Developing Critical Appraisal of Systematic Reviews Reporting Network Meta-Analysis

Andrew William Lee

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN EVIDENCE-BASED HEALTH CARE

KELLOGG COLLEGE, DEPARTMENT OF CONTINUING EDUCATION

UNIVERSITY OF OXFORD

MICHAELMAS 2020

Abstract

The aim of the research presented in this thesis was to advance the critical appraisal of systematic reviews using network meta-analysis (NMA) methods for synthesis. My review of systematic reviews using NMA methodology for synthesis and their characteristics found a rapid increase in publication of such reviews between 2008 and 2012, which has since continued. My survey of the authors of these reviews identified standards for reporting such reviews that were consistent with those subsequently included in the 2015 PRISMA extension statement for reporting of NMA. My programme of research then moved on to summarise how evidence from synthesis using NMA was used in NICE clinical guidelines published or updated during 2015-16. This found that although NMA methods were used far less often than pair-wise meta-analysis, they were used or considered for nearly one quarter of the guidelines reviewed. However, recommendations in the guidelines were more often based on results of NMA conducted by the Guideline Development Group than on results from previously published systematic reviews using NMA methodology. This finding, and reservations expressed by authors in my survey, supported the need to identify a critical appraisal tool to help users identify the quality of systematic reviews that use NMA methodology. After my review of existing critical appraisal tools failed to identify one suitable for systematic reviews using NMA methods for synthesis, I constructed a new tool, using the CASP format and based on items contained in the most current and most widely used related tools. The new tool was tested for inter-rater reliability and the findings of that testing and those from other elements in this programme of research allow the thesis to conclude with implications and some recommendations for further optimisation and validation of the tool.

Acknowledgments

I am grateful to Professor Mike Clarke for providing academic supervision throughout this doctoral thesis. His expertise, enthusiasm and patience have been essential to me completing this thesis. I'm grateful to Professor Tony Ades, Dr Deborah Caldwell and Dr Mark Simmons who also provided supervision at certain stages of it. I also wish to acknowledge Professors Amanda Burls and Carl Heneghan, who both provided encouragement and important feedback at various stages and Nia Roberts (Librarian) for providing search strategy advice. I'm grateful to Professor Alex Sutton for originally inspiring my interest in the use of network meta-analysis and kindly guiding me initially to some sources for further information.

I thank the volunteers from Hull York Medical School staff who participated in reliability testing of the new critical appraisal tool.

My love and thanks go to my family who have supported me in so many ways through many challenging events that coincided with undertaking this research work.

Personal Statement

My introduction to the use of evidence in health care was in the late 1970s as a second-year medical student, at the University of Sheffield Medical School, when I was fortunate enough to be amongst a handful of students invited by the late Professor Eric Wilkes to a talk, given at his home, by a visiting professor from McMaster University. I can't recall the name of the visiting academic, but his talk had a lasting impact on me. He explained that the McMaster approach to undergraduate medical education centred around self-directed learning, with appropriate support, with the aim that students learn how to use critically evidence sources to find answers to any clinical question themselves in order that they should be equipped with these skills to maintain and update their knowledge throughout their subsequent careers. He outlined the evidence that McMaster graduates, overall, achieved higher professional positions and scored more highly on tests of current medical knowledge, long after graduation, than their peers. My medical education had been, as for most medical students at that time, learning from textbooks and expert opinion. This talk opened my mind to a different approach to learning and created my aspiration, for subsequent clinical practice, to be able to access the best evidence to support decisions.

As my work in general practice has continued over the last thirty years, improved access to online evidence sources and the development of context-sensitive links to such resources while using computer medical record systems has fulfilled some of the early ambition I had for being able to use evidence in practice. The ability to access evidence, particularly in critically summarised and synthesised forms, has transformed the way clinicians and others, including commissioners and the public, as consumers of health care, can support their decision making.

I first became aware of network meta-analysis methodology in January 2009, when attending a talk by Alex Sutton, from the University of Leicester, at Oxford as part of the Masters in Evidence-Based Healthcare program. He described a method for combining direct and indirect comparisons using evidence from randomised controlled trials to produce a synthesis, which

linked all the interventions in a network in which each shared at least one direct comparison with another. He also demonstrated that this could produce a ranking of the likelihood of each of the multiple interventions being the most effective. The term used for this method had been 'multiple treatment comparison', though he mentioned that an alternative term, 'network meta-analysis', was gaining more favour. This methodology appeared to be a potential solution to two of the limitations to evidence-based practice I had always experienced as a clinician: firstly, the restriction of evidence synthesis in systematic reviews to pair-wise comparisons, when many clinical decisions involve more than two possible interventions, and secondly the lack of comparisons between all relevant interventions. This inspired me to choose network meta-analysis as the topic of the work summarised in this thesis.

Statement of Contributions

This doctoral thesis is independent and original work of which I am the sole author. Academic supervisors Professor Mike Clarke contributed feedback and guidance on the overall research strategy, methods, presentation and interpretation of results at each stage.

Professor Tony Ades and Dr Deborah Caldwell contributed feedback and guidance during the initial stages of developing the overall research strategy.

Dr Mark Simmons contributed feedback and guidance during the later stages of developing the overall research strategy, when I had to revise it following a pause in my research work. He also advised on and assisted with the statistical calculations of inter-rater reliability in Chapter Seven.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AED	Anti-epileptic drug
AHCPR	Agency for Health Care Policy and Research
AMSTAR	A measurement tool to assess systematic reviews
CASP	Critical appraisal skills programme
CDSR	Cochrane database of systematic reviews
CINAHL	Cumulative index to nursing and allied health literature
CPM	Confidence profile method
DARE	Database of abstracts of reviews of effects
DSU	Decision Support Unit
EMBASE	Excerpta Medica database
GDG	Guideline development group
HTA	Health technology assessment
IC	Indirect comparison
IPD	Individual participant data
IRR	Inter-rater reliability
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
JBI	Joanna Briggs Institute
MA	Meta-analysis
MECIR	Methodological expectations of Cochrane intervention reviews
MEDLINE	Medical literature analysis and retrieval system online
MeSH	Medical subject heading
MOMA	Multiple outcomes multivariate meta-analysis
MTC	Mixed treatment comparison or Multiple treatment comparison
NHS	National Health Service (UK)
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NNT	Number needed to treat
NRSI	Non-randomised study of intervention effects
OQAQ	Overview quality assessment questionnaire
PBAC	Pharmaceutical Benefits Advisory Committee (Australia)
PICO	Population; Intervention; Comparator; Outcome

PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PWMA	Pair-wise meta-analysis
QUOROM	Quality of reporting of meta-analyses
RCT	Randomised controlled trial
RoB	Risk of bias
ROBIS	Risk of bias in systematic reviews
SIGLE	System for information on grey literature
TSD	Technical support document.

CHAPTER ONE

Introduction

1 Introduction

1.1 Overview

This thesis explores trends in the use of network meta-analysis (NMA) methods for synthesis in systematic reviews, current reporting standards and the critical appraisal of such reviews.

1.2 Background

1.2.1 Systematic reviews

Evidence-based health care, whether it is used for decisions about the care of individual patients or in the development of health policy, requires at least narrative and, if appropriate, quantitative synthesis of clinical trials to identify the best available evidence. Systematic reviews of healthcare interventions aim to identify all trials relevant to answering a clinical question and provide a reliable estimate of the relative effects of the interventions included. Some of the key reasons for seeking all relevant trials include gathering together all the available evidence to make it easily accessible, analysing the collected evidence, minimising the risk of bias associated with only considering a sample of what could be relevant, informing further research and, because there is the potential to increase the statistical power and precision of estimates of effect, quantitatively combining data from multiple trials that are sufficiently similar. An essential element of conducting a systematic review is the critical appraisal of included studies in order to summarise their quality, or methodological rigour, sources of potential bias and applicability[1].

Attention to collecting all available research evidence on a clinical topic in a systematic way increased from the 1970s onwards[2]. By 2007, the estimated number of health-related systematic reviews published in one year was 2500[3] and by 2010, close to the time that I started this programme of research, the estimate was 4000[4]. Since then, it has continued to grow, with Page et al estimating 8000 systematic reviews were published during 2014[5] and Gurevitch et al estimating that the total available in the literature was beyond 200,000 by 2018[6].

1.2.2 Meta-analysis

Individual clinical trials rarely have sufficient power to provide definitive answers to clinical questions, so methods of statistical synthesis, known as meta-analysis, have been used to combine data from similar trials to achieve sufficient statistical power and improve precision of the estimate of effect size. Systematic reviews often report meta-analysis of data extracted from trials that meet the inclusion criteria for the review. It is recommended however, to only conduct meta-analysis using data from trials that are judged to be sufficiently similar in characteristics, of sufficient methodological quality and sufficiently free from bias[1].

Possibly the earliest example of the use of statistical techniques to combine data from multiple trials was published in 1904[7] but the term *'meta-analysis'* was first introduced in 1976[8], coinciding with the start of a period of increased attention to systematic reviews. The growth in publication of meta-analyses has since followed a similar trend to that of systematic reviews, with Lee et al[9] reporting that publication of health-related meta-analyses increased from 7 in 1980 to 400 in 2000. The numbers they identified were acknowledged to be an under-estimate of the true annual totals but provided a fair representation of an increasing trend in publication towards the end of the twentieth century, which has continued and gathered considerable pace in the 20 years since then. A search of MEDLINE in November 2019 for records tagged with both 'systematic review' and 'meta-analysis' MeSH terms and limited to human subjects retrieved 6820 records for publications during 2018.

Variability amongst studies included in a meta-analysis is known as heterogeneity. Such heterogeneity may be clinical (variability in the participants, interventions or outcomes studied), methodological (variability in an aspect of study design or other risk of bias) or statistical (variability in the estimates of intervention effects). The latter may be due to either of the former two or a combination of both[10]. Sometimes clinical heterogeneity will be the result of a deliberate choice e.g. to estimate effects on a wider population than has been studied in an individual trial

or to estimate effects of a group of interventions rather than a single one. It is standard practice to accompany the results of a meta-analysis with a statistical measure of heterogeneity e.g. the chi-squared test[11] or the I^2 statistic[12]. Potential reasons for the level of statistical heterogeneity found should be considered when interpreting the results of a meta-analysis[1].

One key decision when conducting meta-analysis is whether to use a fixed-effect or random-effects model. A fixed-effect model assumes that variability in the effects of an intervention across studies are by chance. A random-effects model also assumes that variability in the effects across studies are by chance but additionally that they are across a normal distribution of variation due to heterogeneity in the studies. A fixed-effect result has been described as a 'typical intervention effect' and a random-effects result an 'average intervention effect'[1]. The dependence of the findings of a meta-analysis on the choice to use a fixed-effect or random-effects model can be investigated in a sensitivity analysis. Sensitivity analyses are usually undertaken to assess the robustness of the results of meta-analysis are i.e. how they are affected by choices made in the conduct of statistical synthesis including the criteria for selection of studies for inclusion, the data combined and the analysis methods.

1.2.3 Network meta-analysis

The form of meta-analysis commonly used in systematic reviews since the 1970s is pair-wise meta-analysis. This can only be used to combine data from trials comparing directly the interventions of interest i.e. intervention A v. placebo, or intervention B v. intervention C. Health-related intervention trials commonly compare a new intervention with usual care or a placebo, or with an existing leader in the field. Over time, different interventions may be regarded as the leader or as a valid alternative to the new intervention. Therefore, seeking the answer to a current clinical question may involve combining data from multiple trials that tested more than two interventions. Not all these interventions will have been compared directly with each other and some might no longer be an option for a current decision. Primary research that

compares all the interventions is commonly not available, which leaves health practitioners and makers of health policy with an incomplete series of pair-wise comparisons and important gaps in the available evidence.

However, data from trials can also be synthesized to make indirect comparisons between interventions[13-16]. As with pair-wise meta-analysis, heterogeneity should be routinely considered when assessing the validity of the results of indirect comparison meta-analysis. NMA is an extension of pair-wise meta-analysis with the ability to compare simultaneously more than two interventions, to compare interventions that have not been directly compared in trials and to increase the precision of the estimate of effect size, but it is important that they are used and interpreted appropriately.

NMA can be used to estimate the effects of multiple interventions, by combining both indirect and direct comparisons. This requires a network in which each intervention shares a comparator with at least one other intervention. NMA provides a single synthesis of direct and indirect comparisons between all interventions in the network, potentially providing clinicians and policymakers with answers that make the maximum use of the available evidence, to compare multiple interventions.

Appropriate use of NMA methods is dependent on key assumptions, particularly those of transitivity and consistency, which assume that there are no systematic differences, other than the interventions being compared, between the direct comparisons whose results are being included in a synthesis and that, therefore, there is agreement between direct and indirect comparisons of each pair of interventions within the network. Assessing the validity of these requires input of both content and methodological expertise. These assumptions and details of the history and concept of NMA are described further in Chapter Two.

At the time of commencing the work described in this thesis in 2011, NMA was a relatively new form of statistical synthesis and the most recent review at that time, by Edwards et al, had found

use of NMA methodology was reported in a total of only 7 published systematic reviews up to 2007[17]. By 2015, the estimated number of published systematic reviews using NMA methodology had risen to 150 per year[18].

1.2.4 Critical appraisal

Critical appraisal involves assessing the report of a study for methodological quality, any likelihood of bias affecting the reported results and applicability to the situation the end-user is trying to address. This assessment should be applied both to reports of individual trials and to those of systematic reviews[19]. The importance of critical appraisal is illustrated by examples of the impact on research results that various aspects of study design can have[20-23]. Systematic reviews considering an identical question can produce inconsistent results, which is associated with variation in methodological quality[24]. Some authors have explored the impact of bias specific to systematic reviews and meta-analyses e.g. range of databases searched[25], publication bias[26] and selective reporting[27].

Two internationally recognised centres of excellence for accessing critical appraisal tools and training are the Critical Appraisal Skills Programme (CASP[28]), based in Oxford, UK, and the Joanna Briggs Institute(JBI[29]), in Adelaide, Australia. CASP was developed following the Getting Research into Practice Project in the 1980s, which,

'identified a lack of appreciation among managers and policy makers for the importance of using research evidence to inform decisions'

CASP evolved from initial educational workshops seeking to raise awareness of using research evidence to focusing on systematic reviews as, *'decision- makers' best resource for research evidence'*. It now has branches in multiple countries worldwide, provides training and a range of critical appraisal tools for a wide variety of types of research study.

Systematic reviews of critical appraisal tools (CATs) have concluded that further work is needed to reach a consensus on items to include in such tools[30] and that, for appraisal of systematic

reviews, any CAT that is used should be appropriate for the type of review being considered[31]. It follows therefore that a CAT used to appraise a systematic review reporting synthesis using NMA methods should include items that are particularly relevant to the NMA methodology.

1.2.5 Reporting standards

Critical appraisal depends to a large extent on the quality of reporting because, unless they are going to seek additional information from the authors, an appraiser can only make judgements of how a study was conducted based on what has been registered or reported. Good reporting should not necessarily lead to the conclusion from critical appraisal that a review is of high quality, not significantly affected by bias or that its results are applicable, each of which have to be assessed in their own right, though an association between good reporting and methodological quality has been found[32, 33].

Standards for reporting were developed to encourage transparency regarding the conduct of research in order that,

'the reader can form an opinion about the validity, strengths, weaknesses and biases of the review"[34].

International consensus standards for the reporting of systematic reviews and meta-analyses were first developed and published as the Quality of reporting of meta-analyses (QUOROM) statement in 1999[35] and updated in 2009 to become the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[36]. Network meta-analysis methods had become sufficiently commonly used by the time of the 2009 PRISMA Statement, to warrant mention as a form of meta-analysis that combined direct and indirect comparisons but the statement did not make any recommendations for reporting the specifics of NMA methodology.

In considering the impact of reporting guidelines on the quality of reports, Plint et al[37] demonstrated that the reporting of randomised controlled trials (RCTs) had improved by 2006, subsequent to the 2001 publication of the CONSORT Statement's recommendations for improving the quality of reports of RCTs. Delaney et al[38] found some improvement in reporting

of systematic reviews by 2005 since the publication of QUOROM in 1999, but Moher et al[3] still found inconsistent reporting of systematic reviews in 2007, when they noted that the uptake of QUOROM had been slower than that of CONSORT and hoped for further improvement. In assessing the quality of reporting since the publication of the PRISMA Statement in 2009, a review found that 67% of PRISMA criteria were met in a sample of 23 systematic reviews published during 2013, suggesting there had been some improvement in reporting[39], however concerns about inconsistency in reporting persist[40, 41].

Since 2009, nine extensions to the PRISMA Statement have been developed and a further two are in development, as of November 2019[42]. The Extension Statement covering the reporting of systematic reviews incorporating NMA was published in 2015[43]. In 2018, a protocol for updating the PRISMA Statement was published[44].

1.3 Aim of Thesis

The programme of research presented in this thesis aimed to advance critical appraisal of systematic reviews using NMA methods for synthesis. It was conducted over 8 years as a series of steps, each building on the previous one, as described below.

1.4 Research Timeline

The research reported in this thesis was conducted at the following times:

2012 Review of NMAs in published systematic reviews

2014 Survey of authors of NMAs

2017 Review of the use of network meta-analysis evidence in NICE clinical guidelines

2018 Review of critical appraisal tools

2019 Development and inter-rater reliability testing of new critical appraisal tool

1.5 Research Methodology

The various elements in my research used methods appropriate to the specific objectives and research questions, including literature reviews, a qualitative survey and qualitative synthesis.

1.6 Research Objectives

The research objectives of this thesis were to:

1 Review NMAs in published systematic reviews:

- 1.1 Identify the number of such systematic reviews published, by year up to 2012.
- 1.2 Analyse and summarise the clinical topic areas and selected details of the synthesis in these reviews: the number of interventions compared; the network structure; and the software used to conduct the synthesis.

2 Survey authors of NMA:

- 2.1 Identify a list of reporting requirements that authors who had published NMA consider essential.
- 2.2 Analyse and summarise themes from these authors' responses regarding future use of NMA.

3 Assess use of NMA evidence in NICE clinical guidelines:

- 3.1 Identify the number of instances of results from NMA being directly linked as evidence for recommendations within NICE clinical guidelines.
- 3.2 Summarise references within guidelines to previously published systematic reviews that use NMA methodology for synthesis.
- 3.3 Summarise NMAs undertaken by guideline development groups.

4 Review existing tools that might be used to critically appraise systematic reviews using NMA methodology:

4.1 Identify existing critical appraisal tools that address some or all aspects of quality of systematic reviews using NMA methodology for synthesis.

4.2 Determine whether any existing critical appraisal tool is already suitable for end-users without specialist knowledge of NMA methodology.

4.3 Identify methods for validating a critical appraisal tool in the event that the conclusion of this element in my research is that a new tool should be developed.

5 Develop a new critical appraisal tool:

5.1 Identify appropriate items to include in a critical appraisal tool for systematic reviews using NMA methods for synthesis.

5.2 Conduct an inter-rater reliability test of the new tool.

1.7 Structure of Thesis

1.7.1 Chapter 2: History of network meta-analysis

Chapter Two summarises the history of developments in the methodology and the use of NMA. NMA methods appear to have a role in evidence synthesis that can overcome limitations of pair-wise meta-analysis, but researchers have made relatively little use of them compared to the very widespread use of pair-wise meta-analysis. This discrepancy prompted consideration of whether wider use of this methodology would be appropriate, what factors could be limiting greater use of the methods and how publications had altered up to 2012.

1.7.2 Chapter 3: Network Meta-analyses in Published Systematic Reviews

Chapter Three reports the review that was designed to explore whether the use of this methodology in systematic reviews had increased. It found a substantial increase in the publishing of NMAs in systematic reviews from 2009 to 2012, which has continued since then. Consideration is given as to what has led to the step change and explores this in relation to accessibility and acceptability of NMA methods, and changes in reporting. The increase made it

timely for consensus on standards of conduct and reporting of NMAs to be established and incorporated into guidance on best practice for systematic reviews.

1.7.3 Chapter 4: Survey of authors

Chapter Four describes the survey of authors who had published the systematic reviews using NMA methods for synthesis that I identified in my review. It sought to identify items that might be included in standards for reporting and to identify factors that might limit further increases in the use of NMA methods. Information gathered from the authors in 2014 demonstrated strong support for adoption of standards for the conduct and reporting of NMA. The elements of reporting standards proposed by these authors were consistent with those subsequently included in the 2015 PRISMA extension statement to cover the reporting of NMA. However, my survey also identified some reservation, even amongst existing users of NMA methods, regarding greater use of the methods by researchers and about the ability of publishers and end-users to critically appraise the quality of evidence produced using these methods. This finding supported the need to identify a suitable critical appraisal tool, addressed in Chapters 6 and 7.

1.7.4 Chapter 5: Use of network meta-analysis evidence in NICE clinical guidelines

Chapter Five explores what impact the increasing amount of evidence produced using NMA is having on clinical practice and health service policy in the form of being incorporated into national clinical guidelines. NICE clinical guidelines, published or updated in 2015 or 2016, were reviewed. NMA methods were found to be used far less often than pair-wise meta-analysis but were used or considered for nearly one quarter of the guidelines reviewed. Recommendations in the guidelines were more often based on results of NMA conducted by the Guideline Development Group (GDG) than on results from previously published systematic reviews using NMA methodology. The relatively small number of previously published systematic reviews using NMA methodology that GDGs considered suitable to base recommendations on further

supported the need to identify a suitable critical appraisal tool to help users identify the quality of existing and future systematic reviews that use NMA methodology.

1.7.5 Chapter 6: Review of critical appraisal tools

Chapter Six reports a review undertaken in 2018 to identify existing critical appraisal tools that address some or all aspects of quality of systematic reviews using NMA methodology for synthesis and determine whether any existing one was already suitable for end-users without specialist knowledge of NMA methodology. It also sought to identify methods for validating a critical appraisal tool. Critical appraisal tools were identified containing some items and guidance that address the risk of bias in systematic reviews reporting synthesis using NMA methodology but none that met all my criteria to be user-friendly to the widest group of users. The empirical approaches used to construct tools and inter-rater reliability testing were summarised.

1.7.6 Chapter 7: Critical appraisal tool development

Chapter Seven describes the development of a new tool, based on the CASP systematic review CAT, constructed by sourcing and cross-referencing items and guidance from the most current and most widely used related tools that have been constructed for appraisal of systematic reviews and/or NMA. The new tool is presented. Inter-rater reliability testing of the new tool and feedback from raters is reported.

1.7.7 Chapter 8: Discussion and conclusion

Chapter Eight summarises the thesis in terms of its contribution to knowledge in the fields of application of NMA methods within systematic reviews and critical appraisal of such reviews. The evidence from this thesis is that, although far less commonly used than pair-wise meta-analysis, use of NMA methods continues to increase and is frequently used as the basis for recommendations in UK national clinical guidelines. Reporting standards to cover the use of NMA methods in systematic reviews have been welcomed and authors of the majority of published reviews are now following these standards. The user-friendly critical appraisal tool,

with content aimed at non-specialist users, constructed as part of this thesis, needs further research to determine how it might be optimised or used to contribute to the content and format of an alternative new or revised tool for the 2020s..

1.8 Chapter References

1. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 [updated July 2019]. In: Higgins JPT, Thomas J, Chandler J, et al. (eds): Cochrane, 2019 https://www.training.cochrane.org/handbook (accessed 18 October 2019).

2. Clarke M. History of evidence synthesis to assess treatment effects: personal reflections on something that is very much alive. JLL Bulletin: Commentaries on the history of treatment evaluation, http://www.jameslindlibrary.org/articles/history-of-evidence-synthesis-to-assess-treatment-effects-personal-reflections-on-something-that-is-very-much-alive/ (2015, accessed 4 November 2016).

3. Moher D, Tetzlaff J, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 2007; 4: e78. 2007/03/29. DOI: 10.1371/journal.pmed.0040078.

4. Bastian H, Glasziou P and Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010; 7: e1000326. 2010/09/30. DOI: 10.1371/journal.pmed.1000326.

5. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med 2016; 13: e1002028. 2016/05/25. DOI: 10.1371/journal.pmed.1002028.

6. Gurevitch J, Koricheva J, Nakagawa S, et al. Meta-analysis and the science of research synthesis. Nature 2018; 555: 175-182. 2018/03/09. DOI: 10.1038/nature25753.

7. Pearson K. Report on certain enteric fever inoculation statistics. BMJ 1904; 3: 1243-1246.

8. Glass GV. Primary, Secondary, and Meta-Analysis of Research. Educational Researcher 1976; 5: 3-8. DOI: 10.3102/0013189x005010003.

9. Lee WL, Bausell RB and Berman BM. The growth of health-related meta-analyses published from 1980 to 2000. Eval Health Prof 2001; 24: 327-335. 2001/08/29. DOI: 10.1177/01632780122034948.

10. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994; 309: 1351-1355. 1994/11/19. DOI: 10.1136/bmj.309.6965.1351.

11. Cochran WG. The Combination of Estimates from Different Experiments. Biometrics 1954; 10: 101. DOI: 10.2307/3001666.

12. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557-560. 2003/09/06. DOI: 10.1136/bmj.327.7414.557.

13. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. Oper Res 1989; 37: 210-228. 1989/02/08. DOI: 10.1287/opre.37.2.210.

14. Eddy DM, Hasselblad V and Shachter R. A Bayesian method for synthesizing evidence. The Confidence Profile Method. International Journal of Technology Assessment in Health Care 1990; 6: 31-55. 1990/01/01. DOI: 10.1017/s0266462300008928.

15. Smith TC, Spiegelhalter DJ and Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med 1995; 14: 2685-2699. 1995/12/30. DOI: 10.1002/sim.4780142408.

16. Lu G and Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23: 3105-3124. 2004/09/28. DOI: 10.1002/sim.1875.

17. Edwards SJ, Clarke MJ, Wordsworth S, et al. Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. Int J Clin Pract 2009; 63: 841-854. 2009/06/06. DOI: 10.1111/j.1742-1241.2009.02072.x.

18. Zarin W, Veroniki AA, Nincic V, et al. Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. BMC Med 2017; 15: 3. 2017/01/06. DOI: 10.1186/s12916-016-0764-6.

19. Oxman AD and Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44: 1271-1278. 1991/01/01. DOI: 10.1016/0895-4356(91)90160-b.

20. Sacks HS, Chalmers TC and Smith H, Jr. Sensitivity and specificity of clinical trials. Randomized v historical controls. Archives of Internal Medicine 1983; 143: 753-755. 1983/04/01. DOI: 10.1001/archinte.143.4.753.

21. Savovic J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess 2012; 16: 1-82. 2012/09/20. DOI: 10.3310/hta16350.

22. Colditz GA, Miller JN and Mosteller F. The effect of study design on gain in evaluations of new treatments in medicine and surgery. Drug InfJ 1988; 22: 343-352. DOI: 10.1177/009286158802200307.

23. Chalmers TC, Celano P, Sacks HS, et al. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309: 1358-1361. 1983/12/01. DOI: 10.1056/NEJM198312013092204.

24. Moher D, Soeken K, Sampson M, et al. Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. BMC Pediatrics 2002; 2: 3. 2002/03/27. DOI: 10.1186/1471-2431-2-3.

25. Sampson M, Barrowman NJ, Moher D, et al. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol 2003; 56: 943-955. 2003/10/22. DOI: 10.1016/s0895-4356(03)00110-0.

26. Sutton AJ, Duval SJ, Tweedie RL, et al. Empirical assessment of effect of publication bias on meta-analyses. BMJ 2000; 320: 1574-1577. 2000/06/14. DOI: 10.1136/bmj.320.7249.1574.

27. Page MJ, McKenzie JE, Kirkham J, et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. Cochrane Database of Systematic Reviews 2014: MR000035. 2014/10/02. DOI: 10.1002/14651858.MR000035.pub2.

28. CASP UK. CASP, https://casp-uk.net/ (accessed 12 September 2019).

29. Joanna Briggs Institute. Joanna Briggs Institute, https://joannabriggs.org/ (accessed 12 September 2019).

30. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. BMC Medical Research Methodology 2004; 4: 22. 2004/09/17. DOI: 10.1186/1471-2288-4-22.

31. Crowe M and Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol 2011; 64: 79-89. 2010/12/07. DOI: 10.1016/j.jclinepi.2010.02.008.

32. Tunis AS, McInnes MD, Hanna R, et al. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? Radiology 2013; 269: 413-426. 2013/07/05. DOI: 10.1148/radiol.13130273.

33. Panic N, Leoncini E, de Belvis G, et al. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. PLoS One 2013; 8: e83138. 2014/01/05. DOI: 10.1371/journal.pone.0083138.

34. Harms M. The EQUATOR Network and the PRISMA Statement for the reporting of systematic reviews and meta-analyses. Physiotherapy 2009; 95: 237-240. 2009/11/07. DOI: 10.1016/j.physio.2009.10.001.

35. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354: 1896-1900. 1999/12/10. DOI: 10.1016/s0140-6736(99)04149-5.

36. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339: b2700. 2009/07/23. DOI: 10.1136/bmj.b2700.

37. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Medical Journal of Australia 2006; 185: 263-267. 2006/09/05. DOI: 10.5694/j.1326-5377.2006.tb00557.x.

38. Delaney A, Bagshaw SM, Ferland A, et al. A systematic evaluation of the quality of meta-analyses in the critical care literature. Critical Care 2005; 9: R575-582. 2005/11/10. DOI: 10.1186/cc3803.

39. O'Brien E. Compliance with PRISMA reporting standards in systematic reviews https://www.bazian.com/pdfs/compliance_with_prisma.pdf (2014, accessed 9 November 2019).

40. Pussegoda K, Turner L, Garritty C, et al. Systematic review adherence to methodological or reporting quality. Systematic reviews 2017; 6: 131. 2017/07/20. DOI: 10.1186/s13643-017-0527-2.

41. Page MJ and Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. Systematic reviews 2017; 6: 263. 2017/12/21. DOI: 10.1186/s13643-017-0663-8.

42. PRISMA. PRISMA Extensions, http://www.prisma-statement.org/Extensions/ (accessed 9 November 2019).

43. Hutton B, Salanti G, Caldwell DM, et al. 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: 777-784. 2015/06/02. DOI: 10.7326/M14-2385.

44. Page MJ, McKenzie JE, Bossuyt PM, et al. Updating the PRISMA reporting guideline for systematic reviews and meta-analyses: study protocol. https://osf.io/2v7mk/ (2018, accessed 9 November 2019).

CHAPTER TWO

History of network meta-analysis

2 History of network meta-analysis

2.1 Chapter Aim and Objectives

The aim of this chapter is to summarise the history of developments in the methodology and the use of network meta-analysis (NMA).

2.2 The Evolution of Research Synthesis

NMA is a method of research synthesis, so any history of this specific method must acknowledge it is just one part of the overall history of research synthesis. That wider history is well described by others[1, 2] so this chapter focuses on the origins and development of NMA as a particular method of synthesis, which has also been described as *mixed treatment comparison* or *multiple treatments comparison*[3].

2.3 The Concept of Network Meta-Analysis

NMA is an extension of pair-wise meta-analysis, but with three main advantages claimed for NMA. Firstly, it allows more than two interventions to be compared simultaneously; secondly, interventions can be compared even if they have not been directly compared in trials; and thirdly, it increases precision of the estimate of effect size, by '*borrowing strength*'[4], if the key assumptions described below are valid.

When seeking to compare the effects of interventions A, B and C if there are trials of A versus B and B versus C, but not of A versus C, an estimate of the relative effects of A compared to C can be made by indirect comparison using the results of the trials of A versus B and B versus C. This indirect comparison uses the results of meta-analysis of all trials in each direct comparison, as explained by Bucher[5], to combine, in a single synthesis, all the available direct comparisons and all the indirect comparisons that can be estimated between the selected interventions.

A network diagram can be drawn illustrating the direct comparisons between A, B and C. In more complex networks, multiple combinations of direct comparisons are used to estimate the

same indirect comparison, which are then incorporated in the synthesis. These networks might have more than three interventions and have varying complexity of geometry. Among the simplest is a star, in which the intervention at the centre is the common comparator to all the other interventions. More complex forms of network geometry, when there are common comparators for only a few of a more diverse range of interventions involve multiple conjoined loops and sidearms.

Figure 2.1 shows an example of a network diagram from a NMA comparing six interventions, A-F. The solid lines show where trial results of direct comparisons exist. In this example, there is no single common comparator for all the interventions, but each intervention shares at least one comparator with another intervention in the network. The thickness of each connecting line represents the number of trials between the pair of interventions at either end of the line and the size of each node (letter within a circle) represents the number of participants receiving the intervention. Some network diagrams also include the actual numbers of trials, participants or both.

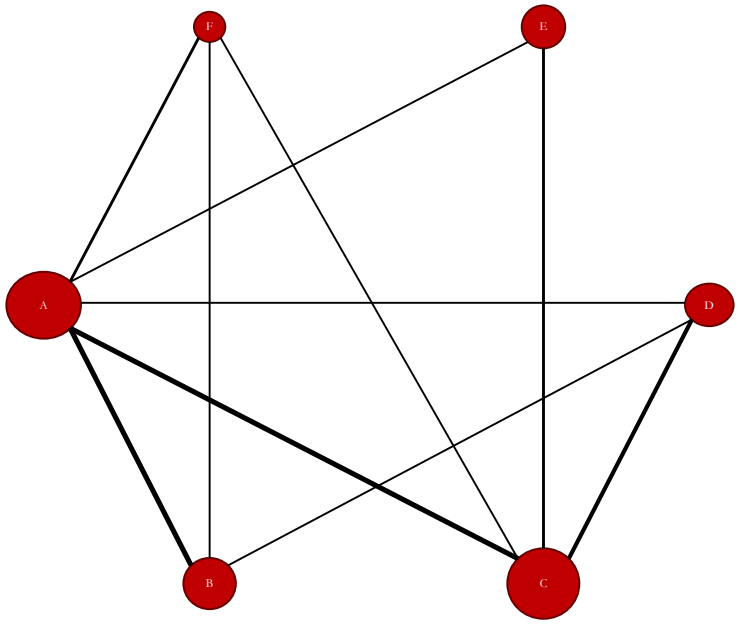

Figure 2.1 Network diagram example

Transitivity and consistency are two key assumptions for these analyses. Transitivity is the assumption that the distribution of effect modifiers (the clinical and methodological characteristics that can affect the outcome) are similar in each set of direct comparisons involving a common comparator (e.g. A versus C and B versus C) used to make an indirect comparison; in this example, A versus B[6]. It cannot be assessed statistically but requires critical interpretation, in relation to each possible indirect comparison, of the effect modifiers in the trials whose results are considered for synthesis. Before a NMA is conducted, those trials must be assessed for significant differences in their populations, interventions, outcomes, methodological features and reporting[7]. It is also important to note that even placebo response has been found to vary over time, which might affect the transitivity assumption when placebo is a common comparator[8].

Consistency is an extension of transitivity. It is the assumption of agreement between the results of direct and indirect comparisons for each pair of interventions. This can be assessed statistically but only when there are both direct and indirect comparisons of one or more pairs of interventions within a network, known as 'closed loops'. The assumption of transitivity needs to be reconsidered if inconsistency is detected. If inconsistency is not detected statistically, however, that does not automatically validate the transitivity assumption[7].

2.4 Origins and Evolution of the Methodology

In 1989, Eddy first described the 'Confidence Profile Method'[9], and the Eddy, Hasselblad and Shachter publication the following year, 'A Bayesian Method for Synthesising Evidence'[10], described:

'a collection of meta-analysis techniques based on Bayesian methods for interpreting, adjusting, and combining evidence to estimate parameters and outcomes important to the assessment of health technologies.'

These techniques were collectively called the *Confidence Profile Method* (CPM) and the article explained indirect comparison with the following example:

'The approach is to use the available evidence to derive probability distributions for the various pairs that have been directly compared. A distribution for the relative effects of other pairs can then be calculated by a series of convolutions. The concept is illustrated by calculating the difference between the test scores of Tom and Bill from knowledge of the differences in scores between Tom and George, and George and Bill.'

Practical application of the methodology was supported by software to conduct the synthesis. Eddy initially used SOFT*PRO, but the commonest software in published systematic reviews reporting NMA[3] is WinBUGS (initially BUGS)[11]. This uses Bayesian methods and such methods are predominant in both software use and methodological developments in NMA, but frequentist approaches and software are also used[12, 13].

A series of methodological publications through the 1990s built on the CPM approach, including notably those by Smith, Spiegelhalter and Thomas[14] and by Higgins and Whitehead[4]. In 1996, the latter wrote about borrowing strength from external trials in a meta-analysis. They argued,

'Many meta-analysis papers include data from three or more treatments, but only consider pairwise comparisons of, say treatment A with control and treatment B with control. There would seem to be little reason not to combine all treatments into one analysis.'

They used the BUGS software to combine the results of trials comparing the effects of beta-blockers versus sclerotherapy, beta-blockers versus control and sclerotherapy versus control on preventing cirrhosis-related bleeding. This was later described by Salanti and Schmid[15] as,

'the first to articulate that relative effects of different treatments can be jointly estimated in a single meta-analysis model to improve power. This landmark paper introduced the basis for the methodology which, now extended and refined, is increasingly known as network meta-analysis.'

In 2002, Lumley published '*Network meta-analysis for indirect treatment comparisons*'[16] in which he:

'presented methods of estimating treatment differences between treatments that have not been directly compared in a randomized trial, and, more importantly, methods of estimating the uncertainty in these differences.'

Lumley acknowledged the limitation of his methods, which was restricted to each trial only having two intervention groups,

'Meta-analyses with large numbers of multi-armed trials present difficulties for network meta-analysis, and extensions to handle multi-armed trials correctly should be investigated.'

Ades subsequently described methods to encompass multi-arm trials and multiple outcomes. In his 2003 article[17], he stated:

'The aim of nearly all meta-analysis has been to summarize evidence comparing one or sometimes more treatments. Usually, only a single outcome is examined, and if there is more than one outcome these are explored in separate meta-analyses, rather than simultaneously. This paper concerns the possibility of combining information from different studies on different, but structurally related, outcomes, and using the data to construct a single model which expresses the relationships between the different kinds of data.'

The following year, together with Lu, Ades published '*Combination of direct and indirect evidence in mixed treatment comparisons*'[18], which was the most frequently cited origin of current network meta-analysis methodology in the survey of published network meta-analysis described later in Chapter Three[18]. They extended the model proposed by Smith, Spiegelhalter and Thomas[14] to encompass trials with more than two intervention groups.

A review of the methods for NMA, with particular emphasis on the issue of inconsistency between direct and indirect evidence, was published in 2008 by Salanti et al[19]. They explained that inconsistency in estimates of intervention effects obtained from direct and indirect comparisons may indicate diversity, bias or a combination of both and described modelling to test for consistency. The review considered potential sources of inconsistency, including genuine diversity in the characteristics of included trials, selection bias, study quality and sponsorship bias. It stresses the importance of planning in advance for investigation of inconsistency, because clinical and epidemiological assessment of inconsistency may be difficult because of factors such as reporting deficiencies or lack of sufficient studies for some comparisons. Salanti et al also highlighted that attention to the geometry (the overall pattern of comparisons amongst interventions) and the asymmetry (the extent to which specific comparisons of interventions are represented more heavily than others in the number of included trials or participants) of networks can be used to inform the design of new trials that would most usefully add to the overall network.

A more recent review of NMA methods was published in 2016 by Efthimiou et al[7] which summarises newer publications on the use of NMA methods. This includes various models for performing NMA, statistical methods for assessing inconsistency, software options, investigating sources of potential bias and reporting results.

The use of individual participant data (IPD) in meta-analyses has many advantages over the use of aggregate data, including improving the quantity and quality of data, which has resulted in it being considered as *'the gold standard in evidence synthesis'* [20]. The use of this approach, initially using PWMA methods, increased between the early 1990s and 2008 to around 50 publications per year[21]. The number of systematic reviews using IPD was found to be 10-22 per year with no particular growth trend in the years leading up to 2015[22]. Gao et al found that the first IPD using NMA methods was published in 2007 and that 21 IPDs using NMA methods were published by June 2019[23]. There are limitations as well as advantages to use of IPD and guidance has been published on the best use of IPD meta-analysis generally[24] and specifically on the use of NMA methods with IPD[20].

Multiple outcomes multivariate meta-analysis (MOMA) is another approach to meta-analysis that has been increasing in recent years[25]. Relevant studies that might be considered for synthesis may not report the same outcomes, which could result in their exclusion from traditional meta-analyses, but MOMA allows for inclusion where outcomes can be regarded as highly correlated. Guidance on conducting this type of synthesis using NMA methods has been published in the recent years[25-27], including the use of IPD[28].

In recent years, interest has developed in creating and maintaining continuously updated meta-analyses using NMA methods[29] and, in 2020, a major project of this kind for Covid-19 related interventions has begun[30].

2.5 Approaches to conducting Network Meta-analysis

A simple meta-regression approach can be used for NMA if there is no multi-arm trial in the network[31]. However, if the network includes multi-arm trials, other methods are more appropriate. Bayesian methods have been used most frequently[32], partly because this approach can most naturally produce estimates of ranking probabilities for the interventions being compared (to give the probability that each intervention is most effective through to least effective)[33]; but frequentist methods to approximate ranking have also been described[34]. The hierarchical model approach is detailed by Lu and Ades[18] and by Salanti[19].

An alternative approach is multi-variate meta-analysis, which can be conducted using Bayesian[35] or Frequentist methods[36]. A further approach, based on graph-theoretical methods, has been described by Rucker[37].

The frequentist approach assumes that the intervention effect has a true value, and the confidence intervals results define the range of values to include that true value with a minimum probability, usually 95%. The Bayesian approach assumes that the intervention effect has a fixed value but within a probability distribution based on a 'prior', which might be a value chosen from existing evidence or might be a 'best guess'. The credible interval results of a Bayesian meta-analysis provide the probability of the range of values within which the fixed value lies, given the data, and this range is the 'posterior' that includes 95% of the probability.

2.6 Use of the NMA Methodology in Published Systematic Reviews

In 1999 Dominici et al[38], used Bayesian methods and data from 46 trials of treatments to prevent migraine headache to produce a ranking of treatments. They stated their aims as,

"In this article we present a meta-analysis of these 46 trials with the goal of synthesizing existing evidence about which treatments are most effective and of quantifying the remaining uncertainty about treatment effectiveness. We hope that the results and methods will be useful in supporting clinical treatment decisions and will help guide the planning of new trials. The critical statistical aspects of this goal are the estimation of treatment effects on a common scale and the relative ranking of treatments,

both within classes and overall. This requires indirect comparisons among treatments that may never have been tested together in the same trial."

They had used data collected by another team working on a systematic review and the first published NMA as part of a systematic review conducted by the authors, seems to be in 2003 by Psaty et al[39]. Their explanation for selecting this method to study treatments for high blood pressure summarises well the commonest problem in using only pair-wise meta-analysis to inform clinical decision making:

'the clinical trials in hypertension have provided a patchwork of evidence about the health benefits of antihypertensive agents. Some trials used placebo or untreated controls, and others used active-treatment comparison groups. Among the latter, the choice of treatment and comparison therapies has varied from one trial to the next. Several approaches to the synthesis of these complex data are possible. The Blood Pressure Trialists, for instance, conducted a prospective series of mini-meta-analyses, but this method left many "unresolved issues" due to multiple comparisons and low power. In this study, we used a new technique, called network meta-analysis, to synthesize the available evidence from placebo-controlled and comparative trials in a single meta-analysis.'

Subsequent uses of network meta-analyses were initially slow to appear in the healthcare literature. Edwards et al[40] identified only seven published systematic reviews reporting NMA, up to July 2007.

2.7 Guidance on the conduct and reporting of NMA

When the first standards for the reporting of systematic reviews and meta-analyses were published in the QUOROM statement[41], the concept of NMA was still in its infancy. However by the time of the PRISMA statement in 2009[42], NMA warranted mention as a form of meta-analysis that combined direct and indirect comparisons but the PRISMA 2009 statement did not make recommendations relating to reporting that address the specifics of NMA methodology. For standards of conduct of systematic reviews, it directed readers to the guidance published by The Cochrane Collaboration[43] and the Centre for Reviews and Dissemination[44]; both of which contained very limited guidance on the use of NMA methodology in systematic reviews. Reporting standards for use of NMA methods were subsequently published as a PRISMA Extension Statement in 2015[45].

The National Institute for Health and Care Excellence (NICE) *Methods Guide* 2008 update[46] included a section on NMA for the first time and the NICE Decision Support Unit's Evidence Synthesis TSD series[47], published initially in 2011, expanded on that guidance. Also in 2011, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) published reports on the interpretation and conduct of network meta-analysis[12, 13], and Ades[48] observed that this,

'seems to represent the first position statement from an academic body on these methods'.

The Cochrane Comparing Multiple Interventions Methods Group (CMIMG) was established in 2010 and has since produced guidance on the use of NMA methods in Cochrane Reviews and promoted training in use of these methods. Version 6 of the Cochrane Handbook for Systematic Reviews of Interventions, published in 2019, contained '*a major new core chapter*', addressing NMA within the handbook, for the first time[49]. This guidance emphasises that NMA is '*more statistically complex than a standard meta-analysis'* and that, consequently, '*close collaboration'* between a statistician with expertise in NMA methods and those with expertise in the clinical content area is essential in the design and conduct of a review to ensure that studies selected for inclusion in NMA fulfil the assumptions of transitivity and consistency[50].

2.8 Conclusions

2.8.1 Summary of the findings of this chapter

NMA is a relatively new form of evidence synthesis. As outlined by Salanti[51], it has faced similar scepticism to that raised originally about pair-wise meta-analysis. Only 7 systematic reviews reporting the use of this method were found up to July 2007, but it appeared to have the potential for greater use and this use has gradually increased through the past decade.

2.8.2 Implications of the findings of this chapter

NMA methods appear to have a role in evidence synthesis that can overcome limitations of pair-wise meta-analysis, but researchers have made relatively little use of them so far. This prompts the question of whether greater use would be appropriate and, if so, what factors could be limiting this greater use. To help address this, I conducted a new review of published systematic reviews reporting NMA to identify whether the use had increased by the time that my programme of research began (2012), to analyse which clinical areas were being studied using this method and to identify what details were being reported of the choices made when using the method. This review is reported in Chapter Three.

2.9 Chapter Summary

- NMA is a relatively new form of evidence synthesis
- An early review found that only 7 systematic reviews reporting use of this method were available to July 2007
- The next step was to conduct an updated review of published systematic reviews reporting NMA to identify whether the rate of publications had altered

2.10 Chapter References

1. Chalmers I, Hedges LV and Cooper H. A brief history of research synthesis. *Eval Health Prof* 2002; 25: 12-37. 2002/03/01. DOI: 10.1177/0163278702025001003.

2. Clarke M. History of evidence synthesis to assess treatment effects: personal reflections on something that is very much alive. *JLL Bulletin: Commentaries on the history of treatment evaluation*, http://www.jameslindlibrary.org/articles/history-of-evidence-synthesis-to-assess-treatment-effects-personal-reflections-on-something-that-is-very-much-alive/ (2015, accessed 4 November 2016).

3. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol* 2014; 67: 138-143. 2013/10/05. DOI: 10.1016/j.jclinepi.2013.07.014.

4. Higgins JP and Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996; 15: 2733-2749. 1996/12/30. DOI: 10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0.

5. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50: 683-691. 1997/06/01. DOI: 10.1016/s0895-4356(97)00049-8.

6. Cipriani A, Higgins JP, Geddes JR, et al. Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine* 2013; 159: 130-137. 2013/07/17. DOI: 10.7326/0003-4819-159-2-201307160-00008.

7. Efthimiou O, Debray TP, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods* 2016; 7: 236-263. 2016/01/13. DOI: 10.1002/jrsm.1195.

8. Julious SA and Wang S-J. How Biased Are Indirect Comparisons, Particularly When Comparisons Are Made Over Time in Controlled Trials? *Drug Inf J* 2008; 42: 625-633. DOI: 10.1177/009286150804200610.

9. Eddy DM. The confidence profile method: a Bayesian method for assessing health technologies. *Oper Res* 1989; 37: 210-228. 1989/02/08. DOI: 10.1287/opre.37.2.210.

10. Eddy DM, Hasselblad V and Shachter R. A Bayesian method for synthesizing evidence. The Confidence Profile Method. *International Journal of Technology Assessment in Health Care* 1990; 6: 31-55. 1990/01/01. DOI: 10.1017/s0266462300008928.

11. Lunn DJ, Thomas A, Best N, et al. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; 10: 325-337. DOI: 10.1023/a:1008929526011.

12. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health* 2011; 14: 417-428. 2011/06/15. DOI: 10.1016/j.jval.2011.04.002.

13. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health* 2011; 14: 429-437. 2011/06/15. DOI: 10.1016/j.jval.2011.01.011.

14. Smith TC, Spiegelhalter DJ and Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995; 14: 2685-2699. 1995/12/30. DOI: 10.1002/sim.4780142408.

15. Salanti G and Schmid CH. Research Synthesis Methods special issue on network meta-analysis: introduction from the editors. *Res Synth Methods* 2012; 3: 69-70. 2012/06/01. DOI: 10.1002/jrsm.1050.

16. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002; 21: 2313-2324. 2002/09/05. DOI: 10.1002/sim.1201.

17. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med* 2003; 22: 2995-3016. 2003/09/16. DOI: 10.1002/sim.1566.

18. Lu G and Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23: 3105-3124. 2004/09/28. DOI: 10.1002/sim.1875.

19. Salanti G, Higgins JP, Ades AE, et al. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17: 279-301. 2007/10/11. DOI: 10.1177/0962280207080643.

20. Debray TP, Schuit E, Efthimiou O, et al. An overview of methods for network meta-analysis using individual participant data: when do benefits arise? *Statistical Methods in Medical Research* 2018; 27: 1351-1364. 2016/08/05. DOI: 10.1177/0962280216660741.

21. Riley RD, Lambert PC and Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; 340: c221. 2010/02/09. DOI: 10.1136/bmj.c221.

22. Simmonds M, Stewart G and Stewart L. A decade of individual participant data meta-analyses: A review of current practice. *Contemp Clin Trials* 2015; 45: 76-83. 2015/06/21. DOI: 10.1016/j.cct.2015.06.012.

23. Gao Y, Shi S, Li M, et al. Statistical analyses and quality of individual participant data network meta-analyses were suboptimal: a cross-sectional study. *BMC Med* 2020; 18: 120. 2020/06/02. DOI: 10.1186/s12916-020-01591-0.

24. Tierney JF, Vale C, Riley R, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLoS Med* 2015; 12: e1001855. 2015/07/22. DOI: 10.1371/journal.pmed.1001855.

25. Riley RD, Jackson D, Salanti G, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* 2017; 358: j3932. 2017/09/15. DOI: 10.1136/bmj.j3932.

26. Achana FA, Cooper NJ, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Medical Research Methodology* 2014; 14: 92. 2014/07/23. DOI: 10.1186/1471-2288-14-92.

27. Efthimiou O, Mavridis D, Cipriani A, et al. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Stat Med* 2014; 33: 2275-2287. 2014/06/12. DOI: 10.1002/sim.6117.

28. Riley RD, Price MJ, Jackson D, et al. Multivariate meta-analysis using individual participant data. *Res Synth Methods* 2015; 6: 157-174. 2015/06/24. DOI: 10.1002/jrsm.1129.

29. Vandvik PO, Brignardello-Petersen R and Guyatt GH. Living cumulative network meta-analysis to reduce waste in research: A paradigmatic shift for systematic reviews? *BMC Med* 2016; 14: 59. 2016/03/31. DOI: 10.1186/s12916-016-0596-4.

30. Boutron I, Chaimani A, Meerpohl JJ, et al. The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19 Pandemic. *Annals of Internal Medicine* 2020 2020/09/16. DOI: 10.7326/M20-5261.

31. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.0 [updated July 2019]. In: Higgins JPT, Thomas J, Chandler J, et al. (eds): Cochrane, 2019 https://www.training.cochrane.org/handbook (accessed 18 October 2019).

32. Nikolakopoulou A, Chaimani A, Veroniki AA, et al. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS One* 2014; 9: e86754. 2014/01/28. DOI: 10.1371/journal.pone.0086754.

33. Neupane B, Richer D, Bonner AJ, et al. Network meta-analysis using R: a review of currently available automated packages. *PLoS One* 2014; 9: e115065. 2014/12/30. DOI: 10.1371/journal.pone.0115065.

34. White IR. Multivariate Random-effects Meta-regression: Updates to Mvmeta. *The Stata Journal* 2011; 11: 255-270. DOI: 10.1177/1536867x1101100206.

35. Mavridis D and Salanti G. A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research* 2013; 22: 133-158. 2012/01/26. DOI: 10.1177/0962280211432219.

36. White IR, Barrett JK, Jackson D, et al. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 2012; 3: 111-125. DOI: 10.1002/jrsm.1045.

37. Rucker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012; 3: 312-324. 2012/12/01. DOI: 10.1002/jrsm.1058.

38. Dominici F, Parmigiani G, Wolpert RL, et al. Meta-Analysis of Migraine Headache Treatments: Combining Information from Heterogeneous Designs. *Journal of the American Statistical Association* 1999; 94: 16-28. DOI: 10.1080/01621459.1999.10473815.

39. Psaty BN, Lumley T, Furberg CD, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents - a network meta-analysis. *JAMA* 2003; 289: 2534-2544. 2003/05/22. DOI: 10.1001/jama.289.19.2534.

40. Edwards SJ, Clarke MJ, Wordsworth S, et al. Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. *Int J Clin Pract* 2009; 63: 841-854. 2009/06/06. DOI: 10.1111/j.1742-1241.2009.02072.x.

41. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999; 354: 1896-1900. 1999/12/10. DOI: 10.1016/s0140-6736(99)04149-5.

42. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700. 2009/07/23. DOI: 10.1136/bmj.b2700.

43. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. In: Higgins JPT and Green S (eds): The Cochrane Collaboration, 2011 http://www.cochrane-handbook.org/ (accessed 4 May 2015).

44. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. CRD, University of York, 2009 https://www.york.ac.uk/inst/crd/index_guidance.htm (accessed 4 May 2015).

45. Hutton B, Salanti G, Caldwell DM, et al. 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: 777-784. 2015/06/02. DOI: 10.7326/M14-2385.

46. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. National Institute for Health and Clinical Excellence, 2008 https://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed 6 September 2012).

47. NICE Decision Support Unit (DSU). Evidence Synthesis TSD series, http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series(2391675).htm (accessed 5 May 2015).

48. Ades AE. ISPOR states its position on network meta-analysis. *Value Health* 2011; 14: 414-416. 2011/06/15. DOI: 10.1016/j.jval.2011.05.001.

49. Higgins J and Thomas J. Cochrane Handbook for Systematic Reviews of Interventions What's New in Version 6. https://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/resources/downloadable_resources/English/Handbook%20-%20Whats%20New%20Flyer_A4-double%20sided.pdf (2019, accessed 23 8 2020).

50. Chaimani A, Caldwell DM, Li T, et al. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, et al. (eds) *Cochrane Handbook for Systematic Reviews of Interventions Version 60 [updated July 2019]*: Cochrane, 2019 https://www.training.cochrane.org/handbook (accessed 18 October 2019).

51. 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. *Res Synth Methods* 2012; 3: 80-97. 2012/06/01. DOI: 10.1002/jrsm.1037.

CHAPTER THREE

Network Meta-analyses in Published Systematic Reviews

3 Network Meta-analyses in Published Systematic Reviews

The work described in this chapter was published in the *Journal of Clinical Epidemiology* and much of the text of that article is repeated verbatim:

Lee, A.W. Review of mixed treatment comparisons in published systematic reviews shows a marked increase since 2009. *Journal of Clinical Epidemiology* 2013; 67: 138-43. doi.org/10.1016/j.jclinepi.2013.07.014[1]

3.1 Background

In 2009, Edwards et al [2] reviewed the use of various methods of indirect comparison in systematic reviews and reported that, other than methods relying on a single common comparator, the mixed treatment comparison (MTC) method was the only one that could estimate the difference between interventions and provide a measure of the uncertainty around that estimate. In addition to MTC, other terms for methods used to compare multiple interventions in a single synthesis include multiple treatment comparison (also known by the same acronym MTC), multiple treatments meta-analysis and network meta-analysis (NMA)[3-5], but network meta-analysis has become the term used in guidance relating to the use, reporting and interpretation of these methods[6-9]. This chapter presents the results of my review of the use of NMA in systematic reviews, which was conducted in 2012 and Chapter Eight describes subsequent trends.

3.2 Chapter Aim and Objectives

The aim of this chapter is to report and investigate trends in the publication of systematic reviews that use NMA methodology for synthesis, and their characteristics in the period to the middle of 2012 (with trends in publication since 2012 discussed in Chapter Eight). The primary objective was to identify the number of such systematic reviews published by year. Secondary objectives were to analyse and summarise the clinical topic areas and selected

details of the synthesis: the number of interventions compared; the network structure; and the software used to conduct the synthesis.

3.3 Methods

Research database searches were undertaken to identify systematic reviews using meta-analysis for synthesis.

3.3.1 Search strategy

The search strategy was developed, with advice from a librarian, to include the various search terms for the synthesis method, in free text, allowing for spelling and plurality variations. The Medical Subject Heading (MeSH) term for meta-analysis was also included for databases that allow searching by MeSH terms, but no MeSH term was available for the specific synthesis method the search aimed to identify (namely, network meta-analysis). The search strategy was modified to suit the requirements of each of the databases searched since some do not allow searching using MeSH terms.

MEDLINE (1946 to 2012), MEDLINE In-Process, Embase (1988 to 2012), CINAHL, DARE, and the Cochrane Database of Systematic Reviews (CDSR; September 2012) were searched for published reviews. SIGLE was searched for reviews reported as conference abstracts. Searches were not restricted by language. These searches were last updated on September 21, 2012. The reference lists of articles retrieved were also checked for potentially eligible articles.

My initial search strategy found the 7 eligible NMAs identified by Edwards[2]. However, checking the reference lists of later NMAs identified one review that had been missed by the electronic database searching. I amended the search strategy, based on the keyword terms attached to this review, to include the MESH term for Bayes Theorem and Bayes as a free-text search term. This amended strategy identified an additional 27 records, from which five

eligible reviews were retrieved. As an example, the final amended search strategy used for MEDLINE is shown in Figure 3.1.

#1	exp Meta-Analysis/
#2	meta*analys$.mp.
#3	mixed treatment comparison$.mp.
#4	multiple treatment comparison$.mp.
#5	network meta-analysis.mp.
#6	mixed treatment.mp.
#7	multiple*treatment$.mp.
#8	network.mp.
#9	exp Bayes Theorem/
#10	bayes$.mp.
#11	3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
#12	1 or 2
#13	11 and 12
#14	limit 13 to humans

Figure 3.1. MEDLINE search strategy

3.3.2 Inclusion criteria

Systematic reviews of healthcare interventions in humans, published up to the end of June 2012 (including articles that were published "online first" before that date but which appeared in print later), in which a meta-analysis combined direct and indirect comparisons among more than two interventions (not counting placebo or usual care) were eligible. Reviews could be of any interventions and report any outcome of benefit or harm. Conference abstracts were included if the comparisons and outcomes of meta-analysis were reported and the review had not been separately reported as a journal article. Methodological articles and cost-effectiveness studies were included if they reported a new NMA using trials identified by a systematic review.

3.3.3 Selection of studies

A single reviewer screened abstracts of the records identified by the searches against the inclusion criteria, with the default to retrieve the full article for further consideration if there was uncertainty based on the abstract. Retrieved full articles were reviewed to make a final decision on eligibility. No assessment was made of quality or risk of bias in the reviews.

3.3.4 Data extraction

Search results from each database were exported to Endnote X5 software (Thomson Reuters [Scientific] Inc., New York, NY). After identifying and removing duplicate results in Endnote, the remaining unique results were exported to an electronic Microsoft Excel spreadsheet. The following data were extracted to the spreadsheet: authors, year of publication, interventions included in NMA, outcomes included in NMA, network structure, software used for data synthesis, and type of publication. One or more clinical topics, selected to match the top level clinical topics used by the CDSR, were assigned to each review to classify the clinical area of each review. This work was done by a single reviewer.

3.4 Results

3.4.1 Articles retrieved

The number of records retrieved by the final search of each source was 625 (MEDLINE and MEDLINE In-Process), 1103 (Embase), 81 (CINAHL), 400 (DARE), 91 (CDSR), and 18 (SIGLE). Other searching identified eight articles. Nearly all articles were published in English, but one published in Chinese was also retrieved. 390 records were judged to be potentially eligible based on the abstract and the full-text articles were retrieved for these. Figure 3.2 summarises the retrieval, exclusion and final inclusion of articles.

3.4.2 Articles included

Authors used various terms when reporting NMAs, including *multiple treatment comparison* and *network meta-analysis*. These terms were also used variably to describe a single synthesis of the results of direct and indirect comparisons or to describe a series of separate direct or indirect pair-wise comparisons. *Mixed treatment comparison* was used most consistently as a term to describe a single synthesis of the results of direct and indirect comparisons in these pre-2013 publications. Consequently, the term *mixed treatment comparison* was used in the published report of this review[1]. However, since that time, a consensus has developed on using *network meta-analysis* as the preferred term. Some articles contained insufficient or conflicting detail, leaving uncertainty about the method of indirect comparison used. In these cases, the decision on whether the article was eligible for inclusion was made on the balance of the information available. Some articles reported an NMA but were excluded because the results analysed were from trials not identified by a systematic review. A high proportion of the articles retrieved that were reviews of evidence for National Institute for Health and Care Excellence (NICE) health technology appraisals (HTAs) did comment on evidence from reviews, including NMAs submitted by manufacturers, but the authors did not provide adequate details for these

reviews to be judged eligible for inclusion, and separate publications were not found for these NMAs.

The full-text articles retrieved included all records appearing to be a report of a NMA and stating in the title or abstract that this was based on a systematic review. The judgement that an article reported a systematic review was based on the description of the methods for the identification of included studies.

A total of 201 articles satisfied the eligibility criteria, 157 were published in full and 44 as conference abstracts, posters, or presentations. See Appendix 1 for references to the included studies and Table A1 listing the study characteristics.

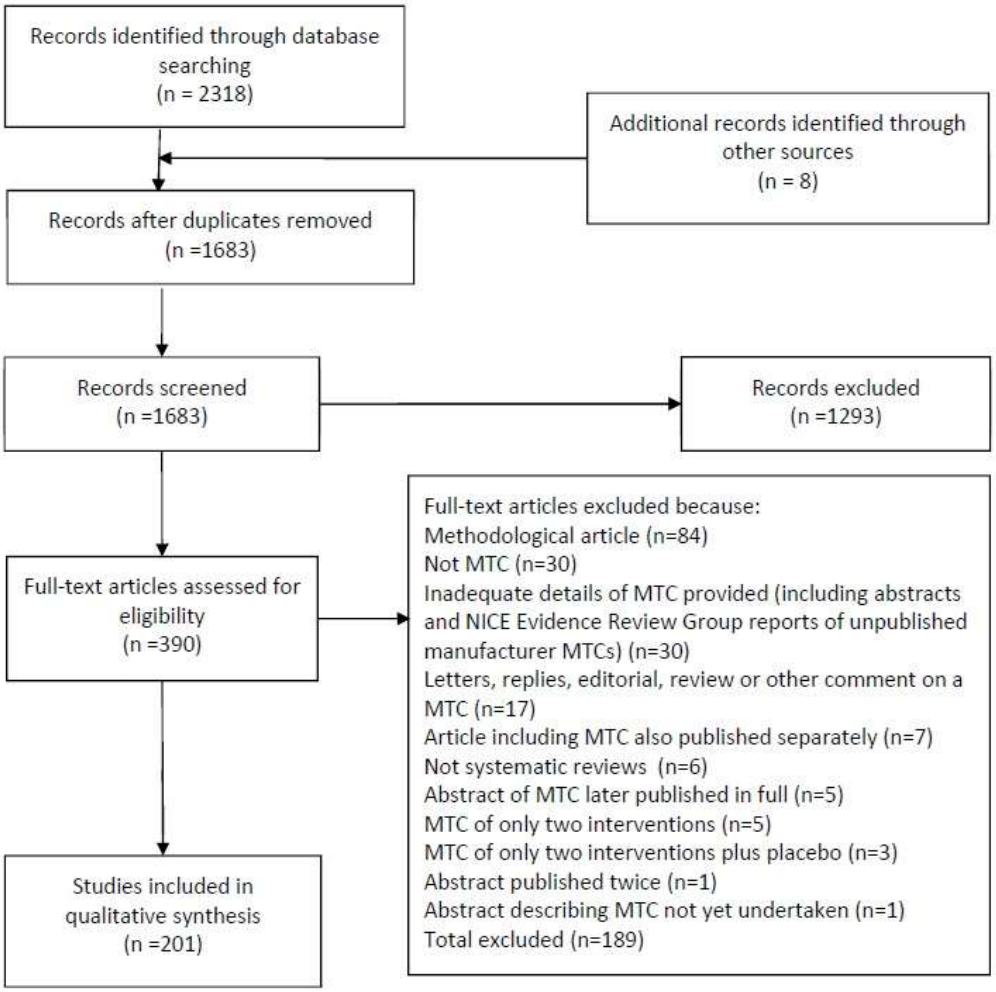

Figure 3.2. Article Flow Diagram

3.4.3 Number of reviews published

Three previous systematic reviews of indirect comparisons in systematic reviews were found[2, 10, 11]. In the two that were most recent to my review, Edwards[2] found seven MTCs in searches up to July 2007 and Song[10] found 18 MTCs in searches up to October 2008. In contrast, I found 201 NMAs published up to the end of June 2012.

Figure 3.3 shows the substantial increase in reporting of NMAs in systematic reviews since Edwards[2] in 2007 and Song[10] in 2008 reviewed indirect comparisons in systematic reviews. The number of reviews for 2012, in Figure 3.3, has been extrapolated by doubling the 6-month total found to the end of June and is represented by the dashed line.

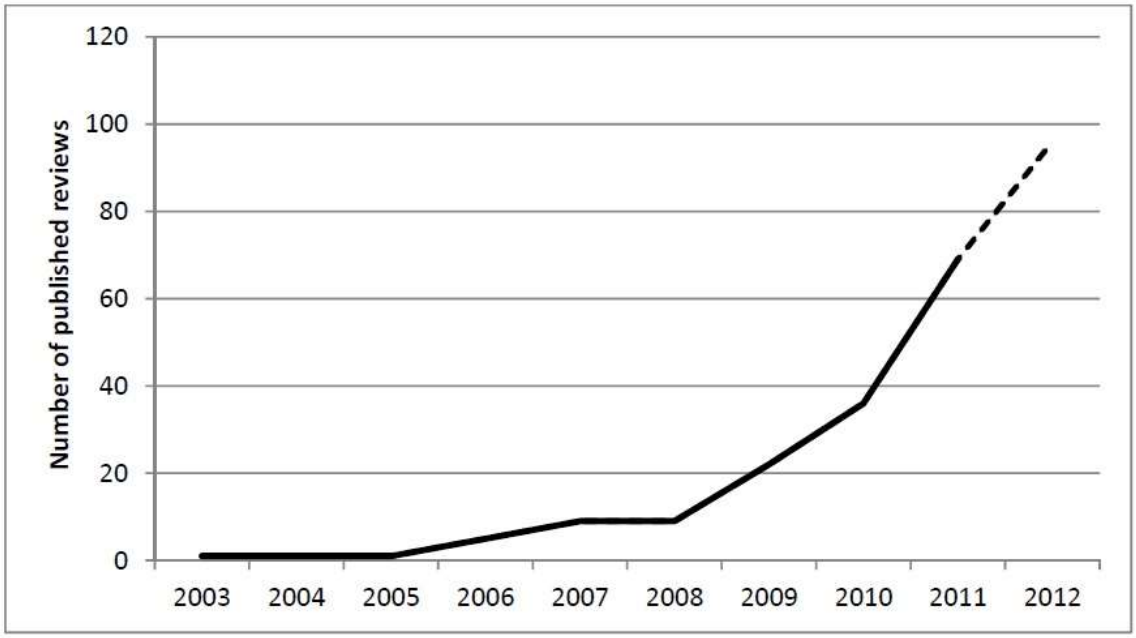

Figure 3.3 Publication of NMAs in systematic reviews, by year

3.4.4 Clinical topic of reviews and scope of use of synthesis

The types of review in which NMAs were conducted range widely in clinical topics and the nature of the comparison. In addition to the comparison of benefits and/or harms of pharmacologic and nonpharmacologic interventions, NMAs were used to provide estimates

of the relative effectiveness of interventions for use in cost-effectiveness studies and to compare outcomes from various widths of surgical resection margins. Another novel use was an early prediction of the effect of a newly introduced technology[12]. The interventions studied include surgical, pharmacologic, psychological, psychosocial, and lifestyle ones. Table 3.1 summarises the clinical topic areas of the reviews. As shown in Figure 3.4, the two most common clinical topic areas are also the ones with the fastest growth in publications since 2010.

Table 3.1 Number of reviews by clinical topic and number of interventions

Clinical Topic	Number of reviews	Number of interventions compared		
		Range	Median	IQR
Heart and circulation	41	3 - 11	6	3
Rheumatology	30	4 - 22	7	4.5
Cancer	25	3 - 41	6	2.5
Endocrine and metabolic	22	3 - 9	6	2
Infectious disease	22	3 - 17	6	4
Anaesthesia and pain control	16	4 - 22	6	3.5
Mental health	16	5 - 14	8	5.5
Lungs and airways	13	4 - 19	6	2
Gastroenterology	10	3 - 13	6	6.25
Neurology	8	4 - 145	6	3.25
Skin	8	3 - 17	8	5
Eyes and vision	6	4 - 30	8	11.25
Child health	6	4 - 8	7	1.75
Complementary and alternative medicine	6	4 - 22	7	3.25
Orthopaedics and trauma	5	4 - 17	6	5
Dentistry and oral health	4	6 - 15	8	3
Kidney disease	4	3 - 5	4	0.5
Urology	4	6 - 14	8	4.25
Blood disorders	3	4 - 41	5	n/a
Tobacco, drugs and alcohol dependence	3	4 - 6	4	n/a
Gynaecology	2	6 - 9	n/a	n/a
Ear, nose and throat	1	6	n/a	n/a

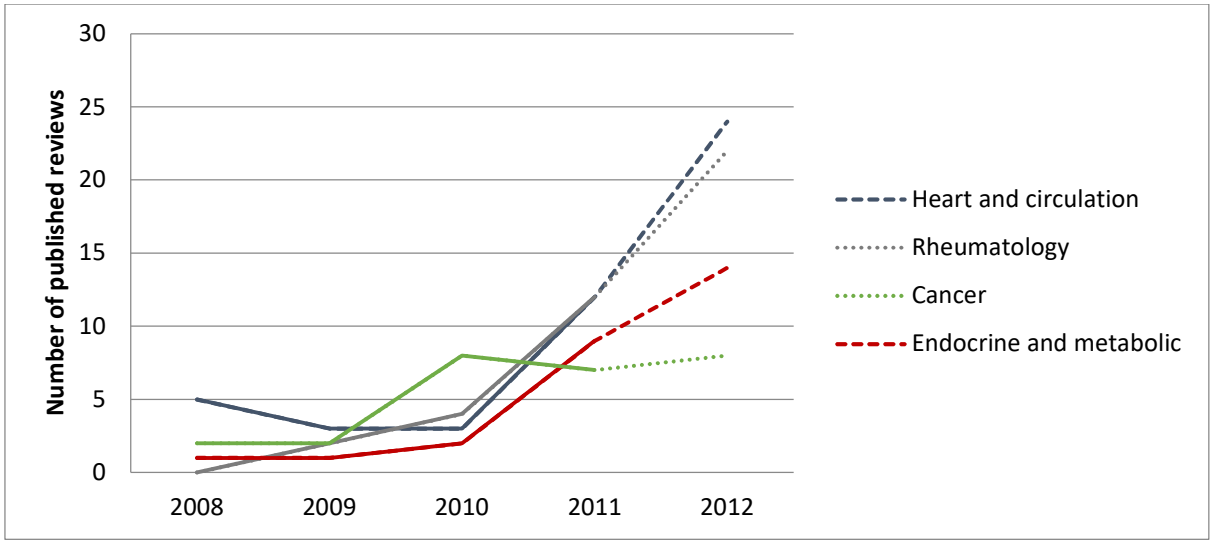

Figure 3.4 Publication of NMAs by clinical topic and year

3.4.5 Reported details of synthesis

Twenty four of the 44 abstracts and 35 of the 157 full publications did not indicate that either placebo or standard care was included in the interventions compared. Eighty-three percent of reviews compared up to nine interventions; 3% compared more than 19 interventions. Zintzaras et al[13] reported the MTC comparing the greatest number of interventions: 145.

Ninety-five percent of full review articles, compared to only 25% of conference abstracts, reported the software used for analysis. WinBUGS (Medical Research Council Biostatistics Unit, Cambridge, United Kingdom) was the most commonly used software (in 76% of the articles that reported software). Stata (StataCorp, College Station, Texas) was the next most common (14%).

Network structures were not reported in most articles. Unsurprisingly, only two conference abstracts included a network diagram but only 41% of full articles included one. Of the network structures provided in full articles, 80% were multi-loop structures, 9% were single loop, and 11% were star structures.

3.5 Discussion

3.5.1 Summary of the findings of this chapter

This review shows the substantial increase in the publishing of NMAs in systematic reviews from 2009 to 2012, when I did the searches for the review. The continued growth in NMA is discussed in Chapter Eight. The clinical topic areas in which there was the fastest growth to 2012 were Heart and circulation, Rheumatology and Endocrine and metabolic.

Although it was beyond the objectives of the review and, therefore, relevant data were not extracted, inconsistencies were observed in the reporting of NMAs in this sample of reviews. The specific methodology used, and the assumptions made were often unclear or not stated.

No MeSH term will specifically identify NMAs, and no alternative method was found to limit searches to specifically identify NMAs rather than pair-wise meta-analyses, in any of the databases searched.

The small number of eligible reviews that I identified for the period up to 2008 is consistent with that found in the previously published reviews on this topic[2, 10, 11]. The substantial increase from 2009 prompted consideration of what had led to this step change, which I explore below in relation to accessibility and acceptability of the NMA method, and changes in reporting.

3.5.2 Accessibility of the NMA method

Input from content and methodological experts is recommended when conducting systematic reviews[14]. Content experts may contribute in various ways, including interpretation of the clinical significance of differences between trials, awareness of unpublished trials and identification of the absence of information on expected outcomes e.g. specific harms. The ability of NMA to compare simultaneously more than two interventions, to compare interventions that have not been directly compared in trials and to increase the precision of

the estimate of effect size would be likely to interest content experts in seeking answers to clinical questions. Increased access to methodological experts with experience in use of NMA methodology might also increase the likelihood of content experts becoming involved in the conduct of systematic reviews using NMA.

Access to methodological experts experienced in use of NMA methods could be influenced by two main factors: software to conduct the analysis and researchers trained in the methodology. No key change in the version or availability of the most commonly used software (Win-BUGS and Stata) is apparent around 2009 that might have contributed to an increased use of NMA methodology. An increasing pool of researchers, however, had been trained. For example, Professor Tony Ades indicated that he and others, from Bristol and Leicester, had been running MTC courses teaching the Bayesian methodology using WinBUGS since 2007, mainly in the United Kingdom but also abroad, to more than 600 attendees in total, with increasing demand (personal communication 2012). This increased availability of methodological experts, working with content experts in review teams, might have been a factor in the observed increase in publications. Allowing for latency from training to publication and presuming that the increased pool of trained methodologists has influenced the number of NMA publications, this suggested that the increase since 2009 would continue, as it has done (see Chapter Eight).

3.5.3 Acceptability of the NMA method

The National Institute for Health and Care Excellence (NICE) publishes a guide to the methods of technology appraisal. The contents of the guide are likely to affect both the pharmaceutical industry's decisions on the research they conduct and commission and research by members of the wider health research community if they see technology appraisal as a potential use of their research. It is also likely that adoption of a method in the guide would affect perceptions of its acceptability. The NICE Methods Guide is updated at

intervals, and the 2008 update[15] included a section on MTCs. The review process before publication of each update to the guide includes distribution of briefings and workshops relating to methods. A briefing on MTCs was provided for the 2008 review process[16]. This briefing and workshop would also raise the profile of MTCs and potentially affect the perceptions of acceptability amongst a wider audience. The NICE Decision Support Unit publishes an Evidence Synthesis Technical Support Documents (TSDs) series[17] that covers many aspects of the conduct of MTCs. The TSDs are commissioned by NICE to provide detailed advice on the methods described in the Methods Guide. There was also an MTC briefing for the review of the Methods Guide that NICE undertook in preparation for their updated 2012 version. This 2011 briefing[18] cited '*the direction of the 2008 Methods Guide*' as one of the potential factors leading to the increased use of MTC methods.

3.5.4 Reporting standards for NMA

My review found that reporting of NMAs in systematic reviews was inconsistent. International consensus standards for the reporting of systematic reviews and meta-analyses were originally published in the form of the QUOROM statement[19], which evolved to become the PRISMA statement[20]. Contemporary to my conduct of this review, the PRISMA statement mentioned the existence of meta-analysis that combines direct and indirect comparisons but did not make recommendations relating to reporting that address the specifics of NMA methodology. Furthermore, the PRISMA statement directed readers to the guidance published by The Cochrane Collaboration[21] and the Centre for Reviews and Dissemination[22] for standards of conduct of systematic reviews, and each contained very limited guidance on the use of NMA methodology in systematic reviews. A basis for standards of conduct of NMAs could be found in the NICE Decision Support Unit's Evidence Synthesis TSDs[17] and in reports on the interpretation and conduct of NMAs

published in 2011 by the International Society for Pharmacoeconomics and Outcomes Research[6, 7].

3.5.5 Strengths and limitations of the research presented in this chapter

My bibliographic review used deliberately broad search terms and included NMAs that had been reported in conference abstracts but not in full journal articles to capture as many of the published reviews reporting NMAs as possible. Comparison with previously published overviews confirmed identification of all previously reported reviews[2, 10, 11]. However, it is still likely that some eligible reviews were not identified by my review because of not searching additional databases, the difficulty in identifying this specific type of review in databases, and human error in the review process, particularly because only one reviewer undertook the work. Additionally, inclusion was restricted to comparisons among more than two interventions (not counting placebo or usual care) because this research was prompted to explore NMA as a means for helping to answer the large number of clinical questions that involve choosing between more than two interventions. This restriction risked failing to identify additional data that would have added to the total number of reviews and risked introducing bias in the selection of included reviews (e.g. by clinical topic) if comparisons of a small number of interventions with placebo or usual care were more common in the NMAs for certain clinical topic areas. Three reviews were excluded because they compared only two interventions or compared two plus placebo, as reported in Figure 3.2. As stated earlier, separate publication was not found for many NMAs that are mentioned in HTAs, so the number of reviews included in this review is an underestimate of the number of NMAs conducted.

It is possible that researchers working on other published systematic reviews had considered undertaking NMA but decided not to proceed because it was not appropriate to do so with

the studies they identified for inclusion in their reviews. However, it was not possible for this review of published systematic reviews to identify or quantify examples of this.

Another limitation is that this review focused on producing a summary of published NMAs by clinical topic and the number of interventions compared. Alternative approaches might have explored other characteristics, such as the type of intervention (e.g. drug versus non-drug).

Finally, I did not formally assess the quality of reporting of the studies included in this review or extract data relating to this but did observe relevant inconsistencies and missing details in the course of extracting other data. This is reinforced by the findings of Coleman et al who published a review of NMAs in 2012[23], shortly after I completed my review, but with a search period limited to between 2006 and July 2011. They identified a smaller number of NMAs (42) than my review, for the comparable period, and did not comment on a trend in publication, but found inconsistency in reporting and concluded that guidelines for conduct and reporting were needed.

3.5.6 Implications of the findings of this chapter

After doing this review, I concluded, in 2013, that the increasing number of NMAs would justify the adoption of a specific MeSH term that would greatly ease the identification of NMAs. I also noted that it would be helpful if databases that do not provide searching by MeSH terms introduced ways to more easily identify NMAs, such as the addition of a relevant term to their indexing systems. However, by the timing of writing this thesis, in 2019, these limitations to identifying NMAs when searching databases persist.

In view of the substantial increase in publication of NMAs by 2012, , it seemed appropriate at that time for a consensus on standards of conduct and reporting of NMAs to be established and incorporated into guidance on best practice for systematic reviews. With this in mind, I conducted a survey of authors who had published systematic reviews reporting NMA to help

in identifying items to include in standards for reporting (Chapter Four). Furthermore, given that the use had remained very low compared with pair-wise meta-analysis despite the increase I found in 2012, the author survey was an opportunity to explore what factors could be limiting greater use.

3.6 Chapter Summary

- The number of published systematic reviews that reported NMA increased rapidly from 2009 and an estimated 90-100 were published in 2012.
- These reviews were difficult to identify because databases offered no focussed search method to separate them from the far more common reviews reporting pair-wise meta-analysis and because authors used various names when reporting NMA
- Previous reviews identified very small numbers of systematic reviews reporting NMAs up to 2008, so this increase over time had not been previously reported
- The reporting of NMAs was inconsistent. The specific methodology used, and the assumptions made were often unclear or not stated in the published reports.
- In view of the increasing use of this method, a consensus on terminology and standards for conduct and reporting seemed timely
- A survey of authors who had published systematic reviews reporting NMA might identify standards for reporting and allow research into the reasons for using NMA methods, reactions to using these methods and to the uptake of NMA findings.

3.7 Chapter References

1. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol 2014; 67: 138-143. 2013/10/05. DOI: 10.1016/j.jclinepi.2013.07.014.

2. Edwards SJ, Clarke MJ, Wordsworth S, et al. Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. Int J Clin Pract 2009; 63: 841-854. 2009/06/06. DOI: 10.1111/j.1742-1241.2009.02072.x.

3. Mills EJ, Bansback N, Ghement I, et al. Multiple treatment comparison meta-analyses: a step forward into complexity. Clin Epidemiol 2011; 3: 193-202. 2011/07/14. DOI: 10.2147/CLEP.S16526.

4. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009; 181: 488-493. 2009/08/06. DOI: 10.1503/cmaj.081086.

5. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21: 2313-2324. 2002/09/05. DOI: 10.1002/sim.1201.

6. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14: 429-437. 2011/06/15. DOI: 10.1016/j.jval.2011.01.011.

7. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14: 417-428. 2011/06/15. DOI: 10.1016/j.jval.2011.04.002.

8. Hutton B, Salanti G, Caldwell DM, et al. 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: 777-784. 2015/06/02. DOI: 10.7326/M14-2385.

9. Ades A.E. CDM, Reken S., Welton N.J., Sutton A.J., Dias S. NICE DSU Technical Support Document 7: Evidence Synthesis Of Treatment Efficacy In Decision Making: A Reviewer's Checklist http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD7-reviewer-checklist.final_.08.05.12.pdf (2012, accessed 12 October 2019).

10. Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 338: b1147. 2009/04/07. DOI: 10.1136/bmj.b1147.

11. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. Health Technol Assess 2005; 9: 1-134, iii-iv. 2005/07/15. DOI: 10.3310/hta9260.

12. Youn JH, Lord J, Hemming K, et al. Bayesian meta-analysis on medical devices: application to implantable cardioverter defibrillators. Int J Technol Assess Health Care 2012; 28: 115-124. 2012/05/09. DOI: 10.1017/S0266462312000037.

13. Zintzaras E, Doxani C, Mprotsis T, et al. Network analysis of randomized controlled trials in multiple sclerosis. Clin Ther 2012; 34: 857-869 e859. 2012/03/27. DOI: 10.1016/j.clinthera.2012.02.018.

14. Cumpston M and Chandler J. Chapter II: Planning a Cochrane Review. In: Higgins JPT, Thomas J, Chandler J, et al. (eds) Cochrane Handbook for Systematic Reviews of Interventions Version 60 [updated July 2019]: Cochrane, 2019 https://www.training.cochrane.org/handbook (accessed 18 October 2019).

15. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. National Institute for Health and Clinical Excellence, 2008 https://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed 6 September 2012).

16. Sutton A, Ades A, Abrams K, et al. Briefing paper for methods review workshop on evidence synthesis (indirect and mixed treatment comparisons). http://www.nice.org.uk/media/4A6/2F/EvidenceSynthesisBriefingPaper.pdf (National Institute for Health and Clinical Excellence, 2007, accessed 14 October 2012).

17. NICE Decision Support Unit (DSU). Evidence Synthesis TSD series, http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series(2391675).htm (accessed 5 May 2015).

18. Trowman R, Stevens A and Tappenden P. Briefing paper for methods review working party on mixed treatment comparisons. http://www.nice.org.uk/media/C67/40/TAMethodsGuideReviewSupportingDocuments.pdf (National Institute for Health and Clinical Excellence, 2011, accessed 14 October 2012).

19. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354: 1896-1900. 1999/12/10. DOI: 10.1016/s0140-6736(99)04149-5.

20. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339: b2700. 2009/07/23. DOI: 10.1136/bmj.b2700.

21. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. In: Higgins JPT and Green S (eds): The Cochrane Collaboration, 2011 http://www.cochrane-handbook.org/ (accessed 4 May 2015).

22. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. CRD, University of York, 2009 https://www.york.ac.uk/inst/crd/index_guidance.htm (accessed 4 May 2015).

23. Coleman CI, Phung OJ, Cappelleri JC, et al. Use of Mixed Treatment Comparisons in Systematic Reviews. Methods Research Report. (Prepared by the University of Connecticut/Hartford Hospital Evidence-based Practice Center under Contract No. 290-2007-10067-I.). https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/mixed-treatment-comparisons_research.pdf (2012, accessed 16 11 2019).

CHAPTER FOUR

Survey of authors

4 Survey of authors

The work described in this chapter was published in Systematic Reviews and much of the text of that article is repeated verbatim:

Lee, A.W. Use of network meta-analysis in systematic reviews: a survey of authors. *Systematic Reviews* (2016) 5:8

doi 10.1186/s13643-015-0174-4[1]

4.1 Background

To explore the themes identified in the review described in Chapter Three, I planned to survey the authors of the systematic reviews included in that review, to identify their perceptions of the use of NMA methods, particularly the standards for reporting they think should apply. Abdelhamid et al[2] surveyed authors of Cochrane Reviews about the use of any indirect comparison method in systematic reviews in 2011. Only 14 % of the Cochrane reviews included in that survey included an indirect comparison analysis, and it was not reported whether any of these were NMA. Additionally, only 23 % of respondents to that survey had ever used an indirect comparison method. The survey reported here was not limited to the authors of Cochrane Reviews and appears to be the first survey of authors who had published systematic reviews that report specifically the use of NMA.

As noted in Chapter Two, international consensus standards for the reporting of systematic reviews and meta-analyses were developed and published as the Quality of reporting of meta-analyses (QUOROM) statement in 1999[3] and updated in 2009 to become the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[4]. The PRISMA statement mentioned meta-analyses that combine direct and indirect comparisons but prior to 2015 there were no PRISMA recommendations for reporting that were specific to NMA methodology. Whilst I was conducting my survey, Hutton[5] reported on the development of an

extension to the PRISMA statement to cover the reporting of NMA and this extension was published in 2015[6]. Both publications cited the published article of my review described in Chapter Three[7]. I decided to complete the survey, to assess the extent to which it validated the items included in the new extension to the PRISMA statement and to explore the other theme of what was limiting the greater use of NMA.

4.2 Chapter aims and objectives

The aims of this chapter are to identify some potential standards for reporting NMA in systematic reviews and to identify potential factors limiting the greater use of NMA, based on a survey of authors that I conducted in 2014. The primary objective was to identify a list of reporting requirements that authors who had published NMA considered essential. The secondary objective was to analyse and summarise themes from these authors' responses regarding future use of NMA.

4.3 Methods

The post-positivist research paradigm was adopted in undertaking this qualitative survey. This is an interpretive approach, which assumes that reality is subjective and, therefore, is an appropriate option for exploring peoples' perceptions, as opposed to the positivist approach that assumes there is an objective reality to be discovered[8]. In keeping with this approach, the responses of authors are quoted extensively in my synthesis of the results. The methods used to identify eligible systematic reviews are described in Chapter Three. The target sample for this survey was one author from each of the 201 systematic reviews included in that synthesis. The survey was conducted by email. For ease of identifying contact details, the corresponding author was contacted but asked to respond with information of any other author they thought would be more appropriate to complete the survey. When notification of an email address no longer in use was received and no alternative contact email address for the corresponding author could be identified, the first named author was contacted as the next preference, then second named, and

so on. When a contacted author was identified to be responsible for more than one review, they were asked to comment on each review separately when answering any question that was specific to a single review.

As shown in Figure 4.1, the survey contained eight questions, which were included in the body of the email, following an introductory explanation regarding the background and purpose of the survey.

1	Why did you choose an indirect comparison method for meta-analysis in your review?
2	If you were repeating the review in future, would you choose an indirect comparison method again? If you would not, please explain why.
3	Please name any reporting guideline(s) you followed when reporting the indirect comparison and the review.
4	What, in your opinion, are the essential details to include for accurate and reliable reporting of indirect comparison in a review?
5	Please name any reporting guideline(s) that, in your opinion, includes the essential details for accurate and reliable reporting of indirect comparison in a review.
6	What is (are), in your opinion, the most important development(s) needed to support the conduct, reporting and use of indirect comparison methods?
7	What would encourage further use of indirect comparison methods by reviewers?
8	What would encourage further use of evidence from reviews reporting indirect comparison methods by decision makers (including policy makers, clinical practitioners and the public)?

Figure 4.1. Survey questions

The email informed authors that their responses could be quoted but would not be attributed and identified the specific reviews each author was being contacted about. Since this was a survey of research authors, using their contact details, which were in the public domain, asking about their research, which is also in the public domain, consent to take part was in the form of their choosing to reply. One reminder was sent to each author who failed to respond and to any author who replied indicating that they would respond but had not done so within four weeks or, when applicable, by the specific time they had indicated. The survey was conducted between March and July 2014. In line with National Health Service (NHS) guidance, application for research ethics approval was not required for this survey and application in advance of the survey for ethics approval from the University of Oxford was not made. I discovered only when submitting an article reporting the survey for publication, that I should have made this application. Following submission of a late application, the Research Ethics Manager, Medical Sciences issued a confirmation letter (Appendix 1) that the study did not raise any concerns regarding the university's policies for ethics approval and approval would have been given if the application had been submitted in advance of the survey.

An Excel spreadsheet was used to record, for each review, approaches to and responses from authors. Alongside each review, I recorded the date of publication, whether it was published in full or as a conference abstract, whether it was reported within a primarily methodological paper. As the survey was conducted, a record was made in the spreadsheet of the date of initial and any subsequent contacts. Authors' responses were also copied into a separate Excel spreadsheet before being imported to QSR Nvivo 10 for qualitative analysis.

The text of authors' responses was coded, in QSR Nvivo 10, by systematically processing each response and highlighting sections of text to create or add to existing nodes, which allowed themes to be identified and grouped under each question in the survey.

4.4 Results

Figure 4.2 shows the number of unique authors identified and contacted and the outcomes of contact. All ten reviews for which a valid email address could not be identified for any author had a contact author responsible for one or more conference abstracts, but none were fully published reviews.

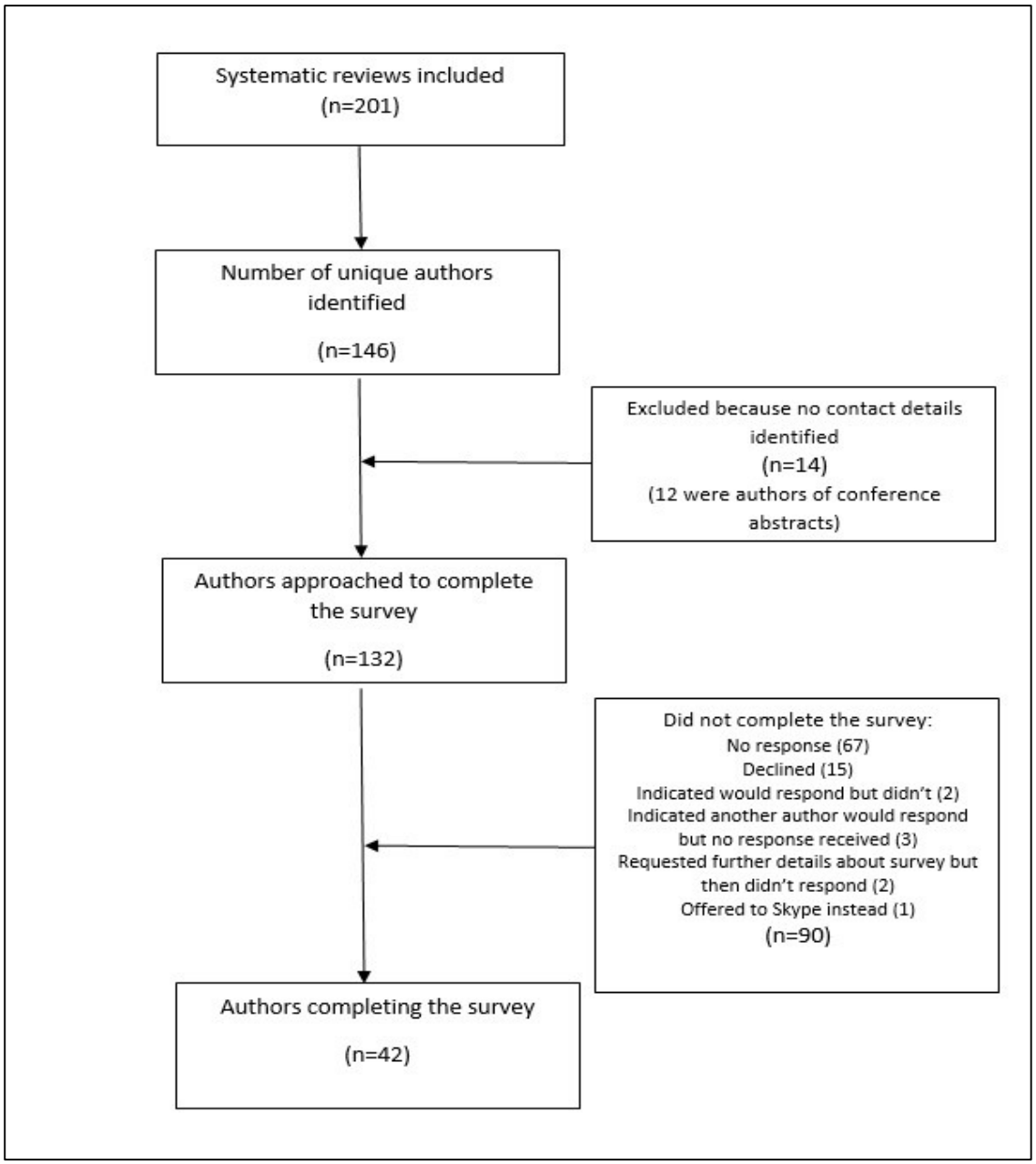

Figure 4.2. Flow diagram of authors contacted and outcomes

The 42 authors completing the survey gives a response rate of 32%. These authors were responsible for 54 (34%) of the 157 fully published systematic reviews and 5 (11%) of the 44 reviews that were only available as conference abstracts. Of the responses, 93% related to reviews published from 2009 onwards; these constituted 88% of all included reviews. None of the completed surveys related to any of the four primarily methodological papers, although one author of two of those papers responded that they declined to complete the survey because their papers were primarily methodological.

4.4.1 Why was a NMA method chosen?

Most responses (28) cited the lack of direct comparison data for all the interventions of interest as the reason for conducting a NMA. Other themes were the desire to rank interventions, the wish to try a new method due to interest generated by other publications, and the view that it improved the strength of the evidence. Comments received included the following:

'lack of evidence directly comparing most of the drugs of interest'

'we wanted to incorporate all available evidence from all trials whether or not they directly compared all treatments simultaneously'

'Since we were comparing several competing interventions, a method for ranking the interventions was intuitively preferable to preparing a series of pairwise comparisons'

'The estimates of indirect comparisons can strengthen and make more accurate the results of direct comparisons when indirect and direct evidence are combined in mixed treatment comparisons.'

4.4.2 Would the author use the same method if repeating the review?

All but one author would use the method again. That author specified the following reason:

'Because the direct result contradicted the network analysis'

Two authors stated they would carry out both direct and indirect meta-analyses,

'we indeed plan to perform a network MA again. But we will first perform a standard MA, as this method is more accepted, and then a NMA.'

and one qualified the decision,

'We would search the literature for direct comparisons, then decide whether an indirect comparison was still necessary'

Others explained their reason for using the method again,

'it enables us to answer the questions we would otherwise not be able to answer, that is, evaluating the comparative effectiveness of these agents there is no other approach that would allow simultaneous comparison of several interventions'

4.4.3 Was any reporting guideline followed?

More than half the responding authors (24) stated that they had not followed any reporting guideline. Most (11) of the other 17 respondents mentioned PRISMA[4] but commented that it does not have content specific to the NMA method. Other responses mentioned publications by ISPOR[9, 10], NICE[11], or the Australian Pharmaceutical Benefits Advisory Committee (PBAC)[12]. Some authors stated that they followed the publishing journal's own guideline or were guided by what other authors had reported when publishing NMA.

4.4.4 What are essential details to report?

Several authors (10) specifically commented that the same details as recommended for pair-wise meta-analysis should be included,

'The paper should include all the elements required by the PRISMA guidelines for standard MA'

Others emphasised the overall purpose of reporting,

'Detailed description of methods used that would allow replication of the MTC'

Some think that data files should be provided,

'All data in a separate file'
'Raw data from the trials so that the analysis can be replicated'

The detailed reporting elements proposed by authors are summarised in Table 4.1 (methodological) and Table 4.2 (results).

Table 4.1 Methodological details that should be reported

Summary of proposed essential methodological details to be reported
A clear research question
Search details – including terms, databases, period, language restrictions
Inclusion/exclusion criteria including whether study quality was assessed and by what method
Assumptions *'The main issue for MTCs to be valid is the consistency assumption. Very often there is not enough information on the included studies to determine whether the consistency assumption is actually reasonable.'*

'Assessment that the studies are comparable. e.g. consistent endpoint definitions, differences in patient populations or study settings are not expected to influence the treatment effect. Explicitly state and discuss assumptions of the analysis.'
Detailed statistical analysis plan including: software; Bayesian/Frequentist method; random/fixed effects model; code; co-variates; handling of missing data; pooling of data; choice of priors; assessment of heterogeneity; assessment of consistency. *'sensitivity analysis of the impact of analyst assumptions is also very important.'*

Table 4.2 Result details that should be reported

Summary of proposed essential result details to be reported
Details of included studies: *'numbers of included studies for each direct comparison; numbers of subjects for each treatment'* *'study bias assessment'* *'the quality assessment of included studies'*
Details relating to the network: *'Network diagram including the number of trials included in each link'* *'Sources of heterogeneity must be assessed and the impact of heterogeneity must be analysed.'* *'Evaluation of the "confidence" in the network (amount of evidence, homogeneity, consistency)'* *'How good a fit the chosen model is to the data set.'*
Details relating to effect estimates: *'point estimates and confidence/credible intervals'* *'95% credible intervals/probability intervals must be included when reporting the effect estimate'* *'absolute effect of each intervention [when reporting input parameters of economic modelling]'* *'reporting of estimates and variances of the direct comparisons that form the indirect comparison'* *'Comparison of results from direct evidence with results from NMA'* *'Sensitivity analyses if necessary, and an explanation of the differences compared with standard MA'* *'The key information needed is to provide separately the direct and indirect estimates, and try to provide the quality of evidence supporting each.'* *'rankings [perhaps rankograms as well] and probabilities of each intervention being best.'* *'For Bayesian mixed treatment comparisons: Probability Rankograms and Surface Under Cumulative Ranking Curve'* *'Well reasoned sensitivity analysis, including/excluding different data sources.'*

4.4.5 Does any reporting guideline contain the essential details?

The most frequent response (11) to this question was that there was no such guideline. A few authors mentioned the planned extension to the PRISMA statement. Of those who named a guideline, the most frequently named, in decreasing order, were ISPOR[9, 10], the NICE TSD series[11], and PRISMA[4].

4.4.6 What developments are needed to support conduct, reporting, and use of the methods?

Most authors (30) stated that development and promotion of such guidelines is needed,

'development of guidelines that include all details necessary to the performance of an indirect comparison'
'Specific guidelines on how to properly report and interpret the results of a network meta-analysis should also be developed'

Each of the following themes was raised by some authors:

4.4.6.1 Methodological

'more evidence that ranking probability is an accurate measure'
'I believe that MTCs should only be presented as comparative effect estimates but never as a ranking of interventions.'
'Risk of bias tools'
'incorporate methods such as I^2 for NMA and the incorporation of inconsistency as in the design-by-treatment model.'
'a way to present results in a simple way (perhaps not in print, but rather using animation) to include absolute and relative effects, risk of bias, for each outcome, with info about precision, heterogeneity and coherence, and overall quality of evidence.'
'objective way of testing for convergence'
'inclusion of evidence from observational studies is probably an area for development to make use of all available evidence'

4.4.6.2 Software

'more easy-to-use software would be good'
'not requiring a higher degree in statistics to make use of it'
'Improvement of available software – at the moment WinBUGS and R are considered by many reviewers too complex to use in routine systematic reviewing.'
'New/improved software needs to be more user-friendly and less time consuming.'

4.4.6.3 Training

'currently few persons are able to carry out the statistical computations, and this is limiting the diffusion of network meta-analyses'
'The statistical methodology should be diffused much more than it is today'

4.4.6.4 Presentation of results

'What is needed the most is a way to communicate the findings to readers/users.'

4.4.6.4 Support

'access to specialist statisticians'

4.4.6.5 Reservations

Several authors (11) expressed concerns about widening the use of NMA methods and evidence produced using those methods:

'I am concerned that accessibility may be at the expense of thinking'
'my concern is the unthinking use of any method without a statistical appreciation of the model(s) and the assumptions'
'Particularly when they use rankings, the information is often misleading because readers tend to focus on the top intervention'
'Nowadays people "believe" in MA as if it were the absolute truth, not understanding that there are good quality and poor quality MA. And same with network MA, but the risk is even higher in this latter case.'
'These methods are becoming increasingly popular, and there are many examples of poorly performed analyses'
'There are many subtleties and underlying assumptions in performing such an analysis, and increasingly there are many "automated" analysis, where data are pulled from papers and fed into computer programs, with poor assessment or even identification of assumptions.'
'Good guidelines have been developed and are being disseminated. They need to be disseminated more widely so that …those publishing indirect comparisons understand how to assess their quality'
'stricter editorial processes to ensure adherence to systematic review, statistical and reporting standards'
'Good guidelines have been developed and are being disseminated. They need to be disseminated more widely so that …those utilizing indirect comparisons in their decision making are able to recognize when they are performed correctly and can be confident in their results'
'The problem of network meta-analyses and indirect comparisons is that the statistical methods are difficult for most readers to understand or reproduce. Therefore, it is a black box.'
'I believe the use of it should be determined via the careful judgement of whether it is really needed, not just 'encouraged' regardless of the research question and the level of parameter heterogeneity.'

4.4.7 What would encourage more use of the methods by systematic reviewers?

Guidelines were again a common theme in response to this question, alongside greater perceived acceptability of use of the NMA methods, particularly through endorsement by key organisations and increased likelihood of publication,

'Guidelines would help promotion of guidelines etc to Cochrane review groups'
'Uptake would also be increased if HTA organisations other than NICE gave explicit statements on the acceptability of indirect comparisons and NMA and issued guidance.'
'Less scepticism by the methods community'
'Increased likelihood of publication'
'Stronger journal policies encouraging its use.'

Other themes were training, software development, and access to statistical expertise:

4.4.7.1 Training

'Reviewers need to be educated in the proper performance and reporting of indirect comparisons. More broadly, they need to be educated that such methods exist and can be used to derive answers unattainable by other methods.'
'The knowledge of the method, of its possibilities, aims and limits would encourage the reviewers to use MTC'
'Understanding of assumptions and pitfalls to give more confidence in the use of the method.'

'Better understanding of indirect comparison methods, and their comparability with 'classical' meta-analysis.'
'The reviewers need to understand the advantages of such methods, and the fact that most limitations of the indirect comparison methods are indeed limitations of the classical direct comparison methods as well.'

4.4.7.2 Software

'Software has to become more user-friendly.'
'easy frequentist software'
'Standard code for different situations'

4.4.7.3 Statistical expertise

'Currently, the statistical expertise necessary is the main limitation'

4.4.8 What would encourage more use of the methods by decision makers?

Training was again a clear theme,

'Again, education is key. Indirect comparisons can be performed well or poorly, and exposure to well performed indirect comparisons and education in how to identify one that has been performed poorly will enable decision makers to broaden the array of evidence that they can use to support their decisions.'

'including tutorial articles explaining the basic premise in non-technical language and individual applications explaining their methods and results in a way that is accessible to wider readers.'

There was mixed opinion regarding the desirability of this aspect of increased use,

'I would not encourage it for the sake of it. I think there are situations where it would be better avoided, e.g. where there is robust direct evidence.'
'the methods are complex and it would be very difficult for someone without extensive technical expertise to identify what the flaws are in any given indirect comparison or network meta-analysis, let alone whether the code used in the analysis was correct.'
'Currently it is seen as a "strange, peculiar" method, and somewhat mistrusted'
'I think it's moving in the right direction really – NICE support them, because they're a 'necessary evil' – there's a well known disconnect between the trial design required for licensing, and that for the evidence base needed for reimbursement and clinical decision making. People maybe still see it as 'voodoo' but in general I think we at least in the UK have a reasonably good and established process for selecting the best evidence available to inform decision making.'
'We currently stand at a crossroads where policy makers are content for indirect comparisons to be undertaken for the purposes of health economic evaluation. However, a fundamental shift in understanding needs to take place so that decision makers (policy makers/clinicians and in time the general public) accept the validity of the estimates of safety/efficacy that come from the analysis in their own right. In particular, indirect comparisons appear to have the same "face validity" to clinicians as pairwise meta-analysis did 15+ years ago.'

4.5 Discussion

4.5.1 Summary of the findings of this chapter

This survey found that systematic review authors who had used NMA methods up to 2012 did so mainly because of the lack of direct trial data or because of the ability to compare and rank multiple interventions. Almost all responding authors who had used these methods were inclined to use them again. Their responses demonstrated strong support for adoption of standards for conduct and reporting of NMA. The elements of reporting standards proposed by these authors are substantially consistent with those reported by Hutton et al[5] as part of the development work for and final publication of the extension of the PRISMA statement[6]. This should augur well for the adoption of that extension, suggesting that it will carry the support of authors publishing this form of meta-analysis. However, support in principle for the content of a reporting guideline does not however necessarily translate into compliance[13]. Whilst several responders to the survey commented that reporting standards should include those required for pair-wise meta-analysis, most reported that they had not followed any reporting guideline.

A notable theme arising from the responses was the tension between, on the one hand, a view that the use of NMA should be more easily accessible, particularly in the form of software tools, and on the other, concerns that there was some inappropriate use of the methods which wider use and greater accessibility might exacerbate. This tension prompts the question as to whether end users (clinicians, commissioners, policy makers) have the ability to critically appraise and interpret the results produced using this methodology. Therefore, in 2015, I concluded that the adoption of standards for conduct and reporting would be significant steps towards clarifying what is appropriate use and what is not, and also that this should be followed by identification of a suitable critical appraisal tool to further support end-users.

4.5.2 Strengths and limitations of the research presented in this chapter

This was the first published survey exploring the attitudes of systematic review authors who had used NMA methods, to reporting standards for such research. The survey focussed mainly on standards for reporting and included only brief mention of standards for conduct.

In adopting the post-positivist approach, I did not assume that the responses would be representative of all authors, all content or methodology experts nor, even, of my selected cohort of authors with experience of conducting and publishing systematic reviews that reported use of NMA up to 2012, because the responders may not be representative of the whole cohort. Perspectives of respondents could be influenced by many factors, including the purpose of conducting their review and their funding source, which I did not attempt to explore. Instead, I sought to identify 'leads' that I could explore further in subsequent elements in my programme of research.

The respondents formed a minority of the authors approached and the views of non-respondents are unknown, but there was no apparent bias in those responding in terms of year of publication or in the number or subject of reviews published. However, I did not collect characteristics of the respondents because it was not the aim of this study to identify whether the responses could be regarded as representative of the whole cohort. Conducting the survey by email afforded limited opportunity to clarify responses and did not allow me to pursue lines of enquiry, so the responses are what authors chose to volunteer and therefore cannot be assumed to represent all their opinions. As a part-time research student, researching reporting standards and critical appraisal criteria for NMA reported in systematic reviews, the survey questions chosen made specific mention of reporting guidelines, which may have influenced the mention of these in responses. My research interests influenced the questions and were likely to have influenced both the selection of responses I quoted and my overall interpretation of all responses; however, I approached this survey without pre-conceived expectations of what the

responses would be but with the hope that the responses might identify specific issues or themes that would be suitable for further exploration.

4.5.3 Implications of the findings of this chapter

There was clearly some reservation, even amongst existing users of NMA methods, regarding greater use of the methods by researchers and about the ability of publishers and end-users to critically appraise the quality of evidence produced using these methods. This finding highlighted a need for definition and adoption of standards for conduct and reporting of NMA, as well as the need to identify a suitable critical appraisal tool.

To further examine trends in the use of NMA methods for synthesis in systematic reviews, before proceeding to identify a suitable critical appraisal tool, there was a further question I wished to explore: what impact was the increasing amount of evidence produced using NMA having on health service policy, in the form of being incorporated into national clinical guidelines? To assess this, I decided to undertake a review of recent NICE clinical guidelines to identify how much evidence produced using NMA was included and this is reported in Chapter Five.

4.6 Chapter summary

- The authors of systematic reviews that included a NMA and had been published up to 2012 demonstrated strong support for adoption of standards for conduct and reporting of network meta-analysis
- The elements of reporting standards proposed in this 2014 survey were consistent with those included in the 2015 PRISMA extension statement
- As more widespread use of NMA methods continues to develop, it should be accompanied by an assurance that the use of these methods is appropriate
- Adoption of standards for conduct and reporting would be a significant step towards clarifying what is an appropriate use of NMA methods and what is not
- This should be followed by the development of a critical appraisal tool to support end-users of systematic reviews using NMA methodology
- Since there are some reservations about the ability of end-users to critically appraise evidence produced using NMA, I decided to review recent NICE clinical guidelines to identify how much of that form of evidence synthesis was included

4.7 Chapter references

1. Lee AW. Use of network meta-analysis in systematic reviews: a survey of authors. Systematic reviews 2016; 5: 8. 2016/01/23. DOI: 10.1186/s13643-015-0174-4.

2. Abdelhamid AS, Loke YK, Parekh-Bhurke S, et al. Use of indirect comparison methods in systematic reviews: a survey of Cochrane review authors. Research Synthesis Methods 2011; 3: 71-79. DOI: 10.1002/jrsm.51.

3. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354: 1896-1900. 1999/12/10. DOI: 10.1016/s0140-6736(99)04149-5.

4. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339: b2700. 2009/07/23. DOI: 10.1136/bmj.b2700.

5. Hutton B, Salanti G, Chaimani A, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. PLoS One 2014; 9: e92508. 2014/03/29. DOI: 10.1371/journal.pone.0092508.

6. Hutton B, Salanti G, Caldwell DM, et al. 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: 777-784. 2015/06/02. DOI: 10.7326/M14-2385.

7. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol 2014; 67: 138-143. 2013/10/05. DOI: 10.1016/j.jclinepi.2013.07.014.

8. Wildemuth BM. Post-Positivist Research: Two Examples of Methodological Pluralism. The Library Quarterly 1993; 63: 450-468. DOI: 10.1086/602621.

9. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14: 429-437. 2011/06/15. DOI: 10.1016/j.jval.2011.01.011.

10. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14: 417-428. 2011/06/15. DOI: 10.1016/j.jval.2011.04.002.

11. NICE Decision Support Unit (DSU). Evidence Synthesis TSD series, http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series(2391675).htm (accessed 5 May 2015).

12. Report of the Indirect Comparisons Working Group (ICWG) to the Pharmaceutical Benefits Advisory Committee: Assessing Indirect Comparisons., http://www.pbs.gov.au/industry/useful-resources/PBAC_feedback_files/ICWG%20Report%20FINAL2.pdf (accessed 5 May 2015).

13. Turner L, Shamseer L, Altman DG, et al. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical

journals? A Cochrane review. Systematic reviews 2012; 1: 60. 2012/12/01. DOI: 10.1186/2046-4053-1-60.

CHAPTER FIVE

Use of network meta-analysis evidence in NICE clinical guidelines

5 Use of network meta-analysis evidence in NICE clinical guidelines

5.1 Background

The National Institute for Clinical Excellence (NICE) (as it was then called) was established in April 1999 with the aim to create consistent guidelines to reduce variation in the availability and quality of healthcare delivered by the National Health Service in the United Kingdom. In 2005, NICE changed its name to the National Institute for Health and Clinical Excellence after absorbing the Health Development Agency. Its mission was expanded to include the development of public health guidance to help prevent illness and promote healthier lifestyles[1]. It subsequently became known as the National Institute for Health and Care Excellence, bringing social care into its scope.

In 2002, NICE published its first clinical guideline: CG1 Schizophrenia[2]. Clinical guidelines recommend how healthcare professionals should care for people with specific conditions. In the UK, NICE guidelines are also aimed at health service managers and commissioners of NHS services. They can cover any aspect of a condition and may include recommendations about providing information and advice, prevention, diagnosis, treatment and longer-term management[3]. NICE publishes details of who is involved in producing their guidelines, ranging from multidisciplinary expert teams to patients and carers, and the methodology used to produce the guideline[4]. Standard systematic review techniques are used to search for and review evidence. Previously conducted primary research or systematic reviews may be included in the evidence used. In addition to these sources, the Guideline Development Group (GDG) established for each guideline, may selectively undertake new meta-analyses to combine the results of primary research.

As discussed in Chapters Two and Eight, pair-wise meta-analysis is by far the most commonly used method of synthesis within systematic reviews of healthcare interventions[5] but the use of network meta-analysis (NMA) methods is increasing rapidly[6]. These methods combine direct and indirect evidence to address the frequent absence of randomised trials that directly compare all the interventions of interest. NICE has published advice on the conduct of NMA in its Decision Support Unit's Evidence Synthesis Technical Support Documents (TSDs)[7].

I undertook a review to identify and summarise the use of evidence from synthesis using NMA within recently published or updated NICE clinical guidelines, as of 2016. The time-period chosen was based on the findings of my earlier review[6], reported in Chapter Three, which identified a rapid increase in the publication of systematic reviews using NMA methodology between 2009 and 2012. Allowing for further increased uptake of use and the time taken to produce guidelines, my assumption was that impact on NICE guidelines should have been evident by 2015.

5.2 Chapter aims and objectives

The aim of this chapter is to summarise the use of evidence from synthesis using NMA within NICE clinical guidelines that were published or updated in 2015 or 2016. The primary objective was to identify the number of instances of results from NMA being directly linked as the evidence for recommendations within guidelines. Secondary objectives were to summarise references within guidelines to previously published systematic reviews that had used NMA methodology for synthesis, and to summarise NMAs undertaken by the GDG for the guideline.

5.3 Methods

5.3.1 Search

The Guidance section of the NICE website[8] was searched by month, using filters to select 'Clinical guidelines' from January 2015 to December 2016. These filters select any clinical

guideline first published or updated between those dates. The search was conducted on 8 March 2017.

5.3.2 Inclusion and exclusion criteria

Only guidelines covering clinical interventions were included. In each case, the full guideline and any addendum were included. Guidelines that were updated by NICE during the review and before the documents were retrieved were excluded.

References to excluded studies in the clinical guidelines were not included or screened to identify systematic reviews or NMA.

5.3.2 Data extraction

The following data were extracted to an electronic spreadsheet: guideline title, guideline reference number, month published, month last updated, total number of references (guideline and any addendum), number of references that are systematic reviews, references that are systematic reviews using NMA methodology, pair-wise meta-analysis conducted by the GDG, NMA conducted by the GDG, any recommendation in the guideline referenced to evidence from NMA.

Guideline and addendum documents were searched using keywords – 'systematic', 'meta-analys' (to detect both meta-analysis and meta-analyses) and 'network' to initially identify systematic reviews and meta-analyses referred to in the narrative, tables and reference lists of the documents. Reference lists were then screened by title and abstract initially, with full text retrieved in case of uncertainty to identify systematic reviews and whether a NMA was conducted.

5.4 Results

77 clinical guidelines were identified as published or updated by NICE from January 2015 to December 2016. Three guidelines were excluded from this review because they were not clinical

guidelines covering clinical interventions: NG47[9] was a cancer service guideline, NG40[10] was a service delivery guideline, and NG26[11] did not cover any clinical intervention. Two further guidelines, NG28[12] and CG61[13], were excluded because the version of these guidelines available was no longer from within the searched period, due to documentation being updated by NICE before I retrieved the guideline version from within the qualifying period. The references for one guideline, CG64[14], were not available from the NICE website but were obtained by personal contact with NICE.

Data were extracted from 72 clinical guidelines and from addendums to 16 of those. See Appendix 3 for references to the included guidelines and Table A2 for the guideline characteristics.

A total of 27,831 references from these guidelines were identified and screened: 1276 of these references were systematic reviews and 15 of these systematic reviews used NMA methodology.

Nearly all systematic reviews referenced in the guidelines were identified by the initial document keyword searches. Only a small number of systematic reviews were identified by subsequent screening of the reference lists of each guideline.

NICE clinical guidelines included in this review presented references in varying styles. References were sometimes numbered sequentially in the order they occur within the text and other times alphabetically by first author. Sometimes, references for the whole guideline were presented collectively near the end of the full guideline or as a separate appendix, and other times they appeared at the end of each chapter of the guideline.

70 of the 72 guidelines referenced previously published systematic reviews. In 12 of those guidelines, at least one of the systematic reviews referenced included use of NMA methods for synthesis.

68 of the 72 guidelines reported meta-analysis undertaken by GDGs. Figure 5.1 shows the number of reviews in which GDGs undertook pair-wise meta-analysis (PWMA) only, both pair-wise meta-analysis and NMA (PWMA+NMA) or did not undertake any form of meta-analysis (MA).

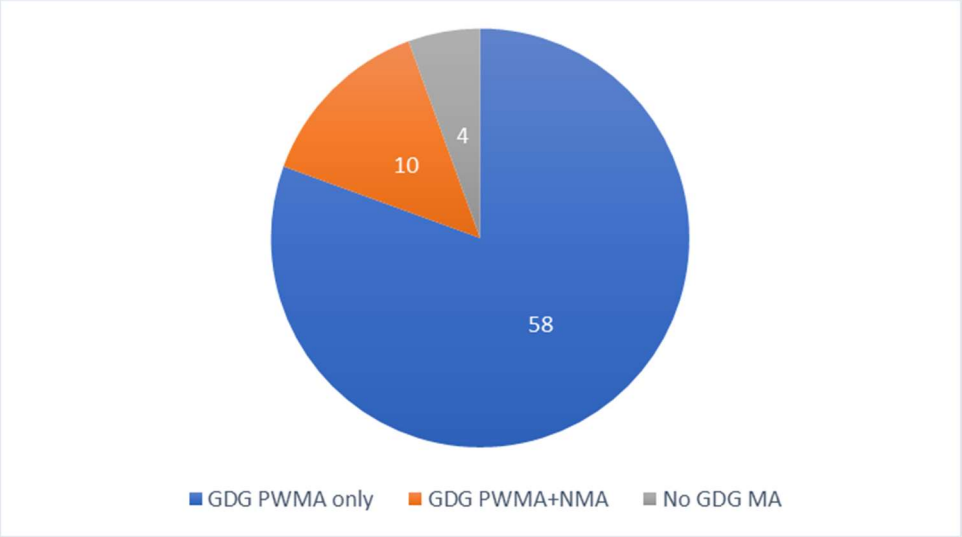

Figure 5.1 NICE Guidelines: use of GDG meta-analysis

In addition to the 10 guidelines that reported use of NMA conducted by the GDG, 7 guidelines stated that conducting NMA was considered by the GDG, but not undertaken for specified reasons. Two of the 12 guidelines that referenced systematic reviews using NMA methodology also reported additional NMAs conducted by the GDG.

5.4.1 Guidelines referencing reviews using NMA methodology with the results not linked as evidence for recommendations

Eight of the 10 guidelines that referenced systematic reviews using NMA methodology, without additional NMAs being conducted by the GDG, did not reference the NMA results as evidence supporting recommendations.

CG126[15] *Stable angina* referenced a systematic review by Stettler et al[16], which reported NMA of the outcomes associated with drug eluting and bare metal stents. The guideline referenced it only as introductory context to show that there is evidence that these stents,

'reduce the risk of restenosis and need for repeat revascularisation procedures, but have no impact on mortality'

but did not use it as evidence to support any recommendation.

CG140[17] *Opioids in palliative care* referenced Bekkering et al's systematic review[18], which reported NMA of strong opioids, but the guideline's recommendations on choice of first-line maintenance therapy were to offer oral sustained-release morphine as first-line maintenance, whereas Bekkering concluded there was no clear superiority of morphine over other opioids. The same guideline also referenced a systematic review by Vissers et al[19] which used NMA to compare the efficacy of selected short-acting strong opioids, of which intranasal Fentanyl provided the greatest reduction in pain. However, the recommendation in the guideline was to not offer short-acting Fentanyl on the basis of higher cost.

NG9[20] *Bronchiolitis* referenced a systematic review, using NMA methodology, by Hartling et al[21]. The GDG decided to exclude this from their clinical review *'because the methods were unclear'* but decided to use results from the NMA as input to the economic analyis they undertook. Confusingly, although the guideline narrative correctly cited the review as Hartling 2011 (or once, in error, Harling), it only included a Cochrane Review by Fernandes et al[22], which Hartling co-authored, in the list of references; but that review did not report NMA.

CG181[23] *Lipid Modification* referenced a systematic review by Ara et al[24], in which clinical effectiveness of statins was assessed from a NMA of 28 trials, but the evidence cited to support the recommendations was the economic evaluation reported in this review, rather than the NMA results.

NG59[25] *Low back pain and sciatica in over 16s* referenced a systematic review by Lewis et al[26], in which the clinical effectiveness of interventions for sciatica was assessed by conducting several

NMAs, but NG59 reports that this review was only considered as evidence for economic evaluation and was excluded due to '*a combination of applicability and methodological limitations'*.

CG90[27] *Depression* referenced a systematic review by Cipriani et al[28] which reported a NMA of 12 antidepressants and the GDG used the results in an economic analysis but concluded there was

'sufficient doubt about the clinical importance of the differences to not justify the development of recommendations for specific drugs'.

NG14[29] *Melanoma*, regarding the method, frequency and duration of follow up, mentioned taking survival following treatment for distant recurrence from a review by Dequen et al[30], that used NMA to assess the effect of various systemic interventions on survival following treatment for distant recurrence of melanoma. However, the section of the guideline making recommendations about systemic anticancer therapy did not reference Dequen's NMA.

NG35[31] *Myeloma:diagnosis and management* included recommendations on using a bisphosphonate to prevent bone disease and referenced a Cochrane Review by Mhaskar et al[32], which assessed the effect of bisphosphonates on risk of vertebral fractures, pain and overall survival for patients with myeloma. The NMA reported by Mhaskar, however, did not identify significant difference in effectiveness or harm between the different bisphosphonates compared and the guideline recommendation of Zoledronic acid as the first choice of bisphosphonate was not based on the NMA results.

5.4.2 Guidelines referencing reviews using NMA methodology with the results linked as evidence for recommendations

Two guidelines referenced previously published NMA results as evidence supporting recommendations but only one of these was a case of NMA providing evidence of significant difference in effectiveness between interventions.

CG137[33] *The Epilepsies* referenced a systematic review by Tudur Smith et al[34]. This systematic review included a NMA of individual patient data from randomised trials of anti-epileptic drugs (AEDs). The results of the NMA were used by the GDG in an economic analysis to assess cost-effectiveness of AEDs used as monotherapy for adults with newly diagnosed focal epilepsy but were not directly cited in the guideline's recommendations for this form of epilepsy. However, the results were directly cited as evidence for recommendations made regarding choice of AEDs for generalised tonic-clonic seizures.

NG36[35] *Cancer of the upper aerodigestive tract* referenced two systematic reviews reporting NMA, by Chen et al[36] and Yan et al[37]. The guideline included recommendations on treatment of locally advanced nasopharyngeal cancer based on results of NMAs in both reviews, which showed that there was uncertainty about the effectiveness of using adjuvant or neo-adjuvant chemotherapy.

5.4.3 Guidelines referencing reviews using NMA methodology and also NMA conducted by the GDG

Two guidelines referenced previously published NMA results and also undertook new NMA.

NG25[38] *Preterm labour and birth* referenced Haas et al[39] but conducted a new NMA *'structured around the database of…'* that review. The guideline's recommendations regarding use of tocolysis used evidence from the NMA conducted by the GDG rather than from the NMA of Haas et al.

CG185[40] *Bipolar disorder* referenced several systematic reviews by Cipriani et al, two of which reported NMAs[28, 41]. The guidelines' recommendations on managing mania and hypomania used evidence from Cipriani's NMA of antimanic medication[41] in addition to results of new NMA undertaken by the GDG. The other review by Cipriani et al[28] was mentioned as,

'evidence of a clinically significant degree of differences in both efficacy and tolerability among antidepressants in unipolar disorder'

but was not referenced as evidence for recommendations, because the guideline relates to bipolar disorder.

5.4.4 Guidelines with no references to reviews using NMA methodology but including NMA conducted by the GDG

Eight guidelines did not reference any previously published NMA results but included newly conducted NMAs.

NMAs were conducted by the GDG for CG152[42] *Crohn's disease*, as part of economic analysis, but the results of the NMAs were not directly linked to any recommendations.

NG17[43] *Type 1 diabetes in adults* included recommendations on preferred insulin regimens based on evidence from NMA conducted by the GDG to rank various long-acting insulin regimens for change in HbA1c level and for the risk of major hypoglycaemic events.

For NG33[44] *Tuberculosis*, the GDG conducted multiple NMAs of treatments assessing effectiveness and adverse events. The results of these NMAs were cited in the recommendations for treatment of latent tuberculosis.

CG92[45] *Venous thromboembolism* included recommendations regarding prophylaxis based on multiple NMAs and economic analyses conducted by the GDG. NMAs estimated the risk of deep vein thrombosis, symptomatic pulmonary embolism and major bleeding with selected prophylactic interventions or no prophylaxis after selected surgical procedures. The results of NMA were also used as the inputs to economic analysis.

CG150[46] *Headaches* reported multiple NMAs conducted by the GDG to assess the effectiveness of treatments for acute migraine and for prophylaxis of migraine. The results of the NMA were also used as inputs to economic modelling conducted by the GDG. Recommendations on acute

and prophylactic treatment for migraine were based on results from the NMAs and economic analysis.

For CG171[47] *Urinary incontinence in women*, NMAs were conducted to assess the effectiveness of antimuscarinic drugs for treating the symptoms of overactive bladder and the results of these were used in the economic analysis also conducted by the GDG. The GDG concluded that the NMA results did not support making a recommendation to use or exclude preferentially any of the antimuscarinic drugs studied and instead based their recommendations on the economic analysis.

NG23[48] *Menopause* also used results from NMAs conducted by the GDG as inputs to economic analysis. NMAs were conducted to estimate the effects of selected treatments on vasomotor symptoms, discontinuation of treatment and vaginal bleeding. The NMA results and economic anaylsis were cited as evidence for the recommendations.

NG24[49] *Transfusion* includes recommendations on alternatives to blood transfusion in surgical patients. The GDG conducted NMAs for the outcomes following selected interventions in high and medium risk patients and the results of the NMA were used as inputs to economic modelling conducted by the GDG. Both the NMA and economic analysis results were cited as evidence for the recommendations.

5.5 Discussion

5.5.1 Summary of the findings of this chapter

The responsible GDG undertook either PWMA alone or both PWMA and NMA for 68 of 72 NICE clinical guidelines that were newly published or updated in 2015 or 2016.

Twelve of these guidelines made recommendations evidenced by results from NMA. In three cases, the evidence was from previously published systematic reviews using NMA methodology and, in one of those three guidelines, evidence was also used from NMA

undertaken by the GDG. In the nine other cases, the evidence was from NMA undertaken by the GDG.

The guidelines reported the GDG undertaking NMA for 10 guidelines and considering but not undertaking NMA for specified reasons in another seven guidelines.

NMA results were used mainly as evidence for identifying a preferred intervention based on ranking of effectiveness and/or adverse effects and those rankings were also used as inputs to economic analyses.

This review did not aim to establish how often recommendations were based on the results of pair-wise meta-analysis in previously published systematic reviews but identified that all but two of the guidelines referenced such reviews. The GDG conducted pair-wise meta-analysis for both those guidelines and for all but four guidelines overall.

Every guideline reported new meta-analysis or referenced previously published meta-analysis.

5.5.2 Strengths and limitations of the research presented in this chapter

This was the first review, as far as I have been able to identify, of the use of evidence from NMA in NICE clinical guidelines.

It is likely that some systematic reviews have not been identified by this review because of human error in the review process, particularly because only one reviewer undertook the work.

5.5.3 Implications of the findings of this chapter

NICE clinical guidelines have a wide impact on the commissioning of health care and on clinical practice. This review confirmed that use of results from meta-analysis is a standard process in production of these clinical guidelines. Whilst pair-wise meta-analysis remains by far the more commonly used method in systematic reviews and was used in all the guidelines studied, this review demonstrated the level of uptake of use of NMA methodology. NMA was

conducted or considered by the GDG for nearly one quarter of the guidelines reviewed and some recommendations were based on the results of NMA in every guideline for which such analysis was conducted.

Results from NMA in previously published systematic reviews, however, were infrequently used as evidence to support recommendations; with this happening in only three out of 72 guidelines.

NMA was used to rank interventions and for inputs to economic analysis but no pattern emerged as to what influenced whether the methodology was used by a GDG or not. Making recommendations between multiple potential interventions and conducting economic analysis were common to guidelines that did not reference NMA results or have NMA conducted by the GDG, as well as to those guidelines that did. It is possible, but unclear, whether prior knowledge and experience of using NMA methodology of members of the GDG (both clinical and research staff) may have influenced the decisions on whether to use this method in their guideline.

For all guidelines, pair-wise meta-analysis was referenced and/or conducted. References to systematic reviews using pair-wise meta-analysis were 85 times more common than references to those using NMA methodology. Despite the amount of previously published systematic reviews, however, the GDG conducted new meta-analysis for almost every guideline. Many guidelines also stated reasons for excluding previously published systematic reviews on grounds of quality, unclear methods or relevance of the clinical question being addressed. These observations raise uncertainty about the level of quality and effectiveness of existing systematic reviews, if GDGs are unable to identify suitable existing evidence despite the large volume of publication of such reviews. This may reflect the wider concerns about research evidence summarised by Heneghan et al[50].

The quality issues identified in some cases, by the GDG, as the reason for excluding previously published systematic reviews and the small number of previously published systematic reviews using NMA methodology that GDGs considered suitable to base recommendations on, reinforces the importance of critical appraisal. This led to the next step in my programme of research, namely be to identify a suitable critical appraisal tool for reviews that include NMA, which is reported in Chapters Six and Seven.

5.6 Chapter summary

- NICE clinical guidelines published or updated in 2015 and 2016 made extensive use of meta-analysis to identify evidence to support their recommendations.
- The GDG conducted new meta-analysis for almost every guideline.
- NMA methods were used far less often than pair-wise meta-analysis but were used or considered for nearly one quarter of the guidelines reviewed.
- Where recommendations in the guidelines were based on results of NMA, these were more often NMAs conducted by the GDG rather than previously published systematic reviews using NMA methodology.
- Only one third of the previously published systematic reviews using NMA methodology that were referenced by guidelines were linked to recommendations in the guidelines.
- In view of the apparent unsuitability of results from NMA, in many reviews, to be used as evidence to base recommendations on, the next step in my programme of research should be to identify a suitable critical appraisal tool for systematic reviews using NMA methodology.

5.7 Chapter references

1. National Institute for Health and Clinical Excellence. History of NICE, https://www.nice.org.uk/about/who-we-are/history-of-nice (2017, accessed 6 October 2017).

2. National Institute for Health and Clinical Excellence. History of NICE, https://www.nice.org.uk/about/who-we-are/history-of-nice (2019, accessed 10 November 2019).

3. National Institute for Health and Clinical Excellence. Types of guideline, https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/types-of-guideline (accessed 6 October 2017).

4. National Institute for Health and Clinical Excellence. Developing NICE guidelines: the manual. *Process and methods [PMG20]*: National Institute for Health and Care Excellence, April 2017 update https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview (accessed 6 October 2017).

5. Moher D, Tetzlaff J, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4: e78. 2007/03/29. DOI: 10.1371/journal.pmed.0040078.

6. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol* 2014; 67: 138-143. 2013/10/05. DOI: 10.1016/j.jclinepi.2013.07.014.

7. NICE Decision Support Unit (DSU). Evidence Synthesis TSD series, http://www.nicedsu.org.uk/Evidence-Synthesis-TSD-series(2391675).htm (accessed 5 May 2015).

8. National Institute for Health and Clinical Excellence. Guidance, https://www.nice.org.uk/guidance (accessed 8 March 2017).

9. National Institute for Health and Care Excellence. Haematological cancers: improving outcomes. National Guideline 47, https://www.nice.org.uk/guidance/ng47/evidence (2016, accessed 8 March 2017).

10. National Institute for Health and Care Excellence. Major trauma: service delivery. National Guideline 40, https://www.nice.org.uk/guidance/ng40/evidence (2016, accessed 8 March 2017).

11. National Institute for Health and Care Excellence. Children's attachment: attachment in children and young people who are adopted from care, in care or at high risk of going into care. National Guideline 26, https://www.nice.org.uk/guidance/ng26/evidence (2016, accessed 8 March 2017).

12. National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. National Guideline 28, https://www.nice.org.uk/guidance/ng28/evidence (2016, accessed 8 March 2017).

13. National Institute for Health and Care Excellence. Irritable bowel syndrome in adults: diagnosis and management. Clinical Guideline 61, https://www.nice.org.uk/guidance/cg61/evidence (2016, accessed 8 March 2017).

14. National Institute for Health and Clinical Excellence. Prophylaxis against infective endocarditis: antimicrobial prophylaxis against infective endocarditis in adults and children undergoing interventional procedures. Clinical Guideline 64, https://www.nice.org.uk/guidance/cg64/evidence (2016, accessed 10 May 2017).

15. National Institute for Health and Care Excellence. Stable angina: management. Clinical Guideline 126, https://www.nice.org.uk/guidance/cg126/evidence (2016, accessed 3 May 2017).

16. Stettler C, Wandel S, Allemann S, et al. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *Lancet* 2007; 370: 937-948. 2007/09/18. DOI: 10.1016/S0140-6736(07)61444-5.

17. National Institute for Health and Care Excellence. Palliative care for adults: strong opioids for pain relief. Clinical Guideline 140, https://www.nice.org.uk/guidance/cg140/evidence (2016, accessed 8 May 2017).

18. Bekkering GE, Soares-Weiser K, Reid K, et al. Can morphine still be considered to be the standard for treating chronic pain? A systematic review including pair-wise and network meta-analyses. *Curr Med Res Opin* 2011; 27: 1477-1491. 2011/06/04. DOI: 10.1185/03007995.2011.586332.

19. Vissers D, Stam W, Nolte T, et al. Efficacy of intranasal fentanyl spray versus other opioids for breakthrough pain in cancer. *Curr Med Res Opin* 2010; 26: 1037-1045. 2010/03/05. DOI: 10.1185/03007991003694340.

20. National Institute for Health and Care Excellence. Bronchiolitis in children: diagnosis and management. National Guideline 9, https://www.nice.org.uk/guidance/ng9/evidence (2015).

21. Hartling L, Fernandes RM, Bialy L, et al. Steroids and bronchodilators for acute bronchiolitis in the first two years of life: systematic review and meta-analysis. *BMJ* 2011; 342: d1714. 2011/04/08. DOI: 10.1136/bmj.d1714.

22. Fernandes RM, Bialy LM, Vandermeer B, et al. Glucocorticoids for acute viral bronchiolitis in infants and young children. *Cochrane Database Syst Rev* 2010: CD004878. 2010/10/12. DOI: 10.1002/14651858.CD004878.pub3.

23. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. Clinical Guideline 181, https://www.nice.org.uk/guidance/cg181/evidence (2016, accessed 12 March 2017).

24. Ara R, Pandor A, Stevens J, et al. Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation. *Health Technol Assess* 2009; 13(34) 2009/07/17. DOI: 10.3310/hta13340.

25. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management. National Guideline 59, https://www.nice.org.uk/guidance/ng59/evidence (2016, accessed 8 March 2017).

26. Lewis R, Williams N, Matar HE, et al. The clinical effectiveness and cost-effectiveness of management strategies for sciatica: systematic review and economic model. *Health Technol Assess* 2011; 15(39) 2011/11/15. DOI: 10.3310/hta15390.

27. National Institute for Health and Care Excellence. Depression in adults: recognition and management. Clinical Guideline 90, https://www.nice.org.uk/guidance/cg90/evidence (2016, accessed 14 May 2017).

28. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746-758. DOI: 10.1016/s0140-6736(09)60046-5.

29. National Institute for Health and Care Excellence. Melanoma: assessment and management. National Guideline 14, https://www.nice.org.uk/guidance/ng14/evidence (2015).

30. Dequen P, Lorigan P, Jansen JP, et al. Systematic review and network meta-analysis of overall survival comparing 3 mg/kg ipilimumab with alternative therapies in the management of pretreated patients with unresectable stage III or IV melanoma. *Oncologist* 2012; 17: 1376-1385. 2012/10/02. DOI: 10.1634/theoncologist.2011-0427.

31. National Institute for Health and Care Excellence. Myeloma: diagnosis and management. National Guideline 35, https://www.nice.org.uk/guidance/ng35/evidence (2016, accessed 21 May 2017).

32. Mhaskar R, Redzepovic J, Wheatley K, et al. Bisphosphonates in multiple myeloma: a network meta-analysis. *Cochrane Database Syst Rev* 2012; 5: CD003188. 2012/05/18. DOI: 10.1002/14651858.CD003188.pub3.

33. National Institute for Health and Care Excellence. Epilepsies: diagnosis and management. Clinical Guideline 137, https://www.nice.org.uk/guidance/cg137/evidence (2016, accessed 17 May 2017).

34. Tudur Smith C, Marson AG, Chadwick DW, et al. Multiple treatment comparisons in epilepsy monotherapy trials. *Trials* 2007; 8: 34. 2007/11/07. DOI: 10.1186/1745-6215-8-34.

35. National Institute for Health and Care Excellence. Cancer of the upper aerodigestive tract: assessment and management in people aged 16 and over. National Guideline 36, https://www.nice.org.uk/guidance/ng36/evidence (2016, accessed 21 May 2017).

36. Chen YP, Wang ZX, Chen L, et al. A Bayesian network meta-analysis comparing concurrent chemoradiotherapy followed by adjuvant chemotherapy, concurrent chemoradiotherapy alone and radiotherapy alone in patients with locoregionally advanced nasopharyngeal carcinoma. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2015; 26: 205-211. 2014/10/31. DOI: 10.1093/annonc/mdu507.

37. Yan M, Kumachev A, Siu LL, et al. Chemoradiotherapy regimens for locoregionally advanced nasopharyngeal carcinoma: A Bayesian network meta-analysis. *Eur J Cancer* 2015; 51: 1570-1579. 2015/06/06. DOI: 10.1016/j.ejca.2015.04.027.

38. National Institute for Health and Care Excellence. Preterm labour and birth. National Guideline 25, https://www.nice.org.uk/guidance/ng25/evidence (2015).

39. Haas DM, Caldwell DM, Kirkpatrick P, et al. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. *BMJ* 2012; 345: e6226. 2012/10/11. DOI: 10.1136/bmj.e6226.

40. National Institute for Health and Care Excellence. Bipolar disorder: assessment and management. Clinical Guideline 185, https://www.nice.org.uk/guidance/cg185/evidence (2016, accessed 17 May 2017).

41. Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* 2011; 378: 1306-1315. 2011/08/20. DOI: 10.1016/S0140-6736(11)60873-8.

42. National Institute for Health and Clinical Excellence. Crohn's disease: management. Clinical Guideline 152, https://www.nice.org.uk/guidance/cg152/evidence (2016, accessed 14 May 2017).

43. National Institute for Health and Care Excellence. Type 1 diabetes in adults: diagnosis and management. National guideline 17, https://www.nice.org.uk/guidance/ng17/evidence (2016).

44. National Institute for Health and Care Excellence. Tuberculosis. National Guideline 33, https://www.nice.org.uk/guidance/ng33/evidence (2016, accessed 10 May 2017).

45. National Institute for Health and Care Excellence. Venous thromboembolism: reducing the risk for patients in hospital. Clinical Guideline 92, https://www.nice.org.uk/guidance/cg92/evidence (2015, accessed 11 June 2017).

46. National Institute for Health and Clinical Excellence. Headaches in over 12s: diagnosis and management. Clinical Guideline 150, https://www.nice.org.uk/guidance/cg150/evidence (2015, accessed 23 May 2017).

47. National Institute for Health and Clinical Excellence. Urinary incontinence in women: management. Clinical Guideline 171, https://www.nice.org.uk/guidance/cg171/evidence (2015, accessed 26 May 2017).

48. National Institute for Health and Clinical Excellence. Menopause: diagnosis and management. National Guideline 23, https://www.nice.org.uk/guidance/ng23/evidence (2015, accessed 26 May 2017).

49. National Institute for Health and Care Excellence. Blood transfusion. National Guideline 24, https://www.nice.org.uk/guidance/ng24/evidence (2015, accessed 23 May 2017).

50. Heneghan C, Mahtani KR, Goldacre B, et al. EBM Manifesto, http://evidencelive.org/manifesto/ (2016, accessed 10 December 2016).

CHAPTER SIX

Review of critical appraisal tools

6 Review of critical appraisal tools

6.1 Background

Critical appraisal is a systematic approach to assessing the quality of a study. This assessment has been described as,

'an estimate of the likelihood that the results are a valid estimate of the truth' [1].

Critical appraisal involves assessing the report of a study for methodological quality and any likelihood of bias and considering whether these might affect the validity of the reported results. It is well established good practice that anyone using published health intervention research to aid decision making on clinical policy, individual patient care or further related research should critically appraise research reports for both the relevance of the research to their own research question and the validity of the reported results.

Two internationally recognised centres of excellence for accessing critical appraisal tools and training are the Critical Appraisal Skills Programme (CASP)[2] in Oxford, UK and the Joanna Briggs Institute (JBI)[3] in Adelaide, Australia.

6.2 Chapter aims and objectives

The aim of this chapter is to try to identify a suitable critical appraisal tool to assess the quality of systematic reviews using network meta-analysis (NMA) methodology for synthesis. The primary objectives were to identify existing critical appraisal tools, as of March 2018, that addressed some or all aspects of quality of systematic reviews using NMA methodology for synthesis and determine whether any existing one was already suitable for end-users without specialist knowledge of NMA methodology. A secondary objective was to identify methods for validating a critical appraisal tool in the event that my conclusion was that a new tool should be developed.

6.3 Methods

6.3.1 Search

The CASP and Joanna Briggs Institute websites were accessed for critical appraisal tools and links to related publications. MEDLINE, MEDLINE In-Process and Embase were searched using free-text terms as follows: *critic* apprais** AND [*method** OR *tool** OR *checklist**] AND develop*. The results were limited to English language articles published between 1990 and the last update of these searches (January 31, 2018). The reference lists of included articles were also checked for potentially eligible articles.

6.3.2 Inclusion criteria

Systematic reviews of critical appraisal tools and articles reporting the development and/or validation and/or comparative testing of a critical appraisal tool for systematic reviews or indirect comparisons of randomised trials of healthcare interventions in humans.

6.3.3 Exclusion criteria

Articles reporting development of and/or validation of critical appraisal tools for anything other than systematic reviews or indirect comparisons of randomised trials of healthcare interventions in humans or restricted to use for a specific clinical topic.

6.3.4 Selection of studies

A single reviewer screened abstracts of the records identified by the search against the inclusion criteria, with the default to retrieve the full article for further consideration if there was uncertainty based on the abstract. Retrieved full articles were assessed by the reviewer to make a final decision on eligibility.

6.3.5 Data extraction

After eliminating articles that could be excluded on the basis of their abstract, search results were exported to Endnote X5 software (Thomson Reuters [Scientific] Inc., New York, NY).

After identifying and removing duplicate results in Endnote, full text was obtained for the remaining unique results. After excluding articles based on reviewing their full text and obtaining full text of any additional relevant articles identified from references, the following data were extracted from the included articles to a Microsoft Word document: authors, year of publication, and the following, according to type of article:

For systematic reviews:

Number of critical appraisal tools identified, number of tools for systematic reviews and/or for indirect comparison (IC)/NMA, summary of characteristics identified, and summary of validation methods.

For articles reporting critical appraisal tool development:

Intended use of the critical appraisal tool/target users (if stated), summary of development, number of items included, items that specifically address synthesis using network meta-analysis, categorisation of items (if stated), response options presented, how guidance for completion is provided, guidance that specifically addresses synthesis using NMA and summary of validation methods.

For articles reporting validation of a critical appraisal tool:

The critical appraisal tool, summary of validation methods, and summary of authors' conclusions.

This extraction work was done by a single reviewer.

6.4 Results

6.4.1 Articles retrieved

A total of 1536 records were retrieved by the searches and this was reduced to 1521 after removing duplicates. After screening abstracts, 59 full-text articles were retrieved. From these,

18 articles were identified that met the inclusion criteria. See Appendix 4 for references to the included studies and Table A3 for the study characteristics.

Three articles were systematic reviews of critical appraisal tools[4-6]. Eleven articles described the development of a critical appraisal tool: eight to assess the quality of systematic reviews[7-13], and three specifically for IC or NMA[14-16]. Four articles reported the validation of one or more of those critical appraisal tools[17-20]. Figure 6.1 summarises the retrieval, exclusion and final inclusion of articles.

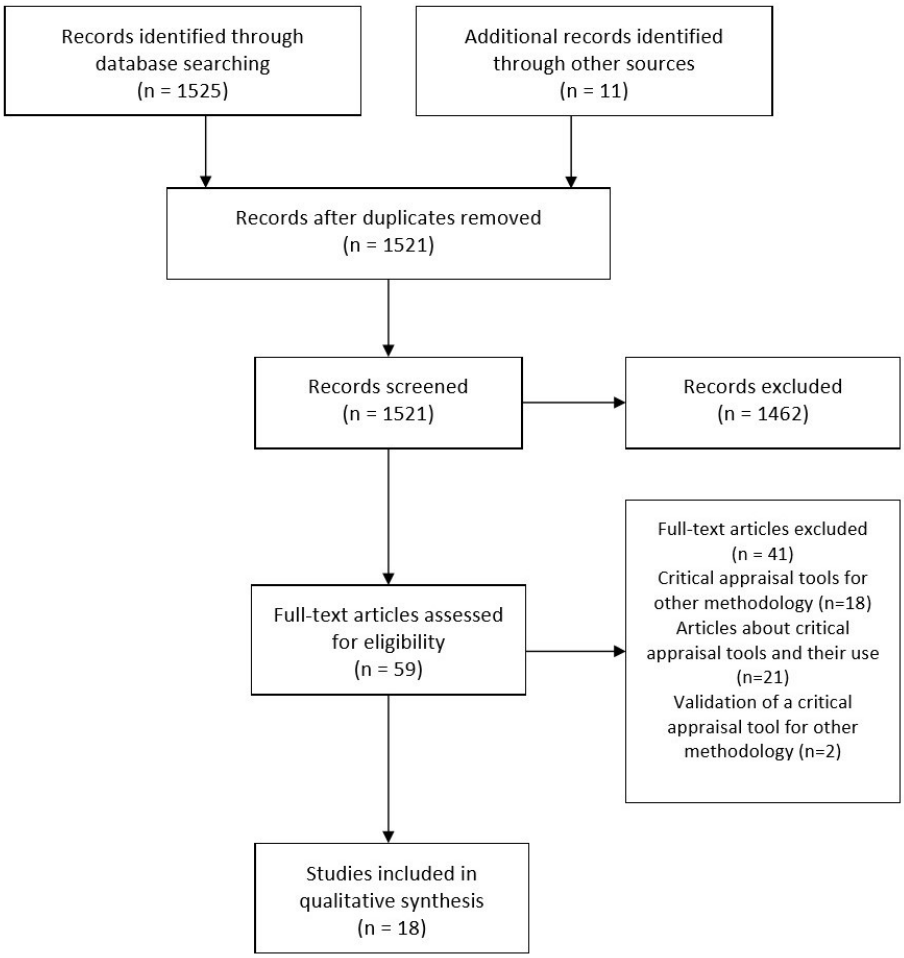

Figure 6.1 Article Flow Diagram

6.4.2 Systematic reviews identified

Table 6.1 Summary of the systematic reviews identified

First Author	Year of publication	Number of critical appraisal tools identified	Number of critical appraisal tools for systematic reviews	Number of critical appraisal tools for IC or NMA
Katrak[4]	2004	121	26	0
Crowe[5]	2011	44	4	0
Zeng[6]	2015	21	6	0

The 26 appraisal tools for systematic reviews identified by Katrak et al (2004)[4] included multiple general guides to critical appraisal, rather than specific tools, some reporting guidelines and some that are no longer accessible. Only 14, of the total 121 tools identified, specified the empirical approach used to construct the tool. A further 31 tools were reported to be based on other source tools, but authors of the included articles had provided no details on how or why they had included/excluded items; 21 of those source tools also did not report the empirical approach used to construct the tool.

Katrak et al commented that although,

'Forty-nine percent (N = 58) of critical appraisal tools summarised the results of the quality appraisal into a single numeric summary score', 'there was no justification provided for any of the scoring systems used.'

Amongst the 26 critical appraisal tools for systematic reviews, Katrak et al reported that 12 were accompanied by, *'guidelines that informed the user of the interpretation of each item contained within them'*.

For three tools, this was in the form of accompanying explanation within the tool and for the other nine tools, it was in the form of a separate handbook or published paper. Katrak et al stated that two tools for systematic reviews had established 'face validity'[13, 21]. One of these was the QUOROM reporting guideline for meta-analyses of randomised trials[21], rather than a critical appraisal tool, and the other was the 'Overview Quality Assessment Questionnaire (OQAQ)', published by Oxman and Guyatt in 1991[13]. The latter was also one of only two

tools for systematic reviews that Katrak et al's review found had undergone inter-rater reliability testing; the other being a tool for meta-analyses evaluating diagnostic tests[22].

Katrak et al reported that,

'*Few critical appraisal tools had documented evidence of their validity and reliability'*

and that they

'*found no gold standard critical appraisal tool for any type of study design*'.

They concluded that critical appraisal tools,

'*should have published evidence of the empirical basis for their construction, validity of items and reliability of interpretation, as well as guidelines for use, so that the tools can be applied and interpreted in a standardized manner.*'

Crowe and Shepherd (2011)[5] undertook a literature review of critical appraisal tools, with a view to developing a comprehensive tool, based on searches conducted in 2008. They identified 44 critical appraisal tools published from 1980 onwards, of which four were to assess systematic reviews. These were the OQAQ[13], AMSTAR (A MeaSurement Tool to Assess systematic Reviews)[9] published in 2007, which was partially based on OQAQ, another checklist, also based on OQAQ, contained in a 1992 review article by Wilson and Henry[23] and a 1997 article by Hunt and McKibbon[24] on locating and appraising systematic reviews, which illustrates the use of OQAQ rather than reporting a new critical appraisal tool.

Crowe and Shepherd reported that most (75%) of all 44 tools they identified had undergone some form of validation, mainly content validation, but most (77%) had not undergone reliability testing, but they reported that both OQAQ and AMSTAR had not been reliability tested, whereas in both cases, reliability testing had been published[13, 18].

Zeng et al (2015)[6] reported a systematic review of critical appraisal tools,

'*for pre-clinical and clinical studies, systematic review and meta-analysis and clinical practice guideline'*,

based on searches conducted in 2014. They identified 21 tools for analysis, with the highest number being from the JBI and from CASP. Six of the tools identified were for appraisal of

systematic reviews. Three of these were a 1987 report of criteria developed by Sacks et al[7], the OQAQ and AMSTAR. The other three were checklists published by CASP[2], JBI[3] and NICE[25].

6.4.3 Critical appraisal tools identified

My searches identified eleven articles describing tools to appraise systematic reviews of healthcare interventions and/or IC or NMA. No article describing the development of any CASP or JBI critical appraisal tool was identified from searches but I decided to include the systematic review tools of each in this review due to their widely recognised status. A published article relating to the JBI tool was subsequently identified after the decision to include the tool[26].

Table 6.2 summarises the characteristics of the tools. Three of these (the 1987 checklist reported by Sacks et al, the OQAQ and AMSTAR), were also identified in one or more of the above-mentioned systematic reviews by Katrak et al, Crowe and Shepherd and Zeng et al.

Table 6.2 Characteristics of the identified critical appraisal tools

First author (Tool name)	Published	Stated scope of the tool	Target users (if stated)	Number of items in tool	Includes items that specifically address IC or NMA	Methods of construction of the tool reported	Validation of the tool reported	Inter-rater reliability reported
Sacks[7]	1987	To assess the quality of meta-analyses	Those who perform meta-analyses and those who use them	23	No	No	No	Yes (but minimal details)
Oxman[8]	1988	To assess the scientific quality of a review	Readers interested in answering a clinical question relevant to their everyday practice	8	No	No	No	No
Oxman[13] (OQAQ)	1991	For evaluating the scientific quality of research overviews	Peer reviewers, clinicians and other decision makers	10	No	Yes	Face and content validity	Yes
Shea[9] (AMSTAR)	2007	For assessing the methodological quality of systematic reviews by building upon empirical data collected with previously developed tools and utilizing expert opinion	Users of systematic reviews	11	No	Yes	Face and content validity	Yes (in separate article[18])

First author (Tool name)	Published	Stated scope of the tool	Target users (if stated)	Number of items in tool	Includes items that specifically address IC or NMA	Methods of construction of the tool reported	Validation of the tool reported	Inter-rater reliability reported
Kung[27] (R-AMSTAR)	2010	To revise an existing tool (AMSTAR) to quantify the quality of systematic reviews	Not stated	11	No	Yes	Face and content validity	Yes
Jansen[14]	2011	To assist decision makers in evaluating a reported network meta-analysis	Policymakers and healthcare professionals	22	Multiple	No	No	No
Ades[15]	2012	For the review of evidence syntheses in the context of decision making to assess whether or not the synthesis meets the requirements set out in the NICE Methods Guide	Reviewers of evidence syntheses and those submitting syntheses for review	41	Multiple	No	No	No
Higgins[10]	2013	To assess the quality of a meta-analysis with more detailed evaluation of the statistical methods than in existing tools	Developers of quality-assessment tools for meta-analyses	43	One item relating to IC	Yes	Face and content validity	Yes
Ortega[16]	2014	For critical appraisal of indirect comparisons of drugs, mainly to be applied in drug evaluation and decision making	Decision makers in drug evaluation and clinicians	20	Multiple	Yes	Face and content validity	Yes
Whiting[11] (ROBIS)	2016	To assess the risk of bias in systematic reviews	Guideline developers, authors of overviews of systematic reviews, and review authors	25	No	Yes	Face and content validity	Yes
Shea[12] (AMSTAR 2)	2017	To increase the value of an existing tool (AMSTAR) as a broad critical appraisal tool for systematic reviews, including non-randomised studies	Health professionals and policy makers who do not necessarily have advanced training in epidemiology	16	No	Yes	Face and content validity	Yes
CASP tool[28]				10	No	No	No	No
JBI tool[29]				11	No	No	No	No

Key components

The content of the identified tools can be summarised into key components. Sacks' tool was the earliest published tool identified and he categorised the criteria in his checklist into six major areas: study design, combinability, control of bias, statistical analysis, sensitivity analysis and applicability of results[7]. Most of these six areas, or key components, were also represented by criteria in the other tools identified. Sacks included consideration of the potential for conflict of interest within the area 'control of bias', whilst several other authors highlighted it separately from other sources of bias and some did not make any mention of it. Therefore, for the purpose of summarising key components of the tools, I treated this as a seventh key component. Most, but not all, tools included some criteria that are reporting standards (i.e. whether an aspect of the review methods or results is reported, rather than an evaluation of what is reported), so I treated this as an eighth component. Table 6.1 summarises the tools identified by the presence (shaded cells)/absence (clear cells) of each of these components.

Table 6.3 Presence/absence of key components in the identified critical appraisal tools

First author (Tool name)	Key components							
	Study design	Combinability	Control of bias	Statistical analysis	Sensitivity analysis	Applicability of results	Potential for conflict of interest	Reporting standard
Sacks[7]	■	■	■	■	■	■	■	
Oxman[8]	■	■	■	■		■		■
Oxman[13] (OQAQ)	■	■	■	■		■		■
Shea[9] (AMSTAR)	■	■	■	■	■	■	■	■
Kung[27] (R-AMSTAR)	■	■	■	■	■	■	■	■
Jansen[14]	■	■	■	■	■	■		■
Ades[15]	■	■	■	■	■			■
Higgins[10]	■	■	■	■	■	■	■	■
Ortega[16]	■	■	■	■	■	■	■	
Whiting[11] (ROBIS)	■	■	■	■	■	■	■	■
Shea[12] (AMSTAR 2)	■	■	■	■	■	■	■	■
CASP SR tool[28]	■	■	■			■		■
JBI SR tool[29]	■	■	■	■	■	■		■

Guidance and response options

Tools varied in the way authors' provided guidance for completion and in the response options to select from. Table 6.4 summarises these characteristics.

Table 6.4 Guidance for completion and response options in the identified critical appraisal tools

First author	Guidance for completion	Response options
Sacks[7]	Within the published article	Not specified. Yes/no implied.
Oxman[8]	Within the published article	Not specified. Yes/no implied.
Oxman[13] (OQAQ)	Within the published article	7-point scale - ranging from 1, 'extensive flaws (very poor)', to 7, 'minimal flaws (exemplary)'
Shea[9] (AMSTAR)	Included alongside each item in the tool	Yes/No/Can't answer/Not applicable
Kung[27] (R-AMSTAR)	Included alongside each item in the tool, with specific criteria for scoring each item	Score from 1 to 4
Jansen[14]	Within the published article	Not specified. Yes/no implied.
Ades[15]	Within the published article (12 pages, excluding the content relating to cost effective analysis)	Not specified. Yes/no implied.
Higgins[10]	Included alongside each item in the tool	Yes/Probably yes/Unsure/Probably no/No
Ortega	Included alongside each item in the tool. For some items specific criteria are provided for whichever of the response options are presented.	Initial three questions, for which the response options are Yes/No/Don't know, followed by a question whether or not to proceed any further with the appraisal. Subsequent items in the tool have response options, relating to reliability, in varying combinations of High/Acceptable/Uncertain/Low.
Whiting[11] (ROBIS)	In a separate 39-page document	For each item: Yes/Probably yes/Probably no/No/No information. After considering the individual items, each domain is then rated as having Low/High/Unclear concern about bias
Shea[12] (AMSTAR 2)	Included alongside each item in the tool	Yes/Partial yes/No Specific criteria for 'yes' or 'partial yes' responses are provided with each item.
CASP SR tool[28]	Included alongside each item in the tool	Initial two questions, for which the response options are Yes/Can't tell/No, followed by a question whether or not to proceed any further with the appraisal. Subsequent items in the tool have either the same response options, or, in relation to results, simply a comments box.
JBI SR tool[29]	Included as 4 pages of guidance in a single document with the tool	Yes/No/Unclear/Not applicable

Content that specifically relates to appraisal of the use of network meta-analysis

Three articles reported checklists specifically for the critical appraisal of IC, including NMA: the 2011 ISPOR Task Force report[14], a 2012 NICE DSU technical support document[15] and Ortega et al (2014)[16]. A fourth tool (Higgins et al[10]) included one item about whether IC was performed appropriately, with brief guidance provided about how IC should and should not be performed. NMA is mentioned as being an extension of IC but with no specific guidance. Table 6.5 summarises the content of the three tools that specifically address the use of NMA.

Table 6.5 Items specific to network meta-analysis in the identified critical appraisal tools

First author	Key Component	Details/explanation in the tool
Jansen[14]	Details of the network of studies	Do the results include a figure showing the network of studies?
	Statistical analysis	Description and justification of statistical model(s) used Whether frequentist/Bayesian approach used Prior distributions for model parameters in Bayesian framework How bias/inconsistency was evaluated Does the study describe an assessment of fit? Are competing models being compared?
	Assumptions	Table/list of studies with information regarding study design and patient characteristics to judge potential similarity/consistency issues Does the discussion address aspects that might violate similarity and consistency assumptions?
Ades[15]	Details of the network of studies	If additional treatments have been added to the set of comparators, is this adequately justified? One reason for adding treatments to the synthesis set might be in order to make a connected network. A network diagram is a useful way of showing the structure of the evidence. The actual data used in the base-case analysis should be set out in a table.
	Connectedness	This should be clear from a network diagram. Approaches used to re-connect disconnected networks require a number of strong assumptions and these must be explained and, if appropriate, justified.
	Statistical analysis	Reviewers should be provided with a precise description of the meta-analytic method used. If a Bayesian analysis is used, details on priors, convergence and number of iterations should also be given. The name of the software module and package used for statistical analysis should be given, and any additional computer code should be provided. It is essential that enough information is provided so that the exact same analysis can be replicated. Evidence in the form of goodness of fit of the chosen model and of alternative models should be presented. When the empirical treatment differences are used as data (e.g. Log Odds Ratios, Log Hazard ratios, etc), the differences in multi-arm trials are correlated and this must be taken into account.
	Assumptions	Are there differences between the trials included in synthesis e.g. population, interventions, outcome measures? If so, how have these been handled in statistical analysis? The network structure should be set out in a diagram and the number of possible inconsistencies set out. If the AB trials tend to have been carried out on systematically different patient populations to the AC trials or the BC trials, there is a high risk that indirect or mixed treatment comparisons will be unreliable. Have adequate checks for inconsistency been made? If inconsistency was detected, what adjustments were made to the analysis, and how was this justified?
Ortega[16]	Connectedness	Are comparators and studies represented graphically? How many intermediate comparators are there between the two alternatives we are most interested in comparing?
	Statistical analysis	Is conducting an IC justified? Is the IC method chosen appropriate? Are the assumptions acceptable? Was any sensitivity analysis done correctly?
	Assumptions	Are details of the studies included in the network similar e.g. patients, interventions, outcome measures, missing patient/data handling? Only consider inconsistency if there is a published direct comparison of the treatments being compared. Where there is discrepancy between direct and indirect results, the reasons must be analysed. Any further statistical analysis conducted to correct for inconsistency must make clinical sense and data to perform it must be available.

Chronology, development and validation of the identified tools

The 1987 Sacks et al checklist[7] was the earliest published tool identified for critical appraisal of systematic reviews. It consisted of items chosen by the authors, presumably based on their expert knowledge of the field. There is no report of how the items were decided on. The only mention of validation is that,

'We found after an initial learning period, two evaluators agreed on more than 90 percent of items scored; the majority of disagreements were due to oversights.'

In 1988, Oxman and Guyatt published their checklist in a special article on reading literature reviews[8]. The checklist consisted of items again chosen by the authors, presumably based on their expert knowledge of the field and there is no report of how the items were decided on or mention of validation.

In 1991, the same authors published the OQAQ[13] tool,

'to measure the extent to which a review is likely to be valid; i.e. the extent to which an overview guards against bias (by which we mean systematic deviation from the truth)'.

Similar to Sacks et al's 'six major areas', Oxman and Guyatt identified 'tasks (dimensions)' that the included items must address: problem formulation, study identification, study selection, validation of studies, data extraction, data synthesis and inference. They developed a list of 25 items *'based on a review of the literature'* and refined by expert consensus, including input from editors and methodological experts. After pilot testing, this was reduced to 10 items by,

'eliminating items that did not discriminate between overviews of high and low scientific quality'.

The article also reports inter-rater reliability (IRR) testing, details of which are described below (section 6.4.4).

In 2007, Shea et al reported the development of the initial version of AMSTAR[9] (see below for AMSTAR-2). They sought to build on and update earlier tools, in particular to include learning from newer methodological research that had shown, for example, the potential for publication and language bias in systematic reviews. They started by compiling a list of the 23 items in the 1987 Sacks et al tool and the 10 from the OQAQ, and adding three items based on newer methodological research findings: language restriction on included studies, publication bias and publication status. Language was included twice, as an addition to the Sacks et al list and as a separate new item. This produced 37 items in the initial tool that was used by two appraisers to assess a set of systematic reviews. The results were subjected to principal components analysis and items below a specified threshold were removed. This

reduced the number of items to 11. The final form of these 11 items was determined by a structured process involving an expert group, including clinicians, methodologists, epidemiologists and *'reviewers who were new to the field'*. The expertise of the latter group is not stated. Validation is reported in separate articles[18, 19]. AMSTAR has become one of the most widely-used tools for appraisal of systematic reviews[12].

Kung et al (2010)[27] developed the Revised AMSTAR tool (R-AMSTAR). This retained the same item content as the original AMSTAR tool, but introduced scores for each item. These item scores were then added to produce an overall score and overall score boundaries were defined to produce an overall rank for the quality of the systematic review, based on that score. IRR testing was assessed from appraisal of 11 systematic reviews by one pair of raters and five systematic reviews by a second pair of raters.

In 2011, the ISPOR Task Force published two reports, one on conducting[30] and the other on interpreting indirect comparisons and network meta-analyses[14]. Within the latter, Jansen et al included a checklist for critical appraisal of such studies, which was produced by an expert group, but no detail was reported of the process used to construct it and no validation of the checklist was reported. The authors acknowledged that their list,

'focuses on reporting quality and does not capture explicit items to judge or score the internal and external validity of a network meta-analysis.'

The 2012 NICE DSU Technical Support Document 7, by Ades et al[15], presents a checklist that,

'is intended to be used for pair-wise meta-analysis, indirect comparisons, and network meta-analysis, without distinction'.

The checklist is presented without details of the process undertaken to construct it and no validation is reported. Ades et al define four categories that the items will address: definition of the decision problem, methods of analysis and presentation of results, issues specific to network synthesis, and embedding the synthesis in a probabilistic cost-effectiveness model. Several items address specific aspects of the synthesis methodology and there is detailed

guidance on multiple aspects of network meta-analysis use. Ades et al state that network meta-analysis does not require,

'extra assumptions of "trial similarity" and "consistency", additional to assumptions that are required in pair-wise meta-analysis',

rather that these assumptions are common to and *'deserve scrutiny'* for both methods. They conclude,

'most items in the proposed checklist apply to both pairwise and network meta-analyses. The only issues that come exclusively under the heading of network synthesis are connectedness of networks, inconsistency, and software implementation.'.

Higgins et al (2013)[10] reported development of a tool, based partially on AMSTAR, designed to make a more '*detailed evaluation of the statistical methods'*. An expert group of medical statisticians contributed to and refined, by consensus, the items to be included in the final version of the tool, which consisted of 43 items grouped into four categories: data sources, analysis of individual studies, the meta-analysis, and reporting and interpretation. One item addresses the use of indirect comparison and NMA is mentioned, in the guidance for this item, as an extension of IC. IRR testing was assessed from appraisals of 26 systematic reviews conducted using the tool by two members of the expert development group; this was only calculated on the four 'summary' questions for each domain, not on all 43 items.

In 2014, Ortega et al[16] published,

'a user-friendly checklist for critical appraisal of indirect comparisons of drugs, considering clinical, methodological/statistical and quality aspects'.

An initial list of potential items was developed by a core group *'using CASP methods'* and informed by a literature review. It is unclear what the reported use of CASP methods means because the citation for this is simply to the Spanish CASP site where I could not find any methods for identifying items for a critical appraisal tool. The list was divided into three categories: clinical, methodology/statistics and quality. Ortega et al state that the items were mostly related to indirect comparisons,

'but including topics related to NMA and multiple treatment comparisons.'

The checklist was refined by an expert core group of hospital pharmacists with input from two external experts and then piloted to calculate inter-rater agreement, using 23 raters to appraise one review. Final revisions were made to the checklist, by consensus of the core group and two external experts, after considering the results and feedback comments from the appraisers. IRR testing of a draft version was reported, but the final version of the tool was not tested. The first section of the tool uses a similar format to CASP tools[2], by concluding with a decision whether or not to proceed any further with the appraisal, dependant on the answers to the initial three questions. Ortega et al acknowledge that their tool includes few items relating specifically to NMA but that,

'some aspects of NMA are common both to meta-analysis of direct comparisons and to other kinds of ICs'.

Whiting et al (2016)[11] developed the ROBIS tool with particular focus on assessing how the risk of bias has been minimised in a review. A steering group of *'experts in the area of systematic reviews'* identified 46 items from the 80 Cochrane MECIR[31] conduct items as relating to bias. Items relating to bias in 40 existing tools for systematic reviews or meta-analyses were also identified. This larger list of items was refined down to the items for inclusion, by means of a Delphi process using input from a wider group of methodologic, systematic review and guideline development experts. Final amendments were made after piloting the tool and considering feedback. The tool has three phases: an optional first to assess relevance, a second to identify concerns with the review process, which is split into four domains, and a third to judge overall risk of bias in the review. The guidance mentions NMA in the introduction to one domain and in relation to one item within that domain (appropriateness of the synthesis) but does not highlight any specific aspects of it for the user to consider and instead advises that statistical advice might be required if NMA has been used. IRR testing is reported involving one pair of appraisers assessing eight systematic reviews. Further IRR testing, with comparison to AMSTAR, was reported separately[20].

In 2017, Shea et al[12] reported an updated version of AMSTAR: AMSTAR 2. They state that this update was partly in response to feedback on the original tool and partly to address the inclusion of non-randomised studies in systematic reviews. They comment that,

'The revisions were not intended to deal with the special requirements of diagnostic test reviews, individual patient data meta-analyses or network meta-analyses, scoping reviews, or realist reviews.'

An expert group, with expertise in conduct of non-randomised studies, appraisal tool development, biostatistics and study design, considered various sources of feedback and updated literature reviews on critical appraisal tools before using a similar structured process to that used in the development of the original AMSTAR tool to produce a revised version. This version was further revised following piloting. Seven of the final 16 items in AMSTAR 2 are designated as 'critical' and carry more weighting in the tool's guidance on rating overall confidence in the results of the review. The options of 'not applicable' and 'can't answer', which were present in AMSTAR, were removed on the basis that,

'all domains are relevant to contemporary systematic reviews of healthcare interventions. If no information is provided to rate an item, the review authors should not be given the benefit of doubt and the item should be rated as a "No."'

but the 'partial yes' option was introduced *'to identify partial adherence to the standard.'* Another change from AMSTAR, is that specific criteria for "yes" or "partial yes" responses are introduced with each item. Details of the IRR testing are included in the next section.

Shea et al stress that AMSTAR 2,

'should not be used to derive an overall score' because *'an overall score may disguise critical weaknesses that should diminish confidence in the results of a systematic review'*

whilst acknowledging that the original AMSTAR was often used in this way and the website *'facilitated'* doing so.

Additional tools included

No article describing the development of any CASP or JBI critical appraisal tool was identified from searches but the tools for systematic reviews in each were included in this review due to their widely recognised status. An article relating to the JBI tool was subsequently identified[26]. The CASP tool for systematic reviews[28] contains items split into three sections: validity of the review, results of the review and local applicability. The first section contains two 'screening' questions, following which the user is asked to decide whether or not to proceed any further with the appraisal.

The JBI tool for systematic reviews states that it should be cited as a published article[26], which makes recommendations for conduct of umbrella reviews which summarise multiple systematic reviews. Contained within that article is a critical appraisal tool for systematic reviews that is described as having been developed and piloted by the umbrella review methodology group but no details of the development process or any validation is reported. The tool in the published article contains 10 items but the tool published on the JBI website[29] contains 11 items; the additional item not included in the published article asks whether there were methods to minimise errors in data extraction. The published article states that the response options are "met", "not met", "unclear" or "not applicable" but the tool published on the JBI website shows the response options as "yes", "no", "unclear" or "not applicable". The characteristics summarised in Tables 6.2, 6.3 and 6.4 are for the version of the tool published on the JBI website.

6.4.4 Articles reporting validation of a critical appraisal tool

I identified four articles that reported validation, mainly in the form of IRR testing of one or more of OQAQ, AMSTAR, or ROBIS[17-20]. Additionally, seven of the articles mentioned above also reported IRR testing[7, 10-13, 16, 27]. Of those, the details of IRR testing of OQAQ, reported by Oxman et al in 1991[13] are included in this section because they were not included

in the article reporting the tool's validation but only in the article reporting its development. The details of IRR testing of AMSTAR 2, reported by Shea et al in 2017[12], are also included in this section for ease of comparison with those for the original AMSTAR.

OQAQ

Oxman et al reported validation of OQAQ and IRR testing in two articles[13, 17]. They assessed validity in three ways. They assessed face validity by having a group of clinical epidemiologists rate the tool, they compared what was done to what was reported by surveying authors of systematic reviews to assess whether that affected the validity of criteria in the tool and they assessed construct validity by testing a set of hypotheses,

'concerning how the instrument should behave if it were really measuring the scientific quality of the overviews.'

Oxman et al found that the mean ratings by 15 clinical epidemiologists for 11 of 13 items included in a survey of satisfaction with the tool were between 4.9 and 6.2, on a scale of 1 (low) to 7 (high). Two aspects had lower ratings: whether information would be available (mean score 3.8) and the need for subjective decisions (mean score 4.0). These were thought to be linked by the same concern about how likely it was that the information required for assessing the appraisal criteria would be included in published reviews. They also found that the appraisers' judgement of the methods used in reviews, using the tool's criteria to assess publications, generally matched what responding authors said they did. The results of testing six of their seven hypotheses confirmed construct validity. The exception was the negative result for a hypothesis that the overall score for a review would correlate positively with the amount of total research training of the authors.

IRR testing of the tool involved nine appraisers, of which six were authors of the tool and three were research assistants in their university department, who appraised 36 published reviews. Oxman et al defined that the intraclass correlation coefficients calculated from the scores selected by the appraisers for each item should be above 0.5 to demonstrate reliability.

Although they stated that the results showed only minor differences in the level of agreement for seven of the ten items in the tool, the mean coefficient for one item and the 95% confidence interval for three others were below 0.5, so four of the ten items in the tool failed to meet the standard that they had defined.

AMSTAR

Shea et al reported validation of AMSTAR in two articles[18, 19]. In 2007, 42 systematic reviews were appraised using AMSTAR, by two appraisers and the agreement score overall and for each item was calculated using the weighted Cohen's kappa. Shea et al stated that the two appraisers applied AMSTAR independently to each review but also that they,

'reached agreement on the assessment results',

which raises uncertainty as to whether the kappa results reported were based on completely independent judgements by each rater or on judgements that may have been adjusted after discussion. Shea et al followed a commonly used interpretation of kappa values, with the degree of agreement indicated by the kappa value being less than chance (any negative value), slight (0-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) or almost perfect (0.81-1.00). The IRR results are summarised in table 6.6. Construct validity was assessed by testing specific hypotheses using total mean scores, for the sample of 42 systematic reviews, as a percentage of the maximum possible total score for AMSTAR and the equivalent scores and percentage of maximum possible total score from different appraisers using OQAQ. As stated earlier, the validity of using AMSTAR to calculate an overall score was later rejected by its authors, when they published AMSTAR 2[12], so their choice of this method to assess construct validity undermines confidence in their conclusions.

In 2009[19], 30 systematic reviews were each appraised, using AMSTAR, OQAQ and the 1987 Sacks' tool[7], by two appraisers and the agreement scores for each tool overall and for each item were calculated using Cohen's kappa, using responses dichotomised into two categories: 'yes' and 'any other'. This time, it was not reported whether the appraisers applied the tool

independently or whether they reached agreement. The authors stated that one of the two reviewers was *'without formal training'* but later seem to contradict this by commenting that,

'the relatively high reliability of total scores for AMSTAR and OQAQ may be partly because of the raters' familiarity with both instruments'.

The IRR results are summarised in table 6.6. Construct validity was tested by assessing convergence using intraclass correlation coefficients based on the scores obtained from each pair of the three tools. However, as with the 2007 report, confidence in their conclusions is undermined by their choice to calculate total mean scores as a percentage of the maximum possible score. They also assessed feasibility by comparing the time taken to complete the three tools and noting where appraisers reported difficulty completing the score for an item. At an average of 14.9 minutes, AMSTAR was the quickest of the three tools to complete. Reporting of the time to complete AMSTAR 2, indicates that this time is in addition to the time taken to read the review[12].

ROBIS

Buhn et al (2017)[20] compared ROBIS with AMSTAR, stating that,

'The major drawback of AMSTAR is that it relies heavily on reporting quality rather than on methodological quality'

and,

'there is a movement from generic quality checklists, such as AMSTAR, toward a more domain-based approach, such as in other critical appraisal tools.'

Sixteen systematic reviews were appraised using ROBIS and AMSTAR, by four appraisers. All reviewers had moderate to high experience of using AMSTAR but no previous experience of using ROBIS. They were familiarised with ROBIS by reading the article describing its development and a guidance document on its use. The degree of agreement between all appraisers for both tools was assessed by calculating Fleiss' kappa. For this assessment, some categories of response used in ROBIS were grouped e.g. 'yes' and 'probably yes'. To maintain comparability with the published validation studies of AMSTAR described above, Buhn et al

dichotomised all AMSTAR responses into 'yes' and 'all others'. IRR for each possible pair of appraisers was calculated using Cohen's weighted kappa, for ROBIS, and Cohen's kappa, for AMSTAR. The IRR results are summarised in table 6.6. Construct validity was assessed by comparing the results with each tool for appraisal of Cochrane versus non-Cochrane reviews, which found use of both tools resulted in a statistically significant higher number of items with 'yes' responses for Cochrane Reviews. The mean time for reaching consensus amongst four reviewers was 10 minutes for AMSTAR and 50 minutes for ROBIS, but the time for applying each tool individually was stated to have not been reliably recorded. Buhn et al concluded that the performance of ROBIS was similar to that of AMSTAR but noted that the AMSTAR kappa results were lower than reported in previous publications. Two thirds of the items in both tools had kappa results in the range of fair agreement (0.21-0.40) or lower, but two items in AMSTAR had negative kappa results, indicating less than chance agreement.

AMSTAR 2

In 2017, Shea et al[12] reported IRR testing of AMSTAR 2, involving three pairs of appraisers and three sets of systematic reviews; one pair of appraisers were developers of the tool. One of the reasons stated for updating the original AMSTAR tool was to address the inclusion of non-randomised studies in systematic reviews, so 34 of the 54 systematic reviews used for the IRR test included only non-randomised trials (18) or a mixture of randomised and non-randomised trials (16). The authors provided less detail of the IRR testing methods than in their 2007 and 2009 reports of testing the original AMSTAR. They did not state whether each pair of raters assessed only one type of review, a mixture of two or of all three types. They acknowledged that the IRR results (0-1) varied widely across items and between pairs of raters but stated that 46 of 50 kappa scores were in the moderate agreement range (0.41 - 0.60) or higher. There are 16 items in the AMSTAR 2 tool, but kappa values were calculated separately for criteria relating to randomised and non-randomised trials for two items, so there were potentially a total of 54 kappa values across the three pairs of reviewers. In addition to Shea et

al's summary of 50 scores, there were three kappa scores of 0 and one score was not calculated. The IRR results are summarised in Table 6.6, showing results for each of the three pairs of appraisers.

Table 6.6 Inter-rater reliability test results – AMSTAR and ROBIS

kappa value	AMSTAR[18] 2007	AMSTAR[19] 2009	AMSTAR[20] 2017	ROBIS[20] 2017	AMSTAR 2 [12] 2017 (NB results are for 3 pairs of appraisers)		
<0	0	0	2	0	0	0	0
0.00 - 0.20	0	0	3	10	1	2	0
0.21 - 0.40	0	1	2	7	3	1	0
0.41 - 0.60	1	3	2	6	2	1	4
0.61 - 0.80	3	5	1	1	9	4	11
0.81 - 1.00	7	2	1	0	3	9	3

(n = number of items with a kappa value within each range)

CASP and JBI tools

No article describing the validation of any CASP or JBI critical appraisal tool was identified.

6.5 Discussion

6.5.1 Summary of the findings of this chapter

The systematic reviews of critical appraisal tools that I identified report that the empirical approach used to construct the tools and validation are not reported for a high proportion of tools. Although it was common for tools to produce a summary score, no justification was provided for any scoring system. Some critical appraisal tools identified in the systematic reviews or mentioned in other included articles were actually reporting guidelines or general guidance on reviewing a paper, rather than a formally designed critical appraisal tool.

My review identified 13 critical appraisal tools for systematic reviews or indirect comparisons of randomised trials of healthcare interventions in humans, published between 1987 and 2017. Three of these tools were developed specifically to include appraisal of synthesis using NMA methodology. The number of items included in the tools ranged from 8 to 43, but only from 8

to 16 in those tools aimed at more general users. The extent to which responses to the items in a tool assess only whether something was reported, as opposed to evaluating the quality, validity or risk of bias ranged from a strong focus on reporting (e.g. Jansen et al[14]) to little focus on reporting (e.g. Ortega et al[16]). Most tools had less focus on reporting. Guidance was included alongside each item in some tools but provided as a separate document or only within the published article for others. Only one tool (R-AMSTAR[27]) was intended to be used to produce a summary score. The authors of the original version of the AMSTAR tool acknowledged it had been used, at times, to produce an overall score but emphasised, when publishing AMSTAR 2, that this should not be done. They referenced examples of the previously published evidence of the hazards of doing so[32, 33] and concluded,

'We strongly recommend that individual item ratings are not combined to create an overall score. Rather, users should consider the potential impact of an inadequate rating for each item.'

Both the empirical approach used to construct the tool and IRR testing were reported for six tools[9-13, 27]. The empirical approach used to construct the tool and IRR testing as part of the development process were reported for one additional tool but testing was not conducted using the final version of the tool[16]. Brief mention of a limited form of IRR testing was reported for another tool[7].

The empirical approach for constructing some tools used either content from earlier ones (e.g. AMSTAR used items included in the 1987 Sacks et al tool and OQAQ; the 2013 Higgins et al tool used items in AMSTAR) or content from other existing sources (e.g. ROBIS used items in the Cochrane MECIR). In each case, an initial higher number of identified items was reduced to a smaller number in the final tool. In other tools, content was devised using a group of experts, sometimes including a range of expertise, e.g. OQAQ, AMSTAR, ROBIS, and other times it was limited to a specific discipline, e.g. statisticians (the 2013 Higgins et al tool) or hospital pharmacists (the 2014 Ortega et al tool).

The arrangements for IRR testing varied in terms of the number of raters, whether or not developers of the tool were included as raters, the number and clinical variety of reviews assessed, whether each item or only a summary question for a category/domain was considered and whether responses were grouped or considered separately.

There were few comparisons of the time taken to complete different tools but these tended to indicate that tools containing fewer items were completed in a shorter time.

Details of the empirical approach to construction or reports of IRR testing were not found for the systematic review tools from CASP or JBI, which are the most widely recognised centres of excellence for accessing critical appraisal tools.

The range of scope and target users of identified tools included those aimed at general users of systematic reviews without specialist expertise (e.g. AMSTAR and AMSTAR 2), and those aimed at very specific and/or expert users (e.g. the 2014 Ortega et al tool for decision makers in drug evaluation and the 2011 Ades et al tool for those assessing whether the NICE Methods guide requirements are met).

Across the three tools identified that included content addressing specific aspects of NMA methodology, whilst details varied, there was substantial common ground as to the components that critical appraisal should consider: connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made.

6.5.2 Strengths and limitations of the research presented in this chapter

This was the first review, as far as I have been able to identify, of critical appraisal tools for systematic reviews including synthesis using NMA methodology.

It is likely that some articles were not identified by my review because of human error in the review process, particularly because only one reviewer did the work.

I included the systematic review tools of CASP and JBI due to their widely recognised status and use by a broad range of users despite not finding any published details of their construction or validation. The lack of those details, compared with the other tools identified, makes it difficult to assess the level of confidence that should be placed in the validity of their content.

My search will not have identified some additional systematic review appraisal tools that are not the subject of published articles but might have content that would have been relevant to consider. Furthermore, consultation with a wider range of experts may also have identified additional tools. However, because of the limitations caused by a lack of published details, I limited the inclusion of such tools to only the CASP and JBI tools.

This review did not include other approaches to quality assessment of evidence produced using meta-analysis, as part of a systematic review. I discuss some of these approaches in Chapter Eight; specifically, approaches that relate to assessing evidence produced using NMA methods.

6.5.3 Implications of the findings of this chapter

This review found no evidence that there is an established consensus on a 'gold standard' for the development, validation and reporting of critical appraisal tools. This does not indicate that there is not identifiable good practice but shows that there is considerable variation in the methods used. I also found that some well-known and widely-used tools lack published details of validation and there is an absence of published information to judge whether good practice was followed in their development.

The earliest published tools for systematic reviews are the 1987 Sacks et al checklist, the 1988 Oxman and Guyatt checklist and the same authors' 1991 OQAQ. Since very few systematic reviews using NMA methodology for synthesis were published before 2009[34], as expected,

these earlier tools did not include any items or guidance that address synthesis using NMA methodology. The 1987 Sacks et al checklist and OQAQ were used as the basis for developing the content of AMSTAR in 2007, which was modified in 2010, as R-AMSTAR, to include a scoring system, and updated in 2017 to AMSTAR 2. None of these three versions of AMSTAR include any items or guidance that address synthesis using NMA methodology and the authors of AMSTAR 2 stated that,

'The revisions were not intended to deal with the special requirements of diagnostic test reviews, individual patient data meta-analyses or network meta-analyses, scoping reviews, or realist reviews.' [12].

The CASP and JBI systematic review tools contain no items or guidance that address synthesis using NMA methodology. ROBIS mentions NMA but has no criteria for assessing its use and the 2013 Higgins et al tool has one item concerning use of IC but without any criteria specific to use of NMA.

The other three tools identified in this review (all published since 2011), each contain some items and guidance that address the use of NMA in detail[14-16]. Whilst details varied between the tools, there was substantial common ground as to the components that critical appraisal should consider: connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made.

As regards the format of a critical appraisal tool, my review suggests that tools with fewer items and with guidance alongside each item are completed in a shorter time, so these could be regarded as user-friendly characteristics. I also identified potential drawbacks in using responses to calculate an overall score.

For the purpose of my research, and from my perspective as a practising clinician, my aim had been to identify a critical appraisal tool that would be user-friendly to a broad range of users without specialist statistical expertise and I concluded from this review that the characteristics of such a tool should,

a) match those of the most widely-used critical appraisal tools identified in this review that were aimed at similar users (AMSTAR, CASP and JBI) i.e. fewer items (10-11); brief guidance that accompanies each item within the tool.

b) not use responses to calculate an overall score because that can,

'disguise critical weaknesses that should diminish confidence in the results of a systematic review'[12]

c) in general, contain items with less focus on reporting and more on evaluation

d) include items and guidance specifically addressing the identified common components for appraisal of the use of NMA methodology, whilst accepting that users with lack of statistical expertise would not be expected to evaluate detailed statistical aspects and consequently there would need to be more focus on assessing reporting for these items

e) have clarity about the empirical process used to construct the tool and the validation undertaken.

Only three tools identified in this review included the relevant NMA methodology content but each one did not meet some of the other criteria listed above[14-16]. Jansen et al had a strong focus on reporting and there were very limited details published of the process used to construct the tool and no published details of other validation[14] Ades et al[15] included 39 items (plus two related to cost effectiveness analysis) and provided extensive guidance within the article but not within the tool. Their tool was developed for use by those assessing whether the NICE Methods guide requirements are met, rather than for a broad range of users without specialist statistical expertise. Ortega et al[16] included 20 items and did provide brief guidance alongside each item within the tool. Their tool was developed for decision makers in drug evaluation but aimed to be user-friendly.

Both the Ades et al and Ortega et al tools reflect current knowledge on research quality and risk of bias and are critical appraisal tools to consider using to assess systematic reviews using NMA methodology for synthesis. However, I did not identify any existing critical appraisal tool that fulfilled all my preferred criteria. The Ortega et al tool was the nearest to meeting my preferred criteria and used part of the CASP format, but it contains twice as many items as the most widely-used appraisal tools for systematic reviews and some items require users to evaluate detailed statistical aspects. Therefore, I decided that the next task in my programme of research should be to construct a new tool, close in format and number of items to those most widely-used tools, reflecting the most up-to-date content from the tools that I had identified. My work towards this objective is described in Chapter Seven.

6.6 Chapter summary

- My review found no evidence that there is an established consensus on a 'gold standard' for the development, validation and reporting of critical appraisal tools but there were some common elements reported that can be regarded as good practice.
- Nearly all relevant critical appraisal tools published since 2011 contain some mention of NMA methodology but only three contain items and guidance that specifically address details of its use
- Across those three tools, there was substantial common ground as to the components that critical appraisal should consider: connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made.
- Neither details of the empirical approach to construction nor reports of IRR testing were found for the systematic review tools of either CASP or JBI, the most widely recognised centres of excellence for accessing critical appraisal tools.
- No items or elements of guidance in the systematic review tools of either CASP or JBI specifically address synthesis with NMA methodology.

- Since I did not identify an existing critical appraisal tool that fulfilled all my preferred criteria, my next task was to construct a tool meeting those criteria.

6.7 Chapter references

1. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clinical Trials 1995; 16: 62-73. 1995/02/01. DOI: 10.1016/0197-2456(94)00031-w.

2. CASP UK. CASP, https://casp-uk.net/ (accessed 12 September 2019).

3. Joanna Briggs Institute. Joanna Briggs Institute, https://joannabriggs.org/ (accessed 12 September 2019).

4. Katrak P, Bialocerkowski AE, Massy-Westropp N, et al. A systematic review of the content of critical appraisal tools. BMC Medical Research Methodology 2004; 4: 22. 2004/09/17. DOI: 10.1186/1471-2288-4-22.

5. Crowe M and Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol 2011; 64: 79-89. 2010/12/07. DOI: 10.1016/j.jclinepi.2010.02.008.

6. Zeng X, Zhang Y, Kwong JS, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. Journal of evidence-based medicine 2015; 8: 2-10. 2015/01/17. DOI: 10.1111/jebm.12141.

7. Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316: 450-455. 1987/02/19. DOI: 10.1056/NEJM198702193160806.

8. Oxman AD and Guyatt GH. Guidelines for reading literature reviews. CMAJ 1988; 138: 697-703. 1988/04/15.

9. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology 2007; 7: 10. 2007/02/17. DOI: 10.1186/1471-2288-7-10.

10. Higgins JP, Lane PW, Anagnostelis B, et al. A tool to assess the quality of a meta-analysis. Res Synth Methods 2013; 4: 351-366. 2013/12/01. DOI: 10.1002/jrsm.1092.

11. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016; 69: 225-234. 2015/06/21. DOI: 10.1016/j.jclinepi.2015.06.005.

12. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017; 358: j4008. 2017/09/25. DOI: 10.1136/bmj.j4008.

13. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. J Clin Epidemiol 1991; 44: 91-98. 1991/01/01. DOI: 10.1016/0895-4356(91)90205-n.

14. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14: 417-428. 2011/06/15. DOI: 10.1016/j.jval.2011.04.002.

15. Ades A.E. CDM, Reken S., Welton N.J., Sutton A.J., Dias S. NICE DSU Technical Support Document 7: Evidence Synthesis Of Treatment Efficacy In Decision Making:

A Reviewer's Checklist http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD7-reviewer-checklist.final_.08.05.12.pdf (2012, accessed 12 October 2019).

16. Ortega A, Fraga MD, Alegre-del-Rey EJ, et al. A checklist for critical appraisal of indirect comparisons. IntJClinPract 2014; 68: 1181-1189. 2014/10/02. DOI: 10.1111/ijcp.12487.

17. Oxman AD and Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44: 1271-1278. 1991/01/01. DOI: 10.1016/0895-4356(91)90160-b.

18. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007; 2: e1350. 2007/12/27. DOI: 10.1371/journal.pone.0001350.

19. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009; 62: 1013-1020. 2009/02/24. DOI: 10.1016/j.jclinepi.2008.10.009.

20. Buhn S, Mathes T, Prengel P, et al. The risk of bias in systematic reviews tool showed fair reliability and good construct validity. J Clin Epidemiol 2017; 91: 121-128. 2017/07/12. DOI: 10.1016/j.jclinepi.2017.06.019.

21. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999; 354: 1896-1900. 1999/12/10. DOI: 10.1016/s0140-6736(99)04149-5.

22. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Annals of Internal Medicine 1994; 120: 667-676. 1994/04/15. DOI: 10.7326/0003-4819-120-8-199404150-00008.

23. Wilson A and Henry DA. Meta-analysis. Part 2: Assessing the quality of published meta-analyses. Medical Journal of Australia 1992; 156: 173-174, 177-180, 184-177. 1992/02/03. DOI: 10.5694/j.1326-5377.1992.tb139704.x.

24. Hunt DL and McKibbon KA. Locating and appraising systematic reviews. Annals of Internal Medicine 1997; 126: 532-538. 1997/04/01. DOI: 10.7326/0003-4819-126-7-199704010-00006.

25. National Institute for Health and Clinical Excellence. Process and methods [PMG10]. Methodology checklist:systematic reviews and meta-analyses: National Institute for Health and Care Excellence, July 2016 update http://www.nice.org.uk/process/pmg10/chapter/appendix-b-methodology-checklist-systematic-reviews-and-meta-analyses (accessed 2 October 2019).

26. Aromataris E, Fernandez R, Godfrey CM, et al. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. International Journal of Evidence-based Healthcare 2015; 13: 132-140. 2015/09/12. DOI: 10.1097/XEB.0000000000000055.

27. Kung J, Chiappelli F, Cajulis OO, et al. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. Open Dent J 2010; 4: 84-91. 2010/11/20. DOI: 10.2174/1874210601004020084.

28. Critical Appraisal Skills Programme (2017) CASP Systematic Review Checklist. (12 9 19), https://casp-uk.net/wp-content/uploads/2018/01/CASP-Systematic-Review-Checklist_2018.pdf (accessed 12 September 2019).

29. Joanna Briggs Institute. Joanna Briggs Institute Checklist for Systematic Reviews and Research Syntheses https://joannabriggs.org/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Systematic_Reviews2017_0.pdf (2017, accessed 12 September 2019).

30. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14: 429-437. 2011/06/15. DOI: 10.1016/j.jval.2011.01.011.

31. Chandler J CR, Higgins J, Lasserson T, Tovey D. Methodological Expectations of Cochrane Intervention Reviews (MECIR): methodological standard for the conduct of new Cochrane Intervention Reviews., https://community.cochrane.org/mecir-manual (2012, accessed 18 October 2019).

32. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. Jama 1999; 282: 1054-1060. 1999/09/24. DOI: 10.1001/jama.282.11.1054.

33. Greenland S and O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics 2001; 2: 463-471. 2003/08/23. DOI: 10.1093/biostatistics/2.4.463.

34. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol 2014; 67: 138-143. 2013/10/05. DOI: 10.1016/j.jclinepi.2013.07.014.

Chapter Seven

Critical appraisal tool development

7 Critical appraisal tool development

7.1 Background

My aim was to fill the gap identified in the review presented in Chapter Six by developing a critical appraisal tool for systematic reviews using NMA methodology that would allow a wide range of end-users of such reviews to critically appraise them. I wished to develop a tool that was simpler in format and with content appropriate to support a generalist appraiser, rather than only for appraisers with pre-existing specialist statistical knowledge of NMA methodology. As described in Chapter Six, my preferred characteristics of a suitable critical appraisal tool, in order for it to be user-friendly to the widest group of users, were that it should:

a) match those of the most widely-used critical appraisal tools identified that were aimed at similar users (AMSTAR, CASP and JBI) i.e. fewer items (10-11); brief guidance that accompanies each item within the tool.

b) not use responses to calculate an overall score because that can,

c) in general, contain items with less focus on reporting and more on evaluation

d) include items and guidance specifically addressing the identified common components for appraisal of the use of NMA methodology (connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made), whilst accepting that users with lack of statistical expertise would not be expected to evaluate detailed statistical aspects and consequently there would need to be more focus on assessing reporting for these items

e) have clarity about the empirical process used to construct the tool and the validation undertaken.

My review of critical appraisal tools, reported in Chapter Six, identified none that fulfilled all those criteria. In that review, I found no compelling evidence base for selection of one particular

format for a new tool, but noted that the most widely-used tools provide brief guidance on completion within the tool alongside each item and that these can be completed in a shorter time. My review found that the items included in tools were developed based on the content of earlier tools or similar resources and/or the opinions of experts, so I decided to combine these approaches. My review also found that an initial higher number of identified source items was then reduced to a smaller number of items included in the final tool and I expected to achieve the same.

Finally, my review found that the Ortega et al tool[1] was the nearest to meeting the preferred characteristics that I identified, and that this tool used part of the CASP format but provided more guidance than the brief hints used in CASP tools. CASP tools use a format with wide international recognition and are designed for generalist users. CASP publish a range of critical appraisal tools covering most study designs but do not currently publish a tool with content relevant to use of NMA methods, whether as part of a SR or not. Based on these considerations, I decided to use the CASP critical appraisal tool format, but to provide guidance within the tool to support users that went beyond the brief hints used in existing CASP tools.

7.2 Chapter aims and objectives

The aim of the research reported in this chapter was to construct and validate a critical appraisal tool for systematic reviews using NMA methodology for synthesis. The primary objectives were to identify appropriate items to include in the tool and to conduct an inter-rater reliability test.

7.3 Methods

To identify items for inclusion in the tool that would have face validity, I selected source tools from those identified in the review reported in Chapter Six. The source tools were selected on the basis that they were tools for generalist use, rather than for appraisal of detailed statistical features, and that both the empirical method for constructing the tool and the inter-rater reliability testing were reported. Additionally, only the most recent iteration of a critical appraisal

tool was included. Items from these sources and any accompanying guidance included in a tool, but not separate from it, were extracted into a Microsoft Excel spreadsheet, with one column for the items from each tool. The items were sorted into rows reflecting overlapping content, matching each item from the existing CASP tool for systematic reviews and including additional rows reflecting content not covered by an existing CASP tool item. Each row was allocated a topic name relating to the aspect of a review that it addresses. Content from the JBI systematic reviews tool was also extracted, on the basis of its international recognition and design for generalist users, similar to the CASP tool. A qualitative synthesis was then undertaken to develop a new tool from the existing CASP tool for systematic reviews, replacing, amending or incorporating new items and adding a separate guidance section, whilst retaining the hints section but amending its content, if indicated.

Because no reports of evidence for the face and content validity and IRR testing of the CASP or JBI tool were identified in the review reported in Chapter Six, although their content was included in the synthesis process for the new tool, I decided not to include any item that could be sourced only from either the CASP or JBI tool but not from any of the other source tools.

The content of the tool I developed was checked for omissions against the PRISMA extension statement on reporting standards for NMAs in systematic reviews. Although this is a reporting standard, it reflects current evidence on the quality elements for such a review, so was considered to be a relevant reference point.

After constructing an initial draft of the tool, a recent set of published systematic reviews was checked for reported methods of minimising bias or any other comments on bias in order to identify any items not addressed by the draft tool. The search strategy reported in Chapter Three was used to search MEDLINE in January 2019 and, starting with those most recently published, abstracts of the retrieved articles were screened to identify at least 50 systematic reviews that reported using NMA methods for synthesis. The full text of these articles was retrieved and

checked for reported methods of minimising bias or any other comments on bias that were not already addressed by any item in the initial version of the draft tool.

I shared the revised draft tool with two experts in systematic reviews and critical appraisal for comment and made further amendments to take account of their feedback.

Inter-rater reliability (IRR) was tested using eight raters, including myself. All raters were staff of Hull York Medical School and were involved in teaching critical appraisal to undergraduate medical students. Each rater appraised three systematic reviews that had used NMA for synthesis. All three reviews were published between 2013 and 2017, on the topic of prophylaxis for migraine[2-4]. No training in using the tool was provided. Raters were asked for general feedback on the tool and to rate their level of confidence to critically appraise a systematic review incorporating NMA before this exercise and their level of confidence using the tool, on a scale of 1 to 5 (1 = low, 5 = high). Fleiss' kappa values were calculated for all raters for each item in the tool to assess the level of agreement between raters above that expected by chance. Additionally, the exact percentage agreement and percent agreement within one ordinal degree (e.g. yes vs partially/unsure) between raters were calculated for each item.

7.4 Results

7.4.1 Source critical appraisal tools included

Items and guidance notes were sourced from the following existing tools: CASP tool for systematic reviews[5], JBI tool for systematic reviews[61], AMSTAR 2[7], ROBIS[8] and Ortega et al (2014)[1].

[1]The version available online from the JBI site was used, not the version in the article cited within the guidance that accompanies it, which is different, as explained in the previous chapter

The 2013 Higgins et al tool[9] was considered because the empirical method for constructing the tool and the inter-rater reliability testing were reported, but it was not included because it was specifically developed to assess detailed statistical content and consequently contains a large number (43) of items.

The 2011 ISPOR Task Force's tool (Jansen et al)[10] and the 2012 NICE DSU Technical Support Document 7 tool (Ades et al)[11] were also considered, because both contain items and detailed guidance that address various aspects of the use of network meta-analysis for synthesis, but were not included because the former had a strong focus on reporting and the latter contains 41 items and is aimed at users with more specialised knowledge. Additionally, the common components specific to NMA (connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made) that I identified in my review reported in Chapter Six, were already represented in Ortega's tool.

7.4.2 Identification of items

24 distinct topics were identified by sorting the items extracted from the included source tools. Each distinct topic was labelled with a name relating to the aspect of a review it addresses. Some items from included tools appear under more than one topic if an item corresponds to more than one CASP item, or to more than one item from other included tools when there is no corresponding CASP item in each case. Since a preferred characteristic for a new tool was to have fewer items, multiple items from included tools appear under the same topic if they correspond to a single CASP item or to a single item from another included tool when there is no corresponding CASP item. Some items from the CASP tool appear under more than one topic if they correspond to items that are addressed separately in other included tools but, when considered together, do not correspond fully to the CASP item. Only content specific to randomised trials and to intervention reviews was included from ROBIS, which contains some

content specific to non-randomised trials and has different options for initial consideration of the review question.

No additional items were included after checking for omissions against the PRISMA extension statement on reporting standards for NMAs in systematic reviews.

Table 7.1 lists the distinct topics identified, the corresponding numbered items in the new tool (if any), and which source tools contain items related to each topic.

Table 7.1 Summary of which source tools contain items related to each topic identified

Topic number	Topic	Item(s) in new tool	CASP	JBI	AMSTAR 2	ROBIS	Ortega et al
1	Review question	1	✓	✓	✓	✓	✓
2	Adherence to protocol	1,8			✓	✓	
3	Inclusion criteria	2	✓	✓	✓	✓	
4	Search methods	3	✓	✓	✓	✓	✓
5	Selection of studies	4			✓	✓	
6	Details of excluded studies	3			✓		
7	Appraisal of included studies	4	✓	✓	✓	✓	✓
8	Extraction of data			✓	✓	✓	
9	Reliability of studies data						✓
10	Details of included studies	5	✓		✓	✓	
11	Selection of studies for synthesis	5	✓		✓	✓	✓
12	Synthesis choices	5,6,7	✓	✓	✓	✓	✓
13	Heterogeneity	6			✓	✓	
14	Inconsistency	6					✓
15	Publication bias	7		✓	✓		
16	Results - general	8	✓			✓	
17	Results - precision	8	✓				
18	Conflict of interest	3,7			✓		✓
19	Conclusions - internal validity	9		✓	✓	✓	
20	Conclusions - External applicability	10	✓				✓

Topic number	Topic	Item(s) in new tool	CASP	JBI	AMSTAR 2	ROBIS	Ortega et al
21	Conclusions - other information required	10	✓				✓
22	Conclusions - benefits v harms		✓				
23	Conclusions- new research			✓			
24	Conclusions - overall risk of bias					✓	

Table 7.2 lists the distinct topics identified, the corresponding numbered items in the new tool (if any), and the items and hints/comments/guidance, extracted from each of the included tools, which relate to each topic. Hints, comments or guidance accompanying items in a tool are shown in italics. For practicality, the contents of the comments column from the tool published by Ortega et al were not included in the table, because they are, in many instances, as lengthy as the guidance provided for other tools in a separate document.

Table 7.2 Summary of source items related to each topic identified

Topic 1	**Review question**	
	Corresponding item(s) in new tool	1
CASP	Did the review address a clearly focused question? *An issue can be 'focused' in terms of: the population studied; the intervention given; the outcome considered.*	
JBI	Is the review question clearly and explicitly stated?	
AMSTAR 2	Did the research questions and inclusion criteria for the review include the components of PICO? *For yes: Population; Intervention; Comparator group; Outcome. Optional (recommended): Timeframe for follow-up. Response options: Yes; No.*	
ROBIS	Does the question addressed by the review match the target question? *Patients/Population(s); Intervention(s); Comparator(s); Outcome(s) for Target question (e.g. overview or guideline) and for Review being assessed.*	
Ortega et al	Was the search strategy for study selection adequate? *Was the search question correctly established? (PICO). Are all relevant studies included? Is the comparator relevant?*	

Topic 2	**Adherence to protocol**	
	Corresponding item(s) in new tool	1,8
AMSTAR 2	Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol? *For Partial Yes: The authors state that they had a written protocol or guide that included ALL the following: review question(s); a search strategy; inclusion/exclusion criteria; a risk of bias assessment. For Yes: As for partial yes, plus the protocol should be registered and should also have specified: a meta-analysis/synthesis plan, if appropriate and: a plan for investigating causes of heterogeneity; justification for any deviations from the protocol. Response options: Yes; Partial yes; No*	
ROBIS	Did the review adhere to pre-defined objectives and eligibility criteria?	
Topic 3	**Inclusion criteria**	
	Corresponding item(s) in new tool	2
CASP	Did the authors look for the right type of papers? *The best sort of studies would address the review's question; have an appropriate study design (usually RCTs for papers evaluating interventions)*	
JBI	Were the inclusion criteria appropriate for the review question?	
AMSTAR 2	Did the review authors explain their selection of the study designs for inclusion in the review? *For Yes, the review should satisfy ONE of the following: Explanation for including only RCTs; OR Explanation for including only NRSI; OR Explanation for including both RCTs and NRSI. Response options: Yes; No.*	
ROBIS	1) Were the eligibility criteria appropriate for the review question? 2) Were eligibility criteria unambiguous? 3) Were any restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)? 4) Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)?	
Topic 4	**Search methods**	
	Corresponding item(s) in new tool	3
CASP	Do you think all the important, relevant studies were included? *Look for: Which bibliographic databases were used; follow up from reference lists; personal contact with experts; search for unpublished as well as published studies; search for non-English language studies*	
JBI	1) Was the search strategy appropriate? 2) Were the sources and resources used to search for studies adequate?	
AMSTAR 2	Did the review authors use a comprehensive literature search strategy? *For Partial Yes (all the following): searched at least 2 databases (relevant to research question); provided key word and/or search strategy; justified publication restrictions (eg, language). For Yes, should also have (all the following): searched the reference lists/bibliographies of included studies; searched trial/study registries; included/consulted content experts in the field; where relevant, searched for grey literature; conducted search within 24 months of completion of the review. Response options: Yes; Partial Yes; No.*	

ROBIS	1) Did the search include an appropriate range of databases/electronic sources for published and unpublished reports? 2) Were methods additional to database searching used to identify relevant reports? 3) Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible? 4) Were restrictions based on date, publication format, or language appropriate?	
Ortega et al	1) Was the search strategy for study selection adequate? *Was the search question correctly established? (PICO). Are all relevant studies included? Is the comparator relevant? 2)* Search strategy. *Are systematic methods used to find all relevant studies? (Were the different databases as well as the search strategy included? Were references checked? Were strict limits used? Was publication bias evaluated?). Response options: Yes = high reliability; Potential bias and its potential effects are analysed in some fashion = Acceptable; No, and bias is not analysed = Low reliability.* 3) Criteria for selection of studies. *Are they stated? Are they reasonable? Are they systematically applied? Yes = high reliability; Acceptable, however some aspects could be improved = Acceptable; No, the answer to one of the three questions is NO, or the item is done in an unacceptable manner (risk of bias) = Low reliability.*	
Topic 5	**Selection of studies**	
	Corresponding item(s) in new tool	4
AMSTAR 2	Did the review authors perform study selection in duplicate? *For Yes, either ONE of the following: at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include; OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 per cent), with the remainder selected by one reviewer. Response options: Yes;*	
ROBIS	Were efforts made to minimise error in selection of studies?	
Topic 6	**Details of excluded studies**	
	Corresponding item(s) in new tool	3
AMSTAR 2	Did the review authors provide a list of excluded studies and justify the exclusions? *For Partial Yes: provided a list of all potentially relevant studies that were read in full text form but excluded from the review. For Yes, must also have: Justified the exclusion from the review of each potentially relevant study. Response options: Yes; Partial Yes; No*	
Topic 7	**Appraisal of included studies**	
	Corresponding item(s) in new tool	4
CASP	Did the review's authors do enough to assess quality of the included studies? *The authors need to consider the rigour of the studies they have identified. Lack of rigour may affect the studies' results. ("All that glitters is not gold" Merchant of Venice - Act II Scene 7)*	
JBI	1) Were the criteria for appraising studies appropriate? 2) Was critical appraisal conducted by two or more reviewers independently?	

AMSTAR 2	Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? *For Partial Yes, must have assessed RoB from: unconcealed allocation and; lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all cause mortality). For Yes, must also have assessed RoB from: allocation sequence that was not truly random, and; selection of the reported result from among multiple measurements or analyses of a specified outcome. Response options: Yes; Partial Yes; No.*	
ROBIS	1) Was risk of bias (or methodological quality) formally assessed using appropriate criteria? 2) Were efforts made to minimise error in risk of bias assessment?	
Ortega et al	Study quality. *What is the quality of the studies being combined? Are the results within each study reported? (transparency is essential for evaluation of applicability). Response options: Good quality RCT = High reliability; There may be a cohort study with good quality, but there is at least one clinical trial for each comparison. The weight of epidemiological studies (number of patients) is lower than that of the clinical trials = Uncertain; The answer does not fit in any of the above situations = Low reliability.*	
Topic 8	**Extraction of data**	
	Corresponding item(s) in new tool	None
JBI	Were there methods to minimize errors in data extraction?	
AMSTAR 2	Did the review authors perform data extraction in duplicate? *For Yes, either ONE of the following: at least two reviewers achieved consensus on which data to extract from included studies; OR two reviewers extracted data from a sample of eligible studies <u>and</u> achieved good agreement (at least 80 per cent), with the remainder extracted by one reviewer. Response options: Yes; No.*	
ROBIS	1) Were efforts made to minimise error in data collection? 2) Were all relevant study results collected for use in the synthesis?	
Topic 9	**Reliability of studies data**	
	Corresponding item(s) in new tool	None
Ortega et al	Reliability of studies data. *Are data obtained from the studies? Or was there a previous extrapolation or any manipulation of the data? Response options: Yes, data are the same as published = High reliability; Data have been manipulated or extrapolated = Low reliability.*	
Topic 10	**Details of included studies**	
	Corresponding item(s) in new tool	5
CASP	If the results of the review have been combined, was it reasonable to do so? *Consider whether: The results were similar from study to study; The results of all the included studies are clearly displayed; The results of different studies are similar; The reasons for any variations in results are discussed*	
AMSTAR 2	Did the review authors describe the included studies in adequate detail? *For Partial Yes (ALL the following): described populations; described interventions; described comparators; described outcomes; described research designs. For Yes, should also have ALL the following: described population in detail; described intervention and comparator in detail (including doses where relevant); described study's setting; timeframe for follow-up. Response options: Yes; Partial Yes; No.*	
ROBIS	Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	

Topic 11	**Selection of studies for synthesis**	
	Corresponding item(s) in new tool	5
CASP	If the results of the review have been combined, was it reasonable to do so? *Consider whether: The results were similar from study to study; The results of all the included studies are clearly displayed; The results of different studies are similar; The reasons for any variations in results are discussed*	
AMSTAR 2	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis? *For yes: included only low risk of bias RCTs; OR, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed analyses to investigate possible impact of RoB on summary estimates of effect. Response options: Yes; No; No meta-analysis conducted.*	
ROBIS	1) Did the synthesis include all studies that it should? 2) Were biases in primary studies minimal or addressed in the synthesis?	
Ortega et al	1) Study similarity. *Is there anything that makes the studies not comparable? Response options: No = High reliability; There are some differences, but they are not very relevant = Uncertain; There are differences that make the studies not comparable = Low reliability.* 2) Common outcome measure. *Do studies share the same outcome measure (defined in the same manner, including same time horizon and reported in the same way)? Response options: Yes = high reliability; No, but differences are discussed and not relevant = Uncertain; No = Low reliability.* 3) Clinical similarity, patient similarity. *Are inclusion/exclusion criteria similar between studies? Were randomised patients within each study similar? Response options: Yes, if there was a difference, it was not relevant = High reliability; Some differences, but most patients were similar. If a subgroup of patients was compared, patients of the subgroups were also similar = Acceptable; Patients or subgroups compared were very different between studies = Low reliability.* 4) Common comparator (control intervention). *Is the common comparator used at the same dose and for the same duration? And are the same concomitant treatments that can act as effect modifiers permitted? Response options: Yes = High reliability; Some differences with little influence on results = Acceptable; Relevant differences = Low reliability.* 5) Outcome results in the control group or common comparator group. *Is this similar in the studies being compared? Response options: Yes or few irrelevant differences = High reliability; No, they are different, but the difference is small and clinically acceptable = Acceptable; No, there are important differences. This is probably a consequence of important differences in patient populations and/or interventions and/or outcome measures = Low reliability.* 6) Missing patient/data. *In studies being combined, are missing patients/data similar? Were they handled in the same way? Is data analysis similar (per protocol, intention to treat - ITT)? Response options: Yes = High reliability; Differences are not relevant = Acceptable; There is not enough information or there are relevant differences, but they do not invalidate the study. ITT was conducted in all studies = Uncertain; The answer does not fit in any of the above situations = Low reliability.*	

Topic 12	**Synthesis choices**	
	Corresponding item(s) in new tool	5,6,7
CASP	If the results of the review have been combined, was it reasonable to do so? *Consider whether: The results were similar from study to study; The results of all the included studies are clearly displayed; The results of different studies are similar; The reasons for any variations in results are discussed.*	

JBI	Were the methods used to combine studies appropriate?
AMSTAR 2	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results? *For yes: The authors justified combining the data in a meta-analysis AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present AND investigated the causes of any heterogeneity. Response options: Yes; No; No meta-analysis conducted.*
ROBIS	1) Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies? 2) Were the findings robust, e.g. as demonstrated through funnel plot or sensitivity analyses? 3) Were biases in primary studies minimal or addressed in the synthesis?
Ortega et al	1) Is it justified to conduct an indirect comparison (IC)? *Is there a direct comparison? In general, direct comparisons are regarded as having less bias than indirect comparisons. If there is a direct and an indirect comparison, it is interesting to contrast their conclusions. If both conclusions are similar, our reliability in results will increase; if not, we should think about the reasons for the discrepancy. Some criteria must be fulfilled to conduct an IC: Are the objective(s), inclusion criteria, methods (assumptions included) and results clear? (If there is no transparency in these points, IC is impossible or very difficult); Are the studies clinically comparable? (patients, general conditions, co-interventions, drug dosing, treatment duration, variables etc., that act as modifiers); Are outcome variables comparable in definition and presentation? Are outcome variables clinically relevant? Response options: Yes; No.* 2) Common comparator. *Are comparators and studies represented graphically? How many intermediate comparators are there between the two alternatives (A and B) we are most interested in comparing? Response options: 1 = high reliability; 2 = uncertain; more than 2 = low reliability. 3)* Outcome relevance*. Is the outcome measure clinically relevant? Is the main outcome variable included in the studies? ,Response options: Yes = high; The most relevant but not the main outcome variable included in the studies = Acceptable; It is not the most relevant but is acceptable = Uncertain; No variable was clinically relevant. Another outcome variable should have been measured in the studies = Low reliability.* 4) Adjusting by control group. *Was an adjusted IC method (Bucher, NMA) conducted? Are relative effects to common comparator of group of each study (or meta-analysis for each comparison) compared? Response options: Yes = High reliability; No = Low reliability.* 5) Is the IC methodology appropriate? *Non-adjusted indirect comparisons that do not compare relative effect, but take one arm of a study without considering the comparator are not appropriate; Methods for indirect comparisons are: Bucher, Network meta-analysis, mixed treatment comparisons. Response options: Yes; No.* 6) Data/statistical analysis conducted in the IC or NMA. *Is it adequate? Are assumptions acceptable? When a sensitivity analysis is necessary, was it done correctly? Response options: Yes. When only two alternatives are compared with a common comparator and there is just one study per comparison, the Bucher method is adequate. In the remaining situations, a network meta-analysis is appropriate. Methods must be clearly stated and adequate and assumptions must be reasonable = High reliability; No = Low reliability.*

Topic 13	**Heterogeneity**	
	Corresponding item(s) in new tool	6
AMSTAR 2	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review? *For Yes: There was no significant heterogeneity in the results; OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review. Response options: Yes; No.*	
ROBIS	Was between-study variation (heterogeneity) minimal or addressed in the synthesis?	
Topic 14	**Inconsistency**	
	Corresponding item(s) in new tool	6
Ortega et al	Is there inconsistency? *(Answer this question only if there is a published direct comparison of the treatments being compared). Are results from direct and indirect comparison different? Response options: No. Inconsistency has been tested and is not statistically significant. There are some methods to test inconsistency such as standardised difference or Bayesian methods such as node splitting = High reliability; Yes = Low reliability.*	
Topic 15	**Publication bias**	
	Corresponding item(s) in new tool	7
JBI	Was the likelihood of publication bias assessed?	
AMSTAR 2	If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review? *For Yes: performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias. Response options: Yes; No; No meta-analysis conducted.*	
Topic 16	**Results - general**	
	Corresponding item(s) in new tool	8
CASP	1) What are the overall results of the review? *Consider: If you are clear about the review's 'bottom line' results? What are these (numerically if appropriate)? How were the results expressed (NNT, odds ratio etc.)?*	
ROBIS	1) Were all pre-defined analyses reported or departures explained? 2) Did the reviewers avoid emphasising results on the basis of their statistical significance?	
Topic 17	**Results - precision**	
	Corresponding item(s) in new tool	8
CASP	How precise are the results? *Look at the confidence intervals, if given*	

Topic 18	**Conflict of interest**	
	Corresponding item(s) in new tool	3,7
AMSTAR 2	1) Did the review authors report on the sources of funding for the studies included in the review? *For Yes: Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information but it was not reported by study authors also qualifies. Response options: Yes; No.* 2) Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review? *For Yes: The authors reported no competing interests OR The authors described their funding sources and how they managed potential conflicts of interest. Response options: Yes; No.*	
Ortega et al	Conflict of interest. *Is there any conflict stated? Is the study independent? Response options: Yes, a conflict of interest statement has been made and no manipulation can be observed = High reliability; There is no conflict of interest statement included or a manipulation of results is observed because of an important conflict of interest = Low reliability*	
Topic 19	**Conclusions - internal validity**	
	Corresponding item(s) in new tool	9
JBI	Were recommendations for policy and/or practice supported by the reported data?	
AMSTAR 2	Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review? *For Yes: included only low risk of bias RCTs; OR if RCTs with moderate or high RoB, the review provided a discussion of the likely impact of RoB on the results. Response options: Yes; No.*	
ROBIS	Was the relevance of identified studies to the review's research question appropriately considered?	
Topic 20	**Conclusions - External applicability**	
	Corresponding item(s) in new tool	10
CASP	1) Can the results be applied to the local population? *Consider whether: The patients covered by the review could be sufficiently different to your population to cause concern; your local setting is likely to differ much from that of the review*	
Ortega et al	External validity and applicability. *Were conclusions obtained from the results? Can the results be extrapolated to our patients? Has the relevant alternative for our practice been analysed? Response options: Yes = High reliability; No = Low reliability.*	
Topic 21	**Conclusions - other information required**	
	Corresponding item(s) in new tool	10
CASP	Were all important outcomes considered? *Consider whether: Is there other information you would like to have seen*	
Ortega et al	External validity and applicability. *Were conclusions obtained from the results? Can the results be extrapolated to our patients? Has the relevant alternative for our practice been analysed? Response options: Yes = High reliability; No = Low reliability.*	

Topic 22	**Conclusions - benefits v harms**	
	Corresponding item(s) in new tool	None
CASP	Are the benefits worth the harms and costs? *Consider: Even if this is not addressed by the review, what do you think?*	
Topic 23	**Conclusions- new research**	
	Corresponding item(s) in new tool	None
JBI	Were the specific directives for new research appropriate?	
Topic 24	**Conclusions - overall risk of bias**	
	Corresponding item(s) in new tool	None
ROBIS	Did the interpretation of findings address all of the concerns identified in domains 1 to 4?	

My screening of the abstracts of the 200 most recently published articles found by searching MEDLINE in January 2019 identified 58 systematic reviews using NMA methods for synthesis. The full text of each of these reviews was obtained and reviewed for reporting of any aspect of potential bias. This identified one item which was not already explicitly addressed in the first draft of the tool: the small study effect. This effect was included in AMSTAR 2, within an item about quantitative synthesis (see topic 15 in Table 7.1 above). The comments alongside item 7 were amended to include mention of small study effect as one form of risk of bias within the distribution of studies included in the synthesis.

The feedback received from experts welcomed the increased amount of explanation provided in comments, compared with the existing CASP systematic review tool, but suggested retaining the CASP section names, which had each been changed in initial drafts. Additional comments were received on the wording of individual items. As a result of this feedback, Sections A and B were revised to match the CASP names but Section C remained 'Are the results applicable to answering my own question?', rather than CASP's 'Will the results help locally?'. Amendments were also made to the wording and order of some items, the latter to reflect reverting to the first two CASP section headings.

7.4.3 The new critical appraisal tool

The three pages of the new critical appraisal tool are shown in figures 7.1, 7.2 and 7.3

Critical appraisal tool for Systematic Reviews that use Network Meta-Analysis for synthesis

Section A: Are the results of the review valid?			
Screening questions:			
1 Did the review address a clearly focused and relevant question?	Yes		The review should specify the population studied, the interventions compared, and the outcome(s) considered (PICO). The outcome(s) should be clinically relevant. Ideally the review question should be articulated in a published protocol.
	Partially/unsure		
	No		
Hints: Are the · Population, interventions and outcomes clearly defined? · outcomes clinically relevant?			
Comments:			
2 Did the authors look for the right type of papers?	Yes		The search strategy, inclusion and exclusion criteria should be suitable to identify and select studies that would address the review's question and have an appropriate study design e.g. usually randomised trials for evaluating the effects of interventions.
	Partially/unsure		
	No		
Hints: The right type of papers would · address the review's question · have appropriate study design e.g. usually randomised trials to evaluate the effects of interventions			
Comments:			
Is it worth continuing?			
Detailed questions:			
3 Were all the important, relevant studies included?	Yes		There should be evidence of a comprehensive search strategy, with multiple electronic databases searched; reference lists of included studies checked; searching for grey literature, "unpublished" studies and studies published in languages other than English. Consider whether the search strategy was suitable for identifying studies covering all interventions in the proposed network. There should be a summary of the number of studies identified, included and excluded (with reasons for exclusion that demonstrate consistency with the stated criteria). There should be a statement of any conflicts of interest. Consider whether any such conflict might have influenced the selection of studies.
	Partially/unsure		
	No		
Hints: Consider whether · multiple databases were searched, reference lists were checked, unpublished and studies published in languages other than English were searched for · the search strategy was suitable for identifying studies covering all interventions in the proposed network · a summary of the number of studies identified, included and excluded, is provided · reasons for exclusion are given · authors' conflicts of interest might have influenced the inclusion or exclusion of studies			
Comments:			
4 Did the review's authors do enough to assess the quality of the included studies?	Yes		Each study included in a network meta-analysis should be critically evaluated for risk of bias. There should be details of how this was conducted, including the criteria used and whether it was conducted by two or more reviewers independently.
	Partially/unsure		
	No		
Hints: Each study identified should be critically evaluated for risk of bias. Bias in studies included in any meta-analysis may affect the results.			
Comments:			

Page 1

Figure 7.1 Critical appraisal tool – page 1

5 Was it reasonable to combine the results of the included studies?	Yes		The reason(s) for conducting NMA should be stated (e.g. lack of direct comparison, to add power). There should be adequate similarity between the studies included in the network e.g. characteristics of the patients, concomitant interventions allowed, outcome measurement and definition, length of follow-up, differential loss to follow-up, differences in baseline risks and placebo responses. Explanation for the choice of model used for the meta-analysis (random/fixed effects) should be reasonable.
	Partially/unsure		
	No		
Hints: Consider whether · the included studies were similar enough in design · the choice of methodology used to combine the results was reasonable			
Comments:			
6 Did the authors take appropriate steps to confirm the validity of the network meta-analysis results?	Yes		The assumptions (e.g. homogeneity, transitivity, consistency) made when conducting the network meta-analysis should be stated and assessed to demonstrate whether they are reasonable. Assessment methods should be qualitative and quantitative[1]. If statistical heterogeneity is present, the reason(s) should be considered, and correction made, if possible and if clinically appropriate. If there are available direct comparisons, any inconsistency between indirect and direct comparison results must be analysed and reasons for any identified inconsistency must be considered. Bias in indirect comparisons may be due to bias in the direct comparisons that were used to create the indirect comparison or to lack of clinical similarity between the studies being combined. Subgroup analysis or meta-regression may be done to correct inconsistency but must make clinical sense and relevant data must be available. Sensitivity analysis can assess the robustness of results.
	Partially/unsure		
	No		
Hints: Consider whether · the assumptions for network meta-analysis (e.g. homogeneity, transitivity, consistency) were adequately assessed (qualitatively and quantitatively) · the reasons for and any attempted adjustment for discrepancy from the assumptions are discussed and reasonable · inconsistencies between direct and indirect comparisons are analysed and discussed · sensitivity analysis demonstrated the robustness of the results			
Comments:			
7 Can you identify likely significant bias due to the distribution of included studies?	Yes		The authors should assess the geometry of the network included in the meta-analysis and potential biases related to it. They should also assess the likelihood of publication bias, including the potential effect of small sample size studies and bias due to selective reporting of outcomes. Consider whether any declared conflict of interest might have influenced the inclusion or exclusion of specific treatments (as nodes in the network).
	Partially/unsure		
	No		
Hints: Consider the potential for bias due to · the geometry of the network · inclusion or exclusion of specific treatments as nodes in the network · publication of studies or selective reporting of outcomes · overestimated effects in results of small studies			
Comments:			

[1] Qualitative assessment is usually done by comparing the characteristics of studies
Quantitative statistic/method examples:
Heterogeneity – I^2, Chi squared (χ^2), Tau squared(τ^2), Q test
Consistency/Coherence - node splitting, loop specific comparison, net heat plot

Figure 7.2 Critical appraisal tool – page 2

Section B: What are the results?			
8 Are the results clearly stated?	Yes		The key results in relation to the review question should be clearly presented[2], clinically relevant and consistent with the protocol, if there is one. Results should include estimates of relative and absolute treatment effects with a measure of precision (confidence intervals/credible intervals), probabilities of each intervention being the most effective treatment and comparison with pairwise meta-analysis estimates.
	Partially/unsure		
	No		
Hints: Identify · estimates of relative and absolute treatment effects with a measure of precision e.g. credible intervals (CrI) · probabilities of being the most effective treatment · comparison with pairwise meta-analysis estimates			
Comments and summary of the results:			
9 Were the authors' conclusions or recommendations for policy supported by their results?	Yes		There should be a clear and reasonable link between any conclusions or recommendations and the results of the review. Comment should be made on the robustness and precision of the results, the quality of the included research and the validity of the network meta-analysis. Consider the likely direction of impact on the estimates of effect of any bias that you or the review authors identified and whether its removal would alter the conclusion.
	Partially/unsure		
	No		
Hints: Consider whether · the results are based on good quality studies · the results are likely to have been affected by bias due to the distribution of included studies · the results are sufficiently precise · the validity of the network meta-analysis results has been adequately established · the authors' conclusions are supported by the results			
Comments:			
Section C: Are the results applicable to answering my own question?			
10 Can the results be applied to the population I am interested in?	Yes		If there are significant differences between the participants in the included studies and your own population, consider whether the review's results are applicable. Consider whether all the interventions or outcomes that are important to your question were included in the review.
	Partially/unsure		
	No		
Hints: Consider whether · the population and setting(s) included in the review are sufficiently similar to those you are concerned with · all interventions that are relevant to your question were included · all the outcomes important to you were included			
Comments:			

[2] A guide to interpreting network meta-analysis results can be found on pages 424-426 of www.sciencedirect.com/science/article/pii/S1098301511014045#sec5

Figure 7.3 Critical appraisal tool – page 3

7.4.4 Inter-rater reliability testing

Table 7.2 shows the Fleiss kappa values, the percent exact agreement and agreement within one ordinal degree for all raters for each item in the tool.

Table 7.2 Inter-rater reliability results – new tool

Item number	Fleiss' kappa	Exact agreement	Agreement within one ordinal degree
1	0.211	33.3%	66.7%
2	0.643	66.7%	66.7%
3	0.176	0%	100%
4	0.216	33.3%	66.7%
5	0.029	0%	33.3%
6	0.204	0%	66.7%
7	0.041	0%	66.7%
8	0.103	66.7%	66.7%
9	0.089	0%	100%

A commonly used interpretation of kappa values[12-14], is that the degree of agreement indicated by the kappa value is less than chance (<0), slight (0-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) or almost perfect (0.81-1.00). The Fleiss' kappa values for all raters for each item in the tool ranged from 0.029 (item 5) to 0.643 (item 2). For comparison, the Fleiss' kappa values for all items ranged from 0.03 to 0.69 for ROBIS and from -0.09 to 1 for AMSTAR in IRR testing reported by Buhn et al in 2017[14].

Table 7.3 shows the number of items with Fleiss kappa values in each degree of agreement range for the new tool compared to the 2017 results for AMSTAR and ROBIS.

Table 7.3 Inter-rater reliability test results – new tool, AMSTAR and ROBIS

kappa value	New Tool	AMSTAR[14] 2017	ROBIS[14] 2017
<0	0	2	0
0.00 - 0.20	7	3	10
0.21 - 0.40	1	2	7
0.41 - 0.60	0	2	6
0.61 - 0.80	1	1	1
0.81 - 1.00	0	1	0

(n = number of items with a kappa value within each range)

The IRR testing involved seven raters and me. Six of the seven raters indicated their level of confidence to critically appraise a systematic review incorporating NMA before the exercise, and answered either 1 or 2, with an average score of 1.17. Five raters indicated their level of confidence using the tool, with a range from 2 to 4 and an average score of 3. The sixth rater did not indicate a score but stated,

'I have some more insight than before. But I would probably feel more competent if I was provided with this tool in a workshop training event.'

The seventh rater did not indicate scores but stated,

'My confidence fell because I realised how poor was my understanding of how to assess beforehand.'

General feedback included,

'Tool is very helpful …gave me some confidence in highlighting the papers that were poor'

'I like the tool and its similarity to the CASP means that it's a familiar format and easy enough to follow. I thought I had some knowledge of systematic reviews with network meta-analyses because I have been on NICE committees and they use NMAs a lot. However, I really needed the guidance you provided in the tool as I'm not too familiar with how they're conducted and what would be indicators for reviewing them. I initially thought that the guidance was clunky and a bit repetitive and that it could be removed from the tool and placed on an accompanying sheet. But I found that it's really handy to have at each stage…'

'I found this a difficult exercise because of my lack of understanding of Network Meta Analysis. The tool provides a good framework for assessing NMA reviews but without understanding of the complex methodology and statistics it is difficult to be critical of these elements and assess their relevance to clinical practice.'

Feedback also suggested amendments to some specific items, including:

- that conflict of interest (item 3) should be considered separately from other issues relating to inclusion of studies in a review because it conflates evidence for poor or biased searching with factors that might cause this.
- that the appraiser's and authors' opinions as to whether it was reasonable to conduct synthesis (item 5) are separate issues and the item conflates the two.

7.5 Discussion

7.5.1 Summary of the findings of this chapter

I identified 24 distinct items from the included source tools and condensed these to 10 items for the new critical appraisal tool. Five items identified from the source tools were not specifically covered by any items in the new tool.

Two related source items that were not covered by any items in the new tool were minimising errors in extraction of data and ensuring that the data extracted were reported in an original study and have not been extrapolated or manipulated in any way. Two of the source tools mention minimising errors in data extraction in general terms[6, 8] and their guidance mentions various ways to achieve this. Only AMSTAR 2[7] more specifically states that two reviewers should reach consensus on data extraction. Ensuring that data is the original was only specified by Ortega et al[1]. Item 4 in the new tool guides the user to the value of involving two reviewers, in that instance for assessing the quality of studies and I decided that this was a sufficient prompt without including an additional item on data extraction. Whilst including data in synthesis that were not original trial data would certainly introduce risk of bias, I decided that this specific detail did not warrant including an additional item but did not fit within any other item.

The remaining three items from the source tools that were not covered by any items in the new tool were each an aspect of considering the overall conclusions of a review. The CASP tool[5] includes an item on considering whether the reported benefits outweigh the harms and costs. Since that should be a universal assessment in considering application of research findings and was not included in any other source tool, I decided not to include it. The JBI tool[6] includes an item on whether directives for new research were appropriate. This covers a specific example of the conclusions that the review authors might make but it was not included in any other source tool and can be considered as addressed, not specifically but in general terms, as part of item 9 in the new tool. Finally, the ROBIS tool[8] includes a tool-specific item so could not directly

correspond to an item in the new tool, but I concluded that its intentions are addressed by the guidance comments for item 9.

The three components identified in the review reported in Chapter Six that specifically address the use of NMA methodology (connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made) are covered in items 5,6 and 7 of the new tool, with a focus on assessing the reporting relating to these items.

Based on evidence of the unreliability of overall scoring in critical appraisal tools[15, 16], the tool was constructed so that users assessing a review should consider each item to assess the likelihood of bias and consider whether that might affect the validity of the reported results, rather than use the tool to calculate an overall score.

No direct comparison between two or more source tools was identified in my review described in Chapter Six, so I chose to compare the IRR test results for my new tool with the results of a 2017 comparison between AMSTAR and ROBIS[14]. Several differences in the IRR testing methods should be considered. AMSTAR responses were dichotomised into "yes" and "all other responses", and ROBIS responses were also grouped (e.g. combining "yes" and "probably yes"), which might result in higher kappa values than if each response was treated separately. The authors of the 2017 comparison stated that their AMSTAR results were lower than reported in other articles[12, 13]. Those earlier reports of IRR testing of AMSTAR also used dichotomised responses, but it is unclear whether those higher results were calculated from ratings scored independently or following discussion between raters, as noted in Chapter Six. Also, appraisers were not provided with any training in using my new tool prior to the IRR test, in contrast to the ROBIS and AMSTAR IRR tests, in which appraisers had prior experience of using AMSTAR and were familiarised with ROBIS.

The percent exact agreement for items in the new tool was low except for item 2, concerning identification of the appropriate studies. The percent agreement within one ordinal degree,

however, was substantial, with the exception of item 5, concerning the appropriateness of synthesis. Agreement within one ordinal degree is a nearer comparison to the method used in the inter-rater reliability testing of ROBIS and AMSTAR, in which responses were combined into categories[14]. Substantial agreement but a low kappa value for an item may be due to a skewed distribution of responses i.e. if a high number of responses are the same[13].

All raters who provided a score for their confidence to critically appraise a systematic review incorporating NMA rated this as higher with the new tool compared to before the exercise. One other rater indicated that using the tool had improved their awareness of gaps in their prior knowledge.

Feedback comments from the raters were generally positive. There were indications that it would be helpful to have training in use of the tool before using it and comments on specific items reinforced this, because the purpose of a question and the supporting information had sometimes been misinterpreted.

7.5.2 Strengths and limitations of the research presented in this chapter

This new tool was constructed using a combination of established methods plus additional checking for omissions. Items and accompanying guidance were initially sourced and cross-referenced from the most current and most widely used related tools that have been constructed for appraisal of systematic reviews and/or NMA. This approach of initially sourcing items from other tools and resources and then reducing to a smaller number in the final tool has been used previously in the development of critical appraisal tools, including by the developers of one of the most widely used tools, AMSTAR[7, 17]. Content development of the new tool also used input from experts, which has been used in development of previous critical appraisal tools, including AMSTAR and ROBIS[8]. The content of the new tool was also checked for omissions against both the relevant PRISMA reporting statement and reported methods of minimising bias or any

other comments on bias in a recent set of published systematic reviews using NMA methods and amendment made to include one aspect not already covered.

This is, as far as I have been able to identify, the first critical appraisal tool designed specifically for appraising SRs using NMA methods for synthesis, with guidance provided within the tool alongside each item and with a similar number of items to the most widely used appraisal tools for systematic reviews.

The validity of a tool is the extent to which it measures what it intends to. Aspects of validity include face validity (the content appears appropriate to what it intends to measure), content validity (the content comprehensively covers what it intends to measure) and construct validity (the tool can be demonstrated to measure what it intends to). The former two aspects are usually based on expert consensus or research evidence, the latter on formal testing e.g. of hypotheses. In addition to validity, a tool should be able to demonstrate reliability (it can be demonstrated to produce repeatable results) and feasibility (it can be used within the limits of the resources available, such as the appraisers' experience and the time taken to complete).

One limitation of face and content validity of the new tool is that its content reflects the judgement of a single person, even though the final content was influenced by feedback obtained from two experts in systematic reviews and critical appraisal. Set against this, the process of development has been reported and the content only includes items from sources that have reported how they established face and content validity. Contribution from a wider group of experts and reaching consensus from that group, e.g. using a Delphi process, may have altered the final content and format of the tool and might lead to greater confidence in its face and content validity.

A second limitation is that, although the source tools used included ones chosen because they have already reported evidence for their face and content validity and IRR testing, other source

tools used were the CASP and JBI tools, for which such reports were not identified. The new tool however, included no item sourced only from either the CASP or JBI tool.

No testing of construct validity has been undertaken so far.

Critical appraisal is a systematic approach to assessing the quality of a study. The aims of using a critical appraisal tool for systematic reviews could be defined as to optimise the ability of any specific rater to assess a review and to standardise the rating process if more than one rater is assessing the same review e.g. for the purpose of identifying evidence for an overview of reviews or a clinical guideline. The former is not assessed by IRR testing. Recommended practice for the latter is to have at least two raters and a process for reaching consensus where there is difference in ratings, which means a high degree of agreement using the tool independently is not critical to achieving the aim. Reliability testing of critical appraisal tools is also limited by the absence of a gold standard to compare any tool with. Caution should therefore be exercised in judging critical appraisal tools simply based on their IRR results.

The choice of systematic reviews for my IRR testing was limited to one clinical topic, which restricts the extent to which conclusions can be drawn about wider applicability. No training in use of the tool or calibration was undertaken with raters, so it is not possible to assess what effect those measures would have on reliability. Although limiting the content of the tool to fewer items was one of my preferred characteristics, in order to support feasibility, time taken to complete the tool was not recorded, which means that there are no data to determine whether it is quicker or slower to complete than other tools.

7.5.3 Implications of the findings of this chapter

The IRR test results show better than chance agreement for each item but mainly in the 'slight' to 'fair' range, based on a commonly used interpretation[14]. The percent exact agreement and percent agreement within one ordinal degree results suggest that the inclusion of the 'partially/unsure' response option is unhelpful, at least as regards reliability between raters. This

response option was adopted from the format of the CASP tools. On reflection, whether a rater is unable to find relevant information in a review or identifies that some criteria contained in an item are not met, both are likely to elicit the same choice of response. However, these two situations have different implications for drawing conclusions about the likelihood of bias and the validity of reported results so it would be better not to amalgamate them into one response. The developers of AMSTAR 2 drew the same conclusion:

'If no information is provided to rate an item, the review authors should not be given the benefit of doubt and the item should be rated as a "No." We have provided a "partial Yes" response in some instances where we considered it worthwhile to identify partial adherence to the standard.'

Consultation with a wider group of experts on the content of the tool would lead to greater confidence in its face and content validity. The new tool uses the CASP SR tool format, for reasons explained in Section 7.1, but no published evidence was found of the time taken to complete the CASP SR tool and an assumption was made that the published evidence for tools with a similar number of items could be extrapolated. This assumption requires confirmation by further testing of the new tool.

Further testing of construct validity, reliability and feasibility of this new critical appraisal tool should be undertaken after any revision resulting from consultation with other experts and modifying the response options to provide distinction between whether only some criteria for an item are met or whether information is missing from a review that prevents a conclusion being reached on an item. A training package in use of the tool should also be developed.

7.6 Chapter summary

- A new critical appraisal tool was constructed for appraising systematic reviews using NMA methods for synthesis
- Construction was undertaken using a combination of established methods (existing sources and expert opinion) plus additional checking for omissions
- The tool consists of ten items corresponding to nearly all the distinct items contained in the most current and most widely used related tools
- Guidance for users is provided within the tool alongside each item
- Inter-rater reliability testing of the tool was conducted by eight raters assessing three systematic reviews
- Results showed better than chance agreement for each item in the tool, but mainly in the 'slight' to 'fair' range
- Further testing of construct validity, reliability and feasibility of this new critical appraisal tool should be undertaken after any revision resulting from consultation with other experts and modifying the response options to provide distinction between whether only some criteria for an item are met or whether information is missing from a review that prevents a conclusion being reached on an item.
- A training package in use of the tool should be developed

7.7 Chapter references

1. Ortega A, Fraga MD, Alegre-del-Rey EJ, et al. A checklist for critical appraisal of indirect comparisons. IntJClinPract 2014; 68: 1181-1189. 2014/10/02. DOI: 10.1111/ijcp.12487.

2. Shamliyan TA, Choi JY, Ramakrishnan R, et al. Preventive pharmacologic treatments for episodic migraine in adults. Journal of General Internal Medicine 2013; 28: 1225-1237. 2013/04/18. DOI: 10.1007/s11606-013-2433-1.

3. Jackson JL, Cogbill E, Santana-Davila R, et al. A Comparative Effectiveness Meta-Analysis of Drugs for the Prophylaxis of Migraine Headache. PLoS One 2015; 10: e0130733. 2015/07/15. DOI: 10.1371/journal.pone.0130733.

4. He A, Song D, Zhang L, et al. Unveiling the relative efficacy, safety and tolerability of prophylactic medications for migraine: pairwise and network-meta analysis. J Headache Pain 2017; 18: 26. 2017/02/22. DOI: 10.1186/s10194-017-0720-7.

5. Critical Appraisal Skills Programme (2017) CASP Systematic Review Checklist. (12 9 19), https://casp-uk.net/wp-content/uploads/2018/01/CASP-Systematic-Review-Checklist_2018.pdf (accessed 12 September 2019).

6. Joanna Briggs Institute. Joanna Briggs Institute Checklist for Systematic Reviews and Research Syntheses https://joannabriggs.org/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Systematic_Reviews2017_0.pdf (2017, accessed 12 September 2019).

7. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017; 358: j4008. 2017/09/25. DOI: 10.1136/bmj.j4008.

8. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016; 69: 225-234. 2015/06/21. DOI: 10.1016/j.jclinepi.2015.06.005.

9. Higgins JP, Lane PW, Anagnostelis B, et al. A tool to assess the quality of a meta-analysis. Res Synth Methods 2013; 4: 351-366. 2013/12/01. DOI: 10.1002/jrsm.1092.

10. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14: 417-428. 2011/06/15. DOI: 10.1016/j.jval.2011.04.002.

11. Ades A.E. CDM, Reken S., Welton N.J., Sutton A.J., Dias S. NICE DSU Technical Support Document 7: Evidence Synthesis Of Treatment Efficacy In Decision Making: A Reviewer's Checklist http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD7-reviewer-checklist.final_.08.05.12.pdf (2012, accessed 12 October 2019).

12. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007; 2: e1350. 2007/12/27. DOI: 10.1371/journal.pone.0001350.

13. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009; 62: 1013-1020. 2009/02/24. DOI: 10.1016/j.jclinepi.2008.10.009.

14. Buhn S, Mathes T, Prengel P, et al. The risk of bias in systematic reviews tool showed fair reliability and good construct validity. J Clin Epidemiol 2017; 91: 121-128. 2017/07/12. DOI: 10.1016/j.jclinepi.2017.06.019.

15. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. Jama 1999; 282: 1054-1060. 1999/09/24. DOI: 10.1001/jama.282.11.1054.

16. Greenland S and O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics 2001; 2: 463-471. 2003/08/23. DOI: 10.1093/biostatistics/2.4.463.

17. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology 2007; 7: 10. 2007/02/17. DOI: 10.1186/1471-2288-7-10.

Chapter Eight

Discussion and conclusion

8 Discussion and conclusion

8.1 Discussion

8.1.1 Trends in use and impact of network meta-analysis

Network meta-analysis (NMA) is a relatively new methodology. My review, reported in Chapter Three, demonstrated that publications of systematic reviews using NMA methods for synthesis had increased substantially from 2009 to 2012, rising from only 7 during all the years before 2008 to an estimated 90-100 published in 2012. My search of MEDLINE in early 2019, as part of the development process for the content of the new critical appraisal tool, suggested that 180-200 such systematic reviews were published in 2018, with a total of more than 1000 now available in the literature. Using my estimate for 2018, the increasing publication of NMA is illustrated in Figure 8.1. The solid line indicates the number of reviews reported in Chapter Three for each complete year (Figure 3.3); the dashed line extends to the estimated number for 2018.

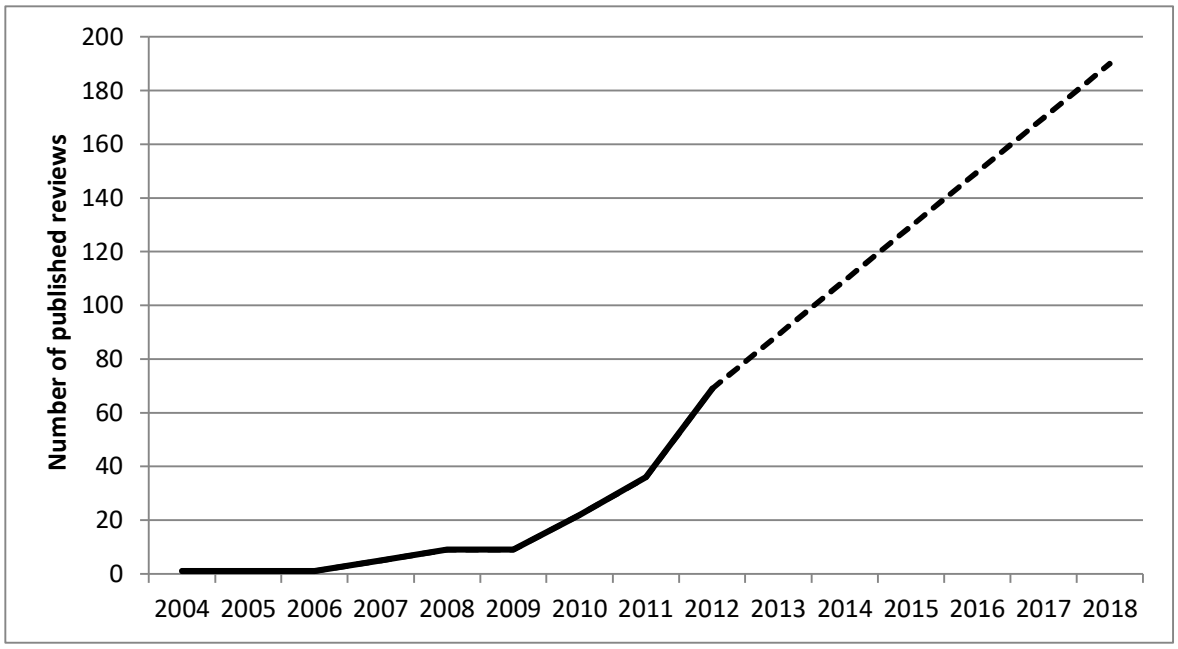

Figure 8.1 Estimate for the number of published NMAs to 2018

The ability of NMA methods to compare simultaneously more than two interventions, to compare interventions that have not been directly compared in trials and to increase the precision of the estimate of effect size is likely to result in continued increased use of these methods. Caldwell explained one reason for the increased popularity of NMA by illustrating the limitations of using various pair-wise meta-analysis options to estimate the relative effectiveness of more than two interventions[1].

End-users might wish to search for NMAs specifically for methodological research such as that reported in this thesis but also if they are seeking evidence to answer a clinical question that estimates the comparative effectiveness/harms of more than two interventions or when planning new research and wishing to identify if an NMA already exists. The challenges to identifying articles reporting use of pair-wise meta-analyses, listed by Lee et al in 2001[2], have been substantially resolved over the past two decades, but similar challenges continue to affect attempts to identify articles reporting the use of NMA, because there is still no MeSH term that will specifically identify NMAs, and databases still do not provide any alternative method to limit searches to specifically identify NMAs rather than pair-wise meta-analyses. However, it is clear that there is increasing use of NMA, with a trajectory of increasing publications that is similar to that for publications reporting use of pair-wise meta-analysis between 1980 and 2000[2]. This makes it important for those conducting and using NMAs to be able to decide when it is appropriate to do them and to have clear reporting of them which can be critically appraised by non-experts.

The most significant impact of evidence synthesis on health care is likely to come from the use of the evidence generated by these research projects in national clinical guidelines. Although NMA methods are far less common than pair-wise meta-analysis in systematic reviews, my review of NICE clinical guidelines (Chapter Five) demonstrated that they were used or considered for nearly one quarter of those guidelines in 2015 and 16, showing that evidence

produced using NMA methods is influencing recommendations for the NHS. My finding that recommendations in the guidelines were more often based on results of NMA conducted de novo by the responsible Guideline Development Group, rather than on results from previously published systematic reviews using NMA methodology, raises uncertainty about the usefulness of those systematic reviews. This highlights the importance of the issues described by Chalmers and Glasziou in 2009[3], including choosing the wrong question for research. Although concerns about their usefulness, quality and effectiveness have prompted debate on whether systematic reviews and meta-analyses are still useful research[4, 5], both sides of this debate acknowledge the role of systematic reviews to inform clinicians, highlight uncertainties and identify what further research is needed. Building on the established use of systematic review and meta-analysis to inform the design of new trials, it seems likely that there will be increasing examples of the results of new trials being included in continuously updated, or living, meta-analyses[5]. Since multiple interventions can be compared simultaneously using NMA, there have been proposals to create and maintain continuously updated meta-analyses using these methods[6] and, in 2020, a substantial project is underway for COVID-19 related interventions[7].

8.1.2 Reporting standards for network meta-analysis

One of the causes of research waste highlighted by Chalmers and Glasziou in 2009 is poor reporting[3]. This adds to research waste because it creates uncertainty regarding the likelihood of bias and the validity of the results of both primary research and evidence synthesis. If recommended good practice in research reporting is followed, this should also limit the inclusion of lower quality research in synthesis by meta-analysis, whether pair-wise meta-analysis, NMA or other synthesis methods are used[8].

Efforts to improve the quality of reporting of systematic reviews have been based on defining and promoting internationally recognised standards. There is evidence that publication of such reporting standards results in improved quality of reporting, based on comparisons before and

after the publication of both QUOROM[9] and the PRISMA Statement, and also that reporting to PRISMA Statement standards is strongly associated with higher study quality, as assessed by a widely used CAT[10]. Panic et al found that endorsement of the PRISMA Statement by journals in their instructions for authors was associated with improved quality of reporting, regardless of whether or not the authors declared following the Statement, and again was associated with higher study quality[11].

In the course of my review, reported in Chapter Three, I observed that the reporting of systematic reviews using NMA methods was inconsistent; with details of specific choices in using the methodology and the assumptions made often being unclear or not stated. Reporting standards for use of NMA methods were subsequently published as a PRISMA Extension Statement[12] and, as reported in Chapter Four, I found evidence suggesting that these standards would be supported and implemented in my survey of authors that began before that Extension Statement was published. Subsequently, in the sample of systematic reviews published in 2018 that I checked as part of the development process for the content of the new critical appraisal tool, I found that 37 of the 58 reviews stated that they had followed the PRISMA reporting guideline and 25 of those 37 specifically referenced the NMA extension. Improved quality of reporting supports critical appraisal, by making relevant details readily available to the person doing the appraisal.

8.1.3 Critical appraisal of network meta-analysis

My survey (Chapter Four) found that some authors who have used NMA methods had reservations about the ability of publishers and end-users to critically appraise the quality of evidence produced using these methods. My early 2018 review (Chapter Six) strengthened the evidence base for addressing those concerns by showing that none of the most widely-used critical appraisal tools for systematic reviews contained content specifically relevant to appraising synthesis using NMA methodology but that three published tools had been developed

specifically to include content for appraising the use of NMA methods. There was common ground between these three tools as to certain components that critical appraisal of NMAs should consider: connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made. However, none of these three tools matched the characteristics that would make a tool suitable for end-users without specialist statistical knowledge of NMA methodology. This led me to construct a new tool, using findings from my review that related to the methods used for development and for drawing face and content validity from existing sources (Chapter Seven).

Training

Although the new tool looks promising in my evaluation of it (Chapter Seven), there are ways in which its use can be optimised, bearing in mind that it is intended to support appraisal by users without specialist statistical knowledge of NMA methodology. Training and guidance provided within the tool should seek to familiarise users with certain concepts and features that are key to interpreting the three common, technical components identified from the published NMA critical appraisal tools: connectedness of the network, performance of the statistical analysis and assessment of validity of the assumptions made. Additionally, users need to understand that there are both statistical and content knowledge elements to assessing the validity of assumptions.

Such training might cover several aspects of connectedness of the network that can be assessed from a network diagram and may require content knowledge relating to the topic of the NMA. These include the choice of interventions to include or exclude as nodes in the network, the risk of small studies overestimating effects and consideration of how many intermediate comparators there are in a network between any two interventions that may be of particular interest to the user. The latter is important because confidence in the estimates of difference in the effects of the two interventions should reduce if they have more than one intermediate comparator[13].

The target users may also need training and guidance to help them to assess detailed aspects of the performance of the statistical analysis because they are unlikely to have the specialist statistical knowledge needed to help them know what to look for and how to judge what they find.

Training in assessment of validity of the assumptions made should cover the concepts of homogeneity, transitivity and consistency and the methods by which each can be assessed, both statistically and based on content knowledge. Some of this assessment is again dependent on statistical analysis, so training should highlight what aspects of reporting should be used to make an assessment.

There is relevant information for interpretation of aspects of reporting in the article by Jansen et al[14], which is referenced within the new tool, but users may benefit from having this explained during the training.

In the current tool, the importance of a published protocol for the systematic review and NMA and the adherence of the review team to it are included within items 1 and 8. These are worded in accordance with the PRISMA statement[15], to allow for the absence of a published protocol but, as with any other criteria in the tool, the acceptability of not meeting a particular standard is likely to vary over time as something becomes standard practice and, consequently, its absence becomes more important when assessing quality. Likewise, it is increasingly an expectation that a review will have been registered in a registry, such as PROSPERO[16] and have had a protocol published in advance. There should be emphasis on adherence to the protocol, rather than just its existence, so revised wording in the tool could better reflect both the change in expectation and the required emphasis on adherence. Users might also benefit from understanding the specific issues that should be included in a protocol for a systematic review using NMA methods, as described by Chaimani et al[17].

Alternatives to creating a new tool

Given that Ades et al[18] state that the only items that need considering exclusively when assessing the use of NMA methods compared to assessing a pair-wise meta-analysis, are '*connectedness of networks, inconsistency and software implementation*', it could be argued that, to create a suitable CAT, an addition could simply be made to an existing tool, such as AMSTAR or the CASP SR tool, to cover those three issues. My findings were that the AMSTAR tool has good evidence of inter-rater reliability, therefore, given that I could not find published inter-rater reliability results for the CASP SR tool, the AMSTAR tool might be considered more suitable to modify than the CASP SR tool. However, although AMSTAR could be updated to cover the exclusive aspects of using NMA methods, this should probably be done using the same methodology that was used for the AMSTAR tool itself. In regard to the CASP SR tool, I could not find published details of the methods used to develop it but, aside from the need to add content to cover use of NMA methods, there were other aspects of its content that do not reflect most current evidence regarding best practice for the conduct and reporting of systematic reviews. This supported my decision to undertake a full synthesis from the sources I identified in order to create a new version of the CASP SR tool, including content appropriate to assessing the use of NMA methods, rather than simply inserting one or more NMA specific items to the existing CASP SR tool.

Critical appraisal for generalist users of research

When CASP evolved in the 1980s to promote awareness of research evidence amongst health decision makers and equip them with critical appraisal skills, the first Cochrane Centre had not yet been founded in Oxford in 1992, and the creation of NICE and its first clinical guidance, a rapid assessment of Zanamivir[19] in 1999 was still more than a decade away. Nearly 40 years on, with more than 7500 published, full Cochrane Reviews and, as of November 2019, 206 NICE clinical guidelines available to health decision makers, along with numerous other forms of

evidence synthesis, this raises the question of whether there is still a need for health decision makers to learn critical appraisal skills and use CATs. However, it seems clear that this is still needed, recalling that critical appraisal has been described as,

'an objective, structured approach that results in a better understanding of a study's strengths and weaknesses'

The promotion of critical appraisal and, indeed, of evidence-based practice generally was a drive to move away from relying on expert opinion to influence decision making towards decision making being based on research evidence. If critical appraisal is only conducted by those involved in producing evidence synthesis, it could be argued that we will have only changed which experts' opinions determine decision making. Continuing to promote training in critical appraisal skills and availability of user-friendly CATs with content derived from the best available evidence of good research conduct and reporting allows the widest range of health decision makers to be able to use this objective, structured approach when considering the evidence available to inform their decisions.

On the other hand, even when a CAT exists and is used, there is still uncertainty as to how well generalist end-users can critically assess aspects of the methods of complex research such as NMA and interpret the results produced using these methods in systematic reviews, compared to the more widely known and standardised methods used for reviews with pair-wise meta-analysis. Without specialist statistical knowledge in use of NMA methods, these users will be unable to assess detailed aspects of the performance of the statistical analysis and have to rely more on aspects of reporting. NMA results have been found to be reported '*heterogeneously*'[20]. Many end-users will be less familiar with these many and varied forms of reporting NMA results than with pair-wise meta-analysis results. Furthermore, although efforts have been made to provide guidance on presenting and interpreting results of NMA[14, 21], new approaches continue to be published[22]. The PRISMA Extension Statement covering the use of NMA methods includes recommendations regarding the presentation of results but it remains to be seen whether that

has resulted in more consistent reporting. The new tool was produced with the aim of assisting generalist end-users, but future, construct validity testing would help in identifying whether it does support them in accurately assessing the quality and validity of systematic reviews using NMA methods.

Other approaches to quality assessment of evidence produced using NMA methods

As described above, there are several existing CATs and, my new one for assessing the overall quality of systematic reviews using NMA methods, but it should also be noted that other approaches have been developed to assess confidence in the evidence produced using NMA methods.

For example, in 2014, the GRADE Working Group reported guidance for establishing the quality of treatment effect estimates obtained from NMA[23]. Their approach involved rating the quality of each direct and indirect effect estimate for each pairwise comparison within the NMA and then rating the NMA effect estimate for each pairwise comparison. In 2018, the GRADE Working Group published recommended modifications to their 2014 guidance with a view to making the process more efficient, acknowledging that the original approach,

'may appear onerous in networks with many interventions'[24].

In 2014, Salanti et al published a modification of the GRADE guidance, in particular, drawing a distinction between rating the effect estimates for each pairwise comparison and rating the ranking of all the interventions within a network[25]. More recently, in 2020, Salanti and others published a new approach: Confidence in Network Meta-Analysis (CINeMA)[26]. They stated that uptake of the earlier 2014 GRADE system and the one by Salanti et al in the same year had been limited by,

'the complexity of the methods and the lack of suitable software'.

CINeMA is also based on the GRADE framework but instead of considering direct and indirect evidence for each pairwise comparison separately, it considers the impact of every study in the network. A web-based application makes it easy to apply to even large networks[27].

A further approach, 'threshold analysis', has been developed specifically to assess confidence in NMA results used in guideline development. The authors argue that their approach is needed because GRADE approaches do not assess the influence of the NMA evidence on a resulting recommendation:

'Threshold analysis quantifies precisely how much the evidence could change (for any reason, such as potential biases, or simply sampling variation) before the recommendation changes, and what the revised recommendation would be. If it is judged that the evidence could not plausibly change by more than this amount, then the recommendation is considered robust; otherwise, it is sensitive to plausible changes in the evidence.' [28]

The above approaches are primarily intended to be used by authors of systematic reviews and guideline developers (to rate confidence in an estimate of effect or confidence that such an estimate adequately supports a specific recommendation) but are relevant to understand by those undertaking critical appraisal and can be used to complement use of a CAT.

8.1.4 Limitations of the research reported in this thesis

My programme of research for this thesis was conducted over eight years in the 2010s, making some elements of it less current in 2021 than others. However, to seek to overcome this, I have attempted to provide updated context for the earlier elements. I conducted the entire programme of research working mainly in isolation, rather than in collaboration with a team of researchers with a range of expertise in systematic reviews, NMA and other statistical methods, although I consulted some experts, as indicated in various chapters. My personal perspective and subjective decision-making are therefore likely to have impacted on various aspects of the research to a greater extent than if I had conducted it in collaboration with a wider team of

people with a range of expertise. Other specific limitations are detailed in each Chapter. For example, the reviews described in Chapters Three, Five and Six were conducted by myself alone, meaning that I may have missed some relevant studies or extracted some data incorrectly. The inclusion criteria used in my review of NMAs (Chapter Three) risked failing to identify additional data that would have added to the total number of reviews and risked introducing bias in the selection of included reviews. That review also did not identify when researchers working on published systematic reviews had considered undertaking NMA but had decided not to proceed with it because it would not be appropriate given the studies they identified for inclusion in their reviews. In my survey of authors (Chapter Four), the respondents formed a minority of the authors approached and the views of non-respondents are unknown, but there was no apparent bias in those responding in terms of year of publication or in the number or subject of reviews published. However, I did not assume that the responses were representative of all authors, all content or methodology experts nor, even, of my selected cohort of authors but instead sought to identify 'leads' that I could explore further in subsequent elements in my programme of research. I included the systematic review tools of CASP and JBI in my review of critical appraisal tools (Chapter Six) despite not finding any published details of their construction or validation. The lack of those details, compared with the other tools identified, makes it difficult to assess the level of confidence that should be placed in the validity of their content, but I have explained the reason for choosing to include them in my review and my restricted use of their content because of these limitations.

The final element of my research was developing the new CAT for systematic reviews using NMA methods (Chapter Seven). I chose to construct a revised version of the CASP SR CAT, for reasons explained in Chapter Seven and earlier in this chapter, but I acknowledge that this was a subjective decision and was not the only justifiable option. Involvement of a wider team of people with relevant expertise might have led to a different decision about which SR CAT to revise, or even whether there was a need for a new CAT at all. I did, however, identify reasons

why the CASP SR CAT could be updated and, as a widely-recognised and used CAT, there is a strong case that it should be updated. My new CAT could contribute to that.

The development of the new CAT was done based on my review of source material, including other CATs and recently published NMAs, and with feedback from a small group of experts. As with all development projects like this, contribution from a wider group of experts (perhaps with the use of a process such as Delphi to reach consensus) might have altered the content and format of the new tool and might lead to greater confidence in its face and content validity. It is also possible that my personal perspective, which was focused on undertaking research into the use of NMA methods, led to decisions on the inclusion of items with an emphasis on the use of NMA methods without appropriate balance with other elements of the overall SR synthesis process. For example, the importance of both methods and content expertise in developing a protocol for a SR could have been more strongly reflected.

Construct validity testing of the new tool was not undertaken and, without that, it remains uncertain whether the new CAT accurately measures what it aims to: the quality and validity of SRs using NMA methods, or indeed whether it is possible for a CAT to support a generalist end-user to do so. Inter-rater reliability testing of the new tool showed better than chance agreement, but mainly only in the slight to fair range. These modest results may be attributable to lack of training before the testing and the lack of distinction in the response options between whether only some criteria for an item are met or whether information is missing from a review that prevents a conclusion being reached on an item. These aspects might also be improved by revisions to the content in the CAT, which might arise from contributions from a wider group of experts. Feasibility was only assessed in the form of raters' confidence and their general feedback but not in terms of the time taken to complete assessments of reviews using the tool, which would be needed to demonstrate that it takes a similar time to complete as other CATs with a similar number of items.

These limitations should also be considered against the backdrop that although the use of NMA methods can overcome some limitations of PWMA, the design and conduct of NMAs requires multidisciplinary input from expert methodologists and clinical topic experts. This raises the need for further research to clarify whether end-users who do not have specialist statistical knowledge can accurately assess the quality and validity of evidence produced in systematic reviews using NMA methods even with an optimised CAT for such studies.

8.2 Recommendations for further research

The research findings presented in this thesis lay important foundations for further research. My recommendations for this are:

- Consultation with a wider group of experts on whether the new critical appraisal tool should be developed further, or whether the content identified during its development would be better used to contribute to modifying another tool or to construct an alternative new tool.
- Consultation with a wider group of experts on updating the content and format of the CASP SR critical appraisal tool, including e.g., appropriate emphasis on adherence to the SR protocol and consideration of modifying the response options.
- Construct, reliability and feasibility validity testing of any new/revised tool, after it has been optimised by consultation with a wider group of experts.
- Assessment of how accurately end-users without specialist training in use of NMA methods can critically appraise use of these methods in systematic reviews and interpret the results produced using these methods.
- Monitoring of future trends in production and use of systematic reviews using NMA methods to identify the need to update content/guidance on their appraisal.

- Development and evaluation of strategies to improve the clarity of reporting of results produced using NMA methods.
- Optimisation of the methods used in the development and validation of critical appraisal tools, perhaps including a consensus on minimum standards for these.
- Assessment of the quality and consistency of reporting of systematic reviews using NMA methods following publication of the PRISMA Extension Statement for NMA methods.

8.3 Conclusions

My research work took place over a period of years in the 2010s during which NMA methods went from a sparingly used and newly evolving synthesis method to an established method used in several hundred published systematic reviews per year, with dedicated guidance from organisations such as Cochrane, GRADE and PRISMA and an influence on recommendations in national guidelines. My research has shown their growing importance through the last decade, and elements of this thesis have already been published in peer-reviewed journals[29, 30] and used by other publications, including the PRISMA Extension Statement to cover the reporting of NMA.

However, my research has also shown how the development of critical appraisal tool content to address the use of NMA methods in systematic reviews has lagged behind the growth in the use of NMA methods. Such content has started to appear and the tool I developed is the first to present that content within the format of a CASP CAT, one of the most user-friendly and widely recognised formats, with content aimed at users without specialist statistical knowledge in the use of NMA methods. Further research is now needed, including consultation with a wider group of relevant experts, to determine how this part of my research work might be optimised or used to contribute to the content and format of an alternative new or revised tool for the 2020s.

8.3 Chapter References

1. Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. *Systematic reviews* 2014; 3: 109. 2014/10/01. DOI: 10.1186/2046-4053-3-109.

2. Lee WL, Bausell RB and Berman BM. The growth of health-related meta-analyses published from 1980 to 2000. *Eval Health Prof* 2001; 24: 327-335. 2001/08/29. DOI: 10.1177/01632780122034948.

3. Chalmers I and Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet (London, England)* 2009; 374: 86-89.

4. Annane D, Jaeschke R and Guyatt G. Are systematic reviews and meta-analyses still useful research? Yes. *Intensive Care Med* 2018; 44: 512-514. 2018/04/18. DOI: 10.1007/s00134-018-5102-3.

5. Moller MH, Ioannidis JPA and Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Med* 2018; 44: 518-520. 2018/04/18. DOI: 10.1007/s00134-017-5039-y.

6. Vandvik PO, Brignardello-Petersen R and Guyatt GH. Living cumulative network meta-analysis to reduce waste in research: A paradigmatic shift for systematic reviews? *BMC Med* 2016; 14: 59. 2016/03/31. DOI: 10.1186/s12916-016-0596-4.

7. Boutron I, Chaimani A, Meerpohl JJ, et al. The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19 Pandemic. *Annals of Internal Medicine* 2020 2020/09/16. DOI: 10.7326/M20-5261.

8. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.0 [updated July 2019]. In: Higgins JPT, Thomas J, Chandler J, et al. (eds): Cochrane, 2019 https://www.training.cochrane.org/handbook (accessed 18 October 2019).

9. Delaney A, Bagshaw SM, Ferland A, et al. A systematic evaluation of the quality of meta-analyses in the critical care literature. *Critical Care* 2005; 9: R575-582. 2005/11/10. DOI: 10.1186/cc3803.

10. Tunis AS, McInnes MD, Hanna R, et al. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? *Radiology* 2013; 269: 413-426. 2013/07/05. DOI: 10.1148/radiol.13130273.

11. Panic N, Leoncini E, de Belvis G, et al. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS One* 2013; 8: e83138. 2014/01/05. DOI: 10.1371/journal.pone.0083138.

12. Hutton B, Salanti G, Caldwell DM, et al. 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: 777-784. 2015/06/02. DOI: 10.7326/M14-2385.

13. Ortega A, Fraga MD, Alegre-del-Rey EJ, et al. A checklist for critical appraisal of indirect comparisons. *IntJClinPract* 2014; 68: 1181-1189. 2014/10/02. DOI: 10.1111/ijcp.12487.

14. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health* 2011; 14: 417-428. 2011/06/15. DOI: 10.1016/j.jval.2011.04.002.

15. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700. 2009/07/23. DOI: 10.1136/bmj.b2700.

16. PROSPERO International prospective register of systematic reviews, https://casp-uk.net/ (accessed 30 October 2020).

17. Chaimani A, Caldwell DM, Li T, et al. Additional considerations are required when preparing a protocol for a systematic review with multiple interventions. *J Clin Epidemiol* 2017; 83: 65-74. 2017/01/16. DOI: 10.1016/j.jclinepi.2016.11.015.

18. Ades A.E. CDM, Reken S., Welton N.J., Sutton A.J., Dias S. NICE DSU Technical Support Document 7: Evidence Synthesis Of Treatment Efficacy In Decision Making: A Reviewer's Checklist http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD7-reviewer-checklist.final_.08.05.12.pdf (2012, accessed 12 October 2019).

19. National Institute for Health and Clinical Excellence. History of NICE, https://www.nice.org.uk/about/who-we-are/history-of-nice (2019, accessed 10 November 2019).

20. Bafeta A, Trinquart L, Seror R, et al. Reporting of results from network meta-analyses: methodological systematic review. *BMJ* 2014; 348: g1741. 2014/03/13. DOI: 10.1136/bmj.g1741.

21. Salanti G, Ades AE and Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; 64: 163-171. 2010/08/07. DOI: 10.1016/j.jclinepi.2010.03.016.

22. Law M, Alam N, Veroniki AA, et al. Two new approaches for the visualisation of models for network meta-analysis. *BMC Medical Research Methodology* 2019; 19: 61. 2019/03/20. DOI: 10.1186/s12874-019-0689-9.

23. Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014; 349: g5630. 2014/09/26. DOI: 10.1136/bmj.g5630.

24. Brignardello-Petersen R, Bonner A, Alexander PE, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol* 2018; 93: 36-44. 2017/10/21. DOI: 10.1016/j.jclinepi.2017.10.005.

25. Salanti G, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014; 9: e99682. 2014/07/06. DOI: 10.1371/journal.pone.0099682.

26. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med* 2020; 17: e1003082. 2020/04/04. DOI: 10.1371/journal.pmed.1003082.

27. CINeMA Confidence in Network Meta-Analysis, https://casp-uk.net/ (accessed 30 October 2020).

28. Phillippo DM, Dias S, Welton NJ, et al. Threshold Analysis as an Alternative to GRADE for Assessing Confidence in Guideline Recommendations Based on Network Meta-analyses. *Annals of Internal Medicine* 2019; 170: 538-546. 2019/03/26. DOI: 10.7326/M18-3542.

29. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol* 2014; 67: 138-143. 2013/10/05. DOI: 10.1016/j.jclinepi.2013.07.014.

30. Lee AW. Use of network meta-analysis in systematic reviews: a survey of authors. *Systematic reviews* 2016; 5: 8. 2016/01/23. DOI: 10.1186/s13643-015-0174-4.

Appendices

Appendix 1

Table A1 Characteristics of included studies for Chapter Three

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Ades [1]	2010	Antipyschotic drugs in schizophrenia	Prevention of relapse; discontinuation caused by intolerable side effects; discontinuation caused by other reasons.	Mental health	Full
Akshintala [2]	2012	Non-steroidal antiinflammatory drugs, glyceryl trinitrate, somatostatin, octreotide, gabexate, allopurinol, ceftazidime, ulinastatin, steroids, and heparin for prophylaxis of post-ERCP pancreatitis	Efficacy (not specified)	Gastroenterology	Conference
Almond [3]	2011	Intravitreal implant, intravitreal injection, grid laser treatment, and observation following branch retinal vein occlusion	Letters gained on the Early Treatment of Diabetic Retinopathy Study (ETDRS) chart.	Eyes and vision	Conference
Anothaisintawee [4]	2011	Drug treatments for chronic prostatitis/chronic pelvic pain syndrome	Total symptom score; pain score; voiding score; quality of life score; response rates.	Urology	Full
Ara [5]	2009	Statins at higher and lower dose in individuals with acute coronary syndrome	Change in LDL-c; Cardiovascular events; Adverse effects	Heart and circulation	Full
Ara [6]	2012	Drug treatments for obesity in primary care	Weight loss (percentage and absolute); BMI change	Endocrine and metabolic	Full
Baker [7]	2009	Drugs for maintenance treatment of chronic obstructive pulmonary disease (COPD)	Exacerbations; mortality; withdrawals.	Lungs and airways	Full
Baldwin [8]	2011	Drug treatments in patients with generalised anxiety disorder	Response and remission (both based on anxiety scale score); withdrawal because of adverse events.	Mental health	Full
Balp [9]	2012	Tobramycin (inhalation powder capsules and inhalation solutions), colistimethate sodium, and aztreonam lysine for inhalation to treat cystic fibrosis patients with chronic pseudomonas aeruginosa infection	Percent change from baseline in FEV1% predicted at 4 weeks	Lungs and airways	Conference
Bangalore [10]	2011	Antihypertensive drugs	Cancers; cancer-related death.	Heart and circulation	Full
Bansback [11]	2009	Systemic treatments for moderate to severe plaque psoriasis	Response score (using Psoriasis Area and Severity Index (PASI)).	Skin	Full
Bash [12]	2012	Amiodarone (IV and oral), diltiazem (IV), flecainide (IV and oral), ibutilide (IV), procainamide (IV), propafenone (IV and oral), sotalol (IV), and vernakalant (IV)	Cardioversion (within 2 hours of treatment initiation; within 8-24 hours of treatment initiation)	Heart and circulation	Full
Bekkering [13]	2011	Strong opioids for cancer-related or non-cancer-related severe chronic pain	Mean change of pain intensity; treatment discontinuation; serious adverse events	Anaesthesia and pain control	Full
Benedict [14]	2010	Antidepressant drugs for major depressive disorder in the elderly	Response; remission; withdrawals.	Mental health	Conference

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Benedict [15]	2010	Rufinamide, topiramate and lamotrigine for Lennox-Gastaut syndrome	Reduction in seizure frequency (drop attacks and total seizures); treatment-limiting adverse effects.	Neurology	Full
Bergman [16]	2010	Biologic agents in patients with rheumatoid arthritis who have inadequate response to disease-modifying antirheumatic drugs	Response (using American College of Rheumatology (ACR) criteria).	Rheumatology	Full
Bianic [17]	2011	Second and third-line treatments in patients with metastatic non-small cell lung cancer	Response; survival.	Cancer	Conference
Blanchard [18]	2011	Surgical, radiotherapy, and chemotherapy treatments (either alone or in combination) in head and neck cancer	Survival	Cancer	Full
Blume [19]	2010	Injectable and oral atypical antipsychotics in Schizophrenia	Prevention of relapse	Mental health	Conference
Borrill [20]	2009	Intravenous proton pump inhibitor drugs for the prevention of gastric or duodenal ulcer re-bleeding after therapeutic endoscopy	Re-bleed events; repeat endoscopy; emergency surgery.	Gastroenterology	Conference
Bounthavong [21]	2011	Linezolid, daptomycin, and vancomycin in methicillin-resistant staphylococcus aureus	Clinical resolution; discontinuation due to adverse reaction.	Skin	Full
Bow [22]	2012	Fluconazole, itraconazole, posaconazole, and voriconazole as primary prophylaxis of invasive fungal disease in allogeneic haematopoietic cell transplant recipients	Proven/probable invasive fungal infection	Blood disorders	Conference
Caldwell [23]	2005	Thrombolytic agents and percutaneous transluminal coronary angioplasty in post acute myocardial infarction	Mortality	Heart and circulation	Conference
Caldwell [24]	2010	Treatments for childhood nocturnal enuresis	Failure to achieve 14 days consecutive dry nights	Child health	Full
Cassidy [25]	2011	Adjuvant chemotherapy regimens for early-stage colon cancer	Disease-free survival; overall survival.	Cancer	Conference
Chang [26]	2012	Focused shock wave therapy (3 different intensity levels) and radial shock wave for managing plantar fasciitis	Success rate; reduction in pain scale.	Rheumatology	Full
Chapell [27]	2009	Rosiglitazone, pioglitazone, and sitagliptin in Diabetes Mellitus.	Change in HbA1C from baseline	Endocrine and metabolic	Full
Chatterjee [28]	2012	Clopidogrel, prasugrel, and ticagrelor	Stent thrombosis; recurrent ischaemic events; all-cause mortality; major non- coronary artery bypass graft bleeding.	Heart and circulation	Conference
Choy [29]	2011	Pharmacological treatments for fibromyalgia	Reduction in pain; responders; fibromyalgia impact questionnaire (FIQ) score; sleep; sleep quality; adverse events.	Rheumatology	Full
Cipriani [30]	2009	Bupropion, citalopram, duloxetine, escitalopram, fluoxetine, fluvoxamine, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, and venlafaxine for the acute treatment of unipolar major depression in adults	Response (assessed on standard scales; early discontinuation of treatment.	Mental health	Full
Cipriani [31]	2011	Aripiprazole, asenapine, carbamazepine, valproate, gabapentin, haloperidol, lamotrigine, lithium, olanzapine, quetiapine, risperidone, topiramate, and ziprasidone for acute mania	Mean change on mania rating scales; response rate; drop out of the allocated treatment.	Mental health	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Cohen [32]	2012	Second-generation antipsychotics in children and adolescents with schizophrenia, bipolar disorder, behavioural impairments comorbid to autism or intellectual disability, Tourette syndrome, and conduct disorder.	Metabolic effects (weight gain; glucose; cholesterol; triglycerides; prolactin); somnolence/sedation; extra-pyramidal syndrome/akathisia.	Mental health	Full
Coleman [33]	2008	Antihypertensive drugs	Cancers	Heart and circulation	Full
Cooper [34]	2006	Stroke prevention treatments for atrial fibrillation	Prevention of stroke; major or fatal bleeding episodes.	Heart and circulation	Full
Cope [35]	2011	Indacaterol and fixed-dose combinations of formoterol + budesonide or salmeterol + fluticasone for the treatment of chronic obstructive pulmonary disease (COPD)	Lung function; health status; breathlessness.	Lungs and airways	Full
Cope[36]	2011	Indacaterol, tiotropium, salmeterol, and formoterol for the treatment of chronic obstructive pulmonary disease (COPD)	Lung function	Lungs and airways	Conference
Corbett [37]	2012	Acupuncture and other physical treatments for chronic pain due to osteoarthritis of the knee	Change in pain (at end of treatment)	Rheumatology	Full
Cummins [38]	2011	Etanercept, adalimumab, infliximab, and palliative care (non-biologic DMARDs) in active and progressive psoriatic arthritis	Response on three rating scales	Rheumatology	Full
Curtin [39]	2004	Clonazepam, lorazepam, haloperidol, and lithium in acute mania	Response (improvement in psychopathology score).	Mental health	Full
Dakin [40]	2010	Adefovir, entecavir, lamivudine, telbivudine, and tenofovir disoproxil fumarate (alone or in combination) in the treatment of chronic hepatitis B	Viral suppression; seroconversion.	Gastroenterology	Full
Danchin [41]	2011	Trimetazidine, dihydropyridines, long-acting nitrates, nicorandil, and ranolazine	Exercise tolerance parameters (e.g. total exercise duration); clinical criteria (e.g. angina attacks per week).	Heart and circulation	Full
Daniels [42]	2012	Second generation ablation techniques in the treatment of heavy menstrual bleeding	Rate of amenorrhoea; rate of heavy bleeding; rate of dissatisfaction with treatment	Gynaecology	Full
Del Santo [43]	2012	Interferon, glatiramer, natalizumab, and fingolimod for relapsing–remitting multiple sclerosis	Freedom from relapse.	Neurology	Full
Devine [44]	2011	Biologic DMARDs with or without methotrexate for rheumatoid arthritis	ACR response criteria	Rheumatology	Full
Dewilde [45]	2012	Aspirin alone (in low/medium/high dose), aspirin and extended-release dipyridamole, aspirin and clopidogrel, and clopidogrel alone following established ischemic stroke	Stroke; combined end point of stroke, MI and vascular death	Heart and circulation	Full
Diels [46]	2011	Telaprevir, boceprevir, and peginterferon/ribavirin in patients with genotype 1 chronic hepatitis C	Sustained viral response	Gastroenterology	Conference

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Djulbegovic [47]	2012	Intravesical agents in patients with pTa and pT1 bladder cancer	Recurrence-free survival	Cancer	Conference
Dunkley [48]	2012	Interventions (lifestyle and pharmacological) in people with metabolic syndrome	Reversal of metabolic syndrome	Endocrine and metabolic	Full
Edwards [49]	2009	Carbapenem, cefepime, and piperacillin/tazobactam in the treatment of hospitalised patients with infection	Clinical response; bacteriological response; all-cause mortality; serious adverse events	Infectious disease	Full
Edwards[50]	2009	Standard- and double-dose proton pump inhibitors for the healing of severe erosive oesophagitis	Endoscopic healing	Gastroenterology	Full
Edwards[51]	2009	Atypical antipsychotics in adults with schizophrenia or bipolar disorder	Health dimension (e.g. anxiety or depression, weight gain, sexual dysfunction); extra-pyramidal symptom related; treatment withdrawals (due to adverse effects and due to lack of efficacy).	Mental health	Full
Edwards[52]	2010	Gefitinib and several platinum-based doublet chemotherapies for first-line treatment of advanced non-small cell lung cancer	Objective response rate (ORR).	Cancer	Conference
Edwards[53]	2010	Gefitinib and several platinum-based doublet chemotherapies for first-line treatment of advanced non-small cell lung cancer	Grades 3/4/5 adverse events	Cancer	Conference
Elliott [54]	2007	Antihypertensive drugs	Development of diabetes.	Heart and circulation	Full
Elliott[55]	2007	Antihypertensive drugs	1-year discontinuation rates	Heart and circulation	Conference
Ellis [56]	2012	Bazedoxifene, alendronate, ibandronate, and risedronate in a high-risk postmenopausal population	Nonvertebral fractures	Rheumatology	Conference
Ernest [57]	2012	30 different methods for the assessment of glaucomatous visual field progression	Incidence of progression	Eyes and vision	Full
Fan [58]	2011	Infliximab and ustekinumab (higher and lower dose) in the treatment of moderate to severe psoriasis	PASI	Skin	Conference
Farris[59]	2010	Anticoagulant therapies in high-risk surgical patients	Venous thromboembolism (VTE); major bleeding.	Heart and circulation	Conference
Freemantle[60]	2011	Dronedarone, amiodarone, sotalol, flecainide, and propafenone for the management of atrial fibrillation (AF)	Mortality; stroke; AF recurrence; incidence of serious adverse events; treatment withdrawals (all cause and due specifically to adverse events); proarrhythmic events.	Heart and circulation	Full
Freemantle[61]	2011	Amphoteracin B (all formulations), itraconazole, voriconazole, caspofungin, micafungin, posaconazole, and anidulafungin for empirical, pre-emptive, and directed treatment strategies for invasive mould disease	Survival; response.	Cancer	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Fretheim[62]	2012	Different classes of antihypertensive drugs for healthy people at risk of cardiovascular disease	Total mortality; myocardial infarction; stroke; angina; heart failure; incidence of diabetes.	Heart and circulation	Full
Fusaro[63]	2012	Paclitaxel, everolimus (EES), sirolimus, zotarolimus eluting stents, and bare metal stents for de novo coronary lesions in subjects with diabetes mellitus	Target vessel revascularisation; death; myocardial infarction; stent thrombosis.	Heart and circulation	Conference
Gartlehner[64]	2008	Second-generation antidepressants for the treatment of depressive disorders in adults	Response; remission; efficacy on specific depression symptoms; adverse events	Mental health	Full
Gartlehner[65]	2011	Second-generation antidepressants for the treatment of depressive disorders in adults	Response; remission; efficacy on specific depression symptoms; adverse events	Mental health	Full
Golfinopoulos[66]	2007	Systemic treatment regimens in advanced colorectal cancer	Death; progression-free survival	Cancer	Full
Golfinopoulos[67]	2009	Systemic treatment regimens for cancer of unknown primary site	Median survival; death.	Cancer	Full
Gray[68]	2012	Orlistat, rimonabant, and sibutramine for the treatment of obesity	Weight loss; BMI change	Endocrine and metabolic	Full
Greenhalgh[69]	2011	Clopidogrel, modified-release dipyridamole (alone or with aspirin), and aspirin (alone) in the prevention of occlusive vascular events in patients with a history of MI, ischaemic stroke/TIA or established peripheral arterial disease	Recurrent stroke; myocardial infarction; vascular death; death from all causes; bleeding	Heart and circulation	Full
Gross[70]	2011	Add-on antihyperglycemic drugs in patients with type 2 diabetes that is not controlled with metformin and a sulfonylurea	HbA1c level change; weight; severe hypoglycaemia.	Endocrine and metabolic	Full
Gurusamy[71]	2011	Intervention methods aimed at decreasing blood loss and/or allogeneic blood transfusion requirements during liver transplantation	Mortality; thromboembolic episodes; serious adverse events; blood loss; transfusion required; hospital stay; intensive therapy unit stay.	Gastroenterology	Full
Guyot[72]	2011	Abatacept (with methotrexate) and other biologic agents in treatment of patients with active rheumatoid arthritis despite methotrexate	Efficacy (assessed by standard rating scales)	Rheumatology	Full
Guyot[73]	2012	Methotrexate plus a biologic (abatacept, infliximab, adalimumab, etanercept, certolizumab pegol, and golimumab) to treat rheumatoid arthritis when methotrexate alone is insufficient	Response (assessed by HAQ and ACR response criteria)	Rheumatology	Full
Hansen [74]	2008	Second-generation antidepressants in social anxiety disorder	Response; adverse events	Mental health	Full
Harrington[75]	2011	Randomized diet therapies for type 2 diabetes mellitus	HbA1c; Lipids.	Endocrine and metabolic	Conference
Hartling[76]	2011	Bronchodilators and steroids, alone or combined, for the acute management of bronchiolitis in children aged less than 2 years	Emergency admission; length of stay.	Lungs and airways	Full
Hawkins[77]	2009	Erlotinib, pemetrexed, docetaxel, and gefitinib in non-small cell lung cancer	Survival	Cancer	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Hopkins [78]	2011	Osteoporosis medications for post-menopausal women	Vertebral and non-vertebral fractures	Orthopaedics and trauma	Full
Ibrahim[79]	2011	Surgical and non-surgical treatments for minimally displaced and undisplaced scaphoid waist fractures	Fracture union; complications; range of motion; grip strength; osteoarthritis of the scaphotrapeziotrapezoid and radiocarpal joints.	Orthopaedics and trauma	Full
Imamura[80]	2010	Non-surgical treatments for women with stress urinary incontinence	Improvement; cure.	Urology	Full
Jansen[81]	2006	Self-monitoring blood glucose, self-monitoring of urine glucose, and interventions without self-monitoring in type 2 diabetes mellitus	HbA1c	Endocrine and metabolic	Full
Jansen[82]	2007	Artemisinin-based Combination Therapies for treatment of uncomplicated P. falciparum malaria	Adequate clinical and parasitological response at day 28	Infectious disease	Full
Jansen[83]	2009	Zoledronic acid, alendronate, ibandronate, and risedronate for the prevention of vertebral fractures in postmenopausal women with osteoporosis	Vertebral fracture	Orthopaedics and trauma	Full
Jansen[84]	2010	Etoricoxib, celecoxib, naproxen, and diclofenac in the initial treatment of ankylosing spondylitis	Standard assessment scores (BASFI and BASDAI); discontinuation due to lack of efficacy.	Rheumatology	Full
Jansen[85]	2011	Zoledronic acid, alendronate, ibandronate, risedronate, and etidronate in the prevention of vertebral, hip, and nonvertebral-nonhip fractures in osteoporosis	Vertebral fractures; hip fractures; all other fractures.	Orthopaedics and trauma	Full
Kessler[86]	2011	Antimuscarinics for treating overactive bladder	Adverse events	Urology	Full
King[87]	2006	Methylphenidate, dexamfetamine, and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents	Core symptoms; quality of life; adverse effects	Child health	Full
Klemp[88]	2011	Aripiprazole, clozapine, olanzapine, risperidone, and haloperidol in schizophrenia	Response on standard scale; weight gain; extrapyramidal symptoms	Mental health	Full
Knight[89]	2010	Varenicline (long and standard course), buproprion, nicotine replacement therapy, and unaided quit for smoking cessation	Quit rate at 52 weeks	Tobacco, drugs and alcohol dependence	Full
Kruidenier[90]	2012	Angioplasty, surgery, exercise therapy, and no treatment for intermittent claudication	Walking distance; quality of life.	Heart and circulation	Full
Kyrgiou[91]	2006	Chemotherapy regimens for ovarian cancer	Survival	Cancer	Full
Lam[92]	2007	Cardiac resynchronisation therapy, implantable cardioverter defibrillator therapy, combined resynchronisation and implantable defibrillator therapy, and medical therapy alone in patients with impaired left ventricular systolic function	All-cause mortality	Heart and circulation	Full
Lang[93]	2012	Tirofiban, abciximab, eptifibatide, and usual care in patients with acute coronary syndrome	Major adverse cardiac events at 30 days	Heart and circulation	Full
Launois[94]	2011	Anti-cytokine agents indicated for the treatment of rheumatoid arthritis	Response (assessed using ACR criteria)	Rheumatology	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Lebmeier[95]	2010	Biologics in patients with rheumatoid arthritis taking concomitant methotrexate (MTX) who have inadequate response to MTX	Response (assessed as change in HAQ score)	Rheumatology	Conference
Lebmeier[96]	2011	Biologics in patients with rheumatoid arthritis taking concomitant methotrexate (MTX) who have inadequate response to MTX	Response (assessed as change in HAQ score)	Rheumatology	Conference
Lerdkiattikorn[97]	2010	Three adjuvant chemotherapy regimens for stage III colon cancer	Survival	Cancer	Conference
LeReun[98]	2010	Raltegravir, nevirapine, efavirenz, lopinavir, and atazanavir in treatment-naive patients with HIV infection	Viral suppression; early discontinuation of treatment.	Infectious disease	Conference
Lewis[99]	2011	Treatment strategies for sciatica	At short, medium, and long term follow up: global effect; pain intensity; condition specific outcome measures.	Orthopaedics and trauma	Full
Logman[100]	2010	Common antimicrobial agents for treatment of complicated skin and soft tissue infections (cSSTI) caused by methicillin-resistant Staphylococcus aureus (MRSA)	Clinical/microbiological treatment success	Infectious disease	Full
Manzoli[101]	2009	Several types of H5N1 vaccine	Seroconversion rate according to haemagglutination–inhibition; microneutralisation.	Infectious disease	Full
Manzoli[102]	2011	Several types of H1N1 vaccines	Seroconversion rate according to haemagglutination-inhibition	Infectious disease	Full
Maund[103]	2012	Treatments for primary frozen shoulder	Pain	Rheumatology	Full
Mauri[104]	2008	22 different regimens involving chemotherapy and/or targeted therapy in advanced breast cancer	Survival	Cancer	Full
McDaid[105]	2010	Paracetamol, non-steroidal anti-inflammatory drugs (NSAIDs), and cyclo-oxygenase 2 (COX-2) inhibitors for the reduction of morphine-related side effects after major surgery	Morphine consumption; nausea and vomiting; sedation	Anaesthesia and pain control	Full
McIntosh[106]	2011	Antihyperglycemic therapies in patients with type 2 diabetes inadequately controlled on metformin monotherapy	HbA1c; body weight; overall hypoglycaemia.	Endocrine and metabolic	Full
McNamara[107]	2010	Trastuzumab/anastrozole, lapatinib/letrozole, anastrazole, and letrozole for the treatment of HER2 positive/hormone receptor positive metastatic breast cancer	Overall survival	Cancer	Conference
Meader[108]	2010	Methadone, buprenorphine, clonidine, and lofexidine for opioid detoxification	Completion of treatment	Tobacco, drugs and alcohol dependence	Full
Mhaskar[109]	2012	Bisphosphonates in the management of patients with MM	Overall mortality; progression-free survival; vertebral fractures; total skeletal related events; pain; hypercalcaemia; treatment-related harms	Cancer	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Migliore[110]	2012	Etanercept, infliximab, and adalimumab for psoriatic arthritis	Response (assessed by ACR criteria)	Rheumatology	Full
Migliore[111]	2012	Etanercept, infliximab, and adalimumab in the treatment of ankylosing spondylitis patients	Response (assessed by ASAS20 criteria)	Rheumatology	Full
Mills[112]	2008	Atorvastatin, fluvastatin, pravastatin, and lovastatin in primary prevention of cardiovascular events	All-cause mortality; cardiovascular disease mortality.	Heart and circulation	Full
Mills[113]	2009	Antifungal treatment for invasive candida infections	Global response; all-cause mortality; fungal-attributable mortality; adverse events	Infectious disease	Full
Mills[114]	2011	Maintenance pharmacotherapies for chronic obstructive pulmonary disease	Event rate of exacerbations	Lungs and airways	Full
Mills[115]	2011	Various statin therapy in patients with and without prior coronary heart disease	Cardiovascular mortality	Heart and circulation	Full
Murad[116]	2012	Alendronate, risedronate, zoledronate, ibandronate, teriparatide, raloxifene, bazedoxifene, denosumab, calcium, vitamin D, and calcium with vitamin D for individuals at risk of developing fragility fracture	Hip fractures; vertebral fractures; non-vertebral fractures.	Rheumatology	Full
Mutebi[117]	2011	Treprostinil, iloprost, bosentan, sitaxsentan, ambrisentan, and sildenafil in pulmonary arterial hypertension	Mortality	Heart and circulation	Conference
Ni[118]	2012	Various probiotics and montmorillonite (alone or in combination) for infantile rotavirus	Symptom improvement (stool frequency; systemic)	Infectious disease	Full
Nixon[119]	2007	Adalimumab, anakinra, etanercept, infliximab, and methotrexate in rheumatoid arthritis	ACR response criteria	Rheumatology	Full
Numthavaj[120]	2011	Aciclovir, valaciclovir (alone or in combination with prednisolone), and prednisolone alone for Bell's palsy	Resolution at 3 and 6 months	Neurology	Full
Orme[121]	2010	Topical hypotensives for primary open-angle glaucoma	Intraocular pressure; hyperaemia-type events	Eyes and vision	Full
Owen[122]	2010	Aspirin (high and low dose) and Warfarin for the primary prevention of stroke in patients with non valvular atrial fibrillation	Stroke; death.	Heart and circulation	Full
Padwal[123]	2011	Bariatric surgeries and standard care for obesity	Change in BMI	Endocrine and metabolic	Full
Palmerini[124]	2012	Different drug-eluting stents and bare-metal stents	Stent thrombosis	Heart and circulation	Full
Pechlivanoglou[125]	2010	Different antifungal drugs used in prophylaxis treatments against invasive fungal infections in high-risk patients	Invasive fungal infection	Blood disorders	Conference
Phung[126]	2010	Noninsulin antidiabetic drugs in patients with type 2 DM not controlled by metformin alone	HbA1c; body weight; hypoglycaemia.	Endocrine and metabolic	Full
Phung[127]	2011	Low-dose unfractionated heparin (twice or three times daily) and low-molecular-weight heparin in hospitalized nonsurgical patients at risk for VTE	DVT; pulmonary embolism; death; major bleed.	Heart and circulation	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Phung[128]	2011	Oral anti-diabetic drugs in patients at high risk for developing Type 2 diabetes	New onset type 2 diabetes	Endocrine and metabolic	Full
Psaty[129]	2003	Antihypertensive therapies used as first-line agents	Coronary heart disease; congestive heart failure; stroke; cardiovascular disease events; cardiovascular disease mortality; blood pressure change; total mortality	Heart and circulation	Full
Puhan[130]	2009	Inhaled maintenance drug regimes in patients with stable COPD	Exacerbations	Lungs and airways	Full
Reich[131]	2012	All biologic agents indicated in the treatment of moderate to severe psoriasis currently available in Europe	Response (assessed on the PASI index)	Skin	Full
Riemsma[132]	2010	Letrozole, anastrozole, and exemestane for hormone sensitive advanced or metastatic breast cancer in post-menopausal women	Objective response rate	Cancer	Full
Riemsma[133]	2011	Tapentadol, transdermal buprenorphine, transdermal fentanyl, hydromorphone, morphine, and oxymorphone in chronic severe pain	Pain intensity; pain relief; quality of sleep; patient global impression of change; quality of life; serious adverse events.	Anaesthesia and pain control	Full
Riemsma[134]	2012	Lapatinib alone, lapatinib plus letrozole, letrozole alone, trastuzumab plus anastrozole, anastrozole alone, tamoxifen, and exemestane for hormone receptor positive (HR+) and HER2+ advanced or metastatic breast cancer	Objective response rate	Cancer	Full
Roskell[135]	2010	Dabigatran (2 dose levels), adjusted-dose vitamin K antagonists (including warfarin), aspirin alone, and aspirin plus clopidogrel for stroke prevention in atrial fibrillation	All stroke; ischaemic stroke; systemic embolism; all-cause mortality; intracranial haemorrhage (excluding haemorrhagic stroke); extracranial haemorrhage (major bleeds); acute myocardial infarction.	Heart and circulation	Full
Roskell[136]	2011	Duloxetine, fluoxetine, gabapentin, milnacipran, pramipexole, pregabalin, either of two tricyclic antidepressants, and tramadol plus paracetamol in the treatment of fibromyalgia	Pain response; discontinuation because of adverse events.	Rheumatology	Full
Roskell[137]	2012	Fingolimod, interferon beta-1a, interferon beta-1b, and glatiramer acetate in relapsing–remitting multiple sclerosis	Annualized relapse rate	Neurology	Full
Rotta[138]	2012	Topical antifungals used in the treatment of dermatomycosis	Mycological cure based on microscopy and/or culture	Skin	Full
Roversi[139]	2012	Apixaban, dabigatran, edoxaban, rivaroxaban, and warfarin for non-valvular atrial fibrillation	Stroke/systemic embolism; major bleeding; discontinuation.	Heart and circulation	Conference
Salanti[140]	2009	Fluoride (in toothpaste, rinse, gel, and varnish) in preventing dental caries	Caries increment	Dentistry and oral health	Full
Sanches[141]	2011	Long-acting insulin analogues (glargine and detemir) and neutral protamine hagedorn insulin in adults with type 1 diabetes	HbA1c change	Endocrine and metabolic	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Schauble[142]	2011	Donepezil, galantamine, rivastigmine (all formulations), and memantine in Alzheimers disease	Cognitive and behaviour scales; adverse events.	Mental health	Conference
Schmitz[143]	2012	Licensed anti-TNF agents in rheumatoid arthritis for methotrexate non-responders	Response (assessed by ACR criteria and HAQ score)	Rheumatology	Full
Sciarretta[144]	2011	Antihypertensives in patients with hypertension or high cardiovascular risk profile for prevention of heart failure	Incidence of heart failure	Heart and circulation	Full
Scott[145]	2011	Exenatide and liraglutide (2 dose levels) for the treatment of type 2 diabetes	HbA1c change	Endocrine and metabolic	Conference
Senior[146]	2012	Rotigotine transdermal patch and both immediate and extended release forms of ropinirole and pramipexole in patients with advanced Parkinson's disease	Incidence of dyskinesia	Neurology	Conference
Siddiqui[147]	2011	Tramadol, oxycodone, hydrocodone, propoxyphene, and codeine in non-malignant pain	Adverse events; withdrawal of treatment due to adverse events	Anaesthesia and pain control	Conference
Singh[148]	2009	Abatacept, adalimumab, anakinra, etanercept, infliximab, and rituximab in patients with rheumatoid arthritis	Response (assessed by ACR criteria); withdrawal due to adverse event.	Rheumatology	Full
Singh[149]	2011	Adalimumab, certolizumab, etanercept, golimumab, infliximab, anakinra, tocilizumab, abatacept, and rituximab therapy in patients with any disease condition except human immunodeficiency disease (HIV/AIDS)	Total adverse events; withdrawal due to adverse events; serious infections; serious adverse events.	Rheumatology	Full
Soares-Weiser[150]	2007	Pharmacological and psychosocial interventions for the prevention of relapse in bipolar disorder	All relapse, manic relapse and depressive relapse	Mental health	Full
Squires[151]	2011	Cilostazol, naftidrofuryl oxalate, pentoxifylline, and inositol nicotinate for the treatment of intermittent claudication in people with peripheral arterial disease	Maximal walking distance; pain-free walking distance.	Heart and circulation	Full
Stam[152]	2012	Etoricoxib, lumiracoxib, celecoxib, non-selective NSAIDs (ibuprofen, diclofenac, naproxen), and acetaminophen (paracetamol) in the treatment of osteoarthritis	Pain; physical function; patient global assessment.	Rheumatology	Full
Steiner[153]	2012	Prasugrel, ticagrelor, and clopidogrel (standard and high-dose) in patients undergoing percutaneous coronary intervention	All-cause death; cardiovascular death; major adverse cardiac events; myocardial infarction; stroke; stent thrombosis; major bleeding; major or minor bleeding.	Heart and circulation	Full
Stettler[154]	2007	Sirolimus-eluting stents, paclitaxel-eluting stents, and bare-metal stents in patients with coronary artery disease	Overall mortality; cardiac death; myocardial infarction; composite of death or myocardial infarction; stent thrombosis; target lesion revascularisation.	Heart and circulation	Full
Stettler[155]	2008	Sirolimus-eluting stents, paclitaxel-eluting stents, and bare-metal stents in people with and without diabetes	Overall mortality; cardiac death; myocardial infarction; composite of death or myocardial infarction; stent thrombosis; target lesion revascularisation.	Heart and circulation	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Stradwick[156]	2011	Ferumoxytol and alternative oral and IV iron replacement therapies for the treatment of iron deficiency anaemia associated with chronic kidney disease	Hb improvement	Kidney disease	Conference
Strassmann[157]	2009	Smoking cessation counselling with or without pharmacotherapy or nicotine replacement therapy for chronic obstructive pulmonary disease (COPD) patients	Prolonged abstinence	Tobacco, drugs and alcohol dependence	Full
Tang[158]	2012	Ondansetron, granisetron, tropisetron, and dolasetron as prophylactic agents for patients with mild to moderate risk of postoperative nausea and vomiting	Postoperative nausea and vomiting; post-operative vomiting.	Anaesthesia and pain control	Full
Tantai[159]	2010	Lamivudine, entecavir, interferon, and interferon plus lamivudine for HBeAg-positive chronic hepatitis B	HBeAg seroconversion	Gastroenterology	Conference
Thijs[160]	2008	Aspirin, aspirin plus dipyridamole, thienopyridines, and combination of aspirin plus thienopyridines in the prevention of serious vascular events after transient ischaemic attack (TIA) or stroke	Serious vascular events (composite of non-fatal stroke, non-fatal myocardial infarction and vascular death)	Heart and circulation	Full
Tosh[161]	2011	Nonbiologic disease-modifying antirheumatic drug strategies in patients with early rheumatoid arthritis	Response (assessed by ACR criteria)	Rheumatology	Full
Trelle[162]	2011	Cardiovascular safety of non-steroidal anti-inflammatory drugs for any medical condition	Myocardial infarction; stroke; cardiovascular death; death from any cause; the Antiplatelet Trialists' Collaboration composite outcome of non-fatal myocardial infarction, non-fatal stroke, or cardiovascular death.	Heart and circulation	Full
Trikalinos[163]	2009	Medical therapy, percutaneous transluminal balloon coronary angioplasty (PTCA), bare-metal stents (BMS), and drug-eluting stents in patients with non-acute coronary artery disease	Death; myocardial infarction; coronary artery bypass grafting; target vessel/lesion revascularisation; any revascularisation.	Heart and circulation	Full
Tropeano[164]	2011	Antihypertensive drugs	Carotid intima-media thickness change	Heart and circulation	Full
Tu[165]	2010	Enamel matrix derivatives (EMD) in conjunction with other regenerative materials and EMD alone in the treatment of infrabony defects 3 mm	Changes in probing pocket depth (PPD); clinical attachment level (CAL); infrabony defect depth.	Dentistry and oral health	Full
Tu[166]	2012	Guided tissue regeneration (GTR) and enamel matrix derivatives (EMD) alone or in conjunction with other regenerative materials	Changes in probing pocket depth (PPD); clinical attachment level (CAL); infrabony defect depth.	Dentistry and oral health	Full
Tudur Smith[167]	2007	Carbamazepine, sodium valproate, phenytoin, phenobarbitone, lamotrigine, oxcarbazepine, gabapentin, and topiramate as monotherapy for epilepsy	Time to treatment failure; time to 12 month remission from seizures; time to first seizure.	Neurology	Full
Turkstra[168]	2011	Abatacept, adalimumab, anakinra, certolizumab, etanercept, golimumab, infliximab, rituximab, and tocilizumab in patients with established rheumatoid arthritis	Response (assessed by ACR criteria)	Rheumatology	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Uthman[169]	2010	Fluoxetine, fluvoxamine, paroxetine, sertraline, and venlafaxine for treating anxiety disorders in children and adolescents	Resonse (assessed by Clinical Global Impressions scale); withdrawal due to treatment emergent adverse events.	Mental health	Full
van de Kerkhof[170]	2011	Topical calcipotriol and betamethasone dipropionate (once or twice daily) and other commonly used topical treatments in psoriasis	Response (assessed by PASI score)	Skin	Full
Van den Bruel[171]	2009	Ophthalmic viscoelastic devices (OVDs) in protecting the cornea during cataract surgery	Loss in endothelial cell density	Eyes and vision	Full
van der Valk[172]	2009	Betaxolol, bimatoprost, brimonidine, brinzolamide, dorzolamide, latanoprost, timolol, and travoprost in patients with primary open-angle glaucoma or ocular hypertension	Mean intraocular pressure (IOP) reduction (at peak and trough)	Eyes and vision	Full
Vansteenkiste [173]	2008	Pemetrexed plus cisplatin compared with recommended first line platinum-based therapies for advanced non-small cell lung cancer	Overall survival; febrile neutropenia; one-year survival; tumour response; toxicity	Cancer	Conference
Vejakama[174]	2012	ACE inhibitor (ACEI)/angiotensin II receptor blocker (ARB) and other antihypertensive drugs in type 2 diabetes.	End stage renal disease; doubling serum creatinine; major micro-vascular complications; macro-albuminuria; micro-albuminuria; albuminuria regression	Kidney disease	Full
Verheggen[175]	2011	Atazanavir/r, darunavir/r, lopinavir/r, and efavirenz in treatment-naïve HIV-1 infected adults	Viral load levels; CD4 cell count; incidence of diarrhoea, nausea and rash.	Infectious disease	Conference
Vieira[176]	2011	Two nucleoside/nucleotide reverse transcriptase inhibitors combined with raltegravir, efavirenz, or protease inhibitors in the management of antiretroviral-naïve HIV adult patients	HIV-1 RNA level; CD4+ cell count	Infectious disease	Full
Virgili[177]	2011	Verteporfin photodynamic therapy, pegaptanib, and ranibizumab for neovascular age-related macular degeneration (AMD)	Visual loss or visual gain of 3 or more lines of visual acuity (15 ETDRS letters)	Eyes and vision	Full
Vissers[178]	2010	Intranasal fentanyl spray, oral transmucosal fentanyl citrate, fentanyl buccal tablet, and oral morphine for the treatment of breakthrough cancer pain	Pain intensity difference	Infectious disease	Full
Vlaar[179]	2011	3 percutaneous coronary intervention (PCI) strategies (culprit vessel only, multi-vessel and staged) in patients with ST-segment elevation myocardial infarction (STEMI) and multivessel disease (MVD)	Mortality (short and long term)	Heart and circulation	Full
Walsh[180]	2010	Fluoride toothpastes of different concentrations in preventing dental caries in children and adolescents	Caries increment	Dentistry and oral health	Full
Wandel[181]	2010	Glucosamine and chondroitin (alone or in combination) in osteoarthritis of the hip or knee	Pain intensity; change in minimal width of joint space.	Rheumatology	Full

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Wang[182]	2010	Different types of central venous catheter	Catheter-related bloodstream infection; catheter colonisation.	Infectious disease	Full
Wang[183]	2012	Various surgical margin thresholds with breast-conserving surgery (with and without radiotherapy) for women with ductal carcinoma in situ	Ipsilateral breast tumour recurrence	Cancer	Full
Welton[184]	2008	Amantadine, oseltamivir, and zanamivir for treatment of influenza A and B	Mean time to end of fever; mean time from end of fever to end of all symptoms; mean time to end of all symptoms.	Infectious disease	Full
Welton[185]	2009	Psychological interventions (usual care, educational, behavioural, cognitive, relaxation, and support) in coronary heart disease	All-cause mortality; cardiac mortality; nonfatal myocardial infarction; total cholesterol; systolic blood pressure; diastolic blood pressure; depression; anxiety.	Heart and circulation	Full
Wiebe[186]	2011	Different sweeteners in obese, diabetic, and healthy populations	2-hour blood glucose response	Endocrine and metabolic	Full
Wojciechowski[187]	2011	Tenofovir, lamivudine, adefovir, and entecavir in the treatment of chronic hepatitis B virus (HBV) infection.	HBV DNA clearance; ALT; histological improvement; serious adverse events	Infectious disease	Conference
Wolff[188]	2010	5% lidocaine medicated plaster, amitriptyline, capsaicin, gabapentin, and pregabalin in painful diabetic peripheral neuropathy	Pain change from baseline	Anaesthesia and pain control	Full
Wolff[189]	2011	5% lidocaine medicated plaster, capsaicin, gabapentin and pregabalin for post-herpetic neuralgia	Pain change from baseline	Anaesthesia and pain control	Full
Wolff[190]	2012	Buprenorphine patch, fentanyl patch, and morphine in patients with chronic moderate to severe pain	Pain intensity; quality of life; quality of sleep; serious adverse events; constipation; nausea; vomiting; somnolence; dizziness; treatment discontinuation.	Anaesthesia and pain control	Full
Woo[191]	2010	Pegylated interferon, lamivudine, adefovir, entecavir, telbivudine, and tenofovir as monotherapy or combination therapy for treatment-naive chronic hepatitis B	Rates of virologic and biochemical response; HBeAg loss; HBeAg seroconversion; HBsAg loss; histologic improvement.	Infectious disease	Full
Woolacott[192]	2006	Supportive care, Etanercept (2 dose levels), Efalizumab, Ciclosporin, Fumaderm, Infliximab, and Methotrexate for the treatment of moderate to severe chronic plaque psoriasis	Response (assessed by PASI score)	Skin	Full
Wu[193]	2011	Angiotensin converting enzyme inhibitors (ACEI), angiotensin II receptor blockers (ARB), and calcium channel blockers in patients with diabetic nephropathy	Doubling of creatinine; requiring dialysis; all-cause deaths.	Kidney disease	Conference
Wu[194]	2012	Angiotensin converting enzyme inhibitors (ACEI), angiotensin II receptor blockers (ARB), calcium channel blockers, and beta-blockers in patients with diabetic nephropathy	Doubling of serum creatinine	Kidney disease	Conference
Wu[195]	2012	Angiotensin converting enzyme inhibitors (ACEI), angiotensin II receptor blockers (ARB), calcium channel blockers, beta-blockers, and diuretics in hypertensive patients with diabetes	All-cause mortality	Endocrine and metabolic	Conference

Author	Year	Interventions compared	Outcome(s) compared	Clinical topic	Publication type
Yi[196]	2010	Bevacizumab plus cisplatin and gemcitabine (BCG), bevacizumab plus carboplatin and paclitaxel (BCP), grouped platinum-based doublet chemotherapy (PLD), and grouped nonplatinum-based doublet chemotherapy (NPLD) for first-line advanced or metastatic non-small cell lung cancer	Progression-free survival	Cancer	Conference
Youn[197]	2012	Conventional drug therapy (including beta-blockers or ACE inhibitors), conventional drug therapy plus class III antiarrhythmic drug therapy (including amiodarone or sotalol), and conventional drug therapy plus conventional intracardiac device therapy for the prevention of sudden cardiac death in high risk patients. Includes predicted effect for conventional drug therapy plus new four-pole connector intracardiac device therapy.	Mortality	Heart and circulation	Full
Zagmutt[198]	2012	Pramipexole, ropinirole, and rasagiline as monotherapy for early stage Parkinson's disease	Total adverse events; Specific adverse events (cognitive, gastrointestinal, sleep/fatigue); dropout rates.	Neurology	Full
Zeppetella[199]	2011	Fast-acting fentanyl formulations (intranasal fentanyl spray, fentanyl pectin nasal spray, oral transmucosal fentanyl citrate, fentanyl buccal tablet, sublingual fentanyl citrate and fentanyl buccal soluble film), and oral morphine in the management of breakthrough cancer pain	Pain intensity difference	Anaesthesia and pain control	Conference
Zintzaras[200]	2012	Pharmacological treatments for relapsing multiple sclerosis	Patients free of relapse; patients without disease progression; patients without MRI progression.	Neurology	Full
Ziogas[201]	2011	42 induction treatments in acute myeloid leukaemia in elderly patients	Complete remission	Cancer	Full

Appendix 1 References

1. Ades AE, Mavranezouli I, Dias S, et al. Network meta-analysis with competing risk outcomes. *Value Health* 2010; 13: 976-983. 2010/09/10. DOI: 10.1111/j.1524-4733.2010.00784.x.

2. Akshintala VS, Hutfless S, Khashab M, et al. A network meta-analysis evaluating the efficacy of pharmacologic prophylaxis for post-ERCP pancreatitis. *Gastrointest Endosc* 2012; 75 (4 Suppl): AB121-122. DOI: 10.1016/j.gie.2012.04.014.

3. Almond C, Hayward E, Freemantle N, et al. The use of a mixed-treatment comparison to assess the costeffectiveness of ozurdex (dexamethasone intravitreal implant in applicator) compared with bevacizumab intravitreal injections for patients with macular oedema following branch retinal vein occlusion. *Value Health* 2011; 14: A506. DOI: 10.1016/j.jval.2011.08.1491.

4. Anothaisintawee T, Attia J, Nickel JC, et al. Management of chronic prostatitis/chronic pelvic pain syndrome: a systematic review and network meta-analysis. *JAMA* 2011; 305: 78-86. 2011/01/06. DOI: 10.1001/jama.2010.1913.

5. Ara R, Pandor A, Stevens J, et al. Early high-dose lipid-lowering therapy to avoid cardiac events: a systematic review and economic evaluation. *Health Technol Assess* 2009; 13(34) 2009/07/17. DOI: 10.3310/hta13340.

6. Ara R, Blake L, Gray L, et al. What is the clinical effectiveness and cost-effectiveness of using drugs in treating obese patients in primary care? A systematic review. *Health Technol Assess* 2012; 16(5) 2012/02/22. DOI: 10.3310/hta16050.

7. Baker WL, Baker EL and Coleman CI. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. *Pharmacotherapy* 2009; 29: 891-905. 2009/07/30. DOI: 10.1592/phco.29.8.891.

8. Baldwin D, Woods R, Lawson R, et al. Efficacy of drug treatments for generalised anxiety disorder: systematic review and meta-analysis. *BMJ* 2011; 342: d1199. 2011/03/15. DOI: 10.1136/bmj.d1199.

9. Balp MM, Capkun-Niggli G, Littlewood K, et al. Efficacy of tobramycin inhalation powder versus other inhaled antibiotics in cystic fibrosis patients with chronic p. aeruginosa infection: a network meta-analysis. *J Cyst Fibros* 2012; 11(Suppl 1): S73. DOI: 10.1016/S1569-1993(12)60236-7.

10. Bangalore S, Kumar S, Kjeldsen SE, et al. Antihypertensive drugs and risk of cancer: network meta-analyses and trial sequential analyses of 324,168 participants from randomised trials. *Lancet Oncol* 2011; 12: 65-82. 2010/12/03. DOI: 10.1016/S1470-2045(10)70260-6.

11. Bansback N, Sizto S, Sun H, et al. Efficacy of systemic treatments for moderate to severe plaque psoriasis: systematic review and meta-analysis. *Dermatology* 2009; 219: 209-218. 2009/08/07. DOI: 10.1159/000233234.

12. Bash LD, Buono JL, Davies GM, et al. Systematic review and meta-analysis of the efficacy of cardioversion by vernakalant and comparators in patients with atrial fibrillation. *Cardiovasc Drugs Ther* 2012; 26: 167-179. 2012/03/16. DOI: 10.1007/s10557-012-6374-4.

13. Bekkering GE, Soares-Weiser K, Reid K, et al. Can morphine still be considered to be the standard for treating chronic pain? A systematic review including pair-wise and network meta-analyses. *Curr Med Res Opin* 2011; 27: 1477-1491. 2011/06/04. DOI: 10.1185/03007995.2011.586332.

14. Benedict A, Naci H, Fahrbach K, et al. Comparative effectiveness of antidepressants in the treatment of the elderly with major depressive disorder : a mixed treatment comparison and meta-regression analysis. *Value Health* 2010; 13: A245. DOI: 10.1016/S1098-3015(11)71872-1.

15. Benedict A, Verdian L and Maclaine G. The cost effectiveness of rufinamide in the treatment of Lennox-Gastaut syndrome in the UK. *Pharmacoeconomics* 2010; 28: 185-199. 2010/02/16. DOI: 10.2165/11313640-000000000-00000.

16. Bergman GJ, Hochberg MC, Boers M, et al. Indirect comparison of tocilizumab and other biologic agents in patients with rheumatoid arthritis and inadequate response to disease-modifying antirheumatic drugs. *Semin Arthritis Rheum* 2010; 39: 425-441. 2010/03/13. DOI: 10.1016/j.semarthrit.2009.12.002.

17. Bianic F, Despiegel N, Cure S, et al. Network meta-analysis of second and third-line treatments on overall response and overall survival in patients with metastatic non-small cell lung cancer. *Eur J Cancer* 2011; 47(Suppl 1): S616-617. DOI: 10.1016/S0959-8049(11)72393-0.

18. Blanchard P, Hill C, Guihenneuc-Jouyaux C, et al. Mixed treatment comparison meta-analysis of altered fractionated radiotherapy and chemotherapy in head and neck cancer. *J Clin Epidemiol* 2011; 64: 985-992. 2011/02/19. DOI: 10.1016/j.jclinepi.2010.10.016.

19. Blume S, Naci H, Green J, et al. The comparative efficacy of injectable and oral atypical antipsychotics in reducing relapses in adult schizophrenia patients: a systematic review and mixed treatment comparison analysis. *Value Health* 2010; 13: A106. DOI: 10.1016/S1098-3015(10)72508-0.

20. Borrill J, Edwards SJ and Gray J. Cost-effectiveness of high-dose intravenous proton pump inhibitors for the prevention of gastric or duodenal ulcer rebleeding after therapeutic endoscopy. *Value Health* 2009; 12: A347. DOI: 10.1016/s1098-3015(10)74706-9.

21. Bounthavong M, Zargarzadeh A, Hsu DI, et al. Cost-effectiveness analysis of linezolid, daptomycin, and vancomycin in methicillin-resistant staphylococcus aureus: complicated skin and skin structure infection using bayesian methods for evidence synthesis. *Value Health* 2011; 14: 631-639. 2011/08/16. DOI: 10.1016/j.jval.2010.12.006.

22. Bow E, Vanness D, Cordonnier C, et al. Primary prophylaxis of invasive fungal disease in allogeneic haematopoietic cell transplant recipients - a mixed treatment comparison of randomised clinical trials. *Bone Marrow Transplant* 2012; 47(Suppl 1s): S78-S79. DOI: 10.1038/bmt.2012.36.

23. Caldwell D, Ades T and Higgins J. Mixed treatment comparison analysis of seven treatments for acute myocardial infarction. *Cochrane Colloquium Abstracts Journal* 2005: O50.

24. Caldwell DM, Welton NJ and Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J Clin Epidemiol* 2010; 63: 875-882. DOI: DOI 10.1016/j.jclinepi.2009.08.025.

25. Cassidy J, Schmoll H, Chu E, et al. Comparative clinical efficacy of adjuvant chemotherapy regimens in randomized controlled trials of early-stage colon cancer: systematic review and meta-analysis. *J Clin Oncol* 2011; 29 (4 Suppl): 498. DOI: 10.1200/jco.2011.29.4_suppl.498.

26. Chang KV, Chen SY, Chen WS, et al. Comparative effectiveness of focused shock wave therapy of different intensity levels and radial shock wave therapy for treating plantar fasciitis: a systematic review and network meta-analysis. *Arch Phys Med Rehabil* 2012; 93: 1259-1268. 2012/03/17. DOI: 10.1016/j.apmr.2012.02.023.

27. Chapell R, Gould AL and Alexander CM. Baseline differences in A1c explain apparent differences in efficacy of sitagliptin, rosiglitazone and pioglitazone. *Diabetes, obesity & metabolism* 2009; 11: 1009-1016. 2009/07/21. DOI: 10.1111/j.1463-1326.2009.01084.x.

28. Chatterjee S, Nerella N, Guha G, et al. Comparing newer oral anti-platelets prasugrel and ticagrelor - a cumulative network meta-analysis. *J Am Coll Cardiol* 2012; 13(Suppl 1): E481. DOI: Doi 10.1016/S0735-1097(12)60482-3.

29. Choy E, Marshall D, Gabriel ZL, et al. A systematic review and mixed treatment comparison of the efficacy of pharmacological treatments for fibromyalgia. *Semin Arthritis Rheum* 2011; 41: 335-345. 2011/08/27. DOI: 10.1016/j.semarthrit.2011.06.003.

30. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746-758. DOI: 10.1016/s0140-6736(09)60046-5.

31. Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* 2011; 378: 1306-1315. 2011/08/20. DOI: 10.1016/S0140-6736(11)60873-8.

32. Cohen D, Bonnot O, Bodeau N, et al. Adverse effects of second-generation antipsychotics in children and adolescents: a bayesian meta-analysis. *J Clin Psychopharmacol* 2012; 32: 309-316. 2012/05/01. DOI: 10.1097/JCP.0b013e3182549259.

33. Coleman CI, Baker WL, Kluger J, et al. Antihypertensive medication and their impact on cancer incidence: a mixed treatment comparison meta-analysis of randomized controlled trials. *J Hypertens* 2008; 26: 622-629. DOI: 10.1097/Hjh.0b013e3282f3ef5e.

34. Cooper NJ, Sutton AJ, Lu G, et al. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Arch Intern Med* 2006; 166: 1269-1275. 2006/06/28. DOI: 10.1001/archinte.166.12.1269.

35. Cope S, Capkun-Niggli G, Gale R, et al. Comparative efficacy of indacaterol 150 mug and 300 mug versus fixed-dose combinations of formoterol + budesonide or salmeterol + fluticasone for the treatment of chronic obstructive pulmonary disease-a network meta-analysis. *International journal of chronic obstructive pulmonary disease* 2011; 6: 329-344. 2011/06/24. DOI: 10.2147/COPD.S18759.

36. Cope S, Zhang J, Raparla S, et al. Comparative efficacy of once-daily indacaterol 75g in COPD in terms of forced expiratory volume: a patient level network meta-analysis. *Chest* 2011; 140: 556A. DOI: 10.1378/chest.1119756.

37. Corbett M, Rice S, Slack R, et al. Acupuncture and other physical treatments for the relief of chronic pain due to osteoarthritis of the knee: a systematic review and network meta-analysis. *CRD Reports 40*,

http://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?ID=12012028486 (2012, accessed 27 September 2012).

38. Cummins E, Asseburg C, Punekar YS, et al. Cost-effectiveness of infliximab for the treatment of active and progressive psoriatic arthritis. *Value Health* 2011; 14: 15-23. 2011/01/08. DOI: 10.1016/j.jval.2010.10.016.

39. Curtin F and Schulz P. Clonazepam and lorazepam in acute mania: a bayesian meta-analysis. *J Affect Disord* 2004; 78: 201-208. 2004/03/12. DOI: 10.1016/S0165-0327(02)00317-8.

40. Dakin H, Fidler C and Harper C. Mixed treatment comparison meta-analysis evaluating the relative efficacy of nucleos(t)ides for treatment of nucleos(t)ide-naive patients with chronic hepatitis B. *Value Health* 2010; 13: 934-945. 2010/09/10. DOI: 10.1111/j.1524-4733.2010.00777.x.

41. Danchin N, Marzilli M, Parkhomenko A, et al. Efficacy comparison of trimetazidine with therapeutic alternatives in stable angina pectoris: a network meta-analysis. *Cardiology* 2011; 120: 59-72. 2011/11/30. DOI: 10.1159/000332369.

42. Daniels JP, Middleton LJ, Champaneria R, et al. Second generation endometrial ablation techniques for heavy menstrual bleeding: network meta-analysis. *BMJ* 2012; 344: e2564. 2012/04/25. DOI: 10.1136/bmj.e2564.

43. Del Santo F, Maratea D, Fadda V, et al. Treatments for relapsing-remitting multiple sclerosis: summarising current information by network meta-analysis. *Eur J Clin Pharmacol* 2012; 68: 441-448. DOI: 10.1007/s00228-011-1141-1.

44. Devine EB, Alfonso-Cristancho R and Sullivan SD. Effectiveness of biologic therapies for rheumatoid arthritis: an indirect comparisons approach. *Pharmacotherapy* 2011; 31: 39-51. 2010/12/25. DOI: 10.1592/phco.31.1.39.

45. Dewilde S and Hawkins N. Investigating incoherence gives insight: clopidogrel is equivalent to extended-release dipyridamole plus aspirin in secondary stroke prevention. *J Clin Epidemiol* 2012; 65: 835-845. 2012/06/26. DOI: 10.1016/j.jclinepi.2012.01.019.

46. Diels J, Cure S and Gavart S. The comparative efficacy of telaprevir versus boceprevir in treatment-naive and treatment-experienced patients with genotype 1 chronic hepatitis. *Value Health* 2011; 14: A266. DOI: 10.1016/j.jval.2011.08.195.

47. Djulbegovic M, Djulbegovic B, Zarei M, et al. Advancing comparative effectiveness research through network meta-analysis: intravesical therapy in bladder cancer. *J Urol* 2012; 187(4 Suppl): e29. DOI: 10.1016/j.juro.2012.02.113.

48. Dunkley AJ, Charles K, Gray LJ, et al. Effectiveness of interventions for reducing diabetes and cardiovascular disease risk in people with metabolic syndrome: systematic review and mixed treatment comparison meta-analysis. *Diabetes, obesity & metabolism* 2012; 14: 616-625. 2012/01/31. DOI: 10.1111/j.1463-1326.2012.01571.x.

49. Edwards SJ, Clarke MJ, Wordsworth S, et al. Carbapenems versus other beta-lactams in the treatment of hospitalised patients with infection: a mixed treatment comparison. *Curr Med Res Opin* 2009; 25: 251-261. 2009/02/13. DOI: 10.1185/03007990802633160.

50. Edwards SJ, Lind T, Lundell L, et al. Systematic review: standard- and double-dose proton pump inhibitors for the healing of severe erosive oesophagitis - a mixed

treatment comparison of randomized controlled trials. *Aliment Pharmacol Ther* 2009; 30: 547-556. 2009/06/30. DOI: 10.1111/j.1365-2036.2009.04077.x.

51. Edwards SJ and Smith CJ. Tolerability of atypical antipsychotics in the treatment of adults with schizophrenia or bipolar disorder: a mixed treatment comparison of randomized controlled trials. *Clin Ther* 2009; 31(Pt 1): 1345-1359. 2009/08/25. DOI: 10.1016/j.clinthera.2009.07.004.

52. Edwards SJ, Welton N and Borrill J. Gefitinib compared with doublet chemotherapy for first-line treatment of non-small cell lung cancer: a systematic review and adjusted indirect comparison. *Value Health* 2010; 13: A252-253. DOI: 10.1016/S1098-3015(11)71908-8.

53. Edwards SJ, Welton N and Borrill J. Tolerability of first-line treatments of locally advanced or metastatic non-small cell lung cancer: a systematic review and adjusted indirect comparison. *Value Health* 2010; 13: A250. DOI: 10.1016/S1098-3015(11)71898-8.

54. Elliott WJ and Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet* 2007; 369: 201-207. 2007/01/24. DOI: 10.1016/S0140-6736(07)60108-1.

55. Elliott WJ, Basu S and Meyer PM. One-year discontinuation rates of antihypertensive drugs in clinical practice: a network meta-analysis. *J Clin Hypertens* 2007; 9(5SupplA): A210. DOI: 10.1111/j.1751-7176.2007.tb00068.x.

56. Ellis A, Jansen J, Luo X, et al. Indirect comparison of bazedoxifene vs. oral bisphosphonates for the prevention of nonvertebral fractures in a high-risk postmenopausal osteoporosis population. *Osteoporos Int* 2012; 23(Suppl 2): S64-65. DOI: 10.1007/s00198-012-1923-z.

57. Ernest PJ, Viechtbauer W, Schouten JS, et al. The influence of the assessment method on the incidence of visual field progression in glaucoma: a network meta-analysis. *Acta Ophthalmol* 2012; 90: 10-19. 2010/10/15. DOI: 10.1111/j.1755-3768.2010.01995.x.

58. Fan T, Bennett H, Smith N, et al. Mixed treatment comparison of infliximab with ustekinumab in patients with moderate to severe psoriasis. *Br J Dermatol* 2011; 165: E38-E39. DOI: 10.1111/j.1365-2133.2011.10671.x.

59. Farris M, LeReun C, Lee J, et al. Mixed treatment comparison of prophylaxis regimens for the prevention of venous thromboembolism in total hip replacement, hip fracture, total knee replacement and general surgery. *Value Health* 2010; 13: A517. DOI: 10.1016/S1098-3015(11)73132-1.

60. Freemantle N, Lafuente-Lafuente C, Mitchell S, et al. Mixed treatment comparison of dronedarone, amiodarone, sotalol, flecainide, and propafenone, for the management of atrial fibrillation. *Europace* 2011; 13: 329-345. 2011/01/14. DOI: 10.1093/europace/euq450.

61. Freemantle N, Tharmanathan P and Herbrecht R. Systematic review and mixed treatment comparison of randomized evidence for empirical, pre-emptive and directed treatment strategies for invasive mould disease. *The Journal of antimicrobial chemotherapy* 2011; 66(Suppl 1): i25-35. 2011/01/05. DOI: 10.1093/jac/dkq439.

62. Fretheim A, Odgaard-Jensen J, Brors O, et al. Comparative effectiveness of antihypertensive medication for primary prevention of cardiovascular disease: systematic

review and multiple treatments meta-analysis. *BMC Med* 2012; 10: 33. DOI: 10.1186/1741-7015-10-33.

63. Fusaro M, Amoroso N, Kumar S, et al. Outcomes with drug eluting or bare metal stents in subjects with diabetes mellitus: A mixed treatment comparison analysis of 19,325 patient-years of follow-up from randomized trials. *J Am Coll Cardiol* 2012; 59: E98. DOI: 10.1016/S0735-1097(12)60099-0.

64. Gartlehner G, Gaynes BN, Hansen RA, et al. Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. *Ann Intern Med* 2008; 149: 734-750. 2008/11/20. DOI: 10.7326/0003-4819-149-10-200811180-00008.

65. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med* 2011; 155: 772-785. 2011/12/08. DOI: 10.1059/0003-4819-155-11-201112060-00009.

66. Golfinopoulos V, Pentheroudakis G, Salanti G, et al. Comparative survival with diverse chemotherapy regimens for cancer of unknown primary site: multiple-treatments meta-analysis. *Cancer Treat Rev* 2009; 35: 570-573. 2009/06/23. DOI: 10.1016/j.ctrv.2009.05.005.

67. Golfinopoulos V, Salanti G, Pavlidis N, et al. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: a meta-analysis. *Lancet Oncol* 2007; 8: 898-911. 2007/09/25. DOI: 10.1016/S1470-2045(07)70281-4.

68. Gray LJ, Cooper N, Dunkley A, et al. A systematic review and mixed treatment comparison of pharmacological interventions for the treatment of obesity. *Obesity reviews : an official journal of the International Association for the Study of Obesity* 2012; 13: 483-498. 2012/02/01. DOI: 10.1111/j.1467-789X.2011.00981.x.

69. Greenhalgh J, Bagust A, Boland A, et al. Clopidogrel and modified-release dipyridamole for the prevention of occlusive vascular events (review of technology appraisal No. 90): a systematic review and economic analysis. *Health Technol Assess* 2011; 15(31) 2011/09/06. DOI: 10.3310/hta15310.

70. Gross JL, Kramer CK, Leitao CB, et al. Effect of antihyperglycemic agents added to metformin and a sulfonylurea on glycemic control and weight gain in type 2 diabetes: a network meta-analysis. *Ann Intern Med* 2011; 154: 672-679. 2011/05/18. DOI: 10.1059/0003-4819-154-10-201105170-00007.

71. Gurusamy KS, Pissanou T, Pikhart H, et al. Methods to decrease blood loss and transfusion requirements for liver transplantation. *Cochrane Database Syst Rev* 2011; 12: CD009052. 2011/12/14. DOI: 10.1002/14651858.CD009052.pub2.

72. Guyot P, Taylor P, Christensen R, et al. Abatacept with methotrexate versus other biologic agents in treatment of patients with active rheumatoid arthritis despite methotrexate: a network meta-analysis. *Arthritis research & therapy* 2011; 13: R204. 2011/12/14. DOI: 10.1186/ar3537.

73. Guyot P, Taylor PC, Christensen R, et al. Indirect treatment comparison of abatacept with methotrexate versus other biologic agents for active rheumatoid arthritis despite methotrexate therapy in the United kingdom. *The Journal of rheumatology* 2012; 39: 1198-1206. 2012/04/17. DOI: 10.3899/jrheum.111345.

74. Hansen RA, Gaynes BN, Gartlehner G, et al. Efficacy and tolerability of second-generation antidepressants in social anxiety disorder. *Int Clin Psychopharmacol* 2008; 23: 170-179. 2008/04/15. DOI: 10.1097/YIC.0b013e3282f4224a.

75. Harrington AR and Malone D. Diet therapies in patients with type-2 diabetes: a mixed treatment comparison of randomized control trials. *Value Health* 2011; 14: A92. DOI: 10.1016/j.jval.2011.02.515.

76. Hartling L, Fernandes RM, Bialy L, et al. Steroids and bronchodilators for acute bronchiolitis in the first two years of life: systematic review and meta-analysis. *BMJ* 2011; 342: d1714. 2011/04/08. DOI: 10.1136/bmj.d1714.

77. Hawkins N, Scott DA, Woods BS, et al. No study left behind: a network meta-analysis in non-small cell lung cancer demonstrating the importance of considering all relevant data. *Value Health* 2009; 12: 996-1003. DOI: DOI 10.1111/j.1524-4733.2009.00541.x.

78. Hopkins RB, Goeree R, Pullenayegum E, et al. The relative efficacy of nine osteoporosis medications for reducing the rate of fractures in post-menopausal women. *BMC Musculoskelet Disord* 2011; 12: 209. 2011/09/29. DOI: 10.1186/1471-2474-12-209.

79. Ibrahim T, Qureshi A, Sutton AJ, et al. Surgical versus nonsurgical treatment of acute minimally displaced and undisplaced scaphoid waist fractures: pairwise and network meta-analyses of randomized controlled trials. *The Journal of hand surgery* 2011; 36: 1759-1768 e1751. 2011/11/01. DOI: 10.1016/j.jhsa.2011.08.033.

80. Imamura M, Abrams P, Bain C, et al. Systematic review and economic modelling of the effectiveness and cost-effectiveness of non-surgical treatments for women with stress urinary incontinence. *Health Technol Assess* 2010; 14(40) 2010/08/27. DOI: 10.3310/hta14400.

81. Jansen JP. Self-monitoring of glucose in type 2 diabetes mellitus: a bayesian meta-analysis of direct and indirect comparisons. *Curr Med Res Opin* 2006; 22: 671-681. 2006/05/11. DOI: 10.1185/030079906X96308.

82. Jansen FH, Lesaffre E, Penali LK, et al. Assessment of the relative advantage of various artesunate-based combination therapies by a multi-treatment bayesian random-effects meta-analysis. *Am J Trop Med Hyg* 2007; 77: 1005-1009. DOI: 10.4269/ajtmh.2007.77.1005.

83. Jansen JP, Bergman GJ, Huels J, et al. Prevention of vertebral fractures in osteoporosis: mixed treatment comparison of bisphosphonate therapies. *Curr Med Res Opin* 2009; 25: 1861-1868. 2009/06/18. DOI: 10.1185/03007990903035281.

84. Jansen JP, Gaugris S, Choy EH, et al. Cost effectiveness of etoricoxib versus celecoxib and non-selective NSAIDS in the treatment of ankylosing spondylitis. *Pharmacoeconomics* 2010; 28: 323-344. 2010/03/13. DOI: 10.2165/11314690-000000000-00000.

85. Jansen JP, Bergman GJ, Huels J, et al. The efficacy of bisphosphonates in the prevention of vertebral, hip, and nonvertebral-nonhip fractures in osteoporosis: a network meta-analysis. *Semin Arthritis Rheum* 2011; 40: 275-284e272. 2010/09/11. DOI: 10.1016/j.semarthrit.2010.06.001.

86. Kessler TM, Bachmann LM, Minder C, et al. Adverse event assessment of antimuscarinics for treating overactive bladder: a network meta-analytic approach. *PLoS One* 2011; 6: e16718. 2011/03/05. DOI: 10.1371/journal.pone.0016718.

87. King S, Griffin S, Hodges Z, et al. A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. *Health Technol Assess* 2006; 10(23) 2006/06/27. DOI: 10.3310/hta10230.

88. Klemp M, Tvete IF, Skomedal T, et al. A review and bayesian meta-analysis of clinical efficacy and adverse effects of 4 atypical neuroleptic drugs compared with haloperidol and placebo. *J Clin Psychopharmacol* 2011; 31: 698-704. 2011/10/25. DOI: 10.1097/JCP.0b013e31823657d9.

89. Knight C, Howard P, Baker CL, et al. The cost-effectiveness of an extended course (12+12 weeks) of varenicline compared with other available smoking cessation strategies in the United States: an extension and update to the BENESCO model. *Value Health* 2010; 13: 209-214. 2009/11/17. DOI: 10.1111/j.1524-4733.2009.00672.x.

90. Kruidenier LM, Viechtbauer W, Nicolai SP, et al. Treatment for intermittent claudication and the effects on walking distance and quality of life. *Vascular* 2012; 20: 20-35. 2012/01/25. DOI: 10.1258/vasc.2011.ra0048.

91. Kyrgiou M, Salanti G, Pavlidis N, et al. Survival benefits with diverse chemotherapy regimens for ovarian cancer: meta-analysis of multiple treatments. *J Natl Cancer Inst* 2006; 98: 1655-1663. 2006/11/16. DOI: 10.1093/jnci/djj443.

92. Lam SK and Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: bayesian network meta-analysis of randomised controlled trials. *BMJ* 2007; 335: 925. 2007/10/13. DOI: 10.1136/bmj.39343.511389.BE.

93. Lang SH, Manning N, Armstrong N, et al. Treatment with tirofiban for acute coronary syndrome (ACS): a systematic review and network analysis. *Curr Med Res Opin* 2012; 28: 351-370. 2012/02/02. DOI: 10.1185/03007995.2012.657299.

94. Launois R, Avouac B, Berenbaum F, et al. Comparison of certolizumab pegol with other anticytokine agents for treatment of rheumatoid arthritis: a multiple-treatment bayesian metaanalysis. *The Journal of rheumatology* 2011; 38: 835-845. 2011/01/18. DOI: 10.3899/jrheum.100665.

95. Lebmeier M, Pericleous L, Guyot P, et al. Indirect treatment comparison to compare efficacy in health assessment questionnaire (HAQ) score for biologic agents with methotrexate in patients with rheumatoid arthritis and active disease despite methotrexate therapy. *Value Health* 2010; 13: A316. DOI: 10.1016/S1098-3015(11)72224-0.

96. Lebmeier M, Pericleous L, Taylor PC, et al. Mixed treatment comparison of biologic agents in patients with rheumatoid arthritis who have responded inadequately to methotrexate therapy in the United Kingdom. *Value Health* 2011; 14: A317. DOI: 10.1016/j.jval.2011.08.465.

97. Lerdkiattikorn P, Chaikledkaew U, Kingkaew P, et al. A systematic review and meta-analysis of adjuvant chemotherapy for stage III colon cancer. *Value Health* 2010; 13: A503. DOI: 10.1016/S1098-3015(11)73059-5.

98. LeReun C, Tilden D, Harvey C, et al. Efficacy and safety of raltegravir in treatment naive HIV+ patients: a mixed treatment comparison approach. *Value Health* 2010; 13: A548. DOI: 10.1016/S1098-3015(11)73289-2.

99. Lewis R, Williams N, Matar HE, et al. The clinical effectiveness and cost-effectiveness of management strategies for sciatica: systematic review and economic model. *Health Technol Assess* 2011; 15(39) 2011/11/15. DOI: 10.3310/hta15390.

100. Logman JF, Stephens J, Heeg B, et al. Comparative effectiveness of antibiotics for the treatment of MRSA complicated skin and soft tissue infections. *Curr Med Res Opin* 2010; 26: 1565-1578. 2010/05/01. DOI: 10.1185/03007995.2010.481251.

101. Manzoli L, Salanti G, De Vito C, et al. Immunogenicity and adverse events of avian influenza A H5N1 vaccine in healthy adults: multiple-treatments meta-analysis. 2009; 9: 482-492. 8.

102. Manzoli L, De Vito C, Salanti G, et al. Meta-analysis of the immunogenicity and tolerability of pandemic influenza A 2009 (H1N1) vaccines. *PLoS One* 2011; 6: e24384. 9. DOI: 10.1371/journal.pone.0024384.

103. Maund E, Craig D, Suekarran S, et al. Management of frozen shoulder: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2012; 16(11) 2012/03/13. DOI: 10.3310/hta16110.

104. Mauri D, Polyzos NP, Salanti G, et al. Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *J Natl Cancer Inst* 2008; 100: 1780-1791. 2008/12/11. DOI: 10.1093/jnci/djn414.

105. McDaid C, Maund E, Rice S, et al. Paracetamol and selective and non-selective non-steroidal anti-inflammatory drugs (NSAIDs) for the reduction of morphine-related side effects after major surgery: a systematic review. *Health Technol Assess* 2010; 14(17) 2010/03/30. DOI: 10.3310/hta14170.

106. McIntosh B, Cameron C, Singh SR, et al. Second-line therapy in patients with type 2 diabetes inadequately controlled with metformin monotherapy: a systematic review and mixed-treatment comparison meta-analysis. *Open Med* 2011; 5: e35-48. 2011/11/03.

107. McNamara S, Moore L and Ray J. A cost-utility analysis on the use of trastuzumab plus anastrazole compared to lapatinib plus letrozole, letrozole monotherapy or anastrazole monotherapy in the treatment of HER2+/hormone receptor positive (HR plus) metastatic breast cancer from the perspective of the UK National Health Service (NHS). *Value Health* 2010; 13: A269.

108. Meader N. A comparison of methadone, buprenorphine and alpha(2) adrenergic agonists for opioid detoxification: a mixed treatment comparison meta-analysis. *Drug and alcohol dependence* 2010; 108: 110-114. 2010/01/16. DOI: 10.1016/j.drugalcdep.2009.12.008.

109. Mhaskar R, Redzepovic J, Wheatley K, et al. Bisphosphonates in multiple myeloma: a network meta-analysis. *Cochrane Database Syst Rev* 2012; 5: CD003188. 2012/05/18. DOI: 10.1002/14651858.CD003188.pub3.

110. Migliore A, Bizzi E, Broccoli S, et al. Indirect comparison of etanercept, infliximab, and adalimumab for psoriatic arthritis: mixed treatment comparison using placebo as common comparator. *Clin Rheumatol* 2012; 31: 133-137. DOI: DOI 10.1007/s10067-011-1790-6.

111. Migliore A, Broccoli S, Bizzi E, et al. Indirect comparison of the effects of anti-TNF biological agents in patients with ankylosing spondylitis by means of a mixed treatment comparison performed on efficacy data from published randomised, controlled trials. *J Med Econ* 2012; 15: 473-480. 2012/02/18. DOI: 10.3111/13696998.2012.660255.

112. Mills EJ, Rachlis B, Wu P, et al. Primary prevention of cardiovascular mortality and events with statin treatments: a network meta-analysis involving more than 65,000 patients. *J Am Coll Cardiol* 2008; 52: 1769-1781. DOI: 10.1016/j.jacc.2008.08.039.

113. Mills EJ, Perri D, Cooper C, et al. Antifungal treatment for invasive candida infections: a mixed treatment comparison meta-analysis. *Annals of clinical microbiology and antimicrobials* 2009; 8: 23. 2009/06/30. DOI: 10.1186/1476-0711-8-23.

114. Mills EJ, Druyts E, Ghement I, et al. Pharmacotherapies for chronic obstructive pulmonary disease: a multiple treatment comparison meta-analysis. *Clin Epidemiol* 2011; 3: 107-129. 2011/04/14. DOI: 10.2147/CLEP.S16235.

115. Mills EJ, Wu P, Chong G, et al. Efficacy and safety of statin treatment for cardiovascular disease: a network meta-analysis of 170,255 patients from 76 randomized trials. *QJM* 2011; 104: 109-124. 2010/10/12. DOI: 10.1093/qjmed/hcq165.

116. Murad MH, Drake MT, Mullan RJ, et al. Clinical review. Comparative effectiveness of drug treatments to prevent fragility fractures: a systematic review and network meta-analysis. *The Journal of clinical endocrinology and metabolism* 2012; 97: 1871-1880. 2012/04/03. DOI: 10.1210/jc.2011-3060.

117. Mutebi A and Malone D. A bayesian multiple treatment comparison of pulmonary arterial hypertension drug classes based on the risk of mortality reported in clinical trials. *Value Health* 2011; 14: A35. DOI: 10.1016/j.jval.2011.02.207.

118. Ni RH, Tang HL, Zhai SD, et al. Multiple treatments for infantile rotavirus enteritis: a network meta-analysis. *World Chinese Journal of Digestology* 2012; 20: 438-443.

119. Nixon R, Bansback N and Brennan A. The efficacy of inhibiting tumour necrosis factor alpha and interleukin 1 in patients with rheumatoid arthritis: a meta-analysis and adjusted indirect comparisons. *Rheumatology (Oxford)* 2007; 46: 1140-1147. 2007/05/05. DOI: 10.1093/rheumatology/kem072.

120. Numthavaj P, Thakkinstian A, Dejthevaporn C, et al. Corticosteroid and antiviral therapy for bell's palsy: a network meta-analysis. *BMC Neurol* 2011; 11: 1. 2011/01/07. DOI: 10.1186/1471-2377-11-1.

121. Orme M, Collins S, Dakin H, et al. Mixed treatment comparison and meta-regression of the efficacy and safety of prostaglandin analogues and comparators for primary open-angle glaucoma and ocular hypertension. *Curr Med Res Opin* 2010; 26: 511-528. 2009/12/18. DOI: 10.1185/03007990903498786.

122. Owen A. Antithrombotic treatment for the primary prevention of stroke in patients with non valvular atrial fibrillation: a reappraisal of the evidence and network meta analysis. *Int J Cardiol* 2010; 142: 218-223. 2010/01/12. DOI: 10.1016/j.ijcard.2009.11.045.

123. Padwal R, Klarenbach S, Wiebe N, et al. Bariatric surgery: a systematic review and network meta-analysis of randomized trials. *Obesity reviews : an official journal of the International Association for the Study of Obesity* 2011; 12: 602-621. 2011/03/29. DOI: 10.1111/j.1467-789X.2011.00866.x.

124. Palmerini T, Biondi-Zoccai G, Della Riva D, et al. Stent thrombosis with drug-eluting and bare-metal stents: evidence from a comprehensive network meta-analysis. *Lancet* 2012; 379: 1393-1402. 2012/03/27. DOI: 10.1016/S0140-6736(12)60324-9.

125. Pechlivanoglou P, Le HH, De Vries R, et al. Mixed treatment comparison of antifungal drugs for prophylaxis treatment against invasive fungal infections in patients receiving chemotherapy for hematological malignancies or allogeneic hematopoietic stem cells transplantation. *Value Health* 2010; 13: A245. DOI: 10.1016/S1098-3015(11)71873-3.

126. Phung OJ, Scholle JM, Talwar M, et al. Effect of noninsulin antidiabetic drugs added to metformin therapy on glycemic control, weight gain, and hypoglycemia in type 2 diabetes. *JAMA* 2010; 303: 1410-1418. 2010/04/15. DOI: 10.1001/jama.2010.405.

127. Phung OJ, Kahn SR, Cook DJ, et al. Dosing frequency of unfractionated heparin thromboprophylaxis: a meta-analysis. *Chest* 2011; 140: 374-381. 2011/02/26. DOI: 10.1378/chest.10-3084.

128. Phung OJ, Sood NA, Sill BE, et al. Oral anti-diabetic drugs for the prevention of type 2 diabetes. *Diabetic medicine : a journal of the British Diabetic Association* 2011; 28: 948-964. 2011/03/25. DOI: 10.1111/j.1464-5491.2011.03303.x.

129. Psaty BN, Lumley T, Furberg CD, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents - a network meta-analysis. *JAMA* 2003; 289: 2534-2544. 2003/05/22. DOI: 10.1001/jama.289.19.2534.

130. Puhan MA, Bachmann LM, Kleijnen J, et al. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Med* 2009; 7: 2. 2009/01/16. DOI: 10.1186/1741-7015-7-2.

131. Reich K, Burden AD, Eaton JN, et al. Efficacy of biologics in the treatment of moderate to severe psoriasis: a network meta-analysis of randomized controlled trials. *The British journal of dermatology* 2012; 166: 179-188. 2011/09/14. DOI: 10.1111/j.1365-2133.2011.10583.x.

132. Riemsma R, Forbes CA, Kessels A, et al. Systematic review of aromatase inhibitors in the first-line treatment for hormone sensitive advanced or metastatic breast cancer. *Breast Cancer Res Treat* 2010; 123: 9-24. 2010/06/11. DOI: 10.1007/s10549-010-0974-0.

133. Riemsma R, Forbes C, Harker J, et al. Systematic review of tapentadol in chronic severe pain. *Curr Med Res Opin* 2011; 27: 1907-1930. 2011/09/13. DOI: 10.1185/03007995.2011.611494.

134. Riemsma R, Forbes CA, Amonkar MM, et al. Systematic review of lapatinib in combination with letrozole compared with other first-line treatments for hormone receptor positive(HR+) and HER2+ advanced or metastatic breast cancer. *Curr Med Res Opin* 2012; 28: 1263-1279. 2012/06/29. DOI: 10.1185/03007995.2012.707643.

135. Roskell NS, Lip GY, Noack H, et al. Treatments for stroke prevention in atrial fibrillation: a network meta-analysis and indirect comparisons versus dabigatran etexilate. *Thrombosis and haemostasis* 2010; 104: 1106-1115. 2010/10/23. DOI: 10.1160/TH10-10-0642.

136. Roskell NS, Beard SM, Zhao Y, et al. A meta-analysis of pain response in the treatment of fibromyalgia. *Pain Pract* 2011; 11: 516-527. 2011/01/05. DOI: 10.1111/j.1533-2500.2010.00441.x.

137. Roskell NS, Zimovetz EA, Rycroft CE, et al. Annualized relapse rate of first-line treatments for multiple sclerosis: a meta-analysis, including indirect comparisons versus fingolimod. *Curr Med Res Opin* 2012; 28: 767-780. 2012/04/03. DOI: 10.1185/03007995.2012.681637.

138. Rotta I, Sanchez A, Goncalves PR, et al. Efficacy and safety of topical antifungals in the treatment of dermatomycosis: a systematic review. *The British journal of dermatology* 2012; 166: 927-933. 2012/01/12. DOI: 10.1111/j.1365-2133.2012.10815.x.

139. Roversi S, Malavasi V, D'Ascenzo F, et al. Picking the best novel oral anticoagulant for atrial fibrillation: evidence from a warfarin-controlled network meta-analysis. *J Am Coll Cardiol* 2012; 59(13Suppl1): E598. DOI: 10.1016/s0735-1097(12)60599-3.

140. Salanti G, Marinho V and Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; 62: 857-864. DOI: DOI 10.1016/j.jclinepi.2008.10.001.

141. Sanches ACC, Correr CJ, Venson R, et al. Revisiting the efficacy of long-acting insulin analogues on adults with type 1 diabetes using mixed-treatment comparisons. *Diabetes Res Clin Pract* 2011; 94: 333-339. DOI: DOI 10.1016/j.diabres.2011.09.001.

142. Schauble B, Modha R, Wieffer H, et al. Establishing the comparative efficacy and safety of acetylcholinesterase inhibitors and memantine using mixed treatment comparison. *10th International Conference on Alzheimer's and Parkinson's Diseases*. Barcelona, Spain2011.

143. Schmitz S, Adams R, Walsh CD, et al. A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a bayesian approach. *Ann Rheum Dis* 2012; 71: 225-230. DOI: DOI 10.1136/annrheumdis-2011-200228.

144. Sciarretta S, Palano F, Tocci G, et al. Antihypertensive treatment and development of heart failure in hypertension: a bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. *Arch Intern Med* 2011; 171: 384-394. 2010/11/10. DOI: 10.1001/archinternmed.2010.427.

145. Scott D, Boye KS, Timlin L, et al. A network meta analysis to compare glycaemic control in patients with type 2 diabetes treated with exenatide once weekly or liraglutide. *Value Health* 2011; 14: A472-473. DOI: 10.1016/j.jval.2011.08.1311.

146. Senior E, Dedeken P and Naci H. Dopamine agonists and dyskinesia in advanced parkinson's disease: a network meta-analysis of rotigotine, pramipexole and ropinirole as adjunct therapy to levodopa. *Movement disorders : official journal of the Movement Disorder Society* 2012; 27(Suppl 1): S138. 2012/06/22. DOI: 10.1002/mds.25051.

147. Siddiqui MK, Gupta J, Bhutani M, et al. Opioids in non-malignant pain: Are they equivalent in safety profile? A network meta-analysis. *Value Health* 2011; 14: A59. DOI: 10.1016/j.jval.2011.02.336.

148. Singh JA, Christensen R, Wells GA, et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. *CMAJ* 2009; 181: 787-796. 2009/11/04. DOI: 10.1503/cmaj.091391.

149. Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. *Cochrane Database Syst Rev* 2011: CD008794. 2011/02/18. DOI: 10.1002/14651858.CD008794.pub2.

150. Soares-Weiser K, Bravo Vergel Y, Beynon S, et al. A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder. *Health Technol Assess* 2007; 11(39) 2007/10/02. DOI: 10.3310/hta11390.

151. Squires H, Simpson E, Meng Y, et al. A systematic review and economic evaluation of cilostazol, naftidrofuryl oxalate, pentoxifylline and inositol nicotinate for the treatment of intermittent claudication in people with peripheral arterial disease. *Health Technol Assess* 2011; 15(40) 2011/12/07. DOI: 10.3310/hta15400.

152. Stam WB, Jansen JP and Taylor SD. Efficacy of etoricoxib, celecoxib, lumiracoxib, non-selective NSAIDs, and acetaminophen in osteoarthritis: A mixed treatment comparison. *Open Rheumatol J* 2012; 6: 6-20. DOI: 10.2174/1874312901206010006.

153. Steiner S, Moertl D, Chen L, et al. Network meta-analysis of prasugrel, ticagrelor, high- and standard-dose clopidogrel in patients scheduled for percutaneous coronary interventions. *Thrombosis and haemostasis* 2012; 108: 318-327. 2012/05/26. DOI: 10.1160/TH11-08-0586.

154. Stettler C, Wandel S, Allemann S, et al. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *Lancet* 2007; 370: 937-948. 2007/09/18. DOI: 10.1016/S0140-6736(07)61444-5.

155. Stettler C, Allemann S, Wandel S, et al. Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis. *BMJ* 2008; 337: a1331. DOI: 10.1136/bmj.a1331.

156. Stradwick S, Hartmann J, Morgan A, et al. Comparative effectiveness of investigational compound ferumoxytol for the treatment of iron deficiency anaemia in chronic kidney disease: systematic review and mixed treatment comparison. *Value Health* 2011; 14: A330. DOI: 10.1016/j.jval.2011.08.537.

157. Strassmann R, Bausch B, Spaar A, et al. Smoking cessation interventions in COPD: a network meta-analysis of randomised trials. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2009; 34: 634-640. 2009/04/10. DOI: 10.1183/09031936.00167708.

158. Tang DH and Malone DC. A network meta-analysis on the efficacy of serotonin type 3 receptor antagonists used in adults during the first 24 hours for postoperative nausea and vomiting prophylaxis. *Clin Ther* 2012; 34: 282-294. DOI: DOI 10.1016/j.clinthera.2012.01.007.

159. Tantai N, Chaikledkaew U, Werayingyong P, et al. A systematic review and meta-analysis of literature assessing clinical efficacy of treatments for patients with HbeAg-positive chronic hepatitis B. *Value Health* 2010; 13: A531-532. DOI: 10.1016/S1098-3015(11)73205-3.

160. Thijs V, Lemmens R and Fieuws S. Network meta-analysis: simultaneous meta-analysis of common antiplatelet regimens after transient ischaemic attack or stroke. *Eur Heart J* 2008; 29: 1086-1092. 2008/03/20. DOI: 10.1093/eurheartj/ehn106.

161. Tosh JC, Wailoo AJ, Scott DL, et al. Cost-effectiveness of combination nonbiologic disease-modifying antirheumatic drug strategies in patients with early rheumatoid arthritis. *The Journal of rheumatology* 2011; 38: 1593-1600. 2011/05/17. DOI: 10.3899/jrheum.101327.

162. Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *BMJ* 2011; 342: c7086. 2011/01/13. DOI: 10.1136/bmj.c7086.

163. Trikalinos TA, Alsheikh-Ali AA, Tatsioni A, et al. Percutaneous coronary interventions for non-acute coronary artery disease: a quantitative 20-year synopsis and a network meta-analysis. *Lancet* 2009; 373: 911-918. 2009/03/17. DOI: 10.1016/S0140-6736(09)60319-6.

164. Tropeano AI, Saleh N, Hawajri N, et al. Do all antihypertensive drugs improve carotid intima-media thickness? A network meta-analysis of randomized controlled trials. *Fundam Clin Pharmacol* 2011; 25: 395-404. DOI: 10.1111/j.1472-8206.2010.00832.x.

165. Tu YK, Woolston A and Faggion CM. Do bone grafts or barrier membranes provide additional treatment effects for infrabony lesions treated with enamel matrix derivatives? A network meta-analysis of randomized-controlled trials. *J Clin Periodontol* 2010; 37: 59-79. DOI: DOI 10.1111/j.1600-051X.2009.01499.x.

166. Tu YK, Needleman I, Chambrone L, et al. A bayesian network meta-analysis on comparisons of enamel matrix derivatives, guided tissue regeneration and their combination therapies. *J Clin Periodontol* 2012; 39: 303-314. DOI: DOI 10.1111/j.1600-051X.2011.01844.x.

167. Tudur Smith C, Marson AG, Chadwick DW, et al. Multiple treatment comparisons in epilepsy monotherapy trials. *Trials* 2007; 8: 34. 2007/11/07. DOI: 10.1186/1745-6215-8-34.

168. Turkstra E, Ng SK and Scuffham PA. A mixed treatment comparison of the short-term efficacy of biologic disease modifying anti-rheumatic drugs in established rheumatoid arthritis. *Curr Med Res Opin* 2011; 27: 1885-1897. 2011/08/19. DOI: 10.1185/03007995.2011.608655.

169. Uthman OA and Abdulmalik J. Comparative efficacy and acceptability of pharmacotherapeutic agents for anxiety disorders in children and adolescents: a mixed treatment comparison meta-analysis. *Curr Med Res Opin* 2010; 26: 53-59. 2009/11/13. DOI: 10.1185/03007990903416853.

170. van de Kerkhof P, de Peuter R, Ryttov J, et al. Mixed treatment comparison of a two-compound formulation product containing calcipotriol and betamethasone dipropionate with other topical treatments in psoriasis vulgaris. *Curr Med Res Opin* 2011; 27: 225-238. 2010/12/15. DOI: 10.1185/03007995.2010.541005.

171. Van den Bruel A, Gailly J, Devriese S, et al. The protective effect of ophthalmic viscoelastic devices on endothelial cell loss during cataract surgery: a meta-analysis using mixed treatment comparisons. *Br J Ophthalmol* 2009; 95: 5-10. DOI: 10.1136/bjo.2009.158360.

172. van der Valk R, Webers CA, Lumley T, et al. A network meta-analysis combined direct and indirect comparisons between glaucoma drugs to rank effectiveness in lowering intraocular pressure. *J Clin Epidemiol* 2009; 62: 1279-1283. DOI: 10.1016/j.jclinepi.2008.04.012.

173. Vansteenkiste J, Edelman MJ, de Marinis F, et al. Indirect comparison of pemetrexed plus cisplatin with other platinum-based therapies for first-line advanced non-small cell lung cancer. *Proceedings of the Internal Thoracic Oncology Congress*. Dresden, Germany 2008.

174. Vejakama P, Thakkinstian A, Lertrattananon D, et al. Reno-protective effects of renin-angiotensin system blockade in type 2 diabetic patients: a systematic review and network meta-analysis. *Diabetologia* 2012; 55: 566-578. 2011/12/23. DOI: 10.1007/s00125-011-2398-8.

175. Verheggen B, Vandeloise E, Treur M, et al. Comparing the efficacy and tolerability of anti-retroviral therapy in treatment-naive HIV-1 infected adults: a systematic review of randomized clinical trials and bayesian mixed treatment comparisons including atazanavir/r, darunavir/r, lopinavir/r, and efavirenz. *Value Health* 2011; 14: A265-266. DOI: 10.1016/j.jval.2011.08.191.

176. Vieira MC, Kumar RN and Jansen JP. Comparative effectiveness of efavirenz, protease inhibitors, and raltegravir-based regimens as first-line treatment for HIV-infected adults: a mixed treatment comparison. *HIV Clin Trials* 2011; 12: 175-189. 2011/11/03. DOI: 10.1310/HCT1204-175.

177. Virgili G, Novielli N, Menchini F, et al. Pharmacological treatments for neovascular age-related macular degeneration: can mixed treatment comparison meta-analysis be useful? *Curr Drug Targets* 2011; 12: 212-220. DOI: 10.2174/138945011794182665.

178. Vissers D, Stam W, Nolte T, et al. Efficacy of intranasal fentanyl spray versus other opioids for breakthrough pain in cancer. *Curr Med Res Opin* 2010; 26: 1037-1045. 2010/03/05. DOI: 10.1185/03007991003694340.

179. Vlaar PJ, Mahmoud KD, Holmes DR, Jr., et al. Culprit vessel only versus multivessel and staged percutaneous coronary intervention for multivessel disease in patients presenting with ST-segment elevation myocardial infarction: a pairwise and network meta-analysis. *J Am Coll Cardiol* 2011; 58: 692-703. 2011/08/06. DOI: 10.1016/j.jacc.2011.03.046.

180. Walsh T, Worthington HV, Glenny AM, et al. Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents. *Cochrane Database Syst Rev* 2010: CD007868. 2010/01/22. DOI: 10.1002/14651858.CD007868.pub2.

181. Wandel S, Juni P, Tendal B, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. *BMJ* 2010; 341: c4675. 2010/09/18. DOI: 10.1136/bmj.c4675.

182. Wang H, Huang T, Jing J, et al. Effectiveness of different central venous catheters for catheter-related infections: a network meta-analysis. *The Journal of hospital infection* 2010; 76: 1-11. 2010/07/20. DOI: 10.1016/j.jhin.2010.04.025.

183. Wang SY, Chu H, Shamliyan T, et al. Network meta-analysis of margin threshold for women with ductal carcinoma in situ. *J Natl Cancer Inst* 2012; 104: 507-516. 2012/03/24. DOI: 10.1093/jnci/djs142.

184. Welton NJ, Cooper NJ, Ades AE, et al. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Stat Med* 2008; 27: 5620-5639. 2008/08/06. DOI: 10.1002/sim.3377.

185. Welton NJ, Caldwell DM, Adamopoulos E, et al. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *Am J Epidemiol* 2009; 169: 1158-1165. 2009/03/05. DOI: 10.1093/aje/kwp014.

186. Wiebe N, Padwal R, Field C, et al. A systematic review on the effect of sweeteners on glycemic response and clinically relevant outcomes. *BMC Med* 2011; 9: 123. 2011/11/19. DOI: 10.1186/1741-7015-9-123.

187. Wojciechowski P, Stozek A, Szmyd J, et al. Safety and efficacy of tenofovir as compared to other nucleot(s)ide analogues in the treatment of chronic hepatitis B - a systematic review with mixed treatment comparison. *Value Health* 2011; 14: A265. DOI: 10.1016/j.jval.2011.08.190.

188. Wolff RF, Bala MM, Westwood M, et al. 5% lidocaine medicated plaster in painful diabetic peripheral neuropathy: a systematic review. *Swiss Med Wkly* 2010; 140: 297-306. 2010/05/12. DOI: smw-12995.

189. Wolff RF, Bala MM, Westwood M, et al. 5% lidocaine-medicated plaster vs other relevant interventions and placebo for post-herpetic neuralgia: a systematic review. *Acta Neurol Scand* 2011; 123: 295-309. 2010/11/03. DOI: 10.1111/j.1600-0404.2010.01433.x.

190. Wolff RF, Aune D, Truyers C, et al. Systematic review of efficacy and safety of buprenorphine versus fentanyl or morphine in patients with chronic moderate to severe pain. *Curr Med Res Opin* 2012; 28: 833-845. 2012/03/27. DOI: 10.1185/03007995.2012.678938.

191. Woo G, Tomlinson G, Nishikawa Y, et al. Tenofovir and entecavir are the most effective antiviral agents for chronic hepatitis B: a systematic review and bayesian meta-analyses. *Gastroenterology* 2010; 139: 1218-1229. DOI: DOI 10.1053/j.gastro.2010.06.042.

192. Woolacott N, Hawkins N, Mason A, et al. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health Technol Assess* 2006; 10(46) 2006/11/07. DOI: 10.3310/hta10460.

193. Wu HY, Tu YK and Chien KL. Comparative effectiveness of angiotensin-converting enzyme inhibitors or angiotensin II-receptor blockers for diabetic nephropathy: a bayesian network meta-analysis. *J Hypertens* 2011; 29(e-Suppl B): e22. DOI: 10.1097/01.hjh.0000408043.77355.5a.

194. Wu HY, Peng YS, Huang JW, et al. Comparative efficacy of various renin-angiotensin system blockers and other antihypertensive drugs in diabetic nephropathy: A systematic review and bayesian network meta-analysis. *Nephrol Dial Transplant* 2012; 27(Suppl 2): ii167. DOI: 10.1093/ndt/gfs221.

195. Wu HY, Peng YS, Hung KY, et al. Comparative effectiveness of various antihypertensive agents in diabetic patients: a systematic review and bayesian network meta-analysis. *Nephrol Dial Transplant* 2012; 27(Suppl 2): ii88-ii89. DOI: 10.1093/ndt/gfs214.

196. Yi Y, Chouaid C, Vergnenegre A, et al. Mixed treatment comparison of bevacizumab-based therapies relative to doublet-chemotherapy combinations to estimate the relative efficacy in progression-free survival for treatment of first-line advanced or metastatic non-small cell lung cancer. *Value Health* 2010; 13: A253-254. DOI: 10.1016/S1098-3015(11)71913-1.

197. Youn JH, Lord J, Hemming K, et al. Bayesian meta-analysis on medical devices: application to implantable cardioverter defibrillators. *Int J Technol Assess Health Care* 2012; 28: 115-124. 2012/05/09. DOI: 10.1017/S0266462312000037.

198. Zagmutt FJ and Tarrants ML. Indirect comparisons of adverse events and dropout rates in early parkinson's disease trials of pramipexole, ropinirole, and rasagiline. *The International journal of neuroscience* 2012; 122: 345-353. 2012/02/07. DOI: 10.3109/00207454.2012.660586.

199. Zeppetella G, Davies AN, Rios C, et al. The efficacy of intranasal fentanyl spray and other opioids for the treatment of breakthrough cancer pain. *Eur J Cancer* 2011; 47(Suppl 1): S239. DOI: 10.1016/S0959-8049(11)71130-3.

200. Zintzaras E, Doxani C, Mprotsis T, et al. Network analysis of randomized controlled trials in multiple sclerosis. *Clin Ther* 2012; 34: 857-869 e859. 2012/03/27. DOI: 10.1016/j.clinthera.2012.02.018.

201. Ziogas DC, Voulgarelis M and Zintzaras E. A network meta-analysis of randomized controlled trials of induction treatments in acute myeloid leukemia in the elderly. *Clin Ther* 2011; 33: 254-279. 2011/05/24. DOI: 10.1016/j.clinthera.2011.04.004.

Appendix 2

CUREC Application response

Mr Andrew Lee
3 The Old School Yard
Redbourne
Gainsborough
DN21 4QN

7th September 2015

Dear Andrew

CUREC Application: Use of network meta-analysis in systematic reviews: a survey of authors

Thank you for forwarding your CUREC application for MS IDREC to review.

I have reviewed your application observing the current policies in place at the University of Oxford for ethics approval for studies with human participants. The study does not raise any issues concerning those policies. MS IDREC is unable to give retrospective approval but I am able to confirm that had your study been submitted prior to commencement, approval would have been given.

If further assurance is required by the journal considering your study for publication, please do not hesitate to refer them to this office.

Yours sincerely

Gill Halstead
Research Ethics Manager, Medical Sciences

Appendix 3

Table A3 Characteristics of included guidelines for Chapter Five

Title	Reference number	Addendum screened	Published	Last updated	Number of references screened	Number of Systematic Review references identified	Number of NMA references identified	Pair-wise meta-analysis conducted by Guideline Development Group	NMA conducted by Guideline Development Group	NMA result referenced as evidence for a recommendation
Chronic kidney disease in adults: assessment and management [1]	CG182		Jul-14	Jan-15	430	19	0	Yes	No	
Gastro-oesophageal reflux disease in children and young people: diagnosis and management [2]	NG1		Jan-15	Jan-15	130	4	0	No	No	
Bladder cancer: diagnosis and management [3]	NG2		Feb-15	Feb-15	426	33	0	Yes	No	
Meningitis (bacterial) and meningococcal septicaemia in under 16s: recognition, diagnosis and management [4]	CG102		Jun-10	Feb-15	218	15	0	Yes	No	
Postnatal care up to 8 weeks after birth [5]	CG37	Yes	Jul-06	Feb-15	500	76	0	No	No	
Depression in children and young people: identification and management [6]	CG28	Yes	Sep-05	Mar-15	360	6	0	Yes	No	
Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes [7]	NG5		Mar-15	Mar-15	178	1	0	Yes	No	
Obesity prevention [8]	CG43		Dec-06	Mar-15	1145	54	0	Yes	No	
Challenging behaviour and learning disabilities: prevention and interventions for people with learning disabilities whose behaviour challenges [9]	NG11		May-15	May-15	595	14	0	Yes	No	
Violence and aggression: short-term management in mental health, health and community settings [10]	NG10		May-15	May-15	295	26	0	Yes	No	
Antenatal and postnatal mental health: clinical management and service guidance [11]	CG192		Dec-14	Jun-15	756	37	0	Yes	No	
Suspected cancer: recognition and referral [12]	NG12		Jun-15	Jun-15	257	0	0	Yes	No	
Lower urinary tract symptoms in men: management [13]	CG97	Yes	May-10	Jun-15	353	11	0	Yes	No	
Chronic kidney disease: managing anaemia [14]	NG8		Jun-15	Jun-15	388	12	0	Yes	No*	
Venous thromboembolism: reducing the risk for patients in hospital [15]	CG92	Yes	Jan-10	Jun-15	752	37	0	Yes	**Yes**	**Yes**
Bronchiolitis in children: diagnosis and management [16]	NG9		Jun-15	Jun-15	176	6	**1**	No	No	No
Melanoma: assessment and management [17]	NG14		Jul-15	Jul-15	192	18	**1**	Yes	No	No

Title	Reference number	Addendum screened	Published	Last updated	Number of references screened	Number of Systematic Review references identified	Number of NMA references identified	Pair-wise meta-analysis conducted by Guideline Development Group	NMA conducted by Guideline Development Group	NMA result referenced as evidence for a recommendation
Diabetes in pregnancy: management from preconception to the postnatal period [18]	NG3		Feb-15	Aug-15	696	24	0	Yes	No	
Coeliac disease: recognition, assessment and management [19]	NG20		Sep-15	Sep-15	182	2	0	Yes	No	
Venous thromboembolic diseases: diagnosis, management and thrombophilia testing [20]	CG144	Yes	Jun-12	Nov-15	287	12	0	Yes	No	
Headaches in over 12s: diagnosis and management [21]	CG150	Yes	Sep-12	Nov-15	328	18	0	Yes	**Yes**	**Yes**
Preterm labour and birth [22]	NG25		Nov-15	Nov-15	132	12	**1**	Yes	**Yes**	**Yes**
Blood transfusion [23]	NG24		Nov-15	Nov-15	347	12	0	Yes	**Yes**	**Yes**
Menopause: diagnosis and management [24]	NG23		Nov-15	Nov-15	521	4	0	Yes	**Yes**	**Yes**
Urinary incontinence in women: management [25]	CG171		Sep-13	Nov-15	1027	37	0	Yes	**Yes**	**Yes**
Care of dying adults in the last days of life [26]	NG31		Dec-15	Dec-15	110	8	0	Yes	No	
Rheumatoid arthritis in adults: management [27]	CG79	Yes	Feb-09	Dec-15	431	20	0	Yes	No	
Intravenous fluid therapy in children and young people in hospital [28]	NG29		Dec-15	Dec-15	45	2	0	Yes	No	
Diabetic foot problems: prevention and management [29]	NG19		Aug-15	Jan-16	276	1	0	Yes	No	
Motor neurone disease: assessment and management [30]	NG42		Feb-16	Feb-16	131	9	0	Yes	No*	
Attention deficit hyperactivity disorder: diagnosis and management [31]	CG72	Yes	Sep-08	Feb-16	586	16	0	Yes	No	
Fractures (complex): assessment and management [32]	NG37		Feb-16	Feb-16	101	0	0	Yes	No*	
Fractures (non-complex): assessment and management [33]	NG38		Feb-16	Feb-16	203	7	0	Yes	No*	
Major trauma: assessment and initial management [34]	NG39		Feb-16	Feb-16	145	3	0	Yes	No*	
Spinal injury: assessment and initial management [35]	NG41		Feb-16	Feb-16	116	1	0	Yes	No*	
Epilepsies: diagnosis and management [36]	CG137		Jan-12	Feb-16	483	23	**1**	No	No	**Yes**
Bipolar disorder: assessment and management [37]	CG185		Sep-14	Feb-16	650	28	**3**	Yes	**Yes**	**Yes**
Myeloma: diagnosis and management [38]	NG35		Feb-16	Feb-16	266	6	**1**	Yes	No	No
Cancer of the upper aerodigestive tract: assessment and management in people aged 16 and over [39]	NG36		Feb-16	Feb-16	346	58	**2**	Yes	No	**Yes**
Depression in adults: recognition and management [40]	CG90		Oct-09	Apr-16	786	63	**1**	Yes	No	No
Routine preoperative tests for elective surgery [41]	NG45		Apr-16	Apr-16	95	4	0	Yes	No	

Title	Reference number	Addendum screened	Published	Last updated	Number of references screened	Number of Systematic Review references identified	Number of NMA references identified	Pair-wise meta-analysis conducted by Guideline Development Group	NMA conducted by Guideline Development Group	NMA result referenced as evidence for a recommendation
Tuberculosis [42]	NG33		Jan-16	May-16	330	11	0	Yes	**Yes**	**Yes**
Crohn's disease: management [43]	CG152	Yes	Oct-12	May-16	325	21	0	Yes	**Yes**	No
Familial hypercholesterolaemia: identification and management [44]	CG71		Aug-08	Jul-16	191	14	0	Yes	No	
Sepsis: recognition, diagnosis and early management [45]	NG51		Jul-16	Jul-16	338	5	0	Yes	No	
Non-Hodgkin's lymphoma: diagnosis and management [46]	NG52		Jul-16	Jul-16	353	7	0	Yes	No	
Type 1 diabetes in adults: diagnosis and management [47]	NG17		Aug-15	Jul-16	806	24	0	Yes	**Yes**	**Yes**
Prophylaxis against infective endocarditis: antimicrobial prophylaxis against infective endocarditis in adults and children undergoing interventional procedures [48]	CG64	Yes	Mar-08	Jul-16	267	4	0	Yes	No	
Non-alcoholic fatty liver disease (NAFLD): assessment and management [49]	NG49		Jul-16	Jul-16	237	3	0	Yes	No	
Cirrhosis in over 16s: assessment and management [50]	NG50		Jul-16	Jul-16	249	13	0	Yes	No	
Stable angina: management [51]	CG126		Jul-11	Aug-16	241	7	**1**	Yes	No	No
Acute upper gastrointestinal bleeding in over 16s: management [52]	CG141		Jun-12	Aug-16	202	11	0	Yes	No	
Heavy menstrual bleeding: assessment and management [53]	CG44		Jan-07	Aug-16	624	31	0	Yes	No	
Autism spectrum disorder in adults: diagnosis and management [54]	CG142		Jun-12	Aug-16	726	14	0	Yes	No	
Fertility problems: assessment and treatment [55]	CG156	Yes	Feb-13	Aug-16	1585	86	0	Yes	No	
Palliative care for adults: strong opioids for pain relief [56]	CG140		May-12	Aug-16	56	8	**2**	Yes	No	No
Cardiovascular disease: risk assessment and reduction, including lipid modification [57]	CG181		Jul-14	Sep-16	292	18	**1**	Yes	No	No
Multimorbidity: clinical assessment and management [58]	NG56		Sep-16	Sep-16	249	7	0	Yes	No	
Dementia: supporting people with dementia and their carers in health and social care [59]	CG42	Yes	Nov-06	Sep-16	776	52	0	Yes	No	
Mental health problems in people with learning disabilities: prevention, assessment and management [60]	NG54		Sep-16	Sep-16	357	14	0	Yes	No	
Psychosis and schizophrenia in children and young people: recognition and management [61]	CG155	Yes	Jan-13	Oct-16	259	20	0	Yes	No	
Jaundice in newborn babies under 28 days [62]	CG98	Yes	May-10	Oct-16	317	1	0	Yes	No	
Chest pain of recent onset: assessment and diagnosis [63]	CG95		Mar-10	Nov-16	230	11	0	Yes	No	
Low back pain and sciatica in over 16s: assessment and management [64]	NG59		Nov-16	Nov-16	745	45	**1**	Yes	No	No

Title	Reference number	Addendum screened	Published	Last updated	Number of references screened	Number of Systematic Review references identified	Number of NMA references identified	Pair-wise meta-analysis conducted by Guideline Development Group	NMA conducted by Guideline Development Group	NMA result referenced as evidence for a recommendation
Spasticity in under 19s: management [65]	CG145		Jul-12	Nov-16	82	1	0	Yes	No	
Hypertension in adults: diagnosis and management [66]	CG127		Aug-11	Nov-16	662	47	0	Yes	No	
Diabetes (type 1 and type 2) in children and young people: diagnosis and management [67]	NG18		Aug-15	Nov-16	919	28	0	Yes	No	
Physical health of people in prison [68]	NG57		Nov-16	Nov-16	169	4	0	Yes	No	
Organ donation for transplantation: improving donor identification and consent rates for deceased organ donation [69]	CG135	Yes	Dec-11	Dec-16	220	5	0	Yes	No	
Intravenous fluid therapy in adults in hospital [70]	CG174		Dec-13	Dec-16	126	10	0	Yes	No	
Hypothermia: prevention and management in adults having surgery [71]	CG65	Yes	Apr-08	Dec-16	306	1	0	Yes	No	
End of life care for infants, children and young people with life-limiting conditions: planning and management [72]	NG61		Dec-16	Dec-16	151	14	0	Yes	No	

Bold type is used to highlight guidelines in which NMA references were identified, NMA was conducted by a Guideline Development Group or NMA results were referenced as evidence for a recommendation.

* indicates that the guideline stated that conducting NMA was considered but not undertaken because there was insufficient data available

Appendix 3 References

1. National Institute for Health and Care Excellence. Chronic kidney disease in adults: assessment and management. Clinical guideline 182, https://www.nice.org.uk/guidance/cg182/evidence (2014, accessed 26 June 2017).
2. National Institute for Health and Care Excellence. Gastro-oesophageal reflux disease in children and young people: diagnosis and management. National Guideline 1, https://www.nice.org.uk/guidance/ng1/evidence (2015, accessed 27 June 2017).
3. National Institute for Health and Care Excellence. Bladder cancer: diagnosis and management. National Guideline 2, https://www.nice.org.uk/guidance/ng2/evidence (2015, accessed 25 June 2017).
4. National Institute for Health and Care Excellence. Meningitis (bacterial) and meningococcal septicaemia in under 16s: recognition, diagnosis and management. Clinical Guideline 102, https://www.nice.org.uk/guidance/cg102/evidence (2015, accessed 25 June 2017).
5. National Institute for Health and Care Excellence. Postnatal care up to 8 weeks after birth. Clinical Guideline 37, https://www.nice.org.uk/guidance/cg37/evidence (2015, accessed 27 June 2017).
6. National Institute for Health and Care Excellence. Depression in children and young people: identification and management. Clinical Guideline 28, https://www.nice.org.uk/guidance/cg28/evidence (2015, accessed 14 June 2017).
7. National Institute for Health and Care Excellence. Medicines optimisation: the safe and effective use of medicines to enable the best possible outcomes. National Guideline 5, https://www.nice.org.uk/guidance/ng5/evidence (2015, accessed 14 June 2017).
8. National Institute for Health and Care Excellence. Obesity prevention. Clinical Guideline 43, https://www.nice.org.uk/guidance/cg43/evidence (2015, accessed 25 June 2017).
9. National Institute for Health and Care Excellence. Challenging behaviour and learning disabilities: prevention and interventions for people with learning disabilities whose behaviour challenges. National Guideline 11, https://www.nice.org.uk/guidance/ng11/evidence (2015, accessed 11 June 2017).
10. National Institute for Health and Care Excellence. Violence and aggression: short-term management in mental health, health and community settings. National Guideline 10, https://www.nice.org.uk/guidance/ng10/evidence (2015, accessed 11 June 2017).
11. National Institute for Health and Care Excellence. Antenatal and postnatal mental health: clinical management and service guidance. Clinical Guideline 192, https://www.nice.org.uk/guidance/cg192/evidence (2015, accessed 5 June 2017).
12. National Institute for Health and Care Excellence. Suspected cancer: recognition and referral. National Guideline 12, https://www.nice.org.uk/guidance/ng12/evidence (2015, accessed 11 June 2017).
13. National Institute for Health and Care Excellence. Lower urinary tract symptoms in men: management. Clinical Guideline 97, https://www.nice.org.uk/guidance/cg97/evidence (2015, accessed 11 June 2017).

14. National Institute for Health and Care Excellence. Chronic kidney disease: managing anaemia. National Guideline 8, https://www.nice.org.uk/guidance/ng8/evidence (2015, accessed 11 June 2017).

15. National Institute for Health and Care Excellence. Venous thromboembolism: reducing the risk for patients in hospital. Clinical Guideline 92, https://www.nice.org.uk/guidance/cg92/evidence (2015, accessed 11 June 2017).

16. National Institute for Health and Care Excellence. Bronchiolitis in children: diagnosis and management. National Guideline 9, https://www.nice.org.uk/guidance/ng9/evidence (2015).

17. National Institute for Health and Care Excellence. Melanoma: assessment and management. National Guideline 14, https://www.nice.org.uk/guidance/ng14/evidence (2015).

18. National Institute for Health and Care Excellence. Diabetes in pregnancy: management from preconception to the postnatal period. National Guideline 3, https://www.nice.org.uk/guidance/ng3/evidence (2015, accessed 26 May 2017).

19. National Institute for Health and Care Excellence. Coeliac disease: recognition, assessment and management. National Guideline 20, https://www.nice.org.uk/guidance/ng20/evidence (2015, accessed 26 May 2017).

20. National Institute for Health and Care Excellence. Venous thromboembolic diseases: diagnosis, management and thrombophilia testing. Clinical Guideline 144, https://www.nice.org.uk/guidance/cg144/evidence (2015, accessed 21 May 2017).

21. National Institute for Health and Clinical Excellence. Headaches in over 12s: diagnosis and management. Clinical Guideline 150, https://www.nice.org.uk/guidance/cg150/evidence (2015, accessed 23 May 2017).

22. National Institute for Health and Care Excellence. Preterm labour and birth. National Guideline 25, https://www.nice.org.uk/guidance/ng25/evidence (2015).

23. National Institute for Health and Care Excellence. Blood transfusion. National Guideline 24, https://www.nice.org.uk/guidance/ng24/evidence (2015, accessed 23 May 2017).

24. National Institute for Health and Clinical Excellence. Menopause: diagnosis and management. National Guideline 23, https://www.nice.org.uk/guidance/ng23/evidence (2015, accessed 26 May 2017).

25. National Institute for Health and Clinical Excellence. Urinary incontinence in women: management. Clinical Guideline 171, https://www.nice.org.uk/guidance/cg171/evidence (2015, accessed 26 May 2017).

26. National Institute for Health and Care Excellence. Care of dying adults in the last days of life. National Guideline 31, https://www.nice.org.uk/guidance/ng31/evidence (2015, accessed 21 May 2017).

27. National Institute for Health and Care Excellence. Rheumatoid arthritis in adults: management. Clinical Guideline 79, https://www.nice.org.uk/guidance/cg79/evidence (2015, accessed 21 May 2017).

28. National Institute for Health and Care Excellence. Intravenous fluid therapy in children and young people in hospital. National Guideline 29, https://www.nice.org.uk/guidance/ng29/evidence (2015, accessed 21 May 2017).

29. National Institute for Health and Care Excellence. Diabetic foot problems: prevention and management. National Guideline 19, https://www.nice.org.uk/guidance/ng19/evidence (2016, accessed 21 May 2017).

30. National Institute for Health and Care Excellence. Motor neurone disease: assessment and management. National Guideline 42, https://www.nice.org.uk/guidance/ng42/evidence (2016, accessed 14 May 2017).

31. National Institute for Health and Care Excellence. Attention deficit hyperactivity disorder: diagnosis and management. Clinical Guideline 72, https://www.nice.org.uk/guidance/cg72/evidence (2016, accessed 21 May 2017).

32. National Institute for Health and Care Excellence. Fractures (complex): assessment and management. National Guideline 37, https://www.nice.org.uk/guidance/ng37/evidence (2016, accessed 17 May 2017).

33. National Institute for Health and Care Excellence. Fractures (non-complex): assessment and management. National Guideline 38, https://www.nice.org.uk/guidance/ng38/evidence (2016, accessed 17 May 2017).

34. National Institute for Health and Care Excellence. Major trauma: assessment and initial management. National Guideline 39, https://www.nice.org.uk/guidance/ng39/evidence (2016, accessed 17 May 2017).

35. National Institute for Health and Care Excellence. Spinal injury: assessment and initial management. National Guideline 41, https://www.nice.org.uk/guidance/ng41/evidence (2016, accessed 17 May 2017).

36. National Institute for Health and Care Excellence. Epilepsies: diagnosis and management. Clinical Guideline 137, https://www.nice.org.uk/guidance/cg137/evidence (2016, accessed 17 May 2017).

37. National Institute for Health and Care Excellence. Bipolar disorder: assessment and management. Clinical Guideline 185, https://www.nice.org.uk/guidance/cg185/evidence (2016, accessed 17 May 2017).

38. National Institute for Health and Care Excellence. Myeloma: diagnosis and management. National Guideline 35, https://www.nice.org.uk/guidance/ng35/evidence (2016, accessed 21 May 2017).

39. National Institute for Health and Care Excellence. Cancer of the upper aerodigestive tract: assessment and management in people aged 16 and over. National Guideline 36, https://www.nice.org.uk/guidance/ng36/evidence (2016, accessed 21 May 2017).

40. National Institute for Health and Care Excellence. Depression in adults: recognition and management. Clinical Guideline 90, https://www.nice.org.uk/guidance/cg90/evidence (2016, accessed 14 May 2017).

41. National Institute for Health and Care Excellence. Routine preoperative tests for elective surgery. National Guideline 45, https://www.nice.org.uk/guidance/ng45/evidence (2016, accessed 14 May 2017).

42. National Institute for Health and Care Excellence. Tuberculosis. National Guideline 33, https://www.nice.org.uk/guidance/ng33/evidence (2016, accessed 10 May 2017).

43. National Institute for Health and Clinical Excellence. Crohn's disease: management. Clinical Guideline 152, https://www.nice.org.uk/guidance/cg152/evidence (2016, accessed 14 May 2017).

44. National Institute for Health and Clinical Excellence. Familial hypercholesterolaemia: identification and management. Clinical Guideline 71, https://www.nice.org.uk/guidance/cg71/evidence (2016, accessed 8 May 2017).

45. National Institute for Health and Care Excellence. Sepsis: recognition, diagnosis and early management. National Guideline 51, https://www.nice.org.uk/guidance/ng51/evidence (2016, accessed 8 May 2017).

46. National Institute for Health and Care Excellence. Non-Hodgkin's lymphoma: diagnosis and management. National Guideline 52, https://www.nice.org.uk/guidance/ng52/evidence (2016, accessed 8 May 2017).

47. National Institute for Health and Care Excellence. Type 1 diabetes in adults: diagnosis and management. National guideline 17, https://www.nice.org.uk/guidance/ng17/evidence (2016).

48. National Institute for Health and Clinical Excellence. Prophylaxis against infective endocarditis: antimicrobial prophylaxis against infective endocarditis in adults and children undergoing interventional procedures. Clinical Guideline 64, https://www.nice.org.uk/guidance/cg64/evidence (2016, accessed 10 May 2017).

49. National Institute for Health and Care Excellence. Non-alcoholic fatty liver disease (NAFLD): assessment and management. National Guideline 49, https://www.nice.org.uk/guidance/ng49/evidence (2016, accessed 10 May 2017).

50. National Institute for Health and Care Excellence. Cirrhosis in over 16s: assessment and management. National Guideline 50, https://www.nice.org.uk/guidance/ng50/evidence (2016, accessed 10 May 2017).

51. National Institute for Health and Care Excellence. Stable angina: management. Clinical Guideline 126, https://www.nice.org.uk/guidance/cg126/evidence (2016, accessed 3 May 2017).

52. National Institute for Health and Care Excellence. Acute upper gastrointestinal bleeding in over 16s: management. Clinical Guideline 141, https://www.nice.org.uk/guidance/cg141/evidence (2016, accessed 3 May 2017).

53. National Institute for Health and Care Excellence. Heavy menstrual bleeding: assessment and management. Clinical Guideline 44, https://www.nice.org.uk/guidance/cg44/evidence (2016, accessed 3 May 2017).

54. National Institute for Health and Care Excellence. Autism spectrum disorder in adults: diagnosis and management. Clinical Guideline 142, https://www.nice.org.uk/guidance/cg142/evidence (2016, accessed 3 May 2017).

55. National Institute for Health and Care Excellence. Fertility problems: assessment and treatment. Clinical Guideline 156, https://www.nice.org.uk/guidance/cg156/evidence (2016, accessed 8 May 2017).

56. National Institute for Health and Care Excellence. Palliative care for adults: strong opioids for pain relief. Clinical Guideline 140, https://www.nice.org.uk/guidance/cg140/evidence (2016, accessed 8 May 2017).

57. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. Clinical Guideline 181, https://www.nice.org.uk/guidance/cg181/evidence (2016, accessed 12 March 2017).

58. National Institute for Health and Care Excellence. Multimorbidity: clinical assessment and management. National Guideline 56, https://www.nice.org.uk/guidance/ng56/evidence (2016, accessed 12 March 2017).

59. National Institute for Health and Care Excellence. Dementia: supporting people with dementia and their carers in health and social care. Clinical Guideline 42, https://www.nice.org.uk/guidance/cg42/evidence (2016, accessed 12 March 2017).

60. National Institute for Health and Care Excellence. Mental health problems in people with learning disabilities: prevention, assessment and management. National Guideline 54, https://www.nice.org.uk/guidance/ng54/evidence (2016, accessed 24 April 2017).

61. National Institute for Health and Care Excellence. Psychosis and schizophrenia in children and young people: recognition and management. Clinical Guideline 155, https://www.nice.org.uk/guidance/cg155/evidence (2016, accessed 12 March 2017).

62. National Institute for Health and Care Excellence. Jaundice in newborn babies under 28 days. Clinical Guideline 98, https://www.nice.org.uk/guidance/cg98/evidence (2016, accessed 12 March 2017).

63. National Institute for Health and Care Excellence. Chest pain of recent onset: assessment and diagnosis. Clinical Guideline 95, https://www.nice.org.uk/guidance/cg95/evidence (2016, accessed 8 March 2017).

64. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management. National Guideline 59, https://www.nice.org.uk/guidance/ng59/evidence (2016, accessed 8 March 2017).

65. National Institute for Health and Care Excellence. Spasticity in under 19s: management. Clinical Guideline 145, https://www.nice.org.uk/guidance/cg145/evidence (2016, accessed 11 March 2017).

66. National Institute for Health and Care Excellence. Hypertension in adults: diagnosis and management. Clinical Guideline 127, https://www.nice.org.uk/guidance/cg127/evidence (2016, accessed 11 March 2017).

67. National Institute for Health and Care Excellence. Diabetes (type 1 and type 2) in children and young people: diagnosis and management. National Guideline 18, https://www.nice.org.uk/guidance/ng18/evidence (2016, accessed 11 March 2017).

68. National Institute for Health and Care Excellence. Physical health of people in prison. National Guideline 57, https://www.nice.org.uk/guidance/ng57/evidence (2016, accessed 11 March 2017).

69. National Institute for Health and Care Excellence. Organ donation for transplantation: improving donor identification and consent rates for deceased organ donation. Clinical Guideline 135, https://www.nice.org.uk/guidance/cg135/evidence (2016, accessed 8 March 2017).

70. National Institute for Health and Care Excellence. Intravenous fluid therapy in adults in hospital. Clinical Guideline 174, https://www.nice.org.uk/guidance/cg174/evidence (2016, accessed 8 March 2017).

71. National Institute for Health and Care Excellence. Hypothermia: prevention and management in adults having surgery. Clinical Guideline 65, https://www.nice.org.uk/guidance/cg65/evidence (2016, accessed 8 March 2017).

72. National Institute for Health and Care Excellence. End of life care for infants, children and young people with life-limiting conditions: planning and management. National Guideline 61, https://www.nice.org.uk/guidance/ng61/evidence (2016, accessed 8 March 2017).