the effect of each of the other interventions against this reference.

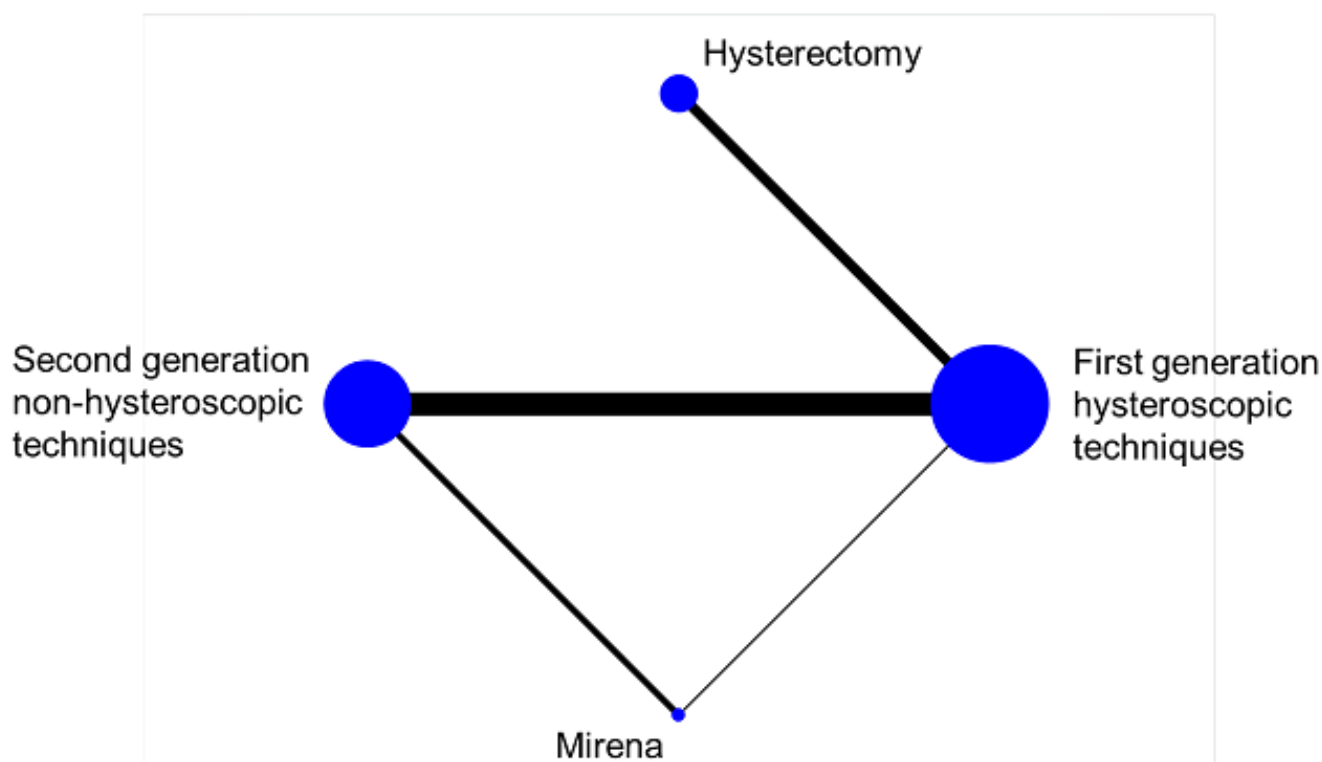


Figure 11.4.a. Network graph of four interventions for heavy menstrual bleeding (Middleton et al 2010). The size of the nodes is proportional to the number of participants assigned to the intervention and the thickness of the lines is proportional to the number of randomized trials that studied the respective direct comparison.

11.4.2 Synthesizing direct and indirect evidence using meta-regression

Network meta-analysis can be performed using several approaches (Salanti et al 2008). The main technical requirement for all approaches is that all interventions included in the analysis form a 'connected' network. A straightforward approach that can be used for many networks is to use meta-regression (see Chapter 10, Section 10.11.4). This approach works as long as there are no multi-arm trials in the network (otherwise, other methods are more appropriate).

We introduced indirect comparisons in Section 11.2.1 in the context of subgroup analysis, where the subgroups are defined by the comparisons. Differences between subgroups of studies can also be investigated via meta-regression. When standard meta-regression is used to conduct a single indirect comparison, a single dummy variable is used to specify whether the result of each study relates to one direct comparison or the other (a dummy variable is coded as 1 or 0 to indicate which comparison is made in the study). For example, in the dietary advice network containing only three intervention nodes (see Section 11.2.1, Figure 11.2.a) the dummy variable might be used to indicate the comparison 'dietitian versus nurse'. This variable takes the value 1 for a study that involves that corresponding comparison and 0 if it involves the comparison 'dietitian versus doctor', and is included as a single covariate in the meta-regression. In this way, the meta-regression model would have an intercept and a regression coefficient (slope). The estimated intercept gives the meta-analytic direct summary estimate for the comparison 'dietitian versus doctor' while the sum of the estimated regression coefficient and intercept gives the direct summary estimate for 'dietitian versus nurse'. Consequently, the estimated coefficient is the indirect summary estimate for the comparison 'doctor versus nurse'.

An alternative way to perform the same analysis of an indirect comparison is to re-parameterize the meta-regression model by using two dummy variables and no intercept, instead of one dummy variable and an intercept. The first dummy variable would indicate the

comparison 'dietitian versus doctor, and the second the comparison 'dietitian versus nurse'. The estimated regression coefficients then give the summary estimates for these two comparisons, and it is convenient to consider these as the two basic comparisons for this analysis. The difference between the two regression coefficients is the summary estimate for the indirect comparison 'doctor versus nurse'.

The coding of each basic comparison using a dummy variable, and the omission of the intercept, proves to be a useful approach for implementing network meta-analysis using meta-regression, and helps explain the role of the coherence equations. Specifically, suppose now that in the dietary advice example, studies that directly compare 'doctor versus nurse' are also available. Because we are already estimating all of the basic comparisons required for three interventions, we do not require a third dummy variable (under coherence, the comparison 'doctor versus nurse' can be expressed as the difference between the other two comparisons: see Section 11.2.3.2). This means that studies comparing 'doctor versus nurse' inform us about the difference between the other two comparisons. Consequently, we need to assign values -1 and 1 to the dummies 'dietitian versus doctor' and 'dietitian versus nurse' respectively. The meta-regression is again fitted including both dummy variables without an intercept. The interpretations of the estimated regression coefficients are the same as for the indirect comparison.

11.4.3 Performing network meta-analysis

We now consider approaches designed specifically for network meta-analysis that can be used when we have multi-arm trials. An overview of methodological developments can be found in (Efthimiou et al 2016).

A popular approach to conducting network meta-analysis is using hierarchical models, commonly implemented within a Bayesian framework (Sobieraj et al 2013, Petropoulou et al 2016). Detailed descriptions of hierarchical models for network meta-analysis can be found elsewhere (Lu and Ades 2004, Salanti et al 2008, Dias et al 2018).

Multivariate meta-analysis methods, initially developed to synthesize multiple outcomes jointly (Jackson et al 2011, Mavridis and Salanti 2013), offer an alternative approach to conducting network meta-analysis. A multivariate meta-analysis approach focuses the analysis on the set of basic comparisons (e.g. each intervention against a common reference intervention) and treats these as analogous to different outcomes. A study can report on one or more of the basic comparisons; for example, there are two comparisons in a three-arm randomized trial. For studies that do not target any of the basic comparisons (e.g. a study that does not include the common reference intervention), a technique known as data augmentation can be used to allow the appropriate parameterization (White et al 2012). The method is implemented in 'network' available for the Stata statistical package (White 2015). A detailed description of the concepts and the implementation of this approach is available (White et al 2012).

Methodology from electrical networks and graphic theory also can be used to fit network meta-analysis and is outlined in by Rücker (Rucker 2012). This approach has been implemented in the R package 'netmeta' (Rucker and Schwarzer 2013).

11.4.3.1 Illustrating example

To illustrate the advantages of network meta-analysis, Figure 11.4.a presents a network of four interventions for heavy menstrual bleeding (Middleton et al 2010). Data are available for four out of six possible direct comparisons. Table 11.4.a presents the results from direct (pairwise) meta-analyses and a network meta-analysis using the meta-regression approach. Network meta-analysis provides evidence about the comparisons 'Hysterectomy versus second generation non-hysteroscopic techniques' and 'Hysterectomy versus Mirena', which no individual randomized trial has assessed. Also, the network meta-analysis results are more precise (narrower confidence intervals) than the pairwise meta-analysis results for two comparisons ('Mirena versus first generation hysteroscopic techniques' and 'Second generation non-hysteroscopic techniques versus Mirena'). Note that precision is not gained for all comparisons; this is because for some comparisons (e.g. hysterectomy versus first generation hysteroscopic techniques), the heterogeneity among studies in the network as a whole is larger than the heterogeneity within the direct comparison, and therefore some uncertainty is added in the network estimates (see Section 11.4.3.2).

Table 11.4.a. Intervention effects, measured as odds ratios of patient dissatisfaction at 12 months of four interventions for heavy menstrual bleeding. Odds ratios lower than 1 favour the column-defining intervention for the network meta-analysis results (lower triangle) and the row-defining intervention for the pairwise meta-analysis results (upper triangle).

Pairwise meta-analysis			
Hysterectomy	-	-	0.38 (0.22 to 0.65)

0.45 (0.24 to 0.82)	Second generation non-hysteroscopic techniques	1.35 (0.45 to 4.08)	0.82 (0.60 to 1.12)
0.43 (0.18 to 1.06)	0.96 (0.48 to 1.91)	Mirena	2.84 (0.51 to 15.87)
0.38 (0.23 to 0.65)	0.85 (0.63 to 1.15)	0.88 (0.43 to 1.84)	First generation hysteroscopic techniques
Network meta-analysis			

11.4.3.2 Assumptions about heterogeneity

Heterogeneity reflects the underlying differences between the randomized trials that directly compare the same pair of interventions (see Chapter 10, Section 10.10). In a pairwise meta-analysis, the presence of important heterogeneity can make the interpretation of the summary effect challenging. Network meta-analysis estimates are a combination of the available direct estimates via both direct and indirect comparisons, so heterogeneity among studies for one comparison can impact on findings for many other comparisons.

It is important to specify assumptions about heterogeneity in the network meta-analysis model. Heterogeneity can be specific to each comparison, or assumed to be the same for every pairwise comparison. The idea is similar to a subgroup analysis: the different subgroups could have a common heterogeneity or different heterogeneities, the latter can be estimated accurately only if enough studies are available in each subgroup.

It is very common to assume that the amount of heterogeneity is the same for every comparisons in the network (Higgins and Whitehead 1996). This has two advantages compared with assuming comparison-specific heterogeneities. First, it shares information across comparisons, so that comparisons with only one or two trials can borrow information about heterogeneity from comparisons with several trials. Likewise, for comparisons with many studies, heterogeneity is estimated more precisely because more data are incorporated, resulting usually in more precise estimates of relative effects. Second, assuming common heterogeneity makes model estimation computationally easier than assuming comparison-specific heterogeneity (Lu and Ades 2009).

The choice of heterogeneity assumption should be based on clinical and methodological understanding of the data, and assessment of the plausibility of the assumption, in addition to statistical properties.

11.4.3.3 Ranking interventions

One hallmark feature of network meta-analysis is that it can estimate relative rankings of the competing interventions for a particular outcome. **Ranking probability**, the probability that an intervention is at a specific rank (first, second, etc.) when compared with the other interventions in the network, is frequently used. Ranking probabilities may vary for different outcomes. As for any estimated quantity, ranking probabilities are estimated with some variability. Therefore, inference based solely on the probability of being ranked as the best, without accounting for the variability, is misleading and should be avoided.

Ranking measures such as the **mean ranks**, **median ranks**, and the **cumulative ranking probabilities** summarize the estimated probabilities for all possible ranks and account uncertainty in relative ranking. Further discussion of ranking measures is available elsewhere (Salanti et al 2011, Chaimani et al 2013, Tan et al 2014, Rücker and Schwarzer 2015).

The estimated ranking probabilities for the heavy menstrual bleeding network (see Section 11.4.3.2) are presented in Table 11.4.b. 'Hysterectomy' is the most effective intervention according to mean rank.

Table 11.4.b. Ranking probabilities and mean ranks for intervention effectiveness in heavy menstrual bleeding. Lower mean rank values indicate that the interventions are associated with less mortality.

Rank	Hysterectomy	Second generation non-hysteroscopic techniques	Mirena	First generation hysteroscopic techniques

Probabilities	1	96%	1%	4%	0%	
	2	4%	46%	40%	9%	
	3	0%	46%	19%	35%	
	4	0%	7%	37%	56%	
Mean Rank		1	3	3	4	

11.4.4 Disagreement between evidence sources (incoherence)

11.4.4.1 What is incoherence?

Incoherence refers to the violation of the coherence assumption in a network of interventions (see Section 11.2.3.2). Incoherence occurs when different sources of information for a particular relative effect are in disagreement (Song et al 2003, Lu and Ades 2006, Salanti 2012). In much of the literature on network meta-analysis, the term **inconsistency** has been used, rather than incoherence.

The amount of incoherence in a closed loop of evidence in a network graph can be measured as the absolute difference between the direct and indirect summary estimates for any of the pairwise comparisons in the loop (Bucher et al 1997, Song et al 2011, Veroniki et al 2013). We refer to this method of detecting incoherence as the 'loop-specific approach'. The obtained statistic is usually called **incoherence factor** or **inconsistency factor** (IF). For example, in the dietary advice network the incoherence factor would be estimated as:

$$IF = |\text{direct MD(BvsC)} - \text{indirect MD(BvsC)}|$$

IF measures the level of disagreement between the direct and indirect effect estimates.

The standard error of the incoherence factor is obtained from

$$\text{Variance [IF]} = \text{Variance [direct MD(BvsC)]} + \text{Variance [indirect MD(BvsC)]}$$

and can be used to construct a 95% confidence interval for the IF:

$$[IF \pm 1.96 \times SE (IF)]$$

Several approaches have been suggested for evaluating incoherence in a network of interventions with many loops (Donegan et al 2013, Veroniki et al 2013), broadly categorized as **local** and **global** approaches. Local approaches evaluate regions of network separately to detect possible 'incoherence spots', whereas global approaches evaluate coherence in the entire network.

11.4.4.2 Approaches to evaluating local incoherence

A local approach, that we term SIDE (Separating Indirect from Direct Evidence) evaluates the IF for every pairwise comparison in a network by contrasting a direct estimate (when available) with an indirect estimate; the latter being estimated from the entire network once the direct evidence has been removed. The method was first introduced by (Dias et al 2010) under the name 'node-splitting'. The SIDE approach has been implemented in the 'network' macro for the Stata statistical package (White 2015) and the 'netmeta' command in R (Schwarzer et al 2015). For example, Table 11.4.c presents the incoherence results of a network that compares the effectiveness of four active interventions and placebo in preventing serious vascular events after transient ischaemic attack or stroke (Thijs et al 2008). Data are available for seven out of ten possible direct comparisons and none of them was found to be statistically significant in terms of incoherence.

In the special case where direct and several independent indirect estimates are available, the 'composite Chi2 statistic' can be used instead (Caldwell et al 2010).

The loop-specific approach described in Section 11.4.4.1 can be extended to networks with many interventions by evaluating incoherence separately in each closed loop of evidence. The approach can be performed using the 'ifplot' macro available for the

Stata statistical package (Chaimani and Salanti 2015). However, unlike the SIDE approach, this method does not incorporate the information from the entire network when estimating the indirect evidence.

Tests for incoherence have low power and therefore may fail to detect incoherence as statistically significant even when it is present (Song et al 2012, Veroniki et al 2014). This means that the absence of statistically significant incoherence is not evidence for the absence of incoherence. More preferably, review authors should consider the confidence interval of incoherence factors and decide whether the confidence interval includes values that are sufficiently large to suggest clinically important discrepancies between direct and indirect evidence. It should be noted that statistical incoherence can be assessed only for parts of the network with available direct evidence. Considerations upon the presence of incoherence are discussed in Section 11.4.4.4.

Table 11.4.c. Results based on the SIDE approach to evaluating local incoherence. P values less than 0.05 suggest statistically significant incoherence.

Comparison	Direct		Indirect		Incoherence factor		
	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error	<i>P value</i>
A versus C	-0.15	0.05	-0.21	0.10	0.07	0.12	<i>0.56</i>
A versus D	-0.45	0.07	-0.32	0.11	-0.14	0.13	<i>0.28</i>
A versus E	-0.26	0.14	-0.23	0.07	-0.03	0.16	<i>0.85</i>
B versus C	0.18	0.11	0.13	0.08	0.05	0.14	<i>0.70</i>
B versus E	0.07	0.07	0.12	0.12	-0.05	0.14	<i>0.70</i>
C versus D	-0.23	0.06	-0.35	0.12	0.12	0.13	<i>0.38</i>
C versus E	-0.06	0.05	-0.11	0.10	0.05	0.11	<i>0.66</i>

411.4.4.3 Approaches to evaluating global incoherence

Global incoherence in a network can be evaluated and detected via incoherence models. These models differ from the coherence models described in Section 11.4.3.1 by relaxing the coherence equations (see Section 11.2.3.2) and allowing intervention effects to vary when estimated directly and indirectly (Lu and Ades 2006). The models add additional terms, equivalent to the incoherence factors (IFs) defined in Section 11.4.4.1, to the coherence equations. For example, in the dietary advice network the coherence equation given in Section 11.2.3.2 would be modified to:

$$\text{'true' indirect MD (BvsC)} = \text{'true' direct MD (AvsC)} - \text{'true' direct MD (AvsB)} + \text{IF}_{ABC}$$

The quantity IF_{ABC} measures incoherence in the evidence loop 'dietitian-doctor-nurse'. Obviously, complex networks will have several IFs. For a network to be coherent, all IF need to be close to zero. This can be formally tested via a Chi² statistic test which is available in Stata in the 'network' macro (White 2015). An extension of this model has been suggested where incoherence measures the disagreement when an effect size is measured in studies that involve different sets of interventions (termed 'design incoherence') (Higgins et al 2012).

Measures like the Q-test and the I² statistic, which are commonly used for the evaluation of heterogeneity in a pairwise meta-analysis (see Chapter 10, Section 10.10.2), have been developed for the assessment of heterogeneity and incoherence in network meta-analysis (Krahn et al 2013, Rucker and Schwarzer 2013, Jackson et al 2014). These have been implemented in the package 'netmeta' in R (Schwarzer et al 2015).

11.4.4.4 Forming conclusions about incoherence

We suggest review authors use both local and global approaches and consider their results jointly before making inferences about incoherence. The approaches presented in Sections 11.4.4.2 and 11.4.4.3 for evaluating incoherence have limitations. As for tests for statistical heterogeneity in a standard pairwise meta-analysis, tests for detecting incoherence often lack power to detect incoherence when it is present, as shown in simulations and empirical studies (Song et al 2012, Veroniki et al 2014). Also, different

assumptions and different methods in the estimation of heterogeneity may impact on the findings about incoherence (Veroniki et al 2013, Veroniki et al 2014). Empirical evidence suggests that review authors sometimes assess the presence of incoherence, if at all, using inappropriate methods (Veroniki et al 2013, Nikolakopoulou et al 2014, Petropoulou et al 2016).

Conclusions should be drawn not just from consideration of statistical significance but by interpreting the range of values included in confidence intervals of the incoherence factors. Researchers should remember that the absence of statistically significant incoherence does not ensure transitivity in the network, which should always be assessed before undertaking the analysis (see Section 11.2.2.3).

Once incoherence is detected, possible explanations should be sought. Errors in data collection, broad eligibility criteria, and imbalanced distributions of effect modifiers may have introduced incoherence. Possible analytical strategies in the presence of incoherence are available (Salanti 2012, Jansen and Naci 2013).

11.5 Evaluating confidence in the results of a network meta-analysis

The GRADE approach is recommended for use in Cochrane Reviews to assess the confidence of the evidence for each pairwise comparison of interventions (see Chapter 14). The approach starts by assuming high confidence in the evidence for randomized trials of a specific pairwise comparison and then downgrades the evidence for considerations of five issues: study limitations, indirectness, inconsistency, imprecision and publication bias.

Rating the confidence in the evidence from a network of interventions is more challenging than pairwise meta-analysis (Dumville et al 2012). To date, two frameworks have been suggested in the literature to extend the GRADE system to indirect comparisons and network meta-analyses: Salanti and colleagues (Salanti et al 2014) and Puhan and colleagues (Puhan et al 2014). Section 11.5.1 describes the principles of each approach, noting similarities and differences.

11.5.1 Available approaches for evaluating confidence in the evidence

The two available approaches to evaluating confidence in evidence from a network meta-analysis acknowledge that the confidence in each combined comparison depends on the confidence in the direct and indirect comparisons that contribute to it, and that the confidence in each indirect comparison in turn depends on the confidence in the pieces of direct evidence that contribute to it. Therefore, all GRADE assessments are built to some extent on applying GRADE ideas for direct evidence. The two approaches diverge in the way they combine the considerations when thinking about an indirect or combined comparison.

More specifically, since indirect and combined comparisons are estimated by combining the information on two or more direct comparisons (See Sections 11.2 and 11.4), the confidence in each direct piece of evidence involved may be used to rate the confidence in the indirect evidence for this comparison. Then, they can be integrated to rate an indirect comparison following two possible ways which are illustrated in Table 11.5.a using the dietary advice example.

At the time of writing, no formal comparison has been performed to evaluate the degree of agreement between these two methods. Thus, at this point we do not prescribe using one approach or the other. However, when indirect comparisons are built on existing pairwise meta-analyses, which have already been rated with respect to their confidence, it may be reasonable to follow the Puhan et al approach. On the other hand, when the body of evidence is built from scratch, or when a large number of interventions are involved, it may be preferable to consider the Salanti and colleagues approach whose application is facilitated via the online tool CINeMA (Confidence in Network Meta-Analysis, <http://cinema.ispm.ch/>).

The framework by Salanti and colleagues is driven by the ability to express each estimated intervention effect from a network meta-analysis as a weighted sum of all the available direct comparisons (see Section 11.4) (Lu et al 2011, König et al 2013, Krahn et al 2013). The weight is determined, under some assumptions, by the **contribution matrix**, which has been implemented in the 'netweight' macro (Chaimani and Salanti 2015) available for the Stata statistical package and programmed in CINeMA. The matrix contains the percentage of information attributable to each direct comparison estimate and can be interpreted as the **contributions** of the direct comparison estimates. Then, the confidence in an indirect or combined comparison is estimated by combining the confidence assessment for the available direct comparison estimates with their contribution to the combined (or network) comparison. This approach is similar to the process of evaluating the likely impact of a high risk of bias study by looking at its weight in a pairwise meta-analysis to decide whether to downgrade or not in a standard GRADE assessment.

As an example, in the dietary advice network (Figure 11.2.a) suppose that most of the evidence involved in the indirect comparison (i.e. the trials including dietitians) is at low risk of bias, and that there are studies of 'doctor versus nurse' that are mostly at high risk of bias. If the direct evidence on 'doctor versus nurse' has a very large contribution to the network meta-analysis estimate of the

same comparison, then we would judge this result to be at high risk of bias. If the direct evidence has a very low contribution, we might judge the result to be at moderate, or possibly low, risk of bias. This approach might be preferable when there are indirect or mixed comparisons informed by many loops within a network, and for a specific comparison these loops lead to different risk of bias assessments. The contributions of the direct comparisons and the risk of bias assessments may be presented jointly in a bar graph with bars proportional to the contributions of direct comparisons and different colours representing the different judgments. The bar graph for the heavy menstrual bleeding example is available in Figure 11.5.a, which suggests that there are two comparisons (first generation hysteroscopic techniques versus Mirena and second generation non-hysteroscopic techniques versus Mirena) for which a substantial amount of information comes from studies at high risk of bias.

Regardless of whether a review contains a network meta-analysis or a simple indirect comparison, Puhan and colleagues propose to focus on so-called 'most influential' loops only. These are the connections between a pair of interventions of interest that involve exactly one common comparator. This implies that the assessment for the indirect comparison is dependent only on confidence in the two other direct comparisons in this loop. To illustrate, consider the dietary advice network described in Section 11.2 (Figure 11.2.a), where we are interested in confidence in the evidence for the indirect comparison 'doctor versus nurse'. According to Puhan and colleagues, the lower confidence rating between the two direct comparisons 'dietitian versus doctor' and 'dietitian versus nurse' would be chosen to inform the confidence rating for the indirect comparison. If there are also studies directly comparing doctor versus nurse, the confidence in the combined comparison would be the higher rated source between the direct evidence and the indirect evidence. The main rationale for this is that, in general, the higher rated comparison is expected to be the more precise (and thus the dominating) body of evidence. Also, in the absence of important incoherence, the lower rated evidence is only supportive of the higher rated evidence; thus it is not very likely to reduce the confidence in the estimated intervention effects. One disadvantage of this approach is that investigators need to identify the most influential loop; this loop might be relatively uninfluential when there are many loops in a network, which is often the case when there are many interventions. In large networks, many loops with comparable influence may exist and it is not clear how many of those equally influential loops should be considered under this approach.

Since network meta-analysis produces estimates for several intervention effects, the confidence in the evidence should be assessed for each intervention effect that is reported in the results. In addition, network meta-analysis may also provide information on the relative ranking of interventions and any concerns about confidence in the evidence also pertain to this output. Consequently, confidence in the evidence should also be considered in interpreting the relative ranking results when these are reported. Salanti and colleagues addressed this based on the contributions of the direct comparisons to the entire network as well as on the use of measures and graphs that aim to assess the different GRADE domains in the network together (e.g. measures of global incoherence (see Section 11.4.4).

The two approaches modify the standard GRADE domains to fit network meta-analysis to varying degrees. These modifications are briefly described in Box 11.5.1.a; more details and examples are available in the original articles (Puhan et al 2014, Salanti et al 2014).

Table 11.5.a. Steps to obtain the overall confidence ratings (across all GRADE domains) for every combined comparison of the dietary advice example. A ✓ or x indicates whether a particular step is needed in order to proceed to the next step.

Direct comparisons	GRADE domains	Step 1		Step 2				Step 3	
		Domain-specific ratings for direct comparisons		Overall rating across domains for direct comparisons		Domain-specific ratings for combined comparisons		Overall rating across domains for combined comparisons	
		Salanti et al	Puhan et al	Salanti et al	Puhan et al	Salanti et al	Puhan et al	Salanti et al	Puhan et al
Dietitian vs nurse	Study limitations	✓	✓	x	✓	✓	x	✓	✓
	Indirectness	✓	✓			✓	x		
	Inconsistency	✓	✓			✓	x		
	Imprecision	-	-			✓	x		
	Publication bias	✓	✓			✓	x		
Dietitian vs doctor	Study limitations	✓	✓	x	✓	✓	x	✓	✓
	Indirectness	✓	✓			✓	x		
	Inconsistency	✓	✓			✓	x		
	Imprecision	-	-			✓	x		
	Publication bias	✓	✓			✓	x		
	Study limitations	✓	✓	x	✓	✓	x	✓	✓
Nurse vs doctor	Indirectness	✓	✓			✓	x		
	Inconsistency	✓	✓			✓	x		
	Imprecision	-	-			✓	x		
	Publication bias	✓	✓			✓	x		

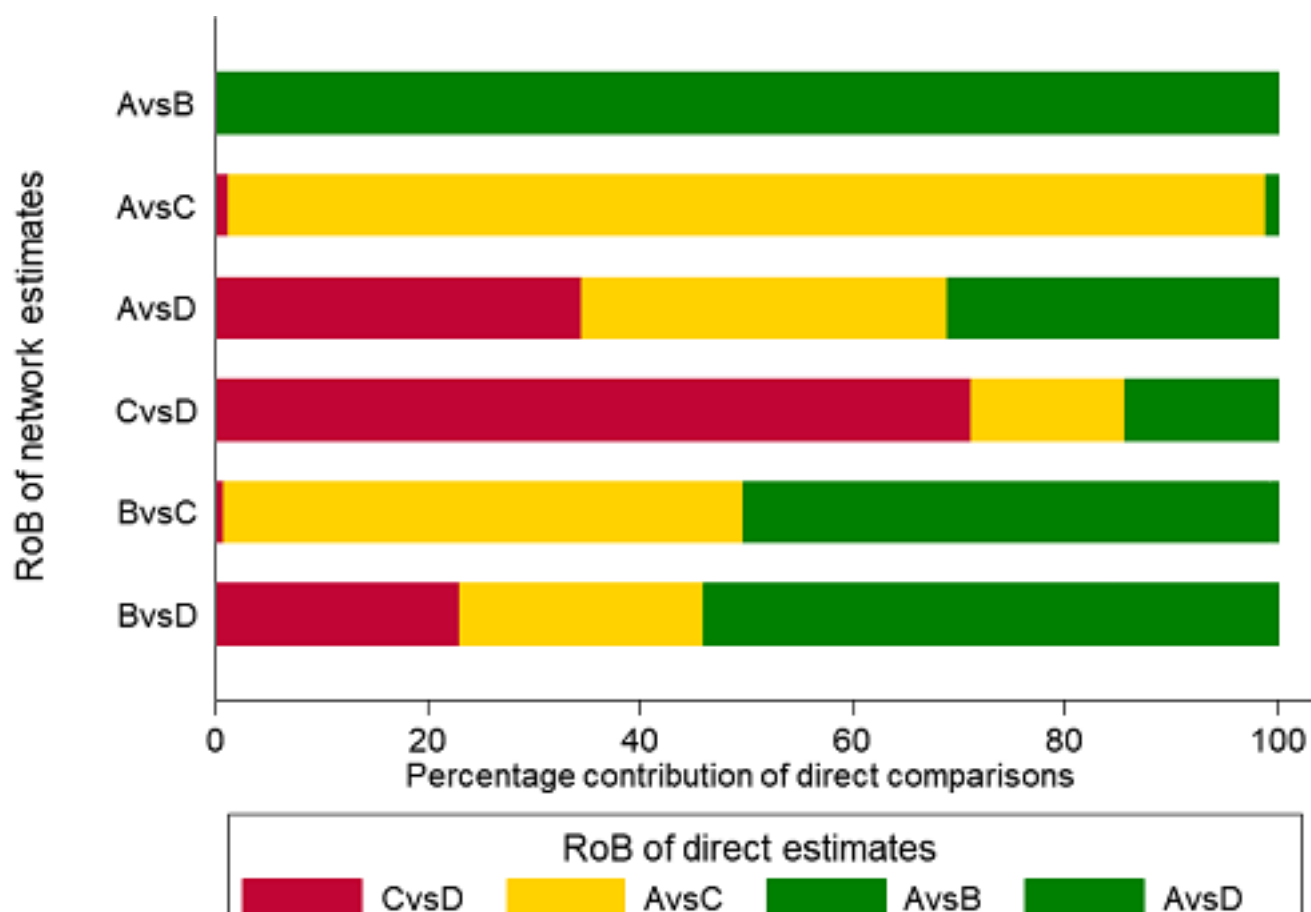


Figure 11.5.a. Bar graph illustrating the percentage of information for every comparison that comes from low (dark grey), moderate (light grey), or high (black) risk of bias studies with respect to both randomization and compliance to treatment for the heavy menstrual bleeding network (Middleton et al 2010). The risk of bias of the direct comparisons was defined based on Appendix 3 of the original paper. The intervention labels are A: first generation hysteroscopic techniques, B: hysterectomy, C: second generation non-hysteroscopic techniques, D: Mirena.

Box 11.5.1.a. Modifications to the five domains of the standard GRADE system to fit network meta-analysis.

Study limitations (i.e. classical risk of bias items): Salanti and colleagues suggest a bar graph with bars proportional to the contributions of direct comparisons and different colours representing the different confidence ratings (e.g. green, yellow, red for low, moderate or high risk of bias) with respect to study limitations (Figure 11.5.a). The decision about downgrading or not is then formed by interpreting this graph. Such a graph can be used to rate the confidence of evidence for each combined comparison and for the relative ranking.

Indirectness: The assessment of indirectness in the context of network meta-analysis should consider two components: the similarity of the studies in the analysis to the target question (PICO); and the similarity of the studies in the analysis to each other. The first addresses the extent to which the evidence at hand relates to the population, intervention(s), comparators and outcomes of interest and the second relates to the evaluation of the transitivity assumption. A common view of the two approaches is that they do not support the idea of downgrading indirect evidence by default. They suggest that indirectness should be considered in conjunction with the risk of intransitivity.

Inconsistency: Salanti and colleagues propose to create a common domain to consider jointly both types of inconsistency that may occur: heterogeneity within direct comparisons and incoherence. More specifically, they evaluate separately the presence of the two types of variation and then consider them jointly to infer whether downgrading for inconsistency is appropriate or not. It is usual in network meta-analysis to assume a common heterogeneity variance. They propose the use of prediction intervals to facilitate the assessment of heterogeneity for each combined comparison. Prediction intervals are the intervals expected to include the true intervention effects in future studies (Higgins et al 2009, Riley et al 2011) and they incorporate the extent of between-study variation; in the presence of important heterogeneity they are wide enough to include intervention effects with different implications for practice. The potential for incoherence for a particular comparison can be assessed using existing approaches for evaluating local and global incoherence (see Section 11.5). We may downgrade for one or two levels due to the presence of heterogeneity or incoherence, or both. The judgment for the relative ranking is based on the magnitude of the

common heterogeneity as well as the use of global incoherence tests (see Section 11.4).

Imprecision: Both approaches suggest that imprecision of the combined comparisons can be judged based on their 95% confidence intervals. Imprecision for relative treatment ranking is the variability in the relative order of the interventions. This is reflected by the overlap in the distributions of the ranking probabilities; i.e. when all or some of the interventions have similar probabilities of being at a particular rank.

Publication bias: The potential for publication bias in a network meta-analysis can be difficult to judge. If a natural common comparator exists, a 'comparison-adjusted funnel plot' can be employed to identify possible small-study effects in a network meta-analysis (Chaimani and Salanti 2012, Chaimani et al 2013). This is a modified funnel plot that allows putting together all the studies of the network irrespective of the interventions they compare. However, the primary considerations for both the combined comparisons and relative ranking should be non-statistical. Review authors should consider whether there might be unpublished studies for every possible pairwise comparison in the network.

11.6 Presenting network meta-analyses

The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions should be considered when reporting the results from network meta-analysis (Hutton et al 2015). Key graphical and numerical summaries include the network plot (e.g. Figure 11.4.a), a league table of the relative effects between all treatments with associated uncertainty (e.g. Table 11.4.a), and measures of heterogeneity and incoherence.

11.6.1 Presenting the evidence base of a network meta-analysis

Network diagrams provide a convenient way to describe the structure of the network (see Section 11.2.3.4). They may be modified to incorporate information on study-level or comparison-level characteristics. For instance, the thickness of the lines might reflect the number of studies or patients included in each direct comparison (e.g. Figure 11.4.a), or the comparison-specific average of a potential effect modifier. Using the latter device, network diagrams can be considered as a first step for the evaluation of transitivity in a network. In the example of Figure 11.6.a the age of the participants has been considered as a potential effect modifier. The thickness of the line implies that the average age within comparisons A versus D and C versus D seems quite different to the other three direct comparisons.

The inclusion of studies with design limitations in a network (e.g. lack of blinding, inadequate allocation sequence concealment) often threatens the validity of findings. The use of coloured lines in a network of interventions can reveal the presence of such studies in specific direct comparisons. Further discussion on issues related to confidence in the evidence is available in Section 11.5.

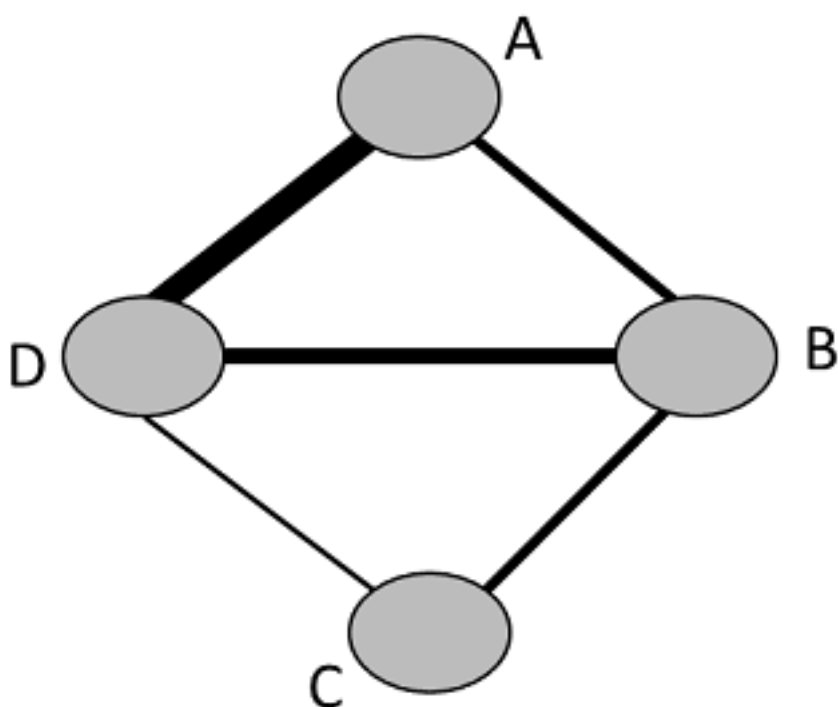


Figure 11.6.a. Example of network diagram with lines weighted according to the average age within each pairwise comparison. Thicker lines correspond to greater average age within the respective comparison.

11.6.2 Tabular presentation of the network structure

For networks including many competing interventions and multiple different study designs, network diagrams might be the most appropriate tool for presenting the data. An alternative way to present the structure of the network is to use a table, in which the columns represent the competing interventions and the rows represent the different study designs in terms of interventions being compared (Table 11.6.a) (Lu and Ades 2006). Additional information, such as the number of participants in each arm, may be presented in the non-empty cells.

Table 11.6.a. Example of table presenting a network that compares seven interventions and placebo for controlling exacerbation of episodes in chronic obstructive pulmonary disease (Baker et al 2009).

Number of studies	Placebo	Fluticasone	Budesonide	Salmeterol	Formoterol	Tiotropium	Fluticasone + Salmeterol	Budesonide + Formoterol
4	x	x		x			x	
4	x	x						
2	x		x		x			x
2	x			x		x		
2	x			x			x	
8	x			x				
2	x				x			
10	x					x		
1	x						x	

1				x		x		
1				x			x	
1					x	x		
1						x	x	

11.6.3 Presenting the flow of evidence in a network

Another way to map the evidence in a network of interventions is to consider how much each of the included direct comparisons contributes to the final combined effect estimates. The percentage information that direct evidence contributes to each relative effect estimated in a network meta-analysis can be presented in the contribution matrix (see Section 11.4), and could help investigators understand the flow of information in the network (Chaimani et al 2013, Chaimani and Salanti 2015).

Figure 11.6.b presents the contribution matrix for the example of the network of interventions for heavy menstrual bleeding (obtained from the 'netweight' macro in Stata). The indirect treatment effect for second generation non-hysteroscopic techniques versus hysterectomy (B versus C) can be estimated using information from the four direct relative treatment effects; these contribute information in different proportions depending on the precision of the direct treatment effects and the structure of the network. Evidence from the direct comparison of first generation hysteroscopic techniques versus hysterectomy (A versus B) has the largest contribution to the indirect comparisons hysterectomy versus second generation non-hysteroscopic techniques (B versus C) (49.6%) and hysterectomy versus Mirena (B versus D) (38.5%), for both of which no direct evidence exists.

		Direct comparisons in the network (% contribution)			
		A-B	A-C	A-D	C-D
Network meta-analysis estimates	Mixed estimates				
	A-B	100.0			
	A-C		97.8	1.1	1.1
	A-D		34.5	31.1	34.5
	C-D		14.4	14.4	71.2

	Indirect estimates				
	B-C	49.6	48.9	0.7	0.7
	B-D	38.5	23.0	15.5	23.0
Entire network		31.4	36.4	10.5	21.7
Included studies		5	11	1	3

Figure 11.6.b. Contribution matrix for the network on interventions for heavy menstrual bleeding presented in Figure 11.4.a. Four direct comparisons in the network are presented in the columns, and their contributions to the combined treatment effect are presented in the rows. The entries of the matrix are the percentage weights attributed to each direct comparison. The intervention labels are A: first generation hysteroscopic techniques, B: hysterectomy, C: second generation non-hysteroscopic techniques, D: Mirena.

11.6.4 Presentation of results

Unlike pairwise meta-analysis, the results from network meta-analysis cannot be easily summarized in a single figure, such as a standard forest plot. Especially for networks with many competing interventions that involve many comparisons, presentation of findings in a concise and comprehensible way might be challenging.

Summary statistics of the intervention effects for all pairs of interventions are the most important output from network meta-analysis. Results from a subset of comparisons is sometimes presented due to space limitations and the choice of the findings to be reported is based on the research question and the target audience (Tan et al 2013). In such cases, the use of additional figures and tables to present all results in detail is necessary. Additionally, review authors might wish to report the relative ranking of interventions (see Section 11.4.3.3) as a supplementary output, which provides a concise summary of the findings and might facilitate decision making. For this purpose, joint presentation of both relative effects and relative ranking is recommended (see Figure 11.6.c or Table 11.4.a of Section 11.4.3.1).

In the presence of many competing interventions, the results across different outcomes (e.g. efficacy and acceptability) might be contradicting with respect to which interventions work best. To avoid drawing misleading conclusions, review authors may consider the simultaneous presentation of results for outcomes in these two categories.

Interpretation of the findings from network meta-analysis should always be considered with the evidence characteristics: risk of bias in included studies, heterogeneity, incoherence, and selection bias. Reporting results with respect to the evaluation of incoherence and heterogeneity (such as I² statistic for incoherence) is important for drawing meaningful conclusions.

11.6.4.1 Presentation of intervention effects and ranking

A table presenting direct, indirect, and network summary relative effects along with their confidence ratings is a helpful format (Puhan et al 2014). In addition, various graphical tools have been suggested for the presentation of results from network meta-analyses (Salanti et al 2011, Chaimani et al 2013, Tan et al 2014). Summary relative effects for pairwise comparisons with their confidence intervals can be presented in a forest plot. For example, Figure 11.6.c shows the summary relative effects for each intervention versus a common reference intervention for the 'heavy menstrual bleeding' network.

Ranking probabilities for all possible ranks may be presented by drawing probability lines, which are known as 'rankograms', and show the distribution of ranking probabilities for each intervention (Salanti et al 2011). The rankograms for the 'heavy menstrual bleeding' network example are shown in Figure 11.6.d. The graph suggests that 'Hysterectomy' has the highest probability of being the best intervention, 'First generation hysteroscopic techniques' have the highest probability of being worst followed by 'Mirena' and 'Second generation non-hysteroscopic techniques' have equal chances of being second or third.

The relative ranking for two (competing) outcomes can be presented jointly in a two-dimensional scatterplot (Chaimani et al 2013). An extended discussion on different ways to present jointly relative effects and relative ranking from network meta-analysis is available in (Tan et al 2013).

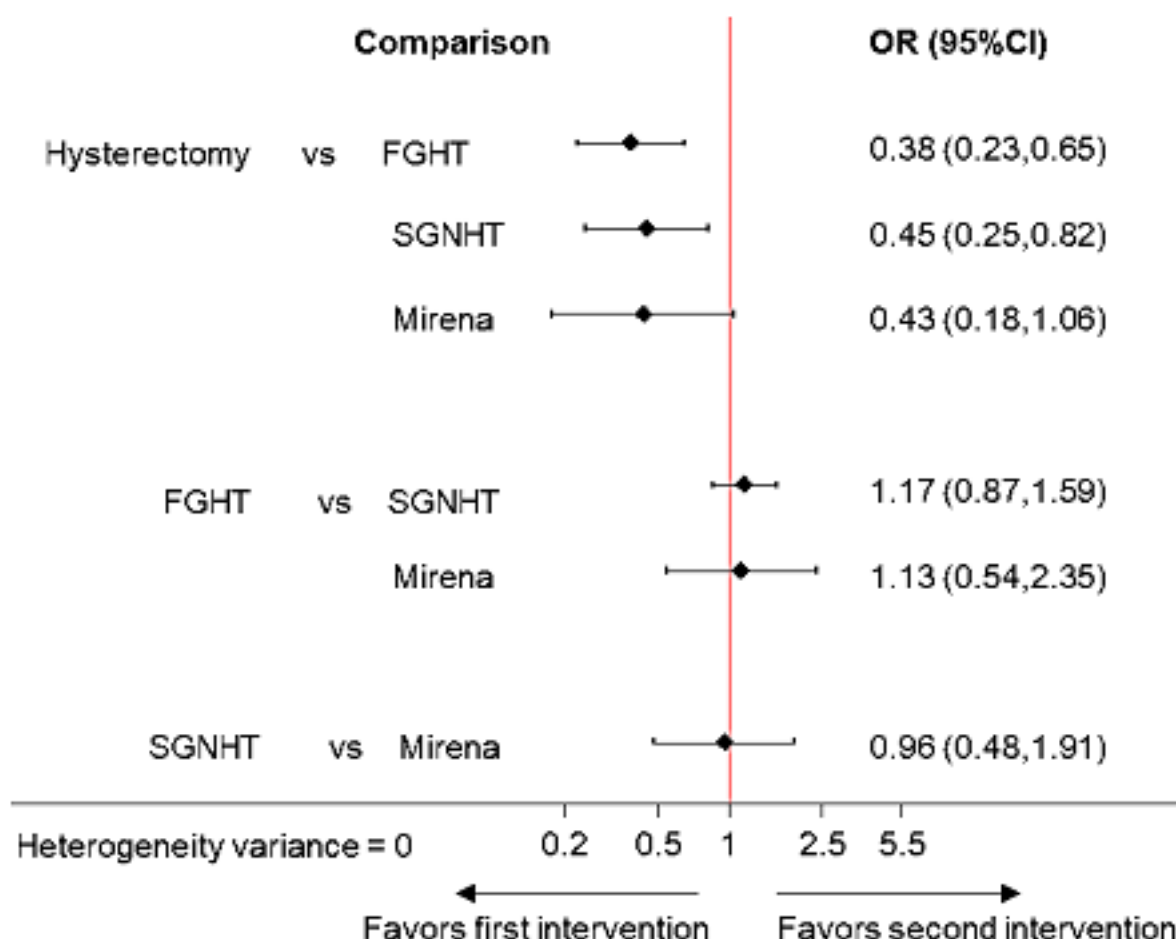


Figure 11.6.c. Forest plot for effectiveness in heavy menstrual bleeding between four interventions. FGHT: first generation hysteroscopic techniques, SGNHT: second generation non-hysteroscopic techniques.

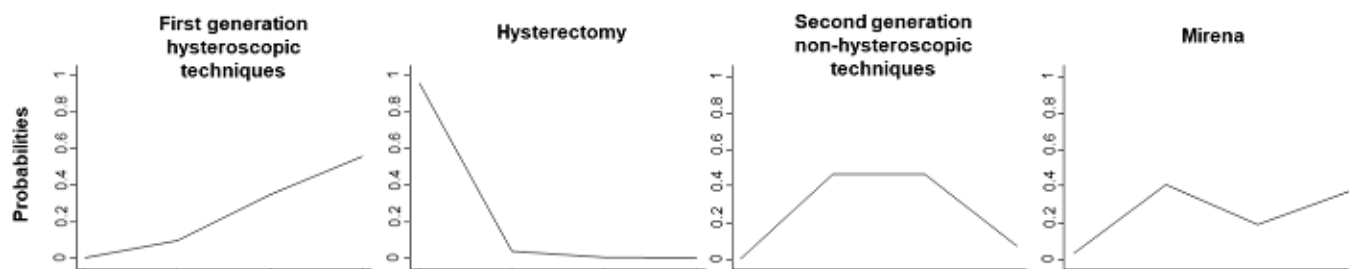


Figure 11.6.d. Ranking probabilities (rankograms) for the effectiveness of interventions in heavy menstrual bleeding. The horizontal axis shows the possible ranks and the vertical axis the ranking probabilities. Each line connects the estimated probabilities of being at a particular rank for every intervention.

11.6.4.2 Presentation of heterogeneity and incoherence

The level of heterogeneity in a network of interventions can be expressed via the magnitude of the between-study variance τ^2 , typically assumed to be common in all comparisons in the network. A judgment on whether the estimated τ^2 suggests the presence of important heterogeneity depends on the clinical outcome and the type of interventions being compared. More extended discussion on the expected values of tau-squared specific to a certain clinical setting is available (Turner et al 2012, Nikolakopoulou et al 2014).

Forest plots that present all the estimated incoherence factors in the network and their uncertainty may be employed for the presentation of local incoherence (Salanti et al 2009, Chaimani et al 2013). The results from evaluating global incoherence can be summarized in the P value of the Chi² statistic incoherence test and the I² statistic for incoherence (see Chapter 10, Section 10.10.2).

11.6.4.3 'Summary of findings' tables

The purpose of 'Summary of findings' tables in Cochrane Reviews is to provide concisely the key information in terms of available data, confidence in the evidence and intervention effects (see Chapter 14). Providing such a table is more challenging in reviews that compare multiple interventions simultaneously, which very often involve a large number of comparisons between pairs of interventions. A general principle is that the comparison of multiple interventions is the main feature of a network meta-analysis, so is likely to drive the structure of the 'Summary of findings' table. This is in contrast to the 'Summary of findings' table for a pairwise comparison, whose main strength is to facilitate comparison of effects on different outcomes. Nevertheless, it remains important to be able to compare network meta-analysis results across different outcomes. This provides presentational challenges that are almost impossible to resolve in two dimensions. One potential solution is an interactive electronic display such that the user can choose whether to emphasize the comparisons across interventions or the comparisons across outcomes.

For small networks of interventions (perhaps including up to five competing interventions) a separate 'Summary of findings' table might be produced for each main outcome. However, in the presence of many (more than five) competing interventions, researchers would typically need to select and report a reduced number of pairwise comparisons. Review authors should provide a clear rationale for the choice of the comparisons they report in the 'Summary of findings' tables. For example, they may consider including only pairwise comparisons that correspond to the decision set of interventions; that is the group of interventions of direct interest for drawing conclusions (see Section 11.3.2.1). The distinction between the decision set and the wider synthesis comparator set (all interventions included in the analysis) should be made in the protocol of the review. If the decision set is still too large, researchers may be able to select the comparisons for the 'Summary of findings' table based on the most important information for clinical practice. For example, reporting the comparisons between the three or four most effective interventions with the most commonly used intervention as a comparator.

11.7 Concluding remarks

Network meta-analysis is a method that can inform comparative effectiveness of multiple interventions, but care needs to be taken using this method because it is more statistically complex than a standard meta-analysis. In addition, as network meta-analyses generally ask broader research questions, they usually involve more studies at each step of systematic review, from screening to analysis, than standard meta-analysis. It is therefore important to anticipate the expertise, time, and resource required before embarking on one.

A valid indirect comparison and network meta-analysis requires a coherent evidence base. When formulating the research question and deciding the eligibility criteria, populations and interventions in relation to the assumption of transitivity need to be considered.

Network meta-analysis is only valid when studies comparing different sets of interventions are similar enough to be combined. When conducted properly, it provides more precise estimates of relative effect than a single direct or indirect estimate. Network meta-analysis can yield estimates between any pairs of interventions, including those that have never been compared directly against each other. Network meta-analysis also allows the estimation of the ranking and hierarchy of interventions. Much care should be taken when interpreting the results and drawing conclusions from network meta-analysis, especially in the presence of incoherence or other potential biases.

11.8 Chapter information

Authors: Anna Chaimani, Deborah M Caldwell, Tianjing Li, Julian PT Higgins, Georgia Salanti.

Acknowledgements: Lorne Becker contributed important insights in the discussion of separating Overview from Intervention Reviews with network meta-analysis. Gordon Guyatt provided helpful comments on earlier version of the Chapter and Jeroen Jansen provided helpful contribution on Section 2.

11.9 References

- Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. *Medical Decision Making* 2013;33: 679-91.
- Baker WL, Baker EL, Coleman CI. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. *Pharmacotherapy* 2009;29: 891-905.
- Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology* 1997;50: 683-91.
- Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331: 897-900.
- Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *Journal of Clinical Epidemiology* 2010;63: 875-82.
- Caldwell DM, Dias S, Welton NJ. Extending treatment networks in health technology assessment: how far should we go? *Value in Health* 2015;18: 673-81.
- Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Research Synthesis Methods* 2012;3: 161-76.
- Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013;8: e76654.
- Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: the network graphs package. *Stata Journal* 2015;15: 905-50.
- Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Additional considerations are required when preparing a protocol for a systematic review with multiple interventions. *Journal of Clinical Epidemiology* 2017;83: 65-74.
- Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine* 2013;159: 130-7.
- Cooper NJ, Peters J, Lai MC, Juni P, Wandel S, Palmer S, Paulden M, Conti S, Welton NJ, Abrams KR, Bujkiewicz S, Spiegelhalter D, Sutton AJ. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value in Health* 2011;14: 371-80.
- Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* 2010;29: 932-44.
- Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network meta-analysis for decision-making*. Chichester (UK): Wiley; 2018.

Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: a review of reporting and methodological quality. *PLoS One* 2010;5: e11054.

Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Research Synthesis Methods* 2013;4: 291-323.

Dumville JC, Soares MO, O'Meara S, Cullum N. Systematic review and mixed treatment comparison: dressings to heal diabetic foot ulcers. *Diabetologia* 2012;55: 1902-10.

Efthimiou O, Debray TPA, Valkenhoef G, Trelle S, Panayidou K, Moons KGM, Reitsma JB, Shang A, Salanti G, GetReal Methods Review Group. GetReal in network meta-analysis: a review of the methodology. *Research Synthesis Methods* 2016;7: 236-63.

Giovane CD, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Statistics in Medicine* 2013;32: 25-39.

Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ. Indirect comparisons of competing interventions. *Health Technology Assessment* 2005;9: 1-iv.

Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996;15: 2733-49.

Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society Series A, (Statistics in Society)* 2009;172: 137-59.

Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* 2012;3: 98-110.

Hughes S. First "comparison" of prasugrel and ticagrelor. *International Journal of Cardiology* 2010.

Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JPA, Straus S, Thorlund K, Jansen JP, Mulrow C, Catalá-López F, Gøtzsche PC, Dickersin K, Boutron I, Altman DG, Moher D. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine* 2015;162: 777-84.

Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Statistics in Medicine* 2011;30: 2481-98.

Jackson D, Barrett JK, Rice S, White IR, Higgins JPT. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Statistics in Medicine* 2014;33: 3639-54.

James A, Yavchitz A, Ravaud P, Boutron I. Node-making process in network meta-analysis of nonpharmacological treatment are poorly reported. *J Clin Epidemiol* 2018;97: 95-102.

Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Medicine* 2013;11: 159.

Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: An ISPOR-AMCP-NPC good practice task force report. *Value in Health* 2014;17: 157-73.

König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Statistics in Medicine* 2013;32: 5414-29.

Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Medical Research Methodology* 2013;13: 35.

Li T, Dickersin K. Citation of previous meta-analyses on the same topic: a clue to perpetuation of incorrect methods? *Ophthalmology* 2013;120: 1113-9.

Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004;23: 3105-24.

Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*

2006;101: 447-59.

Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009;10: 792-805.

Lu G, Welton NJ, Higgins JPT, White IR, Ades AE. Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Research Synthesis Methods* 2011;2: 43-60.

Madan J, Stevenson MD, Cooper KL, Ades AE, Whyte S, Akehurst R. Consistency between direct and indirect trial evidence: is direct evidence always more reliable? *Value in Health* 2011;14: 953-60.

Madan J, Chen Y-F, Aveyard P, Wang D, Yahaya I, Munafo M, Bauld L, Welton N. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2014;177: 295-314.

Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research* 2013;22: 133-58.

Mawdsley D, Bennetts M, Dias S, Boucher M, Welton N. Model-based network meta-analysis: a framework for evidence synthesis of clinical trial data. *CPT: Pharmacometrics & Systems Pharmacology* 2016;5: 393-401.

Middleton LJ, Champaneria R, Daniels JP, Bhattacharya S, Cooper KG, Hilken NH, O'Donovan P, Gannon M, Gray R, Khan KS, International Heavy Menstrual Bleeding Individual Patient Data Meta-analysis Collaborative G, Abbott J, Barrington J, Bhattacharya S, Bongers MY, Brun JL, Busfield R, Sowter M, Clark TJ, Cooper J, Cooper KG, Corson SL, Dickersin K, Dwyer N, Gannon M, Hawe J, Hurskainen R, Meyer WR, O'Connor H, Pinion S, Sambrook AM, Tam WH, Zon-Rabelink IAA, Zupi E. Hysterectomy, endometrial destruction, and levonorgestrel releasing intrauterine system (Mirena) for heavy menstrual bleeding: systematic review and meta-analysis of data from individual patients. *BMJ (Clinical research ed)* 2010;341: c3929.

Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS One* 2014;9: e86754.

Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 2015;18: 116-26.

Page MJ, Higgins JPT, Clayton G, Sterne JAC, Hrobjartsson A, Savovic J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. *PLoS One* 2016;11: 7.

Petropoulou M, Nikolakopoulou A, Veroniki AA, Rios P, Vafaei A, Zarin W, Giannatsi M, Sullivan S, Tricco AC, Chaimani A, Egger M, Salanti G. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *Journal of Clinical Epidemiology* 2016;82: 20-8.

Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, Kessels AG, Guyatt GH, others. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349: g5630.

Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342: d549.

Rucker G. Network meta-analysis, electrical networks and graph theory. *Research Synthesis Methods* 2012;3: 312-24.

Rucker G, Schwarzer G. netmeta: An R package for network meta-analysis. The R project website 2013.

Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Medical Research Methodology* 2015;15: 58.

Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008;17: 279-301.

Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of Clinical Epidemiology* 2009;62: 857-64.

Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 2011;64: 163-71.

- Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* 2012;3: 80-97.
- Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9: e99682.
- Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. *Statistics in Medicine* 2013;32: 2935-49.
- Schwarzer G, Carpenter JR, Rücker G. *Meta-analysis with R*. Cham: Springer; 2015.
- Soares MO, Dumville JC, Ades AE, Welton NJ. Treatment comparisons for decision making: facing the problems of sparse and few data. *Journal of the Royal Statistical Society Series A, (Statistics in Society)* 2013.
- Sobieraj DM, Cappelleri JC, Baker WL, Phung OJ, White CM, Coleman CI. Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: a systematic review. *BMJ Open* 2013;3.
- Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326: 472.
- Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, Holland R, Chen YF, Glenny AM, Deeks JJ, Altman DG. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ* 2011;343: d4909.
- Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Medical Research Methodology* 2012;12: 138.
- Tan SH, Bujkiewicz S, Sutton A, Dequen P, Cooper N. Presentational approaches used in the UK for reporting evidence synthesis using indirect and mixed treatment comparisons. *Journal of Health Services Research and Policy* 2013;18: 224-32.
- Tan SH, Cooper NJ, Bujkiewicz S, Welton NJ, Caldwell DM, Sutton AJ. Novel presentational approaches were developed for reporting network meta-analysis. *Journal of Clinical Epidemiology* 2014;67: 672-80.
- Thijs V, Lemmens R, Fieuws S. Network meta-analysis: simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. *European heart journal* 2008;29: 1086-92.
- Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* 2012;41: 818-27.
- Veroniki AA, Vasiliadis HS, Higgins JPT, Salanti G. Evaluation of inconsistency in networks of interventions. *International Journal of Epidemiology* 2013;42: 332-45.
- Veroniki AA, Mavridis D, Higgins JPT, Salanti G. Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. *BMC Medical Research Methodology* 2014;14.
- White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 2012;3: 111-25.
- White IR. Network meta-analysis. *Stata Journal* 2015;15: 951-85.

Handbook information