NETWORK META-ANALYSIS: STRATEGIES TO IMPROVE INTERPRETATION

NETWORK META-ANALYSIS: STRATEGIES TO IMPROVE INTERPRETATION

By MARK R PHILLIPS, B.Sc *(he/him)*

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements

for the Degree Doctor of Philosophy

Supervisor: Dr. Mohit Bhandari

Number of pages: 226

Thesis Abstract

Network meta-analysis (NMA) is a relatively new methodology that allows for the comparison of multiple interventions, with particular benefit in allowing for the indirect comparison of interventions that have never been compared directly within randomized trials. There lacks a consensus with regard to many NMA methodology considerations, often leading to inadequately conducted NMAs. This provides an opportunity for novel research initiatives to develop and promote improved NMA methodologies; which will translate to improved information provided to clinicians for evidence-based decision making. This thesis highlights the need for better standards in NMA methodology and reporting, developed and user-tested a novel tool for effectively presenting NMA results to clinicians and key stakeholders, and utilized this tool within an NMA of surgical treatment options for displaced femoral neck fractures – an area of great interest for orthopaedic clinicians and researchers due to the numerous available treatment options and considerable equipoise that exists in this area. The NMA methods landscape will continue to evolve, and the components of this thesis provide the groundwork to solidify and advance the implementation of NMA results into clinical practice.

Acknowledgements

There is a vast list of individuals that need acknowledgement for their impact on this thesis, and my life throughout my Ph.D. studies. First and foremost, the Department of Health Research Methods, Evidence, and Impact at McMaster has demonstrated why they continue to run a premier health research methodology program, as they exceeded all expectations in every aspect of my time in the program. I am grateful to have had the opportunity to learn in such a collaborative and supportive environment.

The supportive environment of the HRM program allowed for me to collaborate with numerous colleagues throughout this thesis, and to all of those individuals I thank you. I would like to especially thank Kim Madden, Tahira Devji, Alonso Carrasco-Labra, Sofia Bzovsky, Behnam Sadeghirad, Anila Qasim, and Herman Johal for their support and guidance through the Ph.D. process. I have had many excellent opportunities to learn from these individuals, and greatly appreciate all I have gained from their collaboration.

I have also had the opportunity to learn from truly world-class mentors, of whom I can only hope to continue learning from for years – and decades – to come. Gordon Guyatt, Jason Busse, Lehana Thabane, Sheila Sprague, and Raveendhara Bannuru have been invaluable mentors throughout this thesis and beyond, as they have supported me in multiple aspects of my development. It is a privilege to learn from leaders in the field, and the lessons I have taken from them will shape my future career.

To Mohit Bhandari, your truly exceptional guidance has provided me clarity in my career decisions, and your brilliant teaching abilities have provided me with constant advancements in my understanding of health research concepts and methodologies. Although I have learned a great deal about research methodology from you, it pales in comparison to the life lessons that you have passed on to me. I am forever grateful for the opportunities you have provided for me, and I will continue to thrive under your one-of-a-kind mentorship.

I have been fortunate to have a close and supportive family that has shaped my character throughout my entire life. To my parents Debbie and Gary, and brother Steve, you have supported me through this PhD process – but more importantly through the 20 years of my life prior. I cannot thank you enough for everything you have done for me.

Finally, it is difficult to describe in words how thankful I am for my wife Tina, son Danny, and our newborn that is expected 1 week prior to this thesis defence. Our family has become the single most important aspect of my life, and I appreciate all of the sacrifices made to allow me the opportunity of completing this Ph.D.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

AE: Adverse event

ASA: American Society of Anesthesiologists

CV: Cardiovascular

EQ-5D: Euroqol-5 Dimension

GI: Gastrointestinal

GRADE: Grading of Recommendations, Assessment, Development and Evaluations

HA: Hyaluronic acid

HHS: Harris Hip Score

IF: Internal Fixation

IOP: Intra-ocular pressure

NMA: Network meta-analysis

OA: Osteoarthritis

POAG: Primary open angle glaucoma

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT: Randomized controlled trial

SHS: Sliding hip screw

SUCRA: Surface under the cumulative ranking

THA: Total hip arthroplasty

Declaration of Academic Achievement

Statement of Funding: Mark Phillips was supported financially throughout the course of this PhD in part by an Ontario Graduate Scholarship, part-time employment with the McMaster University Department of Health Research Methods, Evidence, and Impact (HEI), Global Research Solutions Inc, and teaching assistant roles within the McMaster Health Research Methodology program. No portion of this thesis was done as work under employment by these entities.

Chapter 1: Introduction

This chapter is unpublished. MP is the sole author.

Chapter 2: Improving the understanding of NMA methodology

Editorial: A Clinician's Guide to Network Meta-Analysis. This chapter is published in EYE (2022 Aug;36(8):1523-1526. doi: 10.1038/s41433-022-01943-5). MRP and VC developed a first draft of the manuscript. All authors reviewed, edited and approved the manuscript.

Chapter 3: Current NMA reporting methods: A case study in knee OA

The quality of network meta-analysis methods for pharmacological management of chronic pain in knee osteoarthritis requires improvement: A systematic survey. This chapter is prepared for submission to JBJS Reviews. MRP, JWB, RRB, LT, and MP conceptualized the study. MRP, CK, and AP collected and analyzed data. MRP developed a first draft of the manuscript. All authors reviewed, edited and approved the manuscript.

Chapter 4: Developing an improved methodology for reporting NMA results

Part A: Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study. This study is published in BMJ Open (2022 Jun 10;12(6): e056400. doi:10.1136/bmjopen-2021-056400). MRP, BS, JWB, RPB, CC, FKN, RRB, LT, MB, and GHG conceptualized the study. MRP, BS, JWB, and GHG recruited participants for the study. MRP, YJG, and SB collected and analyzed data. MRP, BS, JWB, RPB, CC, FKN, and GHG acted as the steering committee to interpret and implement data from participants. MRP and GHG developed a first draft of the manuscript. All authors reviewed, edited and approved the manuscript.

Part B: User Testing of a Novel Network Meta-Analysis Results Presentation Table: A Qualitative Description Protocol. This protocol is unpublished. The protocol was submitted to the Hamilton Integrated Research Ethics Board for Ethics approval for Part 2 of this chapter.

Chapter 5: Implementing NMA reporting methodology in the surgical management of displaced femoral neck fractures

Part A: Surgical Management of Displaced Femoral Neck Fractures: A Systematic Review and Network Meta-Analysis. This chapter is prepared for submission to JAMA Surgery. MRP, JWB, RRB, LT, and MP conceptualized the study. MRP, CK, and VS collected and analyzed data. MRP developed a first draft of the manuscript. All authors reviewed, edited and approved the manuscript.

Part B: Surgical Management of Displaced Femoral Neck Fractures: A Protocol for a Systematic Review and Network Meta-Analysis. This protocol is published on the ResearchGate pre-print server. DOI: 10.13140/RG.2.2.18428.41602

Chapter 6: Discussion and opportunities for future research

This chapter is unpublished. MP is the sole author.

Chapter 1: Introduction

Network meta-analysis (NMA) is a novel advancement in research methodology, which allows for the comparison of multiple treatment options – even when direct evidence comparing those options does not exist. The publication of NMAs has rapidly increased in occurrence over the past decade, with continual advancements and new considerations developed by methodologists to further advance this analysis approach. Due to the fast moving and changing landscape in the infancy of NMA methodology, there is a lack of consistency and standardization for this approach.

Although NMA allows for a comprehensive evaluation of all available treatment options for a given condition, a concern of this relatively new analysis method is the receptiveness and appropriate interpretation of results to inform clinical practice.[1,2] NMA results often encompass a large number of comparisons, making results lengthy and complex – and potentially at risk of difficult interpretation. Some analysis options, such as surface under the cumulative ranking curve (SUCRA), were primarily developed to simplify interpretation of NMA results by ranking treatments from "best" to "worst" for a given outcome. This form of analysis has more recently been deemed to be an over-simplification of results that forgoes consideration of both the magnitude of effects, as well as the quality of the evidence. For this reason, more advanced methods of ranking treatments for improved interpretation have been developed.[3,4] Notably, the partially and minimally contextualized approaches to interpreting NMA results provide guidance on ranking treatments into "best" to "worst" treatment categories, while simultaneously considering the quality of the evidence and magnitudes of effect.[3,4] While these approaches are

an excellent improvement to previous NMA methods, there remain limitations in the ability for NMA authors to present their results in an easily interpretable manner.

NMA interpretation is further hindered by a lack of standardization within NMA methodology, which has resulted in a large variety of analysis approaches, tables, and figures to be used by authors – some of which are suboptimal and could lead to poor or misleading interpretation of NMA findings. There have been some publications aimed at helping to educate clinicians on the methodology used within an NMA, as well as how to appropriately interpret the results; however, there has not been an attempt to identify clinician perspectives on NMAs to jointly inform optimal reporting methods for enhanced interpretation.[1,2,5] NMA reporting methodologies have advanced considerably in recent years, yet there remains a lack of option to report NMA results for multiple outcomes in a single, digestible format for readers. Reporting and methodology standardization efforts from key groups; such as PRISMA, Cochrane, and GRADE, have become more available, but a major gap exists in the form of an optimal presentation format for multiple outcomes within an NMA.[3,4,6–8]

The main objective of this thesis is to investigate optimal NMA methods of results presentation to allow for improved uptake of results into clinical practice. This thesis identifies current trends in NMA methodology, and highlights the current reporting methods used within NMAs published to date. We propose a novel reporting tool that clearly presents NMA results across multiple outcomes. The methodological advancements in NMA reporting were user-tested for feedback and improvements by clinicians and other stakeholders, and finally demonstrated in an example NMA conducted for surgical management of displaced femoral neck fractures. The end result of this thesis is a collation of works that identify past NMA methodology

and reporting trends, provide a novel tool to better present NMA results for multiple outcomes, and utilize this tool to inform displaced femoral neck fracture management.

Chapter 2 of this thesis includes a published editorial that was invited from a prominent journal in Ophthalmology – *EYE.* With a predominantly clinical audience, the journal wanted to provide a high-level overview of NMA for their audience. This chapter explains core concepts of NMA, with guidance for clinicians to efficiently dissect and interpret their results.

Chapter 3 includes a systematic survey of NMAs published on injectables for knee osteoarthritis (OA). This topic area was chosen due to the relatively large number of NMAs that have been published in this particular field. The objective of this chapter was to illustrate the aforementioned inconsistency and lack of standardization in NMA reporting and methodology, as these numerous NMAs conducted on the same topic used a variety of approaches. The results of this chapter demonstrate how differences in methodology can impact an NMA, as well as the overarching need for improvements to NMA methods for maximizing understanding and interpretability.

Chapter 4 includes both the protocol and published study that was conducted to create and refine an NMA results presentation tool. Through iterative and comprehensive user-testing with relevant stakeholders, an NMA results presentation tool was developed that was demonstrated to be easily interpretable by clinicians and other important audiences; perhaps most importantly by individuals with minimal prior NMA understanding and knowledge. This NMA results tool enables presentation of results across multiple outcomes, while also giving insight into the quality of the evidence underpinning those results.

Once the presentation tool was developed, we conducted an NMA to implement the novel tool; providing an example of its utility for future NMA authors. Chapter 5 includes a protocol and NMA study on the surgical management of displaced femoral neck fractures. This NMA provides an important summary of treatment effects in this clinical area – of which there is considerable equipoise around decision-making in this vulnerable patient population. The results of this NMA are not only important for clinical decision-making; they provide an example of how the NMA results presentation tool from Chapter 4 can be utilized by future NMA authors.

The main objective of this thesis is to investigate optimal NMA methods of results presentation to allow for improved uptake of results into clinical practice. This thesis identifies current trends in NMA methodology, and highlights the current reporting methods used within NMAs published to date. We propose a novel reporting tool that clearly presents NMA results across multiple outcomes. The methodological advancements in NMA reporting were user-tested for feedback and improvements by clinicians and other stakeholders, and finally demonstrated in an example NMA conducted for surgical management of displaced femoral neck fractures. The end result of this thesis is a collation of works that identify past NMA methodology and reporting trends, provide a novel tool to better present NMA results for multiple outcomes, and utilize this tool to inform displaced femoral neck fracture management.

References

1. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. *BMJ*. 2013;346.

2. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-171. doi:10.1016/j.jclinepi.2010.03.016

3. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ*. 2020;371:m3900. doi:10.1136/bmj.m3900

4. Brignardello-Petersen R, Izcovich A, Rochwerg B, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ*. 2020;371:m3907. doi:10.1136/bmj.m3907

5. 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(4):417-428. doi:10.1016/j.jval.2011.04.002

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. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385

7. 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. doi:10.1016/j.jclinepi.2017.10.005

8. Yepes-Nuñez JJ, Li SA, Guyatt G, et al. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol*. 2019;115:1-13. doi:10.1016/j.jclinepi.2019.04.018

Chapter 2: Improving the understanding of NMA methodology in the clinical audience

A Clinician's Guide to Network Meta-Analysis

Mark Phillips[1], David H Steel[2,3], Charles C. Wykoff[4,5], Jason W. Busse[1,6,7], Raveendhara R. Bannuru[8], Lehana Thabane[1,9], Mohit Bhandari[1,10], Varun Chaudhary[1,10] for the Retina Evidence Trials InterNational Alliance (R.E.T.I.N.A.) Study Group

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
2. Sunderland Eye Infirmary, Sunderland, UK
3. Biosciences Institute, Newcastle University, Newcastle Upon Tyne, UK
4. Retina Consultants of Texas (Retina Consultants of America), Houston, TX, USA
5. Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA
6. Michael G. DeGroote National Pain Center, McMaster University, Hamilton, ON, Canada
7. Department of Anesthesia, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
8. Center for Treatment Comparison and Integrative Analysis, Division of Rheumatology, Tufts Medical Center, Boston, MA, USA
9. Biostatistics Unit, St. Joseph's Healthcare-Hamilton, Hamilton, ON, Canada
10. Department of Surgery, McMaster University, Hamilton, ON, Canada

The Evolution of Evidence Synthesis

Increasing interest in promoting evidence-based clinical practice has led to methodological advancements in evidence syntheses.[1,2] Narrative reviews have been superseded by systematic reviews, which may include meta-analysis – statistical pooling of treatment effect estimates across similar trials to improve precision.[3,4,5] Systematic reviews minimize the risk of selection bias by considering all evidence relevant to a clinical question; however, an important limitation of conventional meta-analyses is that they only inform treatments that have been directly compared in clinical trials. Moreover, many trials compare active interventions against placebo, usual or standard care, whereas patients and clinicians are typically concerned with the relative effectiveness of competing interventions. Network meta-analysis (NMA) has emerged to address these limitations by allowing for calculation of the comparative effects of more than two competing interventions, even when they have not been directly compared in clinical trials.[6,7]

What is Network Meta-Analysis?

NMA requires the same steps as a conventional meta-analysis which include a systematic search of the literature, assessment of risk of bias among eligible trials, statistical pooling of reported pair-wise comparisons for all outcomes of interest, and assessment of the overall certainty of evidence on an outcome-by-outcome basis. This provides the 'direct' evidence for treatments that have been compared against each other, which is graphically represented by a network map. An NMA then identifies all interventions that are connected by virtue of a common comparator. For example, two different active treatments may have been compared against placebo in different trials. An NMA allows for a theoretical trial to be created that compares these

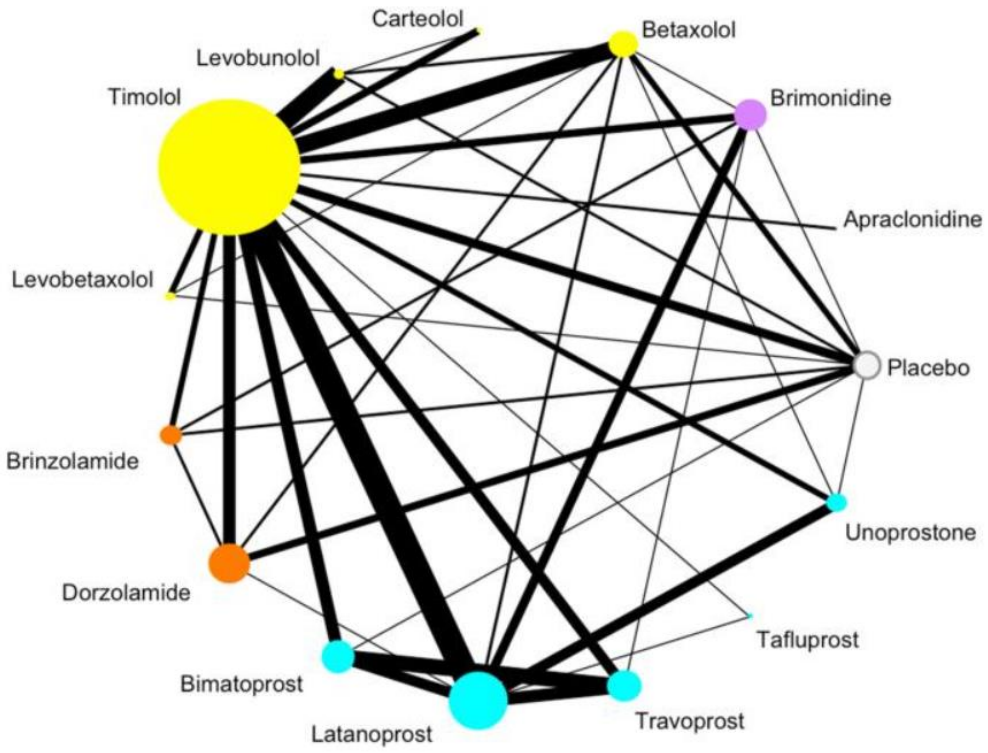
active treatments against each other, based on their effect against a common comparator (placebo), which provides 'indirect' evidence. Indirect comparisons provide an opportunity to fill knowledge gaps within the available evidence, providing a more comprehensive understanding of treatment options for the clinician. The network estimate is the pooled result of the direct and indirect evidence for a given comparison, or only the indirect evidence if no direct evidence is available.[6,8,9] Once all treatments have been compared within a network, there are different methods for ranking treatments to convey their relative net effectiveness. Limitations and advancements in the ranking methodology will be discussed in greater detail within the example provided below.

Network Meta-Analysis in Practice

An example network map on first-line medications effects on intra-ocular pressure (IOP) for primary open angle glaucoma (POAG) is shown in **Figure 1**, which represents all pharmacologic treatments that have been directly evaluated in 114 clinical trials for this condition.[10] Traditional meta-analysis would be limited in comparing two of these treatments at a time, and could not inform effectiveness of treatments that have not been directly compared; however, this NMA provides the relative effectiveness of all 15 treatments in a single investigation, even when no RCT is available to make a direct comparison between two treatments. The network map uses circles, or nodes, for each included treatment, that increase in size relative to the number of patients treated with that medication within included RCTs. The lines connecting different treatments are weighted by the number of RCTs comparing them (i.e., thicker lines convey more direct trials).[10] In this particular study, the authors colour coded their treatment nodes by drug class to improve interpretation. The network is specific to one outcome,

in this case IOP, and the network assumes that the baseline characteristics of patients enrolled across trials are similar.

Figure 1: Network diagram from Li et al (2016)[10] comparing medications for POAG



Size of nodes represents the number of patients, line thickness represents number of trials

Reprinted from: Comparative Effectiveness of First-Line Medications for Primary Open-Angle Glaucoma A Systematic Review and Network Meta-analysis, 123/1, Tianjing Li, Kristina Lindsley, Benjamin Rouse, Hwanhee Hong, Qiyuan Shi, David S. Friedman, Richard Wormald, Kay Dickersin, Ophthalmology: Journal of the American Academy of Ophthalmology, Pages No. 129-140, Copyright (2016), with permission from Elsevier.

As **Figure 1** demonstrates, there are many RCTs assessing pharmacotherapy for POAG. Some treatments, such as Timolol or Latanoprost, have large bodies of evidence, while many

others have far fewer – and smaller – trials assessing their efficacy.[10] This network enables the comparison of 14 active medications, as well as placebo, for POAG.

While the ability to summarize large bodies of evidence is also possible for traditional meta-analyses, NMAs provide comparative effectiveness data between competing treatments. It is important to note that the evidence provided by an NMA is subject to the limitations of the individual RCTs included within the network.[11] In addition, the ranking of interventions by NMAs using methods such as the Surface Under the Cumulative Ranking Curve (SUCRA) approach is problematic – despite this currently being the most common form of treatment ranking in NMAs. This approach ranks all treatments within a network from "best" to "worst" for each analyzed outcome, but only considers the effect estimate and not the associated precision or the certainty of evidence.[12] Thus, interventions supported by small, low-quality trials that report large effects are ranked highly. Minimally or partially contextualized approaches, instead, consider the magnitude of effect in the context of patient-importance as well as the certainty of evidence.[13,14]

How Can You Have Certainty in the Findings of an NMA?

Like all study designs, there are considerations when evaluating the credibility of the findings of an NMA. These include the same issues that should be considered when evaluating a traditional pairwise meta-analysis, such as the rigor of the literature search, risk of bias among included trials, consistency of effect estimates contributing to pooled effects (heterogeneity), precision of the pooled effect estimate, publication bias, and directness of the included evidence in relation to the primary research question.[8,9,15,16] However, there are two additional considerations that are specific to NMAs: incoherence and transitivity.[8,9,15,17]

Incoherence exists when the direct and indirect estimates for a comparison are not consistent with one another.[6] A meta-epidemiological study of 112 published NMAs found inconsistent direct and indirect treatment effects in 14% of the comparisons made.[18] This means that while in most cases it is appropriate to combine indirect and direct evidence, this is not always the case, and review authors should formally explore this issue. In the presence of incoherence, the higher certainty evidence should be presented rather than the network estimate. If the direct and indirect effects are both supported by the same certainty of evidence, then the network estimate can be used but should be downgraded one level for incoherence. The GRADE approach is increasingly used for rating the certainty in evidence for network estimates, which incorporates these aforementioned criteria[11,15–17] A GRADE rating can assign high, moderate, low, or very low certainty in the evidence.[11,15–17] Clinicians should take the certainty of the evidence in consideration when determining the impact findings would have on their clinical practice, as lower certainty evidence provides less confidence in the results.

Transitivity refers to the similarity between study characteristics that allows indirect effect comparisons to be made with assurance that there are limited factors that could modify treatment effects, aside from the intervention under investigation.[6,15] Essentially, transitivity refers to the inclusion of studies that fundamentally address the same research questions within the same population.[6] Intransitivity can result in biased indirect estimates, which would then impact the overall findings of the network estimates.[15,17] As previously discussed, incoherence exists when discrepancies between direct and indirect estimates is present, thus, transitivity is a common cause of incoherence.[17]

Clinicians cannot be expected to evaluate transitivity and incoherence within an NMA and authors should clearly report on these two important aspects. Indeed, absence of reporting should lead readers to question the findings. **Table 1** provides an example and overview of the core items for readers to identify for critical appraisal of published NMAs, as applied to the Li et al. (2016) POAG study.[10,19] These criteria are based on the Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice.[19]

Conclusion

Rigorously conducted and reported NMA may provide helpful information for advancing evidence-based ophthalmology, specifically in the common scenario in which multiple treatment options exist. However, clinicians should appraise the quality of NMAs before accepting the results, and even rigorously conducted NMAs cannot provide high certainty evidence if the primary trials eligible for review are flawed.

Table 1: Example Appraisal of the Li et al (2016) POAG NMA

Item	Considerations	Example from Li et al (2016) POAG NMA[10]
Systematic Review Processes		
Study eligibility	Systematic review with clearly defined, explicit eligibility criteria should be included. A pre-published protocol should summarize the planned conduct of the NMA (ex. PROSPERO registration)	The methods has the section "Eligibility criteria for considering studies for this review", which provides explanation of their eligibility criteria.
Literature search	Reproduceable, systematic search that has confidently retrieved all relevant literature on the topic.	Within the methods, "Search methods for identifying studies" summarizes the systematic approach used for the literature search.
Study selection and assessment	Systematic screening process is used to identify and select all relevant literature from the search that was conducted. Additionally, a risk of bias should be conducted for each included study.	The "Study selection" subheading summarizes the study screening process.
Between-study differences	Did the study plan for subgroup, sensitivity, and/or meta-regression analyses to address hypotheses on between-study differences?	Sensitivity analyses that evaluated specific concentrations of Bimatoprost and Timolol were conducted. Justification for this is provided in the section "Measures of association".
NMA Analysis and Results		
Amount of evidence	For each comparison, consider the number of RCTs informing the effect estimate. Estimates based on a small amount of evidence may be less reliable than those informed by a large number of RCTs.	Table 1 provides an overview of the number of studies included within each comparison.
Certainty in each comparison	For each comparison in the network, authors should provide an evaluation of the certainty of evidence; which is most commonly the GRADE assessment. The GRADE assessment includes assessment of: Risk of bias, inconsistency, indirectness, imprecision, and publication bias – as well as incoherence between direct and indirect estimates (see below).	They did not provide an evaluation of the certainty of evidence. Certainty of evidence ratings is an important aspect of ensuring proper interpretation of future NMA results. [11,15]
Incoherence	An NMA should provide exploration of the direct and indirect estimates to identify any instances of incoherence. Explanations of incoherence, such as transitivity, should be discussed in the context of the results.	They provided multiple versions of statistical assessment of incoherence under the "Evaluation of the assumption for network meta-analysis" section in the methods, and inconsistency portion of the results.

Treatment rankings	As previously discussed, older methods such as SUCRA rankings have major concerns and limitations, more modern methods in minimally or partially contextualized approaches can give more informative rankings of the included treatments.[13,14]	A key limitation is the utilization of SUCRA scores to draw key conclusions within their abstract and manuscript. While it may be warranted to include SUCRA scores within an NMA, it is problematic to use these scores for drawing conclusions over the effect estimates observed.
Robust results across sensitivity assumptions	Often, NMA authors will conduct sensitivity analyses to determine the robustness of the results across different assumptions. If results remain consistent across these sensitivity analyses, greater confidence may be put on the overall findings.	The sensitivity analyses conducted demonstrated comparable results to the main analysis – although minor changes in SUCRA rankings were observed. As previously stated, this should not drive clinical decision making.
Application of Results to Clinical Practice		
Included outcomes	Whenever possible, patient-important outcomes should be directly assessed, opposed to a surrogate outcome.	For POAG, intra-ocular pressure (IOP) is a core outcome, as it is the primary modifiable risk factor for this condition. It is important to also consider patient-reported outcome measures, when applicable. The Li et al (2016) NMA only provides insight into IOP, although an attempt was made to also evaluate visual field.
Included treatment options	When an NMA does not include all potential treatment options, it leaves uncertainty for clinicians in the potential effects of those omitted treatments. This does not give a full picture of the possible treatment options to consider in a clinical scenario.	They compared a comprehensive list of the available treatment options for POAG, increasing the applicability of the findings.
Subgroup effects	When subgroup effects have been explored, these results can inform the specific clinical scenarios in which particular effects may be observed.	NA – as they did not conduct subgroup analyses.
Overall quality and limitations	Once all of these considerations have been made, clinicians may have a greater sense of the quality of the NMA, and thus can make an informed decision on the implementation of the NMA findings in their clinical practice.	Taking all of these aspects into account, Li et al. conducted a thoughtful NMA on an important topic area, although it was not without some limitations. The primary concerns for interpretation are the omission of a certainty of evidence assessment, as well as the heavy reliance on SUCRA rankings to draw conclusions. Clinicians should come to their own conclusions based on these limitations as to its applicability to practice

References

1. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet Lond Engl*. 2017;390(10092):415-423. doi:10.1016/S0140-6736(16)31592-6

2. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097

3. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ*. 1997;315(7121):1533-1537. doi:10.1136/bmj.315.7121.1533

4. Crowther M, Lim W, Crowther MA. Systematic review and meta-analysis methodology. *Blood*. 2010;116(17):3140-3146. doi:10.1182/blood-2010-05-280883

5. Leucht S, Kissling W, Davis JM. How to read and understand and use systematic reviews and meta-analyses. *Acta Psychiatr Scand*. 2009;119(6):443-450. doi:10.1111/j.1600-0447.2009.01388.x

6. Mills E, Thorlund K, Ioannidis J. Demystifying trial networks and network meta-analysis. *BMJ*. 2013;346:f2914. doi:10.1136/bmj.f2914

7. Neupane B, Richer D, Bonner AJ, Kibret T, Beyene J. Network Meta-Analysis Using R: A Review of Currently Available Automated Packages. *PLOS ONE*. 2014;9(12):e115065. doi:10.1371/journal.pone.0115065

8. Foote CJ, Chaudhry H, Bhandari M, et al. Network Meta-analysis: Users' Guide for Surgeons: Part I – Credibility. *Clin Orthop Relat Res*. 2015;473(7):2166-2171. doi:10.1007/s11999-015-4286

9. Chaudhry H, Foote CJ, Guyatt G, et al. Network Meta-analysis: Users' Guide for Surgeons: Part II - Certainty. *Clin Orthop*. 2015;473(7):2172-2178. doi:10.1007/s11999-015-4287-9

10. Li T, Lindsley K, Rouse B, et al. Comparative Effectiveness of First-Line Medications for Primary Open-Angle Glaucoma: A Systematic Review and Network Meta-analysis. *Ophthalmology*. 2016;123(1):129-140. doi:10.1016/j.ophtha.2015.09.005

11. Puhan MA, Schünemann 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. doi:10.1136/bmj.g5630

12. Mbuagbaw L, Rochwerg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev*. 2017;6(1):79. doi:10.1186/s13643-017-0473-z

13. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ*. 2020;371. doi:10.1136/bmj.m3900

14. Brignardello-Petersen R, Izcovich A, Rochwerg B, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ*. 2020;371:m3907. doi:10.1136/bmj.m3907

15. 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. doi:10.1016/j.jclinepi.2017.10.005

16. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-394. doi:10.1016/j.jclinepi.2010.04.026

17. Brignardello-Petersen R, Mustafa RA, Siemieniuk RAC, et al. GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. *J Clin Epidemiol*. 2019;108:77-85. doi:10.1016/j.jclinepi.2018.11.025

18. Song F, Xiong T, Parekh-Bhurke S, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ*. 2011;343:d4909. doi:10.1136/bmj.d4909

19. Mills E, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. Chapter 24: Network Meta-analysis. In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. McGraw Hill Education; :327-356.

Chapter 3: Current NMA reporting methods: A case study in knee OA

The quality of network meta-analysis methods for pharmacological management of chronic pain in knee osteoarthritis requires improvement: A systematic survey

Mark R Phillips, BSc, PhD(c)[1]; Herman Johal MD[1,2]; Raman Mundi MD[1,2]; Ashaka Patel BSc[3]; Christine Kucava MSc[4]; Raveendhara R Bannuru MD PhD[5]; Jason W. Busse, DC PhD[1,6,7,8]; Lehana Thabane PhD[1,9,10]; Mohit Bhandari MD PhD[1,11]

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. Division of Orthopaedic Surgery, McMaster University, Hamilton, Ontario, Canada
3. Department of Health Sciences, McMaster University, Hamilton, Ontario, Canada
4. Faculty of Nursing, McMaster University, Hamilton, Ontario, Canada
5. Center for Treatment Comparison and Integrative Analysis, Tufts Medical Center, Boston, MA, USA
6. The Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Ontario, Canada
7. Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada
8. The Michael G. DeGroote Centre for Medicinal Cannabis Research, McMaster University, Hamilton, Ontario, Canada
9. Biostatistics Unit, St Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada
10. Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa
11. Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Abstract

Background: A network meta-analysis (NMA) allows for comparisons to be made across three or more treatment options, opposed to the ability to only compare two treatments within a traditional pairwise meta-analysis. The pharmacological management of pain secondary to knee osteoarthritis (OA) has been the focus of numerous NMAs; however, they have reported different conclusions. We systematically evaluated this literature to determine the quality of, and discrepancies among, NMAs for pharmacological management of knee OA pain.

Purpose: This study aims to provide a summary of the quality of reporting in currently published NMAs, the study characteristics and methodology used, result presentation methods, and how differences in methodology impact overall study results in NMAs addressing pharmacological management of knee OA pain.

Methods: We conducted a systematic literature search of the Medline, EMBASE, and Cochrane Database of Systematic Reviews for all published NMAs addressing pharmacological management of knee pain associated with OA. We evaluated the quality of reporting with the 30-item Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist for NMAs, methodological differences between systematic reviews, and how these differences may have contributed to inconsistencies in the findings across NMAs.

Results: Eighteen NMAs were eligible for this review. The evaluated timeframe ranged from four weeks to beyond 52 weeks. Four NMAs included both oral and injectable treatments, four included only oral therapies, and 10 considered only injectable therapies. Seven studies conducted sensitivity analyses to explore the effect of risk of bias on their results. Specific

methodology and results reporting approaches differed greatly across the included studies, with no common form of results presentation used to convey key findings. No NMA evaluated the quality of evidence from their analyses. Adherence to the PRISMA reporting checklist was variable, with scores ranging from 7/30 to 30/30 items reported. Differences in results seen across the NMAs seemed to primarily be attributed to the differences in timeframe assessed, analysis methods, study inclusion criteria, and treatments considered within each NMA.

Conclusions: The reporting quality of NMAs exploring management of knee OA is highly variable, and results are inconsistent, likely due to the scope of interventions included and the timeframe considered. Future NMAs should adhere to reporting checklists, consider all available treatment options, evaluate the certainty of the available evidence, and consider an appropriate timeframe that is relevant and meaningful for patients who are suffering from chronic knee OA pain.

Key messages:

- This survey reviewed the reporting quality and methods used for knee OA pain NMAs, and identified highly variable reporting quality and methods utilized across them.
- The included NMAs presented their results in a variety of ways, many of which did not consider harms outcomes.
- This study has demonstrated the variability in reporting quality, methodology, and results observed across NMAs assessing knee OA pain management.

- The findings of this systematic survey demonstrate the importance of proper NMA reporting, as well as insight into how methodological considerations have important implications for NMA findings.

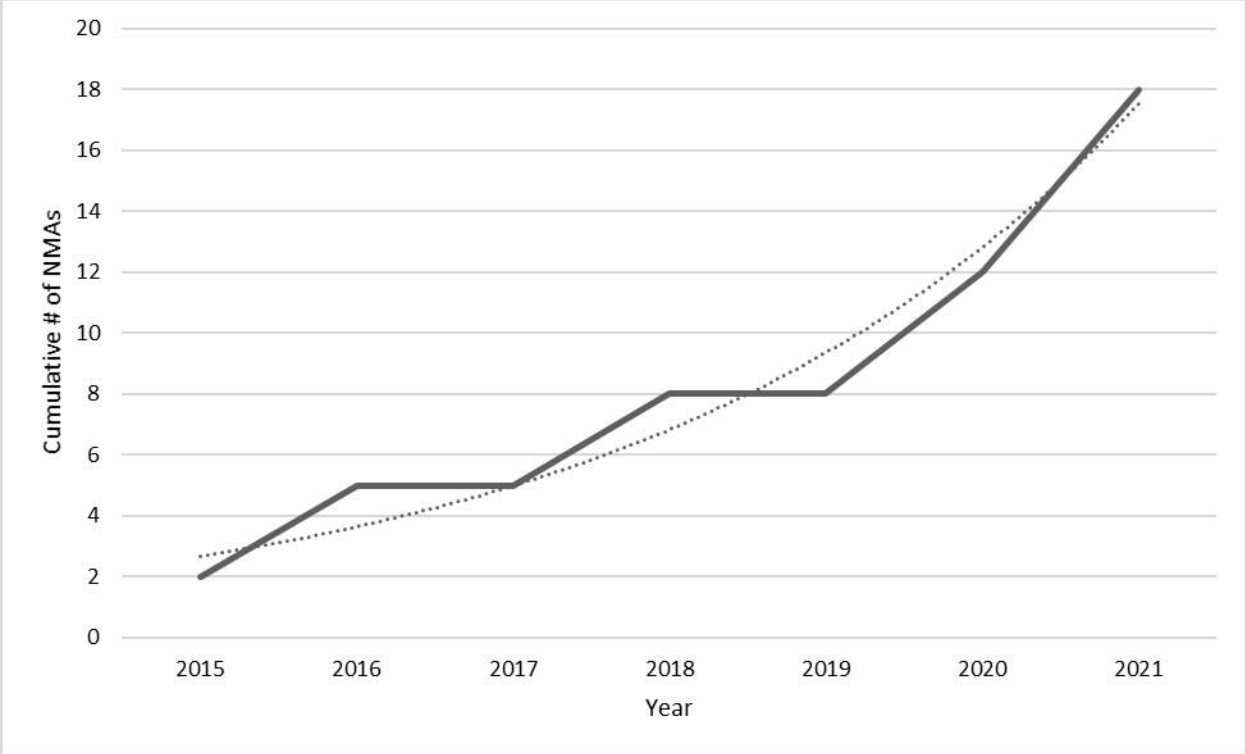
Introduction

Network meta-analysis (NMA) is an increasingly popular methodology to inform comparative effectiveness of available treatments for musculoskeletal conditions, as seen in the knee osteoarthritis (OA). **Figure 1** provides a visualization of the cumulative publication of chronic knee OA pain NMAs year by year. An NMA allows for comparisons to be made across three or more treatment options, in comparison to two treatments within a traditional pairwise meta-analysis. Indirect evidence can be acquired for treatment options that have no direct studies assessing them, given that they have been compared to a common treatment in other studies.[1–3] Establishing the comparative effectiveness of multiple treatments for a given condition is likely to be helpful for both patients and clinicians, given the review is methodologically rigorous and high quality clinical trials are available.[4] NMAs are becoming increasingly influential within orthopedics, particularly as clinical practice guidelines have began adopting this methodology to inform their recommendations.[5–7] The orthopaedic literature has seen a large uptake in the use of NMA, yet understanding how methodology and results have been thoughtfully planned and comprehensively reported when conducting an NMA is needed to ensure that readers can gauge the trustworthiness of their findings.[1,2,8]

Accordingly, one must assess the differences in approach used by NMAs, as methodological considerations may be a contributing factor to discrepant results.[8] The

pharmacological management of chronic knee OA pain has been the recent focus of numerous NMAs to evaluate the wide range of potential treatment options. This study aims to provide a summary of the study characteristics, and methodology used the quality of reporting in currently published NMAs, and how differences in methodology can impact NMAs addressing pharmacological management of knee OA pain.

Figure 1: Cumulative Publication of Included NMAs



Dotted line represents fitted exponential line

Methods

Search Strategy and Criteria

We developed database-specific search strategies (**Appendix A**) and searched Medline, EMBASE, and the Cochrane Database of Systematic Reviews (CDSR), from inception through October 2nd, 2021. In addition, we reviewed the reference lists of relevant studies for any additional eligible reviews. Studies eligible for our review were NMAs exploring pharmacological treatments for pain due to knee OA. Studies that evaluated surgical interventions for knee OA were excluded.

Study Selection

All studies identified by the literature search were assessed for eligibility, independently, by pairs of reviewers (MP, AP, CK) at the title/abstract, and full text stages. Any disagreements in eligibility at the title/abstract screening phase were included for full-text review, while any discrepancies in eligibility at the full-text review stage were resolved by discussion to achieve consensus. A third reviewer was used, if necessary, to resolve any disagreements between the two reviewers at the full-text stage.

Data Collection and Analysis

Data was collected from eligible studies, independently by two reviewers, and discrepancies were resolved by discussion. The following data was collected for each included study: year of publication, number of total patients within the network, total number of trials

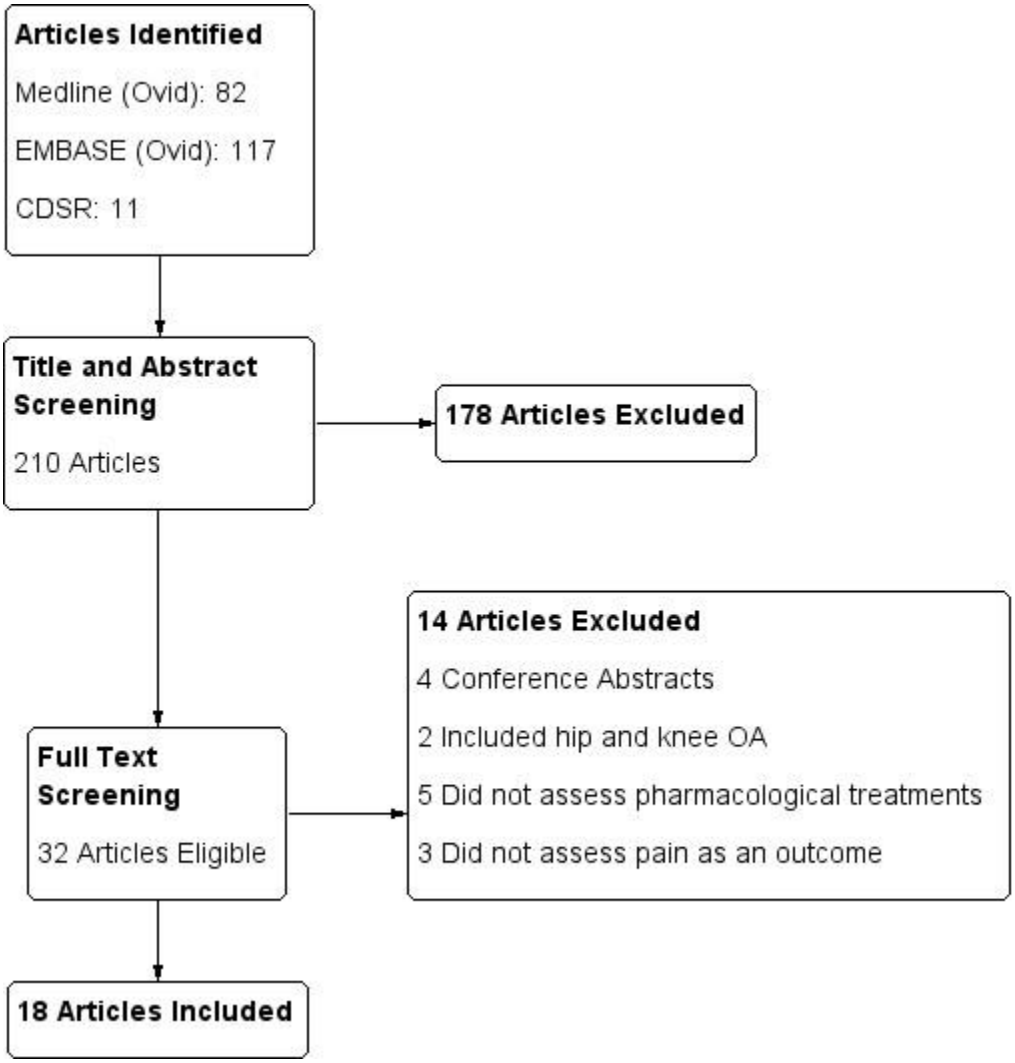
within the network, treatments included in the network, analysis timeframe, frequentist or Bayesian analysis, methods used to present findings, treatment conclusions, the justification for treatment conclusions, use of a minimally important difference (MID) to facilitate interpretation of clinical significance, any sensitivity analyses, sub-group, or meta-regression conducted, and the use of a systematic approach to rate the certainty of NMA evidence.[4,9] Additionally, each study was evaluated using the 30-item PRISMA checklist for NMAs to assess the number of components within the checklist that were successfully reported.[10] Descriptive statistics were reported as count and percentages for dichotomous variables, and means with standard deviations for continuous variables.

Results

Study Selection & Characteristics

The literature search retrieved 210 citations, of which 18 articles were eligible for our review (**Figure 2**).[5,6,11–26] Included studies were published between 2015 and 2021, with over half of the published NMAs being published in the last 2 years (10/18 published in 2020/2021). The cumulative publication trend of included NMAs is provided within **Figure 1**. Reported networks ranged in size from 5 to 137 included trials (883 to 47,133 patients) and 3 to 31 included therapies. The majority of NMAs utilized a Bayesian analysis approach (12 out of 18 studies), 4 of 18 used a frequentist approach, and 2 studies did not report their method of analysis (**Table 1**).

Figure 2: PRISMA Flow Chart



Four NMAs evaluated both intra-articular and oral therapies,[5,11,15,17] whereas 4 considered oral pharmacotherapy,[12,14,16,24] and 10 considered injectable treatments only[6,13,18–23,25,26] (**Table 1**). Among the four NMAs that assessed both oral and injectable treatments, there was a range of 9 to 31 potential treatment options included, and there was not a consistent optimal treatment option concluded. An NMA focussed on early follow-up (closest data to 1 month follow-up) identified corticosteroids as the leading treatment option for pain.[5] Two NMAs focussed on intermediate follow-up (2-6 months and at least 6 months) reported intra-articular

hyaluronic acid (IA-HA) with and without corticosteroids as the best treatment with regard to pain relief.[11,17] The fourth NMA that assessed results beyond 1 year did not identify any treatment with definitive clinical benefit, although glucosamine was shown to be the treatment with the marginally best, yet likely clinically irrelevant, pain relief.[15]

The other 14 NMAs included either only injectables, or only oral therapies. Within these subsets, there was also a lack of consistency regarding the treatment groups included. NMAs that included only injectables ranged from 3 to 15 included treatments, and NMAs of only oral therapies ranged from 4 to 9 included treatments (**Table 1**).

Table 1: Study Characteristics

Author	# Of Patients in Network	# Of Studies in Network	# Of Nodes in Network	Treatments in Network	Outcomes Assessed	Analysis Method
Bannuru 2015	33,243	137	9	Oral and Injectable Therapies	Pain, Function, Stiffness, AEs	Bayesian
Kongtharvonskul 2015	883	8	4	Oral Therapies	Pain, Function, Stiffness, Joint Space Width, AEs	NR
Doros 2016	1,385	5	3	Injectable Therapies	Pain	Bayesian
Smith 2016	5,659	17	4	Oral Therapies	Pain	NR
Trojian 2016	3,391	11	3	Injectable Therapies	Pain, Function, Stiffness, OMERACT-OARSI responders	Frequentist
Gregori 2018	22,037	47	31	Oral and Injectable Therapies	Pain, Function, Joint Space Narrowing	Bayesian
Jevsevar 2018	18,891	56	10	Oral and Injectable Therapies	Pain, Function	Bayesian
Jung 2018	19,045	44	9	Oral Therapies	Pain, Function, Stiffness	Bayesian
Beaudart 2020	15,609	79	26	Oral and Injectable Therapies	Pain, Function, Stiffness, Joint Space Width	Bayesian
Hummer 2020	2,796	14	5	Injectable Therapies	Pain	Bayesian
Phillips 2020	6,712	47	6	Injectable Therapies	Pain, Function, AEs	Frequentist
Chevalier 2020	8,047	42	4	Injectable Therapies	Pain, AEs	Bayesian
Singh 2021	4,604	23	5	Injectable Therapies	Pain, Function	Frequentist
Anil 2021	8,761	79	15	Injectable Therapies	Pain, Function	Frequentist
Zhao 2021	5,575	43	6	Injectable Therapies	Pain, Function, AEs	Bayesian
Zeng 2021	47,133	122	6	Oral Therapies	Pain, Function, GI AEs, CV AEs, Withdrawal due to AE	Bayesian
Migliorini 2021	3,463	30	4	Injectable Therapies	Pain, Function	Bayesian
Han 2021	5,554	43	5	Injectable Therapies	Pain, Function, AEs, Severe AEs	Bayesian

AE: Adverse event, GI: Gastrointestinal, CV: Cardiovascular

NMA Analysis and Methodology

We evaluated all NMAs that evaluated pain as an outcome. The most common additional outcome assessed within the included NMAs was function (14/18). Fewer than half of included NMAs assessed the risk of harms or adverse events (7/18). A summary of additional outcomes assessed within each NMA is provided in **Table 1**. The follow-up time among included studies ranged from 4 weeks to "beyond 52 weeks". The majority of studies assessed timepoints between one and six months after treatment (14/18). There were six studies (6/18) that used direct reference to an MID, all of which reported the average treatment effects relative to the MID threshold. There were seven studies (7/18) that conducted sensitivity analyses to explore the effect of study quality on their results. A summary of the characteristics for each study are included in **Table 2.**

Result Presentation Methods

Included studies presented their results using a number of different methods (**Table 2**). The majority of studies provided forest plots to summarize the network effect estimates (10/18), and half of included studies included a presentation of treatment rankings; p-scores or SUCRA scores for frequentist and Bayesian analyses, respectively (9/18). Fewer NMAs provided league tables to summarize all network comparisons across included treatments (6/18). For studies that evaluated additional outcomes aside from pain, there was no use of a standardized table or figure presentation tool to present network estimates for all outcomes.

Table 2: Summary of Methodology

Author	Analysis Timeframe	Results Considered Against MID?	Certainty of Evidence Assessment	Results Presentation Methods in Main Text	Sensitivity Analysis of Low Risk of Bias Studies	Considerations Regarding Conducted Analyses
Bannuru 2015	Closest to 12 weeks (8 to 26 weeks)	Yes – 20/100-point WOMAC pain	No	**League Tables**	Yes	Sensitivity analyses excluding studies with <50 and <100 patients per treatment arm. Sensitivity based on potential reporting bias and the specific pain scale used included in supplementary material.
Kongtharvonskul 2015	4 to 24 weeks	No	No	**Summary table** of all comparisons	No	Network meta-analysis conducted as a secondary analysis within this study
Doros 2016	18 to 30 weeks	No	No	**Summary table** of all comparisons **Linear trend graph** of mean differences in pain week-by-week	No	Network meta-analysis conducted as a secondary analysis within this study; limited information provided
Smith 2016	Closest to 12 weeks	No	No	**None** (text only)	No	Network meta-analysis conducted as a secondary analysis within this study
Trojian 2016	"at time of best [IA-HA] response" (12 to 26 weeks)	No	No	**Summary table** of all comparisons	No	Only included studies with IA-HA as one of the comparators: Studies of corticosteroid versus IA-saline not included, despite being a comparison in the network.
Gregori 2018	52 weeks or later	No	No	**Forest Plots** (Placebo as reference) **SUCRA Curves** (Treatment ranking)	Yes	Only high-quality studies included in ranking.
Jevsevar 2018	First follow-up at/after 4 weeks (Mean 42 days)	Yes - 19.1 points on 100-point VAS pain	No	**League Tables** **SUCRA Graphs** (Treatment ranking) **SUCRA Table** (Treatment ranking)	No	No range of follow-ups reported. One analysis conducted. For any comparison with 3 or more high quality studies, all lower quality studies were excluded. If a comparison did not have 3 high quality studies, lower quality evidence was included for that comparison.
Jung 2018	Closest to 6 weeks	No	No	**League Table** **P-Score Table** (Treatment ranking)	Yes	Sensitivity analyses excluding studies with <100 patients
Beaudart 2020	A minimum of 6 months	No	No	**Forest Plots** (Placebo as reference) **P-Score in Forest Plot** (Treatment ranking)	Yes	Sensitivity based on removal of any study that was considered high risk of bias for any Cochrane risk of bias domain.
Hummer 2020	Longest reported follow-up	Yes – 0.5 SD units and 8.3	No	**Forest Plots** (Placebo as reference) **League Tables**	No	No clear distinction of the timepoint analyzed. The term quality assessment was used; however, this was not a

		points on WOMAC scale		**Cluster Graph** (Absolute effects) **P-Score Table** (Treatment Ranking) **P-Score Graphs** (Treatment Ranking)		certainty of the body of evidence assessment – it was an evaluation of individual study risks of bias.
Phillips 2020	3 months (+/- 1 month)	Yes – 0.2 SD units	No	**Forest Plots** (Placebo as reference)	Yes	Sensitivity analysis based on risk of allocation concealment bias. Additional sensitivity based on studies with imputed SDs.
Chevalier 2020	1 month; 3 months; 6 months	No	No	**Forest Plots** (Corticosteroid as reference)	No	Certain studies removed from analyses due to having effect estimates, patient age, or gender proportions that were deemed to be "outliers".
Singh 2021	A minimum of 6 months	Yes - 19.1 points on 100-point VAS pain	No	**Forest Plots** (Placebo as reference) **League Tables** **SUCRA Graphs** (Treatment ranking) **SUCRA Table** (Treatment ranking)	No	Studies with fewer than 30 patients per treatment group were excluded from all analyses
Anil 2021	4-6 Weeks; 3 Months; 6 Months; 12 Months	No	No	**Forest Plots** (Placebo as reference) **P-Score Table** (Treatment Ranking)	No	Treatment ranking analysis used heavily to drive conclusions.
Zhao 2021	6 months; 12 months	Yes – 10 points on 100-point VAS/ WOMAC pain score	No	**Forest Plots** (For all comparisons) **SUCRA Tables** (Treatment ranking)	Yes	Sensitivity analysis of only low risk of bias studies to explore heterogeneity. Meta-regression to evaluate impacts publication year, mean age, and sample size of included studies.
Zeng 2021	At/nearest to 4 weeks; At/nearest to 12 weeks	No	No	**Summary tables** for each comparison	No	Also conducted and reported the results of two observational studies that compared oral NSAID to topical NSAID.
Migliorini 2021	3 months; 6 months; 12 months	No	No	**Forest Plots** (Artificially produced null group as reference)	No	5 nodes included, with only 4 treatments. 5th node is labelled as the timeframe of analysis, where treatment effects were defined "0".
Han 2021	at least 4 weeks (range from 6 to 104 weeks)	No	No	**Forest Plots** (For all comparisons) **League Tables** (Treatment ranking) **SUCRA Tables** (Treatment ranking)	Yes	Treatment ranking analysis used heavily to drive conclusions.

No NMAs used presentation methods that provided the certainty of the evidence alongside their effect estimates. All studies presented relative effects, and one included NMA graphically presented absolute effects in the main text (**Table 2**).

Quality of Reporting

An overview of the adherence of included reviews to the PRISMA checklist for NMA reporting is provided in **Table 3**. Overall, studies most frequently reported the study rationale and objectives within their introduction, and the conclusions clearly at the end of the study. With the exception of three NMAs, all studies reported their eligibility, search and selection criteria, and data collection methods for their NMA. The least frequently reported item on the PRISMA checklist was the inclusion of a systematic review protocol and registration; which was only reported by three (17%) reviews. The results of each included trial were infrequently reported (5/18 studies). Other items that were infrequently reported in eligible NMAs were: inclusion of additional analysis methods and results, assessments of inconsistency, and risk of bias across included studies. None of the included NMAs used a systematic approach for rating the certainty of evidence for their reported treatment effects.

Table 3: PRISMA Checklist for NMAs

Study	Bannuru 2015	Kongtharvonskul 2015	Doros 2016	Smith 2016	Trojian 2016	Gregori 2018	Jevsevar 2018	Jung 2018	Beaudart 2020
Introduction									
Rationale	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Objectives	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Methods									
Protocol and registration	No	No	No	No	No	Yes	No	No	Yes
Eligibility Criteria	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Information sources	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Search	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Study Selection	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Data collection process	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Data items	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Geometry of the network	Yes	No	No	No	No	Yes	Yes	Yes	Yes
risk of bias within individual studies	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
summary measures	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes
planned methods of analysis	Yes	No	No	No	No	Yes	Yes	Yes	Yes
assessment of inconsistency	Yes	No	No	No	Yes	No	Yes	Yes	Yes
risk of bias across studies	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes
additional analyses	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Results									
Study selection	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
presentation of network structure	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
summary of network geometry	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
study characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
risk of bias within studies	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes

results of individual studies	No	No	Yes	No	No	No	Yes	No	Yes
synthesis of results	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
exploration for inconsistency	Yes	No	No	No	Yes	No	Yes	Yes	Yes
risk of bias across studies	Yes	No	No	Yes	No	No	Yes	Yes	Yes
results of additional analyses	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Discussion									
summary of evidence	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
limitations	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
conclusions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
funding	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Table 3 (cont'd)

Study	Hummer 2020	Phillips 2020	Chevalier 2020	Anil 2021	Zhao 2021	Zeng 2021	Migliorini 2021	Han 2021	Singh 2021
Introduction									
Rationale	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Objectives	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Methods									
Protocol and registration	No	No	No	No	No	Yes	No	No	No
Eligibility Criteria	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Information sources	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Search	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Study Selection	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Data collection process	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Data items	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geometry of the network	Yes	Yes	Yes	No	No	No	Yes	Yes	No
Risk of bias within individual studies	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Summary measures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Planned methods of analysis	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Assessment of inconsistency	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Risk of bias across studies	No	No	Yes	No	No	No	Yes	Yes	No
Additional analyses	No	Yes	No	No	Yes	Yes	No	Yes	No
Results									
Study selection	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Presentation of network structure	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No
Summary of network geometry	Yes	Yes	No	No	No	Yes	No	Yes	No

Study characteristics	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Risk of bias within studies	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Results of individual studies	Yes	No	No	No	No	Yes	No	No	No
Synthesis of results	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exploration for inconsistency	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Risk of bias across studies	No	No	No	No	No	No	Yes	No	No
Results of additional analyses	No	Yes	No	No	Yes	Yes	Yes	No	No
Discussion									
Summary of evidence	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Limitations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Conclusions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Funding	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No

Discussion

Our review identified a total of eighteen NMAs that explored pharmacological treatment of knee pain due to OA. The reporting quality of these studies was highly variable, with specific methodologic and results reporting approaches also differing greatly across the included studies. The optimal therapy concluded across NMAs was inconsistent, which was likely due to the scope of interventions included within each NMA, the timeframe considered, and the analytic methods chosen. The included studies also used different methods to identify the best treatment options within their network. Many studies used treatment rankings – such as SUCRA rankings – or judgements based on magnitude of effect, to identify the best treatment within their main conclusions. While initially developed and utilized due to their simplicity of interpretation, the use of these ranking methodologies poses a potentially misleading, if not inappropriate, interpretation of NMA findings. SUCRA rankings do not consider the certainty of the evidence or the precision of the estimate.[28] In addition, these ranking techniques provide no context to the importance of differences between ranked treatments – and chance alone may explain the differences between rankings.[28] Recent advancements have been made to provide a more comprehensive and appropriate methodology to rank treatments within NMA, namely the minimally and partially contextualized approaches to drawing NMA conclusions.[28,29] These approaches provide a framework that is based on two main guiding principles: treatments should be categorized into most and least effective or harmful categories, and that the categorization of the treatments should be based on the estimates of effect, the quality of the evidence, and the treatment rankings.[28,29] The main difference between the minimally and partially contextualized

approaches, is that the partially contextualized approach includes value judgements regarding the estimates of effect.[29]

Clinical Insights from Included NMAs

Despite the large disparity in methodology and reporting quality within the included NMAs, general trends of knee OA pain management can be gleaned from the few NMAs that included both oral and injectable therapies, which were generally well-reported.[10] The timeframe of these four analyses provides a synergistic overview of the follow-up period of a knee OA patient, as they evaluated the short-term[5], intermediate term[11,17], and long-term[15] follow-up periods. These three reviews concluded that IA-HA, corticosteroids, and a combination of both these treatments provided meaningful benefits in pain relief in the short to intermediate timeframe, while the fourth depicted uncertainty regarding the meaningfulness of the benefits seen by their top ranked treatment: glucosamine sulfate.[5,11,15,17] The utilization of an MID to determine the clinical importance of effects is, however, more nuanced than it is often implemented. Commonly, the MID is looked at as a threshold in which, when treatment effects are below the MID, they are unimportant clinically. Instead, precision around the estimate that represents the variability in patient response to treatment could indicate that a certain proportion of patients could observe clinically meaningful benefit.[30] In patient populations where values and preferences suggest a willingness to try treatment for any potential relief of their symptoms, there may be clinical justification to provide these treatment options. In order to do so, clinical decisions must be made within the context of both benefits and harms, as the intended benefits must outweigh potential harms in both the clinician and patient's informed judgement. This raises a major concern with the included NMAs, as nearly half of the knee OA

pain NMAs did not provide any assessment of harms outcomes. A previous review demonstrated that chronic pain reviews often do not include all relevant patient-important outcomes.[31] Similarly, the knee OA chronic pain literature also lacks focus on the presentation of all patient-important outcomes. This is important for future knee OA NMAs to adopt in order to provide a holistic overview of the benefits and harms of all treatments.

Methodology Considerations for Future NMAs

The decision of a follow-up time for pooling outcome measures is an important consideration when planning an NMA, which is demonstrated by the differences in results across the included studies that assessed different timepoints. The decision to assess particular timepoints should be decided *a priori* based on the clinical importance and patient values and preferences in favor of that particular analysis timeframe. For patients with chronic pain and treatments that are not disease modifying, patients would likely be most interested in the longest pain relief possible. There are also some scenarios in which a patient may require a fast and Patient values and preferences should be considered in both future trials, as well as evidence synthesis, to ensure that analysis timeframes align with what is important clinically. With respect to injectables, this may drive future research to focus on repeat injection courses, opposed to the current body of evidence supporting the pain relief profile of a single injection course. When treatments have varying durations of effectiveness, it may be advantageous to include multiple timepoints for the outcomes of interest to determine the trajectory of therapeutic effect across the treatment options. When doing so, there would be a need to efficiently report the results of multiple outcomes – and timepoints – in a clear and easily interpretable manner. With many treatments, multiple relevant timepoints, and multiple clinically relevant outcomes, use of an effective and

easily interpretable presentation tool may help aid in the uptake of NMA results into clinical practice.

Similar to differences in the timeframes analyzed by the included studies, there was a great deal of variability in the types and quantity of treatments included within these NMAs. Ideally, an NMA should consider all available treatment options to provide a comprehensive and holistic summary of the available treatments. Empirical evidence has demonstrated that the exclusion of treatments from an NMA can have substantial and important effects on the results. If an NMA omits available treatment options for the condition in question, the results could drastically differ from an NMA that incorporated all treatments into the analysis, thus reducing the usefulness of NMAs that do not include all available treatment options.[32] Recently, NMAs in the knee OA literature have also begun to utilize the NMA approach to evaluate and identify intra-treatment differences within a class of interventions.[18,19] The most frequent use of this approach was to evaluate high molecular weight and low molecular weight HA's separately, as well as different formulations of PRP – such as leukocyte-rich vs leukocyte-poor PRP - and different types of stem cells.[18,19,23] These analyses could provide further insights into the management of knee OA, and provide further justification for the potential benefits that NMA can provide in elucidating complex clinical questions where many different treatments – along with numerous different formulations of those treatments - exist. Through NMA, all of these specific categories within treatment classes could, if warranted when within-class differences are observed, be analyzed separately against all other treatment options, even when direct RCT evidence is not available.[3]

This does not mean that this is always the optimal analysis approach, and careful consideration must be taken when considering the trade-offs of separating similar treatments

into their own nodes versus pooling all treatments within a broad treatment class.[33] While separating individual treatment formulations and variants out could highlight within-class differences - it also creates a thinner network, which could negatively affect the precision and generalizability of the results. Unless compelling differences within a treatment class exist and warrant separation of these treatments in the network, a reliable answer from a broad analysis of the entire class may be more informative than an imprecise and unreliable answer that is specific to treatments within the class.[33]

Future NMAs should carefully consider guidance for the development of high-quality NMA investigations such as the PRISMA reporting checklist, Cochrane Handbook for Systematic Reviews, and the GRADE instrument for rating the certainty of evidence in NMAs. Similarly, journal editors and reviewers should enforce the adherence to these guidelines and instruments. One of the largest concerns gleaned from this body of knee OA NMAs was the uniform lack of certainty/quality of evidence assessments. Assessments of the overall quality of evidence, such as GRADE, are important to consider along with the magnitude of effect.[4,9] It is imperative that readers understand the certainty of the evidence that is informing the results and conclusions of any meta-analysis, including NMAs, as this has direct impact on the clinical implications to be drawn. Clinically relevant treatment effects drawn from relatively low certainty of evidence may be intriguing, but ultimately prompt further investigation before having practice-changing implications, whereas the same treatment effects seen in a body of high to moderate certainty evidence may warrant immediate adoption in practice. Readers of NMAs should also carefully consider the scope of interventions and timeframes analyzed, as these differences can have important influence on the clinical scope and relevance of the results. Future efforts to evaluate

and provide evidence-based guidance using NMA should assess all available treatment options, at clinically relevant follow-up periods, and for all patient-important benefit and harms outcomes, to inform evidence-based clinical care.

Conclusion

The reporting quality of current NMAs of treatment for knee OA is highly variable, and the length of follow-up assessed and competing interventions considered may drastically affect results. Future NMAs should adhere to reporting checklists, consider all available treatment options, evaluate both benefit and harms, and explore timepoints that are meaningful to patient values and preferences. It is necessary for future investigations to also evaluate the certainty of the evidence within these NMA investigations.

References

1. Foote CJ, Chaudhry H, Bhandari M, et al. Network Meta-analysis: Users' Guide for Surgeons: Part I – Credibility. *Clin Orthop Relat Res*. 2015;473(7):2166-2171. doi:10.1007/s11999-015-4286

2. Chaudhry H, Foote CJ, Guyatt G, et al. Network Meta-analysis: Users' Guide for Surgeons: Part II - Certainty. *Clin Orthop*. 2015;473(7):2172-2178. doi:10.1007/s11999-015-4287-9

3. Mills E, Ioannidis JPA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. Chapter 24: Network Meta-analysis. In: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. McGraw Hill Education; :327-356.

4. Puhan MA, Schünemann 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. doi:10.1136/bmj.g5630

5. Jevsevar DS. Treatment of osteoarthritis of the knee: evidence-based guideline, 2nd edition. *J Am Acad Orthop Surg*. 2013;21(9):571-576. doi:10.5435/JAAOS-21-09-571

6. Trojian TH, Concoff AL, Joy SM, Hatzenbuehler JR, Saulsberry WJ, Coleman CI. AMSSM Scientific Statement Concerning Viscosupplementation Injections for Knee Osteoarthritis: Importance for Individual Patient Outcomes. *Clin J Sport Med Off J Can Acad Sport Med*. 2016;26(1):1-11. doi:10.1097/JSM.0000000000000274

7. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthritis Cartilage*. 2019;27(11):1578-1589. doi:10.1016/j.joca.2019.06.011

8. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ Can Med Assoc J J Assoc Medicale Can*. 1997;156(10):1411-1416.

9. 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. doi:10.1016/j.jclinepi.2017.10.005

10. 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. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385

11. Bannuru RR, Schmid CH, Kent DM, Vaysbrot EE, Wong JB, McAlindon TE. Comparative effectiveness of pharmacologic interventions for knee osteoarthritis: a systematic review and network meta-analysis. *Ann Intern Med*. 2015;162(1):46-54. doi:10.7326/M14-1231

12. Kongtharvonskul J, Anothaisintawee T, McEvoy M, Attia J, Woratanarat P, Thakkinstian A. Efficacy and safety of glucosamine, diacerein, and NSAIDs in osteoarthritis knee: a systematic review and network meta-analysis. *Eur J Med Res*. 2015;20:24. doi:10.1186/s40001-015-0115-7

13. Doros G, Lavin PT, Daley M, Miller LE. A method for establishing class III medical device equivalence: sodium hyaluronate (GenVisc 850) for the treatment of knee osteoarthritis. *Med Devices Auckl NZ*. 2016;9:205-211. doi:10.2147/MDER.S104327

14. Smith SR, Deshpande BR, Collins JE, Katz JN, Losina E. Comparative pain reduction of oral non-steroidal anti-inflammatory drugs and opioids for knee osteoarthritis: systematic analytic review. *Osteoarthritis Cartilage*. 2016;24(6):962-972. doi:10.1016/j.joca.2016.01.135

15. Gregori D, Giacovelli G, Minto C, et al. Association of Pharmacological Treatments With Long-term Pain Control in Patients With Knee Osteoarthritis: A Systematic Review and Meta-analysis. *JAMA*. 2018;320(24):2564-2579. doi:10.1001/jama.2018.19319

16. Jung SY, Jang EJ, Nam SW, et al. Comparative effectiveness of oral pharmacologic interventions for knee osteoarthritis: A network meta-analysis. *Mod Rheumatol*. 2018;28(6):1021-1028. doi:10.1080/14397595.2018.1439694

17. Beaudart C, Lengelé L, Leclercq V, et al. Symptomatic Efficacy of Pharmacological Treatments for Knee Osteoarthritis: A Systematic Review and a Network Meta-Analysis with a 6-Month Time Horizon. *Drugs*. 2020;80(18):1947-1959. doi:10.1007/s40265-020-01423-8

18. Hummer CD, Angst F, Ngai W, et al. High molecular weight Intraarticular hyaluronic acid for the treatment of knee osteoarthritis: a network meta-analysis. *BMC Musculoskelet Disord*. 2020;21. doi:10.1186/s12891-020-03729-w

19. Phillips M, Vannabouathong C, Devji T, et al. Differentiating factors of intra-articular injectables have a meaningful impact on knee osteoarthritis outcomes: a network meta-analysis. *Knee Surg Sports Traumatol Arthrosc Off J ESSKA*. Published online January 3, 2020. doi:10.1007/s00167-019-05763-1

20. Chevalier X, Sheehan B, Whittington C, et al. Efficacy and Safety of Hylan G-F 20 Versus Intra-Articular Corticosteroids in People with Knee Osteoarthritis: A Systematic Review and Network Meta-Analysis. *Clin Med Insights Arthritis Musculoskelet Disord*. 2020;13:1179544120967370. doi:10.1177/1179544120967370

21. Singh H, Knapik DM, Polce EM, et al. Relative Efficacy of Intra-articular Injections in the Treatment of Knee Osteoarthritis: A Systematic Review and Network Meta-analysis. *Am J Sports Med*. Published online August 17, 2021:3635465211029659. doi:10.1177/03635465211029659

22. Anil U, Markus DH, Hurley ET, et al. The efficacy of intra-articular injections in the treatment of knee osteoarthritis: A network meta-analysis of randomized controlled trials. *The Knee*. 2021;32:173-182. doi:10.1016/j.knee.2021.08.008

23. Zhao D, Pan JK, Yang WY, et al. Intra-Articular Injections of Platelet-Rich Plasma, Adipose Mesenchymal Stem Cells, and Bone Marrow Mesenchymal Stem Cells Associated With Better Outcomes Than Hyaluronic Acid and Saline in Knee Osteoarthritis: A Systematic Review and Network Meta-analysis. *Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc*. 2021;37(7):2298-2314.e10. doi:10.1016/j.arthro.2021.02.045

24. Zeng C, Doherty M, Persson MSM, et al. Comparative efficacy and safety of acetaminophen, topical and oral non-steroidal anti-inflammatory drugs for knee osteoarthritis: evidence from a network meta-analysis of randomized controlled trials and real-world data. *Osteoarthritis Cartilage*. 2021;29(9):1242-1251. doi:10.1016/j.joca.2021.06.004

25. Migliorini F, Driessen A, Quack V, et al. Comparison between intra-articular infiltrations of placebo, steroids, hyaluronic and PRP for knee osteoarthritis: a Bayesian network meta-analysis. *Arch Orthop Trauma Surg*. 2021;141(9):1473-1490. doi:10.1007/s00402-020-03551-y

26. Han SB, Seo IW, Shin YS. Intra-Articular Injections of Hyaluronic Acid or Steroids Associated With Better Outcomes Than Platelet-Rich Plasma, Adipose Mesenchymal Stromal Cells, or Placebo in Knee Osteoarthritis: A Network Meta-analysis. *Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc*. 2021;37(1):292-306. doi:10.1016/j.arthro.2020.03.041

27. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions, Version 6.3. Chapter 10: Analyzing Data and Undertaking Meta-Analyses. Published 2022. https://training.cochrane.org/handbook/current

28. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ*. 2020;371:m3900. doi:10.1136/bmj.m3900

29. Brignardello-Petersen R, Izcovich A, Rochwerg B, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ*. 2020;371:m3907. doi:10.1136/bmj.m3907

30. Busse JW, Bartlett SJ, Dougados M, et al. Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop. *J Rheumatol*. 2015;42(10):1962-1970. doi:10.3899/jrheum.141440

31. Mulla SM, Maqbool A, Sivananthan L, et al. Reporting of IMMPACT-recommended core outcome domains among trials assessing opioids for chronic non-cancer pain. *Pain*. 2015;156(9):1615-1619. doi:10.1097/j.pain.0000000000000241

32. Mills EJ, Kanters S, Thorlund K, Chaimani A, Veroniki AA, Ioannidis JPA. The effects of excluding treatments from network meta-analyses: survey. *BMJ*. 2013;347:f5195. doi:10.1136/bmj.f5195

33. Gotzsche PC. Why we need a broad perspective on meta-analysis. It may be crucially important for patients. *BMJ*. 2000;321(7261):585-586. doi:10.1136/bmj.321.7261.58

Appendix A: Literature Search Strategies

Medline Search		
1	Network Meta-Analysis/ or network meta analysis.mp	6161
2	NMA.ti,ab.	2658
3	Mixed Treatment.ti,ab.	600
4	1 or 2 or 3	8029
5	knee.ti,ab.	152976
6	Osteoarthritis, Knee/ or osteoarthritis.mp	97976
7	5 and 6	37157
8	4 and 7	82

EMBASE Search		
1	Network Meta-Analysis/ or network meta analysis.mp	8914
2	NMA.ti,ab.	4025
3	Mixed Treatment.ti,ab.	1094
4	1 or 2 or 3	11679
5	knee.ti,ab.	197174
6	knee osteoarthritis/ or osteoarthritis.mp	156140
7	5 and 6	56789
8	4 and 7	117

Cochrane Database of Systematic Reviews Search		
1	(Network Meta Analysis):ti,ab,kw	655
2	(NMA):ti,ab,kw	134
3	(Mixed Treatment):ti,ab,kw	21401
4	1 or 2 or 3	22043
5	knee	33028
6	osteoarthritis	20138
7	5 and 6	13824
8	4 and 7 (Limit to reviews)	11

Chapter 4: Developing an improved methodology for reporting NMA results

Part A: Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study

Mark R Phillips[1], Behnam Sadeghirad[1,2], Jason W Busse[1,2], Romina Brignardello-Petersen[1], Carlos A Cuello-Garcia[1], Fernando Kenji Nampo[3], Yu Jia Guo[4], Sofia Bzovsky[5], Raveendhara R Bannuru[6], Lehana Thabane[1,7], Mohit Bhandari[1,8], Gordon H Guyatt[1]

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada
3. Department of Latin-American Institute of Life and Nature science, Federal University of Latin-American Integration, Foz do Iguacu, Brazil
4. Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
5. Division of Orthopaedics, McMaster University, Hamilton, Ontario, Canada
6. Center for Treatment Comparison and Integrative Analysis, Tufts Medical Center, Boston, Massachusetts, USA
7. Biostatistics Unit, St. Joseph's Healthcare, Hamilton, Ontario, Canada
8. Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Abstract

Objective The Grading of Recommendations, Assessment, Development and Evaluation working group recently developed an innovative approach to interpreting results from network meta-analyses (NMA) through minimally and partially contextualised methods; however, the optimal method for presenting results for multiple outcomes using this approach remains uncertain. We; therefore, developed and iteratively modified a presentation method that effectively summarises NMA results of multiple outcomes for clinicians using this new interpretation approach.

Design Qualitative descriptive study.

Setting A steering group of seven individuals with experience in NMA and design validation studies developed two colour-coded presentation formats for evaluation. Through an iterative process, we assessed the validity of both formats to maximise their clarity and ease of interpretation.

Participants 26 participants including 20 clinicians who routinely provide patient care, 3 research staff/research methodologists and 3 residents.

Main outcome measures Two team members used qualitative content analysis to independently analyse transcripts of all interviews. The steering group reviewed the analyses and responded with serial modifications of the presentation format.

Results To ensure that readers could easily discern the benefits and safety of each included treatment across all assessed outcomes, participants primarily focused on simple information

presentations, with intuitive organisational decisions and colour coding. Feedback ultimately resulted in two presentation versions, each preferred by a substantial group of participants, and development of a legend to facilitate interpretation.

Conclusion Iterative design validation facilitated the development of two novel formats for presenting minimally or partially contextualised NMA results for multiple outcomes. These presentation approaches appeal to audiences that include clinicians with limited familiarity with NMAs.

Strengths and limitations of this study

- Extensive design validation in a targeted audience has validated the network meta-analyses (NMA) presentation approaches within this study; something that has not been done for other presentation formats.
- Structured qualitative research methodology has ensured accurate use of user feedback to develop and refine the NMA presentation formats.
- Limited by the omission of some information within the presentation formats in order to achieve simplicity and interpretability, such as greater detail for individual outcomes, absolute effects or specifics about the certainty of evidence assessments.
- The aforementioned information should still be included in NMA manuscripts, but cannot be feasibly fit within the presentation formats.

Introduction

Network meta-analysis (NMA) provides an increasingly popular approach to evidence synthesis that allows comparison between multiple competing treatment options within a single analysis.[1,2] Although NMA is an important tool for clinicians, patients and other stakeholders, results involve multiple treatments and outcomes, and as a result are complex and difficult to interpret.[3]

Common methods for presenting NMA results include the use of forest plots, league tables and surface under the cumulative ranking curve.[1,4] The key limitation with these options is that they can only provide results of a single outcome.[5] NMAs often compare multiple benefit and harm outcomes, resulting in challenges for NMA authors seeking to avoid presentation methods that are onerous for clinicians to review and challenging for them to understand.[6]

There are a number of novel approaches that have been suggested for presenting NMA results for multiple outcomes[7,8]; however, these approaches lack key information, present challenges to interpretation and have not undergone design validation with their target audiences. While some previously suggested approaches have merit for a limited number of outcomes,[4,6,9-12] although not all taking certainty of evidence into account, they have serious limitations for simultaneous presentation of multiple outcomes.

Recently, the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group has suggested two variations on a new methodology that places interventions in categories from best to worst considering the estimates of effect and certainty of the evidence for each comparison.[13,14] We; therefore, developed interpretable presentation

approaches for NMAs with multiple outcomes that builds on GRADE guidance and effectively summarises results for clinicians and other relevant audiences.

Methods

Study design

A seven-member steering committee (MRP, BS, JWB, RB-P, CAC-G, FKN and GG) oversaw study design and implementation. The committee generated two initial presentation formats and chose a combination of large group sessions and individual design validation interviews to inform iterative modifications of the two initial formats. The presentation format consisted of treatment options in rows and outcomes in columns, with colour-coded shading of cells to identify the magnitude and certainty of the treatment effect in relation to the reference treatment. The steering committee developed the initial versions through a series of internal group discussions, which involved: determining the pertinent information for the presentation format to contain, options for how that information could be shown within a single presentation format, and draft presentation formats that may present this pertinent information. The group believed that the format should provide both relative treatment effects, as well as the certainty in those estimates for all outcomes, within a single presentation tool.

The steering committee developed initial versions of the presentation tool, which they then presented in separate large-group settings to gain outside insight. Initial large group testing with two groups of methodologists, graduate students in health research-focussed programmes and statisticians, as well as presentation at a national conference (2019 Canadian Pain Society annual scientific meeting), provided the foundational feedback for modifications of the initial presentation versions. After making iterative improvements from the group presentation

feedback, the steering committee began one-on-one interviews with clinicians to gain further insights for improvement. The steering committee reviewed input from four rounds of design validation individual interviews, iteratively modifying the formats after each round and presenting updated options of the presentation versions to subsequent participants.

For the user interviews, the committee chose a qualitative descriptive study approach that focuses on creating a close description of the information that participants provide.[15] This is ideal for design validation that, without interpretive direction, aims to optimise the understandability of a tool within the target population. Participants provided informed consent at the beginning of their interview. We followed, when applicable, the consolidated criteria for reporting qualitative research checklist in reporting our findings.[16]

Sampling and recruitment

This study used purposeful sampling to identify participants who could provide information-rich interviews to inform the design validation process.[15,17] Target users for this study included academic and non-academic clinicians, research staff/research methodologists and residents. The steering committee, through their professional contacts, provided a pool of initial possible participants that the principal investigator supplemented using snowball sampling technique.[18] Specifically, we asked individuals who agreed to participate for contact information of any colleagues whom we could approach to interview. Prior to their interviews, each participant received information outlining the purpose of the study. Study recruitment ceased when data collection reached redundancy—the point at which there were no further refinements requested to improve the interpretability of the presentation formats.[18]

Data collection

The principal investigator (MRP) conducted all design validation interviews either in-person or through video teleconferencing. Interviews followed a flexible interview guide to leave the conversation open for participants to explore any topics they felt were relevant and important.[15] Throughout the study, the principal investigator iteratively updated the interview guide to explore areas of importance that emerged. Interviews began with a brief introduction to NMA methods, followed by questions regarding the participant's familiarity and experience with NMA. Participants then viewed the current versions of the NMA presentation formats and provided feedback. YJG or MRP transcribed all interviews verbatim. Transcripts were not returned to participants and interviewers did not conduct follow-up interviews. The steering committee incorporated all feedback to arrive at two final presentation versions.

Patient and public involvement

This study did not include patient or public involvement.

NMA for design validation

The steering committee developed five core criteria to which the example NMA must adhere: (1) variability in quality of evidence (2) variability in magnitudes of effect; (3) assessment of both benefits and harms; (4) inclusion of both continuous and binary outcomes; and (5) including at least five outcomes and five interventions. Based on these criteria the steering committee chose, for design validation, a recent NMA that used a minimally contextualised approach to address acute pain management in patients experiencing non-low back acute musculoskeletal injuries.[19]

Based on the GRADE approach,[13] this NMA categorised, for each benefit outcome, interventions as among those with the largest benefit, those with intermediate benefit, and those with the least benefit. For each harm outcome, they categorised interventions as among the least harmful, intermediate harm and the most harmful. They then categorised interventions as those for which there was high or moderate certainty evidence, and those for which there was low or very low-quality evidence.[19] These results provided the example for design validation.

Data analysis

Two reviewers (MRP and SB) independently conducted data analysis, in duplicate, using a qualitative content analysis approach.[17] The study team recruited participants, collected data and conducted data analysis in parallel. As new data became available, the reviewers coded and grouped similar phrases, patterns and themes.[17] When discrepancies in feedback were identified, these would be noted and further elaborated on within future interviews. The feedback for this discrepancy would then be shared with the steering committee to review and identify if sufficient data had been captured to adequately determine a resolution for the discrepancy through consensus.[17] Data triangulation was used through multiple forms of data collection, as both large group and individual interview sessions were used. Additionally, data triangulation was provided through two forms of data analysis: independent qualitative content analysis, and group deliberation through steering committee meetings.[17,20] The steering committee met four times over a period of 14 months to review the collected data and made iterative changes to the presentation formats as dictated by feedback, initially from large group presentations and subsequently from design validation. When analysis of the data provided actionable feedback, the reviewers presented their findings to the steering committee who ranked feedback as a 'large

change required', 'moderate change required' or 'minor change required' and then revised the presentation format(s) accordingly.

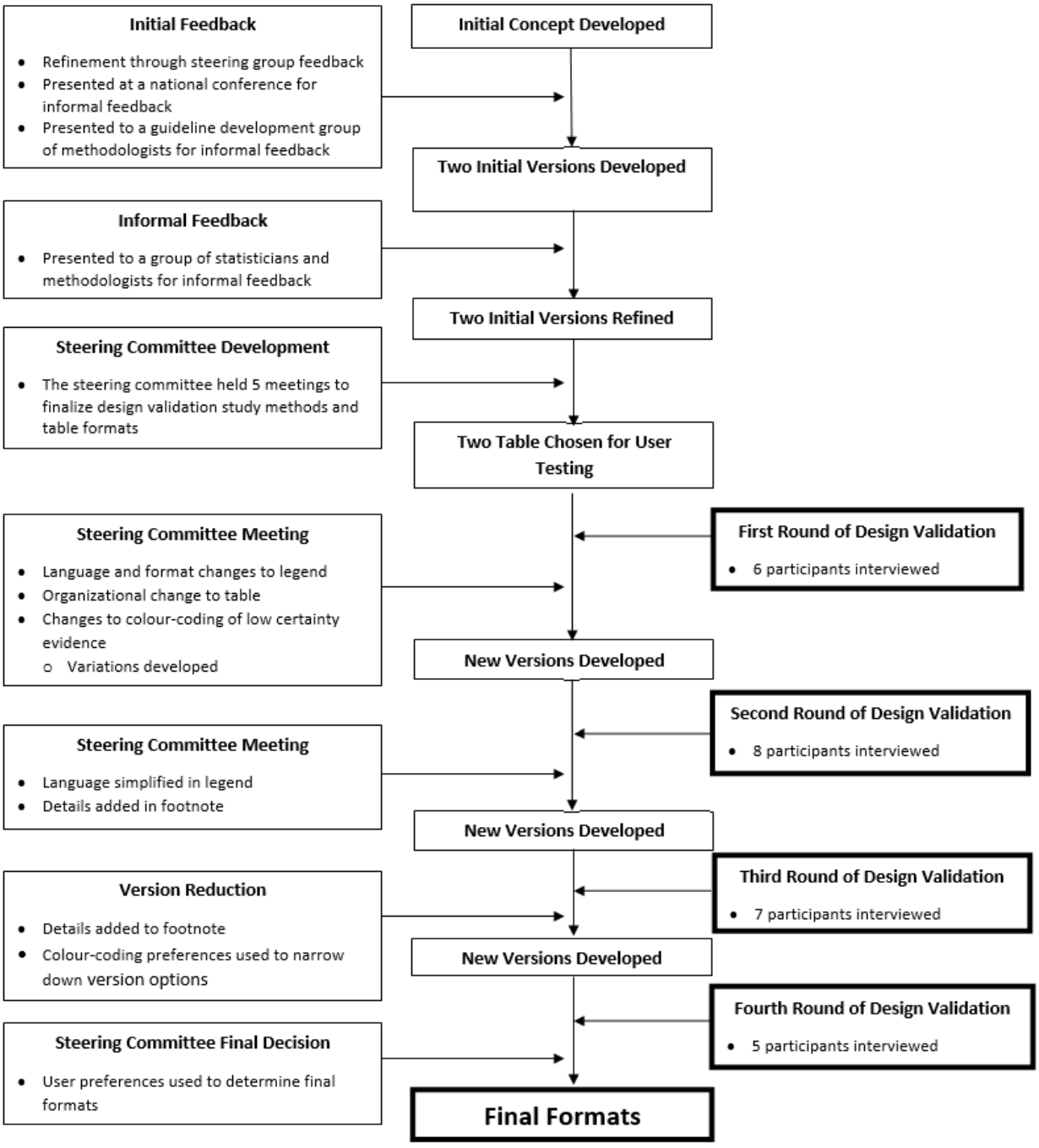
Subsequent participants provided input on the modified versions of the NMA results presentations. Participants commented regarding their interpretation of the data within the presentation format; the team considered study objectives met once participants consistently reported a clear interpretation of the results with no or minimal suggested modifications. Reviewers documented all changes to the presentation format in a study audit trail.[15,20] Reviewers conducted all qualitative analysis using RQDA software (R V.3.5.0).

Results

Study sample

Two focus groups, both of which included methodologists, graduate students and statisticians, participated in the initial large group testing: the first, a critical care guideline development group (GUIDE: https://guidecanada.org/) many of whose members have NMA expertise (65 attendees); the second, a research group (CLARITY: http://www.clarityresearch.ca/) who meet regularly at McMaster University to discuss current methodological and statistical topics (20 attendees).

Figure 1: Study Overview



The design validation portion of this study included 26 participants of mean (SD) age of 47.6 (13.9) years, 20 of whom were clinicians whose primary activity involved direct patient care (77%); 3 research staff/research methodologists (12%) and 3 residents (12%). Typical participants were male (73%) physicians in clinical practice for almost two decades (mean (SD): 19.5 (14.3) years) with no prior involvement with conducting an NMA (58%) (**Table 1**).

Table 1: Participant Demographics: n=26

Demographic	Value
Age (Mean, SD) years	47.6 (13.9)
Gender (Count, %)	
Male	19 (73.1%)
Female	7 (26.9%)
Primary Occupation (Count, %)	
Clinician	20 (76.9%)
Research Staff/ Methodologist	3 (11.5%)
Resident	3 (11.5%)
Highest Degrees Held (Count, %)	
MD	12 (46.2%)
MD, MSc/MPH	8 (30.8%)
PhD	3 (11.5%)
MD, PhD	2 (7.7%)
BSc	1 (3.9%)
Years in Practice (Mean, SD)	19.5 (14.3)
Previous involvement in an NMA? (Count, %)	
Yes	11 (42.3%)
No	15 (57.7%)
Used an NMA to inform practice? (Count, %)	
Yes	17 (65.4%)
No	9 (34.6%)

SD: Standard Deviation, MD: Doctor of Medicine, MSc: Masters of Science, MPH: Masters of Public Health, PhD: Doctor of Philosophy, BSc: Bachelor of Science, NMA: Network Meta-Analysis.

Content analysis themes

Main themes that arose from the content analysis conducted on interview transcripts of participant interviews included 'organisational', 'language/terminology', 'included information' and 'colour options'. Respondents also provided feedback regarding necessary details to include in the presentations' footnote. The following sections provide details regarding the most important feedback and how this feedback informed choices regarding presentation format. The fourth round of design validation resulted in minimal new information, resulting in two presentation versions that participants deemed satisfactory.

Final presentation versions

Ultimately, respondents proved equally enthusiastic about two options; the steering group, therefore, chose to offer both as alternative presentations. **Figure 1** summarises the development process from conceptualisation to the final presentation versions. We will refer to the presentation in **Figure 2** as the 'colour gradient' version and the presentation in **Figure 3** as the 'stoplight' version. Each presentation has a legend and footnote with pertinent information that the design validation process demonstrated necessary to include.

Figure 2: Gradient Colour Variation

Intervention	BENEFIT OUTCOMES					ADVERSE EVENTS		
	Pain ≤ 2 h post-tx	Pain 1 to 7 d post-tx	Physical function	Treatment satisfaction	Symptom relief	GI-related AE's	Neurologic AE's	Dermatologic AE's
	MD (95% CI)	MD (95% CI)	MD (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Topical NSAID	**-1.02 (-1.64,-0.39)**	**-1.08 (-1.40,-0.75)**	**1.66 (1.16,2.16)**	**5.20 (2.03,13.33)**	**6.39 (3.48,11.75)**	1.14 (0.65,2.01)	1.18 (0.51,2.74)	0.78 (0.52,1.15)
Oral NSAID	**-0.93 (-1.49,-0.37)**	**-0.99 (-1.46,-0.52)**	**0.73 (0.17,1.30)**	3.24 (0.43,24.70)	**3.10 (1.39,6.91)**	**1.77 (1.33,2.35)**	1.02 (0.65,1.59)	1.33 (0.43,4.09)
Acetaminophen	**-1.03 (-1.82,-0.24)**	**-1.07 (-1.89,-0.24)**	0.90 (-0.27,2.61)	2.43 (0.18,32.70)	2.73 (0.90,8.27)	0.50 (0.06,4.38)	-	-
Acetaminophen + Diclofenac	**-1.11 (-2.00,-0.21)**	-1.09 (-2.20,0.01)	-	3.45 (0.18,66.96)	**3.72 (1.02,13.52)**	-	-	-
Topical NSAID + Menthol Gel	**-1.68 (-0.27,-3.09)**	-0.89 (-2.33,0.54)	-	-	**13.34 (3.30,53.92)**	2.35 (0.04,124.85)	1.22 (0.02,69.98)	0.53 (0.05,6.29)
TENS	**-1.94 (-2.90,-0.98)**	**-1.18 (-2.09,-0.28)**	0.68 (-0.20,1.57)	-	6.00 (0.78,46.36)	1.25 (0.14,11.01)	1.12 (0.13,9.98)	1.18 (0.13,11.03)
Specific acupressure	**-1.59 (-2.52,-0.66)**	**-2.09 (-3.86,-0.32)**	**1.51 (0.42,2.61)**	0.50 (0.04,6.49)	2.54 (0.52,12.38)	0.80 (0.02,41.67)	0.80 (0.01,42.60)	0.80 (0.01,45.60)
Manipulation	**-1.75 (-2.68,-0.81)**	0.40 (-1.71,2.51)	0.09 (-1.06,0.87)	-	**167.71 (6.67,4217.10)**	0.50 (0.01,31.30)	1.41 (0.03,78.76)	-
Acetaminophen + Chlorzoxazone	-	**-2.92 (-5.41,-0.43)**	-	-	-	0.35 (0.01,10.59)	-	-
Laser therapy	-	-1.04 (-2.28,0.19)	-	-	**32.08 (4.93,208.60)**	0.49 (0.01,24.85)	0.49 (0.01,25.41)	0.49 (0.01,27.21)
Mobilization	-	3.40 (-0.05,6.85)	0.12 (-0.59,0.83)	2.07 (0.07,58.49)	**7.99 (1.29,49.41)**	0.93 (0.02,47.12)	0.93 (0.02,48.18)	0.93 (0.02,51.60)
Acetaminophen + Opioid	-0.52 (-1.47,0.43)	**-1.71 (-2.97, -0.46)**	-	2.50 (0.14,44.86)	1.47 (0.55,3.91)	**5.63 (2.84,11.16)**	**3.53 (1.92,6.49)**	-
Acetaminophen, Ibuprofen + Codeine	**-1.36 (-2.49,-0.23)**	-	-	-	-	-	-	-
Acetaminophen + Ibuprofen	-0.70 (-1.62,0.22)	-1.18 (-2.74,0.38)	-	-	3.62 (0.99,13.14)	-	-	-
Non-Specific Acupressure	-0.05 (-0.99,0.89)	-0.18 (-1.91,1.55)	-0.18 (-1.32,0.96)	0.44 (0.03,5.76)	1.80 (0.36,9.03)	0.85 (0.02,44.76)	0.85 (0.02,45.76)	0.85 (0.01,48.97)
Exercise	-	-0.81 (-2.64,1.02)	-0.43 (-1.00,0.14)	3.50 (0.21,59.42)	0.84 (0.31,2.29)	1.04 (0.06,17.06)	1.08 (0.07,17.95)	1.08 (0.06,18.84)
Cyclobenzaprine	-	-2.03 (-4.11,0.06)	-	-	-	0.64 (0.03,15.74)	1.95 (0.20,18.88)	-
Supervised Rehab	-	0.96 (-0.35,2.27)	0.24 (-0.59,1.07)	2.25 (0.15,34.07)	5.09 (0.84,30.78)	1.06 (0.02,54.49)	1.06 (0.02,55.71)	1.06 (0.02,59.65)
Ibuprofen + Cyclobenzaprine	-1.05 (-2.63,0.53)	-1.51 (-3.06,0.04)	-	5.52 (0.21,147.01)	-	1.10 (0.13,9.42)	**4.91 (1.45,16.61)**	-
Menthol Gel	-	-1.14 (-2.28,0.00)	0.70 (-0.61,2.02)	-	-	-	-	1.00 (0.11,8.91)
Ultrasound	-	-0.40 (-2.46,1.66)	-	-	-	-	-	-
Glucosamine	-	-0.10 (-1.89,1.69)	-	-	-	-	-	-
Phenyramidol	-	-	-	-	-	-	0.32 (0.01,8.45)	-
Massage therapy	-0.70 (-1.90,0.50)	-	-	-	-	-	-	-
Education	-	-	0.10 (-0.67,0.87)	-	0.93 (0.39,2.24)	-	-	-
Acetaminophen, Ibuprofen + Oxycodone	-0.94 (-2.27,0.38)	-	-	-	-	-	-	-
Fentanyl	**-3.52 (-4.99,-2.04)**	-	-	-	-	**59.38 (6.21,567.71)**	**5.73 (1.20,27.47)**	-
Tramadol	0.95 (-0.80,2.70)	-	-	-	-	5.98 (0.33,108.25)	**6.72 (1.24,36.39)**	-

Legend

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	**Better than placebo and some other interventions**	**May be better than placebo and some alternatives**	**No more harmful than placebo**	**May be no more harmful than placebo**
INTERMEDIATE	**Better than placebo, but no better than any other interventions**	**May be better than placebo, but no better than other interventions**	**More harmful than placebo, but no worse than other interventions**	**May be more harmful than placebo, but no worse than other interventions**
AMONG THE WORST	**No better than placebo**	**May be no better than placebo**	**More harmful than placebo and some other interventions**	**May be more harmful than placebo and some alternatives**

Footnote

- : no evidence

Reference Group = Placebo
Bold = statistically significant ($p<0.05$)
MD: Mean Difference
OR: Odds Ratio
CI: Confidence Interval
h: hours
d: days
tx: treatment
AE: adverse event
NSAID: non-steroidal anti-inflammatory drug
TENS: transcutaneous electrical nerve stimulation

Figure 3: Stoplight Colour Version

Intervention	BENEFIT OUTCOMES					ADVERSE EVENTS		
	Pain ≤ 2 h post-tx	Pain 1 to 7 d post-tx	Physical function	Treatment satisfaction	Symptom relief	GI-related AE's	Neurologic AE's	Dermatologic AE's
	MD (95% CI)	MD (95% CI)	MD (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Topical NSAID	-1.02 (-1.64,-0.39)	-1.08 (-1.40,-0.75)	1.66 (1.16,2.16)	5.20 (2.03,13.33)	6.39 (3.48,11.75)	1.14 (0.65,2.01)	1.18 (0.51,2.74)	0.78 (0.52,1.15)
Oral NSAID	-0.93 (-1.49,-0.37)	-0.99 (-1.46,-0.52)	0.73 (0.17,1.30)	3.24 (0.43,24.70)	3.10 (1.39,6.91)	1.77 (1.33,2.35)	1.02 (0.65,1.59)	1.33 (0.43,4.09)
Acetaminophen	-1.03 (-1.82,-0.24)	-1.07 (-1.89,-0.24)	0.90 (-0.27,2.61)	2.43 (0.18,32.70)	2.73 (0.90,8.27)	0.50 (0.06,4.38)	-	-
Acetaminophen + Diclofenac	-1.11 (-2.00,-0.21)	-1.09 (-2.20,0.01)	-	3.45 (0.18,66.96)	3.72 (1.02,13.52)	-	-	-
Topical NSAID + Menthol Gel	-1.68 (-0.27,-3.09)	-0.89 (-2.33,0.54)	-	-	13.34 (3.30,53.92)	2.35 (0.04,124.85)	1.22 (0.02,69.98)	0.53 (0.05,6.29)
TENS	-1.94 (-2.90,-0.98)	-1.18 (-2.09,-0.28)	0.68 (-0.20,1.57)	-	6.00 (0.78,46.36)	1.25 (0.14,11.01)	1.12 (0.13,9.98)	1.18 (0.13,11.03)
Specific acupressure	-1.59 (-2.52,-0.66)	-2.09 (-3.86,-0.32)	1.51 (0.42,2.61)	0.50 (0.04,6.49)	2.54 (0.52,12.38)	0.80 (0.02,41.67)	0.80 (0.01,42.60)	0.80 (0.01,45.60)
Manipulation	-1.75 (-2.68,-0.81)	0.40 (-1.71,2.51)	0.09 (-1.06,0.87)	-	167.71 (6.67,4217.10)	0.50 (0.01,31.30)	1.41 (0.03,78.76)	-
Acetaminophen + Chlorzoxazone	-	-2.92 (-5.41,-0.43)	-	-	-	0.35 (0.01,10.59)	-	-
Laser therapy	-	-1.04 (-2.28,0.19)	-	-	32.08 (4.93,208.60)	0.49 (0.01,24.85)	0.49 (0.01,25.41)	0.49 (0.01,27.21)
Mobilization	-	3.40 (-0.05,6.85)	0.12 (-0.59,0.83)	2.07 (0.07,58.49)	7.99 (1.29,49.41)	0.93 (0.02,47.12)	0.93 (0.02,48.18)	0.93 (0.02,51.60)
Acetaminophen + Opioid	-0.52 (-1.47,0.43)	-1.71 (-2.97, -0.46)	-	2.50 (0.14,44.86)	1.47 (0.55,3.91)	5.63 (2.84,11.16)	3.53 (1.92,6.49)	-
Acetaminophen, Ibuprofen + Codeine	-1.36 (-2.49,-0.23)	-	-	-	-	-	-	-
Acetaminophen + Ibuprofen	-0.70 (-1.62,0.22)	-1.18 (-2.74,0.38)	-	-	3.62 (0.99,13.14)	-	-	-
Non-Specific Acupressure	-0.05 (-0.99,0.89)	-0.18 (-1.91,1.55)	-0.18 (-1.32,0.96)	0.44 (0.03,5.76)	1.80 (0.36,9.03)	0.85 (0.02,44.76)	0.85 (0.02,45.76)	0.85 (0.01,48.97)
Exercise	-	-0.81 (-2.64,1.02)	-0.43 (-1.00,0.14)	3.50 (0.21,59.42)	0.84 (0.31,2.29)	1.04 (0.06,17.06)	1.08 (0.07,17.95)	1.08 (0.06,18.84)
Cyclobenzaprine	-	-2.03 (-4.11,0.06)	-	-	-	0.64 (0.03,15.74)	1.95 (0.20,18.88)	-
Supervised Rehab	-	0.96 (-0.35,2.27)	0.24 (-0.59,1.07)	2.25 (0.15,34.07)	5.09 (0.84,30.78)	1.06 (0.02,54.49)	1.06 (0.02,55.71)	1.06 (0.02,59.65)
Ibuprofen + Cyclobenzaprine	-1.05 (-2.63,0.53)	-1.51 (-3.06,0.04)	-	5.52 (0.21,147.01)	-	1.10 (0.13,9.42)	4.91 (1.45,16.61)	-
Menthol Gel	-	-1.14 (-2.28,0.00)	0.70 (-0.61,2.02)	-	-	-	-	1.00 (0.11,8.91)
Ultrasound	-	-0.40 (-2.46,1.66)	-	-	-	-	-	-
Glucosamine	-	-0.10 (-1.89,1.69)	-	-	-	-	-	-
Phenyramidol	-	-	-	-	-	-	0.32 (0.01,8.45)	-
Massage therapy	-0.70 (-1.90,0.50)	-	-	-	-	-	-	-
Education	-	-	0.10 (-0.67,0.87)	-	0.93 (0.39,2.24)	-	-	-
Acetaminophen, Ibuprofen + Oxycodone	-0.94 (-2.27,0.38)	-	-	-	-	-	-	-
Fentanyl	-3.52 (-4.99,-2.04)	-	-	-	-	59.38 (6.21,567.71)	5.73 (1.20,27.47)	-
Tramadol	0.95 (-0.80,2.70)	-	-	-	-	5.98 (0.33,108.25)	6.72 (1.24,36.39)	-

Legend

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	**Better than placebo and some alternatives**	**May be better than placebo and some alternatives**	**No more harmful than placebo**	**May be no more harmful than placebo**
INTERMEDIATE	**Better than placebo, but no better than any alternatives**	**May be better than placebo, but no better than any alternatives**	**More harmful than placebo, but no worse than any alternatives**	**May be more harmful than placebo, but no worse than any alternatives**
AMONG THE WORST	**No better than placebo**	**May be no better than placebo**	**More harmful than placebo and some alternatives**	**May be more harmful than placebo and some alternatives**

Footnote

- : no evidence

Reference Group = Placebo
Bold = statistically significant, p<0.05
MD: Mean Difference
OR: Odds Ratio
CI: Confidence Interval
h: hours
d: days
tx: treatment
AE: adverse event
NSAID: non-steroidal anti-inflammatory drug
TENS: transcutaneous electrical nerve stimulation

Figure organisation

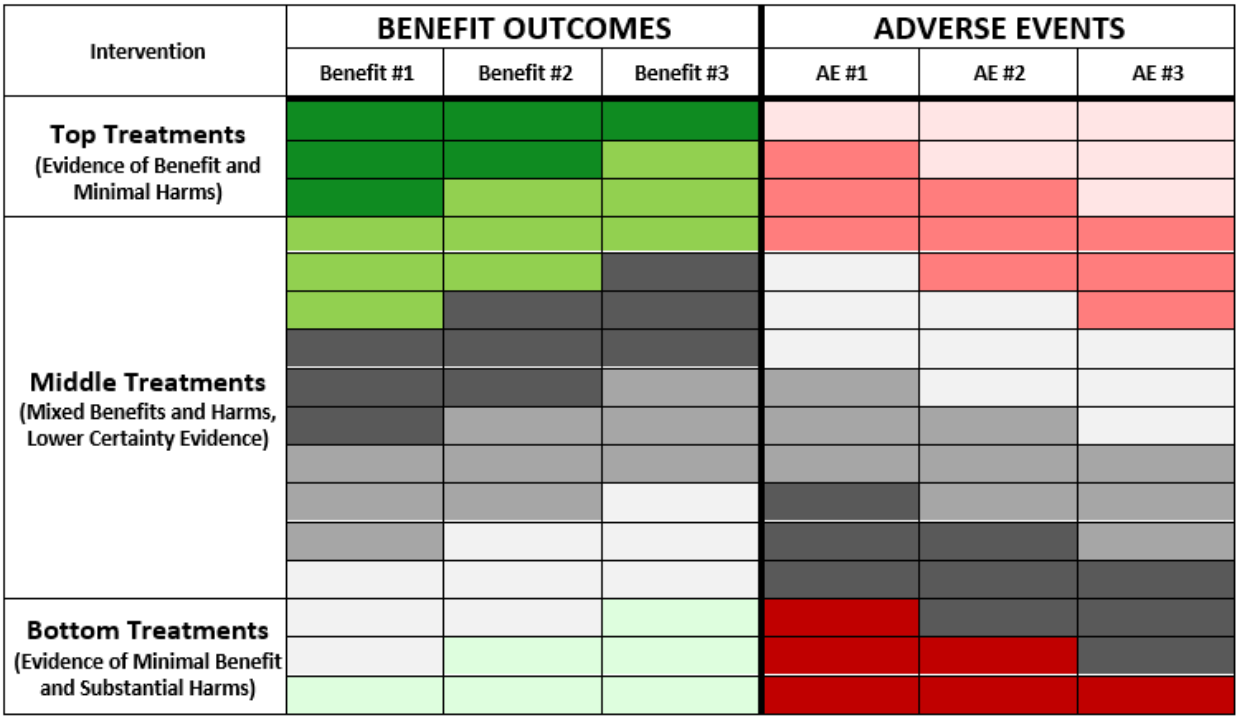
Design validation identified a number of key components that aid in interpreting presentation formats. Within the organisational theme, the use of a bolded vertical line to separate benefit and adverse event outcomes, as well as the header and results data (horizontal), proved desirable. Regarding the ordering of interventions from top to bottom in the rows, participants preferred ordering treatment options at the top with high/moderate certainty evidence of maximal benefit and minimal harm to those with high/moderate certainty evidence of minimal or no benefits and significant harms placed in the bottom rows. Respondents provided mixed feedback regarding the organisation of the presentation within the middle section, with no consistent guidance that could be applied across all NMAs. This leaves the optimal ordering within the middle rows that include treatments that have low/very low certainty evidence, treatments with high/moderate certainty evidence of intermediate effects and treatments with trade-offs between both large benefits and large harms, uncertain (or perhaps there is no single optimal ordering). **Figure 4** provides an overview of guidance regarding intervention order within the rows.

Presentation terminology

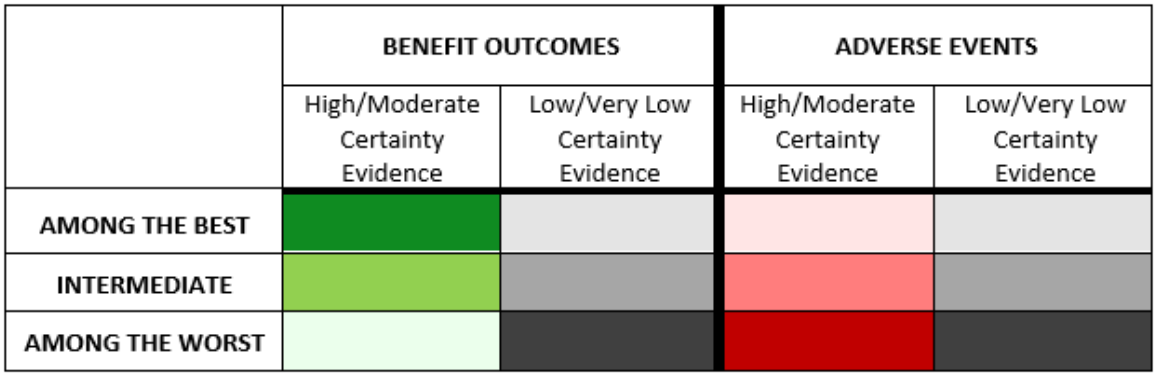
Respondents indicated that the presentation should clearly and succinctly label outcomes with specification of the measure of treatment effect (eg, ORs mean differences) and that the header of each column should include these labels. Participants had no strong preference regarding the terminology of 'benefit' and 'adverse events' outcome categories; options discussed included 'effectiveness/efficacy outcomes' and 'harms outcomes'. Whatever option

investigators choose, the terminology should remain consistent across the presentation, legend and manuscript text.

Figure 4: Intervention Organizational Guide



Legend



Presentation included information

Participants considered the magnitude of treatment effect, CIs/credible intervals, certainty of evidence and statistical significance to be the four important elements that should be included in each comparison cell. Possibilities explicitly discussed but rejected included sample size, patient characteristics and heterogeneity/incoherence estimates. Respondents considered these items as important elements of the NMA, but felt they would be better suited within another section of the manuscript rather than within this summary presentation.

Footnote included information

Participants felt that footnotes should include: an indication of a dash representing no available evidence (-: no evidence); designation of the reference group (eg, reference group: placebo); and labelling of how statistical significance within the presentation is identified (ie, Bold=statistically significant, $p<0.05$); as well as all abbreviations used within the presentation.

Legend organisation

Participants felt that benefit outcomes should be located in the left columns, with a bold vertical line separating the benefit and adverse event outcomes within the legend—similar to the structure of the main presentation. They also suggested a bold horizontal line separating the header from the legend in a similar format as within the main presentation. Within the benefit and adverse event sections, respondents preferred that high/moderate certainty evidence categories should be presented in the left column, and low/very low certainty in the right column. High and moderate certainty evidence, as well as low and very low certainty evidence were grouped together to simplify the presentation format into two groups (high/moderate and

low/very low), as participants perceived these groupings to hold similar weight in clinical decision making.

Legend terminology

Participants encouraged the use of simple language within the legend. Participants preferred legend rows organised from 'among the best' to 'among the worst' vertically down the first column of the legend, with the middle category labelled as 'intermediate'. Terms such as 'better' and 'worse' were clearer to participants than terminology such as 'statistically significant'; specifically, respondents favoured 'better than placebo' over 'statistically significant over placebo'.

The language used for our NMA example, in accordance with the minimally contextualised approach, contained treatments that were 'better than placebo and some other interventions', 'better than placebo, but no better than any other interventions', and 'no better than placebo' for high/moderate certainty evidence of benefit outcomes. For high/moderate certainty evidence of harm outcomes, the corresponding language was 'no more harmful than placebo', 'more harmful than placebo, but no worse than other interventions', and 'more harmful than placebo and some other interventions'. Participants felt that, with respect to category of magnitude of effect low/very low certainty evidence descriptions should be the same as those of the high/moderate certainty evidence categories, with the included qualifier of 'may be' at the beginning of the description of low to very low certainty evidence.

Gradient colour coding

The gradient colour-coding scheme uses three shades of green for the high/moderate certainty benefit outcomes (**Figure 5**: cells 1–3), and three shades of red for the high/moderate

certainty adverse events (**Figure 5**: cells 7–9). The use of three-shade grey gradient for low/very low certainty evidence is consistent for both beneficial outcomes and adverse events (**Figure 5**: cells 4–6, 10–12). Participants preferred dark grey be used for the 'among the worst' category (least beneficial or most harmful) and light grey be used for the 'among the best' category (most beneficial or least harmful), when presenting low/very low certainty of evidence results.

Figure 5: Gradient Colour-Coding Legend

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	1	4	7	10
INTERMEDIATE	2	5	8	11
AMONG THE WORST	3	6	9	12

Presentation terminology

Respondents indicated that the presentation should clearly and succinctly label outcomes with specification of the measure of treatment effect (eg, ORs mean differences) and that the header of each column should include these labels. Participants had no strong preference regarding the terminology of 'benefit' and 'adverse events' outcome categories; options discussed included 'effectiveness/efficacy outcomes' and 'harms outcomes'. Whatever option investigators choose, the terminology should remain consistent across the presentation, legend and manuscript text.

Presentation included information

Participants considered the magnitude of treatment effect, CIs/credible intervals, certainty of evidence and statistical significance to be the four important elements that should be included in each comparison cell. Possibilities explicitly discussed but rejected included sample size, patient characteristics and heterogeneity/incoherence estimates. Respondents considered these items as important elements of the NMA, but felt they would be better suited within another section of the manuscript rather than within this summary presentation.

Footnote included information

Participants felt that footnotes should include: an indication of a dash representing no available evidence (-: no evidence); designation of the reference group (eg, reference group: placebo); and labelling of how statistical significance within the presentation is identified (ie, Bold=statistically significant, $p<0.05$); as well as all abbreviations used within the presentation.

Legend organisation

Participants felt that benefit outcomes should be located in the left columns, with a bold vertical line separating the benefit and adverse event outcomes within the legend—similar to the structure of the main presentation. They also suggested a bold horizontal line separating the header from the legend in a similar format as within the main presentation. Within the benefit and adverse event sections, respondents preferred that high/moderate certainty evidence categories should be presented in the left column, and low/very low certainty in the right column. High and moderate certainty evidence, as well as low and very low certainty evidence were grouped together to simplify the presentation format into two groups (high/moderate and

low/very low), as participants perceived these groupings to hold similar weight in clinical decision making.

Legend terminology

Participants encouraged the use of simple language within the legend. Participants preferred legend rows organised from 'among the best' to 'among the worst' vertically down the first column of the legend, with the middle category labelled as 'intermediate'. Terms such as 'better' and 'worse' were clearer to participants than terminology such as 'statistically significant'; specifically, respondents favoured 'better than placebo' over 'statistically significant over placebo'.

The language used for our NMA example, in accordance with the minimally contextualised approach, contained treatments that were 'better than placebo and some other interventions', 'better than placebo, but no better than any other interventions', and 'no better than placebo' for high/moderate certainty evidence of benefit outcomes. For high/moderate certainty evidence of harm outcomes, the corresponding language was 'no more harmful than placebo', 'more harmful than placebo, but no worse than other interventions', and 'more harmful than placebo and some other interventions'. Participants felt that, with respect to category of magnitude of effect low/very low certainty evidence descriptions should be the same as those of the high/moderate certainty evidence categories, with the included qualifier of 'may be' at the beginning of the description of low to very low certainty evidence.

Gradient colour coding

The gradient colour-coding scheme uses three shades of green for the high/moderate certainty benefit outcomes (**Figure 5**: cells 1–3), and three shades of red for the high/moderate

certainty adverse events (**Figure 5**: cells 7–9). The use of three-shade grey gradient for low/very low certainty evidence is consistent for both beneficial outcomes and adverse events (**Figure 5**: cells 4–6, 10–12). Participants preferred dark grey be used for the 'among the worst' category (least beneficial or most harmful) and light grey be used for the 'among the best' category (most beneficial or least harmful), when presenting low/very low certainty of evidence results.

Participants had clear views regarding the colour shades used in **Figure 5**: cell 3 (among the least beneficial; high/moderate certainty), and **Figure 5**: cell 7 (among the least harmful; high/moderate certainty): because green is intuitively associated with positive results, they suggested caution regarding the use of a green shade for treatments categorised as 'among the worst' in benefit outcomes supported by high/moderate certainty evidence (**Figure 5**: cell 3). Participants strongly suggested that the shade of green used in this cell should, as a result, be a pale and faint green. Similarly, **Figure 5**: cell 7 uses a shade of red, despite being within the 'among the best' category in adverse events supported by high/moderate certainty evidence. Intuitively, participants noted that red is associated with poorer results. In order to avoid this inappropriate association, they suggested **Figure 5**: cell 7 should use a pale and faint shade of red. Other options tested used white for **Figure 5**: cell 3, and **Figure 5**: cell 7; however, participants ultimately believed that faint colouring within the respective colour gradients was most appropriate and did not hinder interpretation.

<u>Stoplight colour coding</u>

Because it dealt with the aforementioned concerns of the gradient colour-coding, participants also expressed enthusiasm for the stoplight colour coding. The use of the same colour scheme across **Figure 6**: cells 1–3 and **Figure 6**: cells 7–9 simplifies the interpretation

based on colour. Although the stoplight colour-coding addressed concerns with the gradient option, some participants preferred the gradient colour coding due to the clear distinction between benefit and harms outcomes. Others also felt that the stoplight colour coding looked distracting due to the inclusion of three bold colours, while the gradient colour coding reserves bold colours that 'stand out' for the comparisons with large benefits or large harms.

Figure 6: Stoplight Colour-Coding Legend

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	1	4	7	10
INTERMEDIATE	2	5	8	11
AMONG THE WORST	3	6	9	12

Discussion

The GRADE working group has developed methodologically coherent and innovative approaches to rating treatments within NMAs, including both benefits and harms, as 'among the best', 'intermediate' and 'among the worst'.[13 14] This may represent an important advance in the interpretation of the results of NMAs for clinicians using findings to guide clinical care. Clinicians, however, need to apply this rating for all outcomes of importance to patients. Rigorously developed, user-friendly, intuitive and tested approaches to simultaneous presentation of rated

treatments across multiple outcomes has thus far been unavailable for either the new GRADE rating approach or prior approaches to enhance interpretability.[4–6 9 12]

This study has addressed existing limitations by developing presentation methods that summarise NMA results for multiple outcomes in clear and interpretable formats. Although previous methods may still be useful in presenting the results of individual outcomes in greater detail with certainty of evidence incorporated,[4–6 9] the current presentation method allows for a clear and succinct summary of all outcomes considered within an NMA in a single presentation that our design validation has found both appealing and understandable to clinicians, many with limited prior exposure to NMAs.[6]

Strengths and limitations

Extensive design validation in a targeted audience has validated our NMA presentation approaches, allowing future NMA's to enhance the ease with which clinicians can interpret their results. Additional strengths of this study include consultation with individuals involved in the process of developing and disseminating systematic reviews and clinical practice guidelines, and extensive design validation that included the careful selection of a study population that reflects the broader clinical audience who will be making use of NMA results. The use of structured qualitative research methods including duplicate data analysis allowed the accurate and appropriate incorporation of user feedback to be incorporated into iterative presentation development.

Our study does have limitations. First, although the simplicity of the developed presentations represents a strength, achieving that simplicity required the omission of data that some audiences may consider important.[6] For instance, the previous development of an NMA

summary of findings table for individual outcomes provides greater detail for each treatment comparison that cannot feasibly fit within a multiple outcome presentation.[6] A particularly important omission may be the absolute effects of interventions that sometimes become crucial in trading off benefits and harms.[8] For this reason, authors may find it most appropriate to include both the multiple outcome presentation from this investigation, as well as additional outcome summaries suggested by other investigators.[4,6–11] This usability of this presentation tool was assessed specifically within the example NMA for pain management, which does not provide insights into the potential differences in usability for different future NMAs. Finally, we did not implement member checking. We did, however, employ data source triangulation to ensure that the findings of our study were robust.

Relation to prior work

Recent publications have addressed the issue of presenting NMA results for multiple outcomes, but have limitations that our proposal has addressed.[7,8] First, and crucially important, other options do not address the certainty of the evidence.[7,8] The Kilim plot provides a measure of the 'strength of statistical evidence', which equates to the magnitude of the p value.[8] Considerations of risk of bias, inconsistency, indirectness, publication bias, intransitivity and incoherence may, however, reduce certainty in treatment effects with low p values (which may or may not represent large effects). Additionally, the lack of design validation precludes confidence in how target users will understand these formats. For these reasons, the presentation versions proposed in the current study represent important improvements on previous tools for reporting NMA results for multiple outcomes.

Choosing a presentation variation

Authors can, based on the appropriateness of the colour-coding and the corresponding categorisation, choose between the two presentation versions in this manuscript. For example, the stoplight colour-coding variation may be most suitable when some treatments are better than the reference for some outcomes, while other treatments are worse for some outcomes. The three categories and explanations for benefit outcomes would then be 'among the best—better than reference (colour: green)', 'intermediate—same as reference (colour: yellow)', 'among the worst—worse than reference (colour: red)'. Intuitively, these descriptions and colours align. **Appendix A** provides an example of this scenario, with suggested details on the appropriate language to use within the legend.

The colour-gradient variation of the presentation may be most appropriate when the reference treatment is the worst (or best) treatment option across all outcomes. This would typically occur when placebo is the reference treatment, as placebo would likely be the worst treatment for benefit outcomes and the best treatment option for adverse event outcomes. The acute pain NMA used for our presentation formats fits this scenario. Although typically occurring with a placebo reference treatment, there may also be NMAs with other reference treatments that would intuitively follow this gradient colour coding. **Appendix B** provides an example with suggested details on the appropriate language to use within the legend.

Additional considerations

There is no single set of legend terminologies that universally apply to all NMAs, so authors must use their discretion to determine the most applicable and intuitive terminology. Authors may use the general guidance provided in this study in conjunction with categorisation recommendations of the minimally or partially contextualised approach.[13,14] The minimally and

partially contextualised approaches to NMA treatment categorisation have the potential for more than three categories, which would require an adaptation to the colour schemes we identified. The appropriate title for this presentation format represents another consideration that this study did not test. We would encourage authors to be explicit in defining the patient population assessed within the presentation.

Methodologists and statisticians have long bemoaned an excessive focus on statistical significance, in particular through the use of p values.[21–24] Notwithstanding, our participants felt it was important to highlight results indicating statistical significance, and our view is that there is considerable merit in the suggestion. Bolding or italics would be two possible ways of such highlighting, and the choice may depend on a journal's particular font suggestions.

A final consideration is the use of colours in the presentation methods. Participants believed that green, yellow, and red were the most intuitive colours for the table colour coding; however, these colours may be problematic for colour-blind individuals. Authors who want to ensure colour-blind accessibility may consider using blue instead of green, and orange instead of red; although this was not specifically tested within this investigation.

Conclusions

This study used end-user design validation to develop easily interpretable presentation formats for reporting NMA results with multiple outcomes, with a focus both on relative magnitude of effects and certainty of evidence. If further empirical study verifies our finding that clinicians, and potentially patients—who are increasingly involved in clinical shared-decision making—who are naïve to NMAs find the presentation understandable and appealing, its wide implementation may enhance the impact and usefulness of NMAs.

Data availability statement

Data sharing not applicable as no datasets generated and/or analysed for this study. No data are available.

Patient consent for publication

Not applicable.

Ethics approval

This study involves human participants but an Ethics Committee exempted this study. After reviewing the protocol, the Hamilton Integrated Research Ethics Board (HiREB) committee and chair, judging the study to be a quality improvement investigation within the methodology and knowledge translation field, provided an exemption from formal ethics approval.

References

1. Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for clinicians. Intern Emerg Med 2017;12:103–11. doi:10.1007/s11739-016-1583-7
2. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. BMJ 2013;346:f2914. doi:10.1136/bmj.f2914
3. Ellis SG. Do we know the best treatment for in-stent restenosis via network meta-analysis (NMA)?: simple methods any interventionalist can use to assess NMA quality and a call for better NMa presentation. JACC Cardiovasc Interv 2015;8:395–7. doi:10.1016/j.jcin.2014.11.012

4. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64:163–71. doi:10.1016/j.jclinepi.2010.03.016
5. Law M, Alam N, Veroniki AA, *et al*. Two new approaches for the visualisation of models for network meta-analysis. BMC Med Res Methodol 2019;19:61. doi:10.1186/s12874-019-0689-9
6. Yepes-Nuñez JJ, Li S-A, Guyatt G, *et al*. Development of the summary of findings table for network meta-analysis. J Clin Epidemiol 2019;115:113. doi:10.1016/j.jclinepi.2019.04.018
7. Daly CH, Mbuagbaw L, Thabane L, *et al*. Spie charts for quantifying treatment effectiveness and safety in multiple outcome network meta-analysis: a proof-of-concept study. BMC Med Res Methodol 2020;20:266. doi:10.1186/s12874-020-01128-2
8. Seo M, Furukawa TA, Veroniki AA, *et al*. The Kilim plot: a tool for visualizing network meta-analysis results for multiple outcomes. Res Synth Methods 2021;12:86-95. doi:10.1002/jrsm.1428
9. Chaimani A, Higgins JPT, Mavridis D, *et al*. Graphical tools for network meta-analysis in STATA. PLoS One 2013;8:e76654. doi:10.1371/journal.pone.0076654
10. Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. BMC Med Res Methodol 2013;13:35. doi:10.1186/1471-2288-13-35
11. Tan SH, Cooper NJ, Bujkiewicz S, *et al*. Novel presentational approaches were developed for reporting network meta-analysis. J Clin Epidemiol 2014;67:672–80. doi:10.1016/j.jclinepi.2013.11.006

12. Mbuagbaw L, Rochwerg B, Jaeschke R, *et al*. Approaches to interpreting and choosing the best treatments in network meta-analyses. Syst Rev 2017;6:79. doi:10.1186/s13643-017-0473-z
13. Brignardello-Petersen R, Florez ID, Izcovich A, *et al*. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. BMJ 2020;371:m3900. doi:10.1136/bmj.m3900
14. Brignardello-Petersen R, Izcovich A, Rochwerg B, *et al*. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. BMJ 2020;371:m3907. doi:10.1136/bmj.m3907
15. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015;25:1212–22. doi:10.1177/1049732315588501
16. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19:349–57. doi:10.1093/intqhc/mzm042
17. Neergaard MA, Olesen F, Andersen RS, *et al*. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009;9:52. doi:10.1186/1471-2288-9-52
18. Saunders B, Sim J, Kingstone T, *et al*. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52:1893–907. doi:10.1007/s11135-017-0574-8
19. Busse JW, Sadeghirad B, Oparin Y, *et al*. Management of Acute Pain From Non-Low Back, Musculoskeletal Injuries : A Systematic Review and Network Meta-analysis of Randomized Trials. Ann Intern Med 2020;173:730–8. doi:10.7326/M19-3601

20. Maher C, Hadfield M, Hutchings M. Ensuring rigor in qualitative data analysis: a design research approach to coding combining NVivo with traditional material methods. Int J Qual Methods 2018.
21. Greenland S, Senn SJ, Rothman KJ, *et al*. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31:337–50. doi:10.1007/s10654-016-0149-3
22. Gagnier JJ, Morgenstern H, Misconceptions MH. Misconceptions, misuses, and misinterpretations of P values and significance testing. J Bone Joint Surg Am 2017;99:1598–603. doi:10.2106/JBJS.16.01314
23. Goodman SN. Toward evidence-based medical statistics. 1: the P value fallacy. Ann Intern Med 1999;130:995–1004. doi:10.7326/0003-4819-130-12-199906150-00008
24. Phillips M. Letter to the editor: editorial: threshold P values in orthopaedic Research-We know the problem. What is the solution? Clin Orthop Relat Res 2019;477:1756–8. doi:10.1097/CORR.0000000000000827

Appendix A: Example Legend When Active Treatment is Reference

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	**Better than reference**	**May be better than reference**	**Less harmful than reference**	**May be less harmful than reference**
INTERMEDIATE	**No better than reference**	**May be no better than reference**	**No more harmful than reference**	**May be no more harmful than reference**
AMONG THE WORST	**Worse than reference**	**May be worse than reference**	**More harmful than reference**	**May be more harmful than reference**

Appendix B: Example Legend When Placebo (Or Any Sham/Null Treatment Effect) is Reference

	BENEFIT OUTCOMES		ADVERSE EVENTS	
	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence	High/Moderate Certainty Evidence	Low/Very Low Certainty Evidence
AMONG THE BEST	**Better than placebo and some other interventions**	**May be better than placebo and some alternatives**	**No more harmful than placebo**	**May be no more harmful than placebo**
INTERMEDIATE	**Better than placebo, but no better than any other interventions**	**May be better than placebo, but no better than other interventions**	**More harmful than placebo, but no worse than other interventions**	**May be more harmful than placebo, but no worse than other interventions**
AMONG THE WORST	**No better than placebo**	**May be no better than placebo**	**More harmful than placebo and some other interventions**	**May be more harmful than placebo and some alternatives**

Part B: User Testing of a Novel Network Meta-Analysis Results Presentation Table: A Qualitative Description Protocol

Mark R Phillips[1], Behnam Sadeghirad[1,2], Jason W Busse[1,2], Romina Brignardello-Petersen[1], Carlos A Cuello-Garcia[1], Fernando Kenji Nampo[3], Raveendhara R Bannuru[4], Lehana Thabane[1,5], Mohit Bhandari[1,6], Gordon H Guyatt[1]

1. Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. Anesthesia, McMaster University, Hamilton, Ontario, Canada
3. Department of Latin-American Institute of Life and Nature science, Federal University of Latin-American Integration, Foz do Iguacu, Brazil
4. Center for Treatment Comparison and Integrative Analysis, Tufts Medical Center, Boston, Massachusetts, USA
5. Biostatistics Unit, St. Joseph's Healthcare, Hamilton, Ontario, Canada
6. Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Introduction

Network meta-analysis (NMA) represent a novel approach to evidence synthesis that allows comparison among numerous treatment options within a single investigation.[1] Publications have aimed at educating clinicians on the methodology used within an NMA as well as addressing the challenging issue of how to interpret the results.[1-4] None of the interpretation strategies suggested thus far have proved satisfactory, raising the opportunity for novel reporting methods to improve the interpretability of NMAs. Such novel methods are now available, but the optimal presentation approaches for these methods have yet to be developed and tested.

Purpose Statement

This study aims to conduct user testing to provide insights into the optimization of a novel table format designed to clearly present NMA results.

Research Question

The research question for this proposed study is "How can the novel table format be optimized to promote interpretability of NMA findings addressing multiple outcomes within the NMA user population?" This research question follows the EPPiC framework for qualitative studies. The emphasis will be on table formatting, while the phenomenon of interest is the interpretability of results. The purposeful sample will include academic and non-academic clinicians, investigators, research staff, and residents/trainees, with the largest proportion of participants being clinicians. The context for the study will be the reporting of NMA results.

Potential Contribution of Findings

The findings of this study will contribute to the development of an easily interpretable method of presenting NMA results for multiple outcomes. Having an optimized method in which all target audience groups can easily interpret the results will enhance the future use of NMA to inform clinical practice and support shared-care decision-making. Although there are multiple ways in which NMA results can be presented[4], the table developed through this rigorous investigation will provide clarity to the complex results that can arise from an NMA.

Research Team and Steering Committee

The research team that will be conducting this study is led by Mark Phillips, a PhD student in the McMaster Health Research Methodology program. The project will be supervised by Dr. Gordon Guyatt, a researcher within the departments of Health Research Methodology, Evidence, and Impact and Medicine at McMaster. A group of collaborators from the team's professional networks will be utilized to identify and help recruit a purposeful sample of participants to provide rich insights into the improvement and interpretability of the NMA results table.

The study's steering committee will consist of Mark Phillips, Gordon Guyatt, Jason Busse, Romina Brignardello-Petersen, Behnam Sadeghirad, Carlos Cuello, and Fernando Kenji Nampo. The feedback themes derived from data collection will be presented to the steering committee to inform the adaptation of the NMA table. This will be done iteratively throughout data analysis, with a steering committee meeting occurring no longer than after 10 interview sessions have been analyzed. The exact timing of steering committee meetings will be determined by Mark Phillips and Gordon Guyatt based on the data collected.

Study Design

This study will be conducted using a qualitative description approach that focuses on providing a close description of the information that participants provide, and relies on minimal interpretation on the part of the research team. This is an ideal approach for the current project, as the user testing aims to optimize the understandability of the table within the target population, which requires direct input and description of the input without interpretive direction of the research team. **Appendix A** of this document includes the initial formats to be included within the user testing sessions. We decided on some core criteria for the NMA to use: variability in quality of evidence, variability in magnitudes of effect, assesses both benefits and harms, includes both continuous and binary outcomes, and contains at least 5 outcomes and at least 5 interventions

Sampling and Recruitment

Academic and non-academic clinicians, investigators, research staff, and residents/trainees, who are willing to provide information-rich accounts will be eligible. These individuals will be identified from the professional contacts and clinical practices of the research team. The core group of our participants will be clinicians with a varying degree of research experience and knowledge.

Potential participants will be contacted for recruitment via email, or by personal contact within the participating clinical sites. The sample size for this study will be determined through data saturation of both the academic and non-academic clinician group–we anticipate that

between 30 to 40 participants will be required to achieve saturation. To reach a greater number of participants, we will supplement our recruitment with snowball sampling techniques: at the end of each interview we will ask participants to provide contact information for any individuals that they believe would provide rich information on the topic of interest. Snowball sampling will continue until the data collected adequately covers the themes and core concepts that have arisen throughout data collection. As data collection reaches redundancy–the point at which there are no further refinements required to improve the interpretability of the NMA results table–recruitment will stop.

Data Collection

Individual Interviews

Data collection will be conducted using multiple interview sessions that will be conducted either in-person, or through a video teleconferencing platform. Participants will complete a demographics questionnaire prior to the interview **(Appendix B).** All interviews will be transcribed verbatim, with additional field notes included from the researcher in order to provide additional contextual information such as nonverbal cues and overall moods and attitudes. Interviews will follow a flexible interview guide (**Appendix C**); however, the researcher will leave the discussion open for participants to explore any topics that they feel are relevant and important. The interview guide will be updated iteratively to explore areas of importance that emerge throughout the study. This approach prevents the researcher from leading the participants to discuss topics of particular interest to the researcher, which could result in a biased assessment of the phenomenon of interest.

Interviews will begin with a brief introduction to NMA, followed by questioning on the participants familiarity and experience with NMA. Participants will then be shown the current version(s) of the NMA tables, and asked to provide feedback. Initially, two table versions will be shared (**Appendix A**), and throughout the course of the study the feedback on these versions will be consolidated into one table version. A complete interview guide is provided in **Appendix C**.

Group Presentations

Workshop/conference presentations of the tables will provide data source triangulation, and input from these sessions will be incorporated into the development of the table format. Initial table formats were informed by feedback provided by a number of presentations. A group of statisticians and methodologists at McMaster University provided feedback that was used to create the current formats in this protocol (**Appendix A**). Results of previous NMAs were also presented at conferences, where the feedback of attendees was used to refine the table formats. The iterations of table formats that are developed throughout this study will continue to be presented at workshops and conferences in order to provide a thorough source of data triangulation. The feedback from these presentations will be recorded and used in conjunction with the data collected from interviews.

Data Analysis Plan

Data will be analyzed using the qualitative content analysis approach.[6] As new data is collected, it will be coded and grouped with similar phrases, patterns, and themes that have arose throughout the data collection process. Data collection, recruitment, and data analysis will be

done iteratively alongside one another. There will be no specific timeframe for when the table format will be iteratively altered according to participant feedback. Instead, the research team will periodically meet to go over the collected data that may be actionable in terms of table format improvements. The actionable feedback will be ranked as a "critical change required", "moderate change required", or "minor change required" by two independent research team members. These rankings will be presented to the steering committee along with the actionable items in order to prioritize the required changes to the draft table versions. Once actionable feedback has been reported by multiple participants, the proposed change will be presented to the research team. After the research team has reviewed the actionable items the table format will be adjusted accordingly, and subsequent participants will provide input on the newly improved iteration of the NMA results table. The final manuscript will provide a thorough explanation of the themes that arose within the study, and describe the changes that were implemented in order to optimize the interpretability of the NMA results table format. If subgroup differences arise between academic and non-academic participants, a detailed explanation of these differences will be provided, along with suggested table variations for the different groups. This will allow for different versions of the table to be used depending on the target audience. Data analysis will be done using RQDA software (R version 3.5.0).

Strategies to Promote Rigor or Trustworthiness

For transparency, we will keep an audit trail with all decisions throughout the project that will provide justification for all study-related decisions.[9]

Ethics

Using a consent form that adheres to the structure and guidelines of the Hamilton Integrated Research Ethics Board (HIREB) template, all participants will provide informed consent **(Appendix D)**. At the beginning of each interview, to ensure that participants understand the study and their rights as participants, the interviewer will review the consent form with the participant. There are three main ethical concerns that the research team will address: maintaining confidentiality and anonymity, data security, and minimizing any potential undue harms to participants.

Confidentiality and anonymity will be discussed explicitly within the consent form. It will be made clear to all participants that all identifiers and personal information will be removed from the data during analysis and not included within the final analysis. Any requests by participants for information to not be used within the study will be respected.

The research team will ensure the appropriate storage of all study data to ensure the security of participant personal information. All data files will be password protected and held on a password protected computer. This password will only be known by the pertinent research staff working on this study. In the event of a potential data breach, the ethics board and all participants will be notified and debriefed about the situation.

The phenomenon of interest within this study may not be seen as a sensitive topic with risks of participant psychological harm; however, this is always a consideration when collecting personal data. The main risk in causing undue harm would be the loss of anonymity of data provided by participants, which will be addressed by the numerous safeguards implemented to limit the risk of data loss or inappropriate sharing of personal information. Participants will be

told during the informed consent process that they are free to leave the study at any point, and none of their data will be used without their direct consent. All study data will be stored for ten years, in accordance with HIREB standards.

Conclusion

The purpose of this study is to aid gain insights into the optimal way to report NMA results for multiple outcomes. The use of the NMA results table will be refined throughout the course of this study in order to develop a table – or tables - that enhance the interpretability of NMA results. This will aid both future NMA authors in articulating their findings in an easily understandable manner, as well as knowledge users who may utilize NMA findings to inform their clinical practice.

References

1. Foote CJ, Chaudhry H, Bhandari M, et al. Network Meta-analysis: Users' Guide for Surgeons: Part I - Credibility. *Clin Orthop Relat Res.* 2015;473:2166-2171.

2. Chaudhry H, Foote CJ, Guyatt G, et al. Network Meta-analysis: Users' Guide for Surgeons: Part II - Certainty. *Clin Orthop Relat Res.* 2015;473:2172-2178.

3. Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *Bmj.* 2013;346:f2914.

4. Salanti G, Ades AE, 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.

5. Morse J. Critical Analysis of Strategies for Determining Rigor in Qualitative Inquiry. *Qualitative Health Research.* 2015;25:1212-1222.

6. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description – the poor cousin of health research? *BMC Medical Research Methodology.* 2009;9:1-5.

7. Koch T, Harrington, A. Reconceptualizing rigour: the case for reflexivity. *Journal of Advanced Nursing.* 1998;28:882-890.

8. Gentles S, Jack, SM., Nicholas, DB., McKibbon, KA. Critical Approach to Reflexivity in Grounded Theory. *The Qualitative Report.* 2014;19:1-14.

9. Rodgers BL, Cowles KV. The qualitative research audit trail: a complex collection of documentation. *Res Nurs Health.* 1993;16:219-226.

Appendix A: Initial Table Formats

Table Presentation #1

	Benefit: Statistically significant difference with placebo and at least one other tx **Harm: no statistical difference with placebo**	**Statistically significant difference with placebo**	**Harm: Statistically significant difference with placebo and at least one other tx** **Benefit: no statistical difference with placebo**
High or moderate certainty evidence	**Among the most effective**	**Inferior to the most effective, but superior to placebo**	**No more effective than placebo**
	No more harmful than placebo	**Less harmful than some alternatives, but more harmful than placebo**	**Among the most harmful**
Low or very low certainty evidence	**May be among the most effective**	**May be inferior to the most effective, but superior to placebo**	**May be no more effective than placebo**
	May be no more harmful than placebo	**May be less harmful than some alternatives, but more harmful than placebo**	**May be among the most harmful**

Intervention	Effectiveness outcomes					Harms outcomes		
	Pain ≤ 2 h post-tx	Pain 2 to 7 d post-tx	Function	Treatment satisfaction	Symptom relief	GI-related AE's	Neurologic-related AE's	Derm-related AE's
	MD (95% CI)	MD (95% CI)	MD (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Topical NSAID	**1.02 (0.38,1.66)**	**1.08 (0.76,1.40)**	**1.66 (1.13,2.19)**	**5.20 (2.03,13.33)**	**32.1 (3.81,269.8)**	1.12 (0.59,2.12)	1.20 (0.49,2.95)	0.76 (0.51,1.14)
Oral NSAID	**0.92 (0.34,1.51)**	**1.03 (0.56,1.49)**	**0.74 (0.14,1.33)**	3.24 (0.43,24.70)	2.66 (0.88,8.05)	**1.77 (1.32,2.36)**	0.96 (0.58,1.61)	1.88 (0.45,7.79)
Acetaminophen & Diclofenac	**1.10 (0.17,2.03)**	**1.13 (0.03,2.23)**	-	3.45 (0.18,66.96)	3.19 (0.54,18.83)	-	-	-
Acetaminophen	**1.01 (0.18,1.84)**	**1.11 (0.29,1.92)**	0.90 (-0.52,2.32)	2.43 (0.18,32.70)	2.34 (0.51,10.77)	0.24 (0.01,4.88)	-	-
TENS	**2.30 (1.10,3.50)**	**1.98 (0.70,3.26)**	-	-	-	1.88 (0.06,57.84)	1.38 (0.04,44.79)	1.93 (0.05,69.49)
Specific acupressure	**1.60 (0.64,2.56)**	**2.09 (0.34,3.84)**	**1.51 (0.35,2.68)**	0.50 (0.04,6.49)	2.54 (0.39,16.67)	0.80 (0.02,41.67)	0.80 (0.01,43.74)	0.80 (0.01,45.94)
Manipulation	**1.74 (0.79,2.70)**	-0.40 (-2.49,1.69)	0.05 (-1.03,1.14)	-	**162.8 (4.98,5321.5)**	0.50 (0.01,31.29)	1.41 (0.02,82.79)	-
Acetaminophen & Chlorzoxazone	-	**2.96 (0.49,5.43)**	-	-	-	0.35 (0.01,10.59)	-	-
Fentanyl	**3.54 (2.00,5.08)**	-	-	-	-	**61.7 (6.44,591.5)**	**6.60 (1.19,36.71)**	-
Ibuprofen & Cyclobenzaprine	1.04 (-0.57,2.66)	**1.55 (0.01,3.09)**	-	5.52 (0.21,147.0)	-	1.10 (0.13,9.42)	**4.69 (1.27,17.37)**	-
Menthol Gel	-	**1.14 (0.01,2.27)**	0.70 (-0.67,2.08)	-	-	-	-	1.00 (0.11,9.03)
Laser therapy	-	0.69 (-0.70,2.07)	-	-	**7.31 (3.22,16.61)**	0.49 (0.01,24.85)	0.49 (0.01,26.09)	0.49 (0.01,27.41)
Mobilization	-	-3.40 (-6.83,0.03)	0.10 (-0.66,0.87)	2.07 (0.07,58.49)	**7.76 (1.04,58.09)**	0.93 (0.02,47.10)	0.93 (0.02,49.38)	0.93 (0.02,51.99)
Acetaminophen & Ibuprofen	0.74 (-0.26,1.74)	1.22 (-0.33,2.76)	-	-	3.10 (0.52,18.29)	-	-	-
Non- Specific Acupressure	0.05 (-0.92,1.02)	0.18 (-1.53,1.89)	-0.18 (-1.39,1.02)	0.44 (0.03,5.76)	1.80 (0.27,12.10)	0.85 (0.02,44.76)	0.85 (0.02,46.98)	0.85 (0.01,49.34)
Exercise	-	0.81 (-1.00,2.62)	0.24 (-0.39,0.86)	3.50 (0.21,59.42)	0.85 (0.22,3.23)	0.95 (0.02,48.96)	0.98 (0.06,16.52)	0.98 (0.06,17.10)
Cyclobenzaprine	-	2.07 (-0.01,4.14)	-	-	-	0.64 (0.03,15.74)	1.86 (0.18,19.52)	-
Supervised Rehab	-	-1.14 (-2.47,0.20)	0.19 (-0.68,1.06)	2.25 (0.15,34.07)	5.17 (0.71,37.67)	1.06 (0.02,54.49)	1.06 (0.02,57.21)	1.06 (0.02,60.10)
Tramadol	-0.93 (-2.74,0.88)	-	-	-	6.22 (0.34,112.78)	-	**7.74 (1.24,48.14)**	-
Acetaminophen & Opioid	0.54 (-0.46,1.54)	0.82 (-0.34,1.98)	-	2.50 (0.14,44.86)	1.43 (0.38,5.36)	**5.85 (2.93,11.71)**	**4.06 (1.81,9.09)**	-

Table Presentation #2

	Statistically significant difference with placebo and at least one other tx	Statistically significant difference with placebo	Statistically no difference with placebo
High or moderate certainty evidence	Among the most effective	Inferior to the most effective, but superior to placebo	No more effective than placebo
	Among the most harmful	Less harmful than some alternatives, but more harmful than placebo	No more harmful than placebo
Low or very low certainty evidence	May be among the most effective (or harmful)	May be inferior to the most effective, but superior to placebo (or less harmful than some alternatives, but more harmful than placebo)	May be no more effective (or harmful) than placebo

Intervention	Effectiveness outcomes					Harms outcomes		
	Pain ≤ 2 h post-tx	Pain 2 to 7 d post-tx	Function	Treatment satisfaction	Symptom relief	GI-related AE's	Neurologic-related AE's	Derm-related AE's
	MD (95% CI)	MD (95% CI)	MD (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Topical NSAID	**1.02 (0.38,1.66)**	**1.08 (0.76,1.40)**	**1.66 (1.13,2.19)**	**5.20 (2.03,13.33)**	**32.1 (3.81,269.8)**	1.12 (0.59,2.12)	1.20 (0.49,2.95)	0.76 (0.51,1.14)
Oral NSAID	**0.92 (0.34,1.51)**	**1.03 (0.56,1.49)**	**0.74 (0.14,1.33)**	3.24 (0.43,24.70)	2.66 (0.88,8.05)	**1.77 (1.32,2.36)**	0.96 (0.58,1.61)	1.88 (0.45,7.79)
Acetaminophen & Diclofenac	**1.10 (0.17,2.03)**	**1.13 (0.03,2.23)**	-	3.45 (0.18,66.96)	3.19 (0.54,18.83)	-	-	-
Acetaminophen	**1.01 (0.18,1.84)**	**1.11 (0.29,1.92)**	0.90 (-0.52,2.32)	2.43 (0.18,32.70)	2.34 (0.51,10.77)	0.24 (0.01,4.88)	-	-
TENS	**2.30 (1.10,3.50)**	**1.98 (0.70,3.26)**	-	-	-	1.88 (0.06,57.84)	1.38 (0.04,44.8)	1.93 (0.05,69.5)
Specific acupressure	**1.60 (0.64,2.56)**	**2.09 (0.34,3.84)**	**1.51 (0.35,2.68)**	0.50 (0.04,6.49)	2.54 (0.39,16.67)	0.80 (0.02,41.67)	0.80 (0.01,43.7)	0.80 (0.01,45.9)
Manipulation	**1.74 (0.79,2.70)**	-0.40 (-2.49,1.69)	0.05 (-1.03,1.14)	-	**162.8 (4.98,5321.5)**	0.50 (0.01,31.29)	1.41 (0.02,82.8)	-
Acetaminophen & Chlorzoxazone	-	**2.96 (0.49,5.43)**	-	-	-	0.35 (0.01,10.59)	-	-
Fentanyl	**3.54 (2.00,5.08)**	-	-	-	-	**61.7 (6.44,591.5)**	**6.60 (1.19,36.7)**	-
Ibuprofen & Cyclobenzaprine	1.04 (-0.57,2.66)	**1.55 (0.01,3.09)**	-	5.52 (0.21,147.0)	-	1.10 (0.13,9.42)	**4.69 (1.27,17.4)**	-
Menthol Gel	-	**1.14 (0.01,2.27)**	0.70 (-0.67,2.08)	-	-	-	-	1.00 (0.11,9.03)
Laser therapy	-	0.69 (-0.70,2.07)	-	-	**7.31 (3.22,16.61)**	0.49 (0.01,24.85)	0.49 (0.01,26.1)	0.49 (0.01,27.4)
Mobilization	-	-3.40 (-6.83,0.03)	0.10 (-0.66,0.87)	2.07 (0.07,58.49)	**7.76 (1.04,58.09)**	0.93 (0.02,47.10)	0.93 (0.02,49.4)	0.93 (0.02,52.0)
Acetaminophen & Ibuprofen	0.74 (-0.26,1.74)	1.22 (-0.33,2.76)	-	-	3.10 (0.52,18.29)	-	-	-
Non- Specific Acupressure	0.05 (-0.92,1.02)	0.18 (-1.53,1.89)	-0.18 (-1.39,1.02)	0.44 (0.03,5.76)	1.80 (0.27,12.10)	0.85 (0.02,44.76)	0.85 (0.02,46.98)	0.85 (0.01,49.34)
Exercise	-	0.81 (-1.00,2.62)	0.24 (-0.39,0.86)	3.50 (0.21,59.42)	0.85 (0.22,3.23)	0.95 (0.02,48.96)	0.98 (0.06,16.52)	0.98 (0.06,17.10)
Cyclobenzaprine	-	2.07 (-0.01,4.14)	-	-	-	0.64 (0.03,15.74)	1.86 (0.18,19.52)	-
Supervised Rehab	-	-1.14 (-2.47,0.20)	0.19 (-0.68,1.06)	2.25 (0.15,34.07)	5.17 (0.71,37.67)	1.06 (0.02,54.49)	1.06 (0.02,57.2)	1.06 (0.02,60.1)
Tramadol	-0.93 (-2.74,0.88)	-	-	-	6.22 (0.34,112.78)	-	**7.74 (1.24,48.1)**	-
Acetaminophen & Opioid	0.54 (-0.46,1.54)	0.82 (-0.34,1.98)	-	2.50 (0.14,44.86)	1.43 (0.38,5.36)	**5.85 (2.93,11.71)**	**4.06 (1.81,9.09)**	-

Appendix B: Demographic Questionnaire Form

We appreciate your participation in this research. Please answer these demographic questions so that we can describe, as a group, the participants in this study.

1. What is your age? _____ years

2. What is your gender? ☐ Female ☐ Male

3. What is your occupation?

☐ Clinician in an academic centre
☐ Clinician in a non-academic centre
☐ Fellow/trainee
☐ Researcher
☐ Other _____

3. What are the highest degrees that you hold?

☐ MD
☐ PhD
☐ MSc
☐ BSc
☐ Other _____

4. How many years have you been in your professional career? _____ years

5. Have you ever been involved in, aided with, or been the author of, a network meta-analysis?

☐ Yes ☐ No

5a. If Yes to Question #5: Please check all of the tasks that you were involved with in your previous network-meta analysis project(s)?

☐ Project Conceptualization
☐ Article Screening/Data extraction
☐ Data Analysis
☐ Manuscript Writing

☐ Other _____

6. Have you ever read or used information from a network meta-analysis to inform your practice/career?

☐ Yes ☐ No

Appendix C: Interview Guide

Part 1: Introductions

1. Introduce all research members in the interview
2. Review the informed consent document

Part 2: Background and Summary of NMA

Summary: We would like to start by briefly reviewing the topic of interest today, which is network meta-analysis (NMA). NMA is a way to take all of the current published randomized trial data on a specific topic and pool it, in order to try and answer a specific research question. For example, an NMA may be used to assess the best treatment option for a certain medical condition. All of the data published on treatment options for the condition are collected, and then pooled together using statistics in order to compare the outcomes of the potential options. The results then provide the effectiveness, and harms, for each of the potential treatment options.

The NMA method is a new advancement in health research, because older methods typically could only compare 2 different treatment options. Since NMA can compare many different treatments, it can help us decide which treatment is optimal amongst a large number of potential treatment options. This does, however, come with some limitations. While this method allows for a thorough analysis of multiple treatment options, the results can often be cumbersome due to the vast amounts of information created by the analysis. Due to this, we are developing a table that can be used to present the results of NMA in an interpretable way.

Results within this table will be reported as either mean differences between the reference treatment and the comparator, or odds ratios. Mean differences identify the difference in mean outcome scores between the two treatments being compared, while odds ratios provide an estimate of the probability of an outcome occurring if you were to receive the treatment being assessed. Both of these measures are provided with confidence intervals, which provide an indication of the variance around each of the outcome estimates.

To begin, we would like to understand your current knowledge of NMA:

3. How familiar are you with NMA?
4. Have you ever been part of an NMA project?
    a. If so, what was your role in the NMA project?
5. Have you ever read an NMA?

Part 3: Review of the table format

The table(s) I am showing you summarizes the results of an NMA that assessed pain management treatment options.

Please think aloud as you interpret this table(s)

Regarding the legend:

6. Please provide any feedback you may have regarding the legend table.

   a. Do you find the language within the legend to be understandable? If not, what is confusing?
   b. Do you have any feedback regarding the format of the legend?
      i. Do you have feedback regarding the coloring used?
      ii. Do you have feedback regarding the language used?
      iii. Do you have feedback regarding the indication of the certainty of evidence component of the legend?

Next, we will review the results table:

7. Now that you have reviewed the legend in more detail, does the legend accurately and completely summarize the results table?
   a. If not, what could be changed?
8. Please provide any feedback you have regarding the results within the table
   a. Are the results easily understandable? If not, what is confusing or could be changed?
9. Do you have any feedback regarding the format of the table?
      i. Do you have feedback regarding the coloring used?
      ii. Do you have feedback regarding the language used?
      iii. Do you have feedback regarding the outcome reporting within the table?
      iv. Do you have feedback regarding the indication of the certainty of evidence component of the results?
10. Please provide any other feedback that you may have regarding the table

<u>Part 4: Assessing Participant Interpretation</u>

As this is an exercise to understand the interpretability of the table, we would like to finish with a discussion about your interpretation of the results.

11. Based on the results within the table, please describe how you interpret the findings?
   a. Based on both the benefits and the harms, which treatment(s) do you consider to be the optimal choice(s)?
   b. Which treatment(s) do you believe are the least optimal choices? What information is important for you in deciding this?
12. How confident are you in your interpretation?
   a. Why are you/aren't you confident in your interpretation?
   b. What would aid in improving your interpretation?

<u>Part 5: Closing Remarks</u>

Thank you for taking the time to participate in this study. Before we end our discussion, we would like to ask if you have any colleagues that may be interested in participating in this study. Following this interview, it would be great if we could connect via email with anyone who you believe may be able to provide valuable insights to this project.

Appendix D: Participation Information and Consent Form

Title of Study: User Testing of a Novel Network Meta-Analysis Results Presentation Table: A Qualitative Description

Principal Investigators: **Dr. Gordon Guyatt**

Mark Phillips

You are being invited to participate in an interview for a research study being conducted by Mark Phillips, under the supervision of Dr. Gordon Guyatt. In order to decide whether or not you want to participate, you should understand what is involved and the potential risks and benefits. This form gives detailed information about the interview. If you agree to participate in the interview, you will be asked to sign a consent form before it begins. Please take your time to make your decision.

WHY IS THIS RESEARCH BEING DONE?

Network meta-analysis (NMA) is a new approach to evidence synthesis that is able to compare numerous treatment options within a single investigation. Unlike pairwise meta-analysis, this allows for all potential treatment options to be included as comparators within the meta-analysis. While this method allows for a thorough analysis of multiple treatment options, the results can often be cumbersome due to the vast amounts of information created by the analysis. This poses an opportunity for novel results reporting methods to improve the interpretability of the valuable

amounts of information obtained from conducting NMA.

WHAT IS THE PURPOSE OF THIS STUDY?

This study aims to conduct user testing to provide insights into the optimization of a novel table format, designed to clearly present NMA results.

WHAT WILL MY RESPONSIBILITIES BE IF I TAKE PART IN THE STUDY?

If you volunteer to participate in this study, you will be asked to attend a one-on-one discussion with an individual from the research team. The discussion will take place in a convenient location for you to attend, or via an online teleconference platform. The discussion will take approximately one hour. Audio recording will be conducted for the interview.

WHAT ARE THE POSSIBLE RISKS?

There are no serious risks to you if you take part in the interview session. All data will be masked and personal information will not be collected. All information will be kept as password protected files on a password protected computer. If you find any of the discussion during the interview upsetting or uncomfortable, you can end your participation in the discussion at any time.

WHAT ARE THE POSSIBLE BENEFITS FOR ME AND/OR FOR SOCIETY?

The findings from this study will help to improve the development of a novel table format to present NMA results. This will allow for an increased uptake of future NMA work, as results will be more clearly reported in an understandable yet thorough manner.

IF I DO NOT WANT TO TAKE PART IN THE STUDY, ARE THERE OTHER CHOICES?

It is important for you to know that you can choose not to take part in the interview. If you decide not to participate, this decision will not have any negative consequences for you.

WHAT INFORMATION WILL BE KEPT PRIVATE?

The interview discussion will be digitally recorded and transcribed. The following steps will be taken to protect your confidentiality:

- All personal information including name and email address will be kept in a secure place, separate from the consent forms, interview recordings and transcripts.
- The digital recording will be downloaded to a secure computer that is password protected and assigned a study ID number.
- The interview transcript will be assigned a study ID number.
- Participants will be assigned a 3 digit study ID number within the transcript
- Any information that could identify you as an individual will be deleted from the transcript.
- The transcript will be accessible only to the researchers and on a secure computer that is password protected.
- If something you said during the interview is quoted in the final manuscript, your identity will be kept confidential.

All information collected during the study will be stored until the completion of the study and the findings have been released.

CAN PARTICIPATION IN THE INTERVIEW END EARLY?

You may decide at any time that you do not want to finish the interview discussion without negative consequences. You also may refuse to answer any questions you don't want to answer and still participate in the interview discussion. Information collected in the discussion up to the point of your withdrawal from the interview will still be used in the study unless specifically requested to be removed.

WILL I BE PAID TO PARTICIPATE IN THIS STUDY?

No, participation in the study is entirely voluntary.

IF I HAVE ANY QUESTIONS OR PROBLEMS, WHOM CAN I CALL?

If you have any questions about the research now or later, please contact Mark Phillips at phillimr@mcmaster.ca.

CONSENT TO PARTICIPATE

I understand that I am being asked to take part in an interview discussion about the optimal presentation of NMA results. I have received a Participant Information Sheet and I have read it thoroughly. I have had the opportunity to ask questions, and all of my questions have been answered to my satisfaction. I understand that:

- I will participate in an interview lasting roughly one hour
- The interview discussion will be digitally recorded and then transcribed with all identifying information removed.
- I can ask to review the digital recording.
- The digital recording will be downloaded to a secure computer that is password protected and assigned a study ID.
- The digital file also will be password protected.
- All identifying information will be kept confidential.
- My answers will be anonymous and my name will not appear in any study reports.
- My participation is entirely voluntary.
- I can refuse to answer specific questions or withdraw from the study even after I agree to participate.
- If I do not want to answer a question or decide to withdraw, this will not affect my participation in future research.
- I will receive a signed copy of this consent form.

If I have any questions or comments about the study, I can contact Mark Phillips at phillimr@mcmaster.ca. I agree to participate in an interview for the study "**User Testing of a Novel Network Meta-Analysis Results Presentation Table: A Fundamental Qualitative Description".**

_____ _____

Name of Participant Signature of Participant

_____

Date

Consent form administered and explained in person by:

_____ _____

Name and title Signature

_____

Date

Principal Investigator's signature:

_____ _____

Name Signature

_____

Date

Chapter 5: Implementing NMA Reporting Methodology

Part A: Surgical Management of Displaced Femoral Neck Fractures: A Systematic Review and Network Meta-Analysis

Mark R Phillips, PhD(c)[1]; Jason W. Busse, DC PhD[1,2,3,4]; Lehana Thabane, PhD[1,5,6]; Raveendhara R Bannuru, MD PhD[7]; Christine Kucava, BScN MSc[8,9], Varun Srikanth, BSc(c)[10], Mohit Bhandari, MD PhD[1,11]

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. The Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Ontario, Canada
3. Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada
4. The Michael G. DeGroote Centre for Medicinal Cannabis Research, McMaster University, Hamilton, Ontario, Canada
5. Biostatistics Unit, St Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada
6. Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa
7. Center for Treatment Comparison and Integrative Analysis, Tufts Medical Center, Boston, MA, USA
8. Hamilton Health Sciences, Hamilton, Ontario, Canada
9. Department of Nursing, McMaster University, Hamilton, Ontario, Canada
10. Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
11. Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Abstract

Importance: There are a number of surgical options currently utilized to manage displaced femoral neck fractures; an injury that poses significant morbidity and mortality risks to patients.

Objective: To compare benefit and harms outcomes between total hip arthroplasty (THA), hemiarthroplasty (HA), screw internal fixation (IF), and sliding hip screw (SHS) for displaced femoral neck fractures within a network meta-analysis (NMA).

Data Sources: MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials (CENTRAL). Reference lists of relevant studies were hand-searched.

Study Selection: Randomized controlled trials that compared at least two of the surgical options of interest were eligible. Articles were screened in duplicate at the title/abstract and full-text stages by independent reviewers.

Data Extraction and Synthesis: This study follows the PRISMA checklist extension for network meta-analysis. This study is registered on PROSPERO and the protocol was published on a pre-print server prior to analysis. The Cochrane risk of Bias and GRADE tools were used to assess the risk of bias and quality of the evidence included.

Outcomes: Outcomes that were planned within a prior study protocol included: mortality, reoperation, hip-related complications, short term and long-term function, and short-term and long-term quality of life.

Results: A total of 44 studies were included in this systematic review and NMA. Cemented HA demonstrated significant improvement in mortality over uncemented HA, while also providing benefits in hip-related complication rates over uncemented HA, Screw IF, and SHS. THA also demonstrated beneficial results in hip-related complications when compared to Screw IF and SHS. THA provided functional outcome improvements; however, the extent of functional improvement may not meet the minimally important difference (MID) threshold for most patients. Limited evidence is available to differentiate the outcomes of cemented and uncemented THA options, as THA was represented almost exclusively by cemented THA evidence in this review.

Conclusions: Cemented HA and THA generally provided the greatest benefit to harm ratios amongst the available surgical treatment options for displaced femoral neck fractures. The advantageous risk profile makes cemented HA an attractive option for patients that are potentially concerns for future complications or death, and cemented THA is an advantageous option when function is a primary focus of the patient's recovery – while retaining an advantageous complication rates.

Introduction

The incidence of hip fractures worldwide was 1.3 million in 1990, which is projected to increase to 2.6 million cases in the year 2025, and 4.5 million by 2050.[1] It has also been projected that the annual costs to the healthcare system as a result of hip fractures will reach $9.8 billion within the United States and $650 million within Canada alone.[2,3] Due to the increasing incidence and large associated costs, the optimal management of hip fracture is a pivotal area of investigation to reduce its burden on both the patient population and healthcare system.[4]

Hip fractures have a significant impact on the patient, as mortality rates are estimated to be between 5-10% one month after fracture and as high as 30% one year after hip fracture.[5] A large risk factor for mortality is the need for a reoperation or revision surgery.[6] Hip fractures also pose a significant impact on patient daily function and quality of life. A cohort study of over 10,000 hip fracture patients suggested that 71% of patients had trouble walking four months after surgery for their hip fracture, and 58% still had these troubles one year after surgery.[7] In addition to troubles walking, 65% of these patients reported hip pain four months after surgery, and 59% reported pain one year after surgery.[6,7]

There are multiple treatment options for displaced femoral neck fractures, broadly categorized as arthroplasty or internal fixation.[6] Arthroplasty options include total hip arthroplasty (THA) or hemiarthroplasty (HA), while the most common forms of internal fixation (IF) rely on cancellous or sliding hip screws (SHS).[8]

There has been an increasing number of randomized controlled trials (RCTs) and meta-analyses published comparing the various surgical interventions available to treat displaced femoral neck fractures; however, a comprehensive network meta-analysis (NMA) remains an underutilized methodology within this population.[9–12] NMA allows for indirect comparisons of interventions that have not been directly compared within an RCT, but have a common comparison group in separate RCTs.[13] Clinical decision-making with regard to the management of displaced femoral neck fractures requires careful consideration of numerous outcomes, as mortality and other harmful outcomes must be weighed alongside the potential function and quality of life implications that each option may have. For this reason, there is an excellent opportunity to conduct an NMA that compares these outcomes for potential treatment options to aid in the clinical decision-making process for displaced femoral neck fractures.

Methods

This study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols checklist extension for network meta-analysis.[14] This study is also registered on PROSPERO (303952), and the protocol has been published on a pre-print server (https://www.researchgate.net/Femoral Neck NMA Protocol).

Eligibility Criteria

Eligible studies must have been randomized controlled trials that compared one of: 1) THA, 2) HA, 3) SHS, or 4) screw IF. Studies must have been published in English, or had an English

translation available for eligibility. Published articles were reviewed and grouped with other articles that were deemed to be derived from the same study.

Information Sources and Article Selection

A systematic literature search of MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials (CENTRAL) was conducted to identify all eligible studies (**Supplementary Material: Section 1**). The last update of the systematic search was conducted on June 4th, 2022. The reference lists of retrieved studies and recent pairwise meta-analyses on the topic were hand-searched to identify additional eligible studies. Articles were screened independently and in duplicate using Covidence software (covidence.org) at the title/abstract and full-text stage, with a third reviewer consulted to resolve any disagreement.

Data Collection and Outcomes

All data was collected in a standardized and pilot-tested data extraction form independently and in duplicate. A total of 7 outcomes were evaluated: mortality, unplanned secondary procedures (re-operation), and hip-related complications up to 5 years, and short-term function (up to six months post-surgery), long-term function (1 year to 5 years post-surgery), short-term quality of life (up to six months post-surgery), and long-term quality of life (1 year to 5 years post-surgery).

Network Geometry

Network plots were presented for each of the NMAs conducted for the primary analysis of all outcomes. Nodes were weighted by the number of patients that were assessed with that

treatment, whereas connections between the nodes were weighted by the number of unique studies that informed that direct comparison. The protocol described two networks: One in which all HA comparisons were included in a single node, and one in which cemented and uncemented HA were analyzed separately. Due to the prevalence of trials comparing cemented and uncemented HA, this comparison formed the dominant indirect first order loop for most comparisons across the outcomes. As a result, the findings of the NMA in which cemented and uncemented HA were separated is presented as the primary analysis of this study, while the secondary analysis combining all HA options was reported as a secondary analysis.

Methods of Analysis and Summary Measures

NMA for each outcome was conducted using a frequentist random-effects model. Analyses were conducted using the netmeta package in R software (v3.6.2). Results were evaluated according to the minimally contextualized framework[15], and presented using a novel NMA presentation tool (**Table 1**).[16] Study results in this framework were presented with screw IF as a reference.

All continuous outcomes were analyzed as weighted mean differences (MD) with 95% confidence intervals (CIs). Function outcomes were reported as the Harris Hip Score (HHS), and quality of life outcomes were reported as EQ-5D. When studies reported function or quality of life using another measurement tool, means and SDs were linearly translated to the HHS or EQ-5D scale as proposed by Thorlund et al.[17] Results were informed by the minimally important difference (MID) of HHS (8 points on the 0 - 100 point HHS scale) and EQ-5D (0.145 points on the 0 - 1 point EQ-5D scale).[10,18,19] Results were considered to illustrate an appreciable number of patients may achieve the MID when CIs surpassed ½ of the MID.[20] The risk difference (RD) of achieving the MID was unable to be calculated due to insufficient data reported for function and quality of life outcomes.[17] Due to the nature of hip fractures, results were reported as final follow-up scores, but current methodology for calculating the RD of achieving the MID requires change from baseline scores.[17] All dichotomous outcomes were reported as Relative Risk (RR) and RD with 95% CIs.

Assessment of Heterogeneity and Inconsistency

Within-design heterogeneity was assessed using the chi-squared Cochran's Q test and I^2 statistic. Incoherence was evaluated using the node splitting method. Incoherence heat plots were provided to visualize specific areas within the network that contributed to network incoherence.[21,22] In addition to loop-specific incoherence, global network incoherence was assessed via design-by-treatment interaction random effects models.

Additional Analyses

Secondary analyses were conducted to evaluate the impact of risk of bias on the results. While the protocol suggested a subgroup analysis based on age, there were insufficient studies that evaluated patients under the age of 60 to perform such an analysis. Thus, all results pertain to the elderly hip fracture population. Additional analysis also includes the results of the secondary network in which cemented and uncemented HA were analyzed in a single HA group. A post-hoc secondary analysis was conducted to assess the results of uncemented and cemented THA being separated as individual nodes.

Risk of Bias and Certainty of Evidence Assessment

Risk of bias was evaluated using a modified version of the Cochrane Risk of Bias assessment tool proposed by Akl et al.[23] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to evaluate the quality of evidence, specifically for NMA, was conducted to determine the certainty in each treatment effect estimate from the analysis.[24,25] The study protocol further describes the specific methodology in which the GRADE approach was conducted.

Results

Included Studies

The systematic literature search retrieved 3090 unique publications, which were grouped with other publications of the same study to represent 3072 individual studies. After screening, a total of 44 studies (consisting of 62 unique publications) were included within this systematic review and NMA. The literature search PRISMA flow chart for article screening is provided in

Supplementary material: Section 2. Included studies were published between 1986 and 2022. Included studies had an average sample size of 207, ranging from 32 to 1495 patients. The average age of participants within included studies was 80, and 35% of all patients were male. A complete overview of study details and reference list of all included studies is included in **Supplementary material: Section 3**.

Risk of Bias and GRADE Assessment

A complete overview of the risk of bias assessments for all included studies is provided within **Supplementary material: Section 4.** The direct, indirect, and network estimates for all outcomes are provided in **Supplementary material: Section 5,** which includes the GRADE assessments for all comparisons. The subsequent sections of the results will provide an overview of the main findings for each outcome, which are summarized in **Table 2**.

Table 1: Legend for Minimally Contextualized Categorization of Interventions

	BENEFIT OUTCOMES		ADVERSE EVENTS	
Among the Best	Better than 2 or more other interventions	May be better than 2 or more other interventions	Less harmful than 2 or more other interventions	May be less harmful than 2 or more other interventions
Intermediate	Better than 1 other intervention	May be better than 1 other intervention	Less harmful than 1 other intervention	May be less harmful than 1 other intervention
No difference	No different than any other intervention	May be no different than any other intervention	No different than any other intervention	May be no different than any other intervention
Among the Worst	Worse than 1 or more intervention	May be worse than 1 or more intervention	More harmful than 1 or more intervention	May be more harmful than 1 or more intervention

Table 2: Summary of NMA Results for All Outcomes

Intervention	BENEFIT OUTCOMES				ADVERSE EVENTS		
	Short-term Function	**Long-term Function**	**Short-term Quality of Life**	**Long-term Quality of Life**	**Mortality**	**Reoperation**	**Hip-Related Complications**
	HHS MD (95% CI)	*HHS MD (95% CI)*	*EQ-5D MD (95% CI)*	*EQ-5D MD (95% CI)*	*RR (95% CI)*	*RR (95% CI)*	*RR (95% CI)*
Cemented HA	**9.56 (3.47, 15.65)**	2.76 (-2.28, 7.80)	**0.08 (0.03, 0.13)**	0.01 (-0.08, 0.09)	0.92 (0.75, 1.11)	**0.25 (0.13, 0.51)**	**0.42 (0.28, 0.64)**
THA	**11.90 (5.65, 18.15)**	**7.83 (2.73, 12,93)**	**0.10 (0.05, 0.16)**	0.07 (-0.03, 0.17)	1.06 (0.85, 1.31)	**0.21 (0.11, 0.40)**	**0.55 (0.36, 0.82)**
Uncemented HA	**7.09 (0.48, 13.69)**	2.61 (-3.19, 8.40)	-00.06 (0.00, 0.12)	-0.03 (-0.12, 0.06)	1.02 (0.83, 1.25)	**0.31 (0.15, 0.65)**	0.65 (0.41, 1.02)
SHS	**-15.80 (-28.32, -3.28)**	-1.79 (-6.80, 3.22)	0.01 (-0.06, 0.08)	0.06 (-0.03, 0.16)	0.88 (0.66, 1.18)	1.13 (0.54, 2.36)	1.20 (0.78, 1.83)

Effect estimate reference: Screw IF
Bold: Statistically significant compared to reference
Colour categorization based on the direct or indirect estimate if they were rated higher quality evidence than the network estimate
MD: Mean difference, RR: Risk ratio, CI: Confidence interval, THA: Total hip arthroplasty, HA: Hemiarthroplasty, SHS: Sliding hip screw, IF: Internal fixation
Functional score measured as the Harris Hip Score (HHS): Minimally important difference = 8 points.
Quality of Life measured as EuroQol 5 Dimension questionnaire (EQ-5D): Minimally important difference = 0.145 points.

Mortality

The minimally contextualized categorization of treatments deemed cemented HA as the most beneficial treatment with regard to mortality (**Table 2**). Moderate quality evidence showed cemented HA has a significantly lower risk of mortality than uncemented HA (RR: 0.90, 95% CI: 0.82 - 0.99, RD: -3% 95% CI: -5% - -0.1%). Although the network estimate between cemented HA and THA was not significant (RR: 0.87, 95% CI: 0.75 - 1.00, moderate quality evidence, RD: -1%, 95% CI: -4% - 1%), the direct RR estimate between cemented HA and THA was significant (RR: 0.80, 95% CI: 0.67 - 0.95, moderate certainty, RD: -2%, 95% CI: -4% - 1%). All other mortality comparisons were non-significant with moderate or low quality of evidence. A complete summary of the mortality analysis is reported in **Supplementary material: Section 6.**

There were no concerns of heterogeneity, as well as loop-specific or global incoherence ($Q = 7.94$, $p = 0.6345$) within the mortality NMA. Assessment of heterogeneity and incoherence is included within **Supplementary material: Section 7**.

Reoperation

All three arthroplasty options (THA, cemented HA, and uncemented HA) had significantly reduced reoperation rates over screw IF and SHS with moderate certainty evidence (**Table 2; Supplementary material: Section 5**). Moderate quality evidence suggested no difference between THA, cemented HA, and uncemented HA with regard to reoperation, with the exception of THA vs uncemented HA being no different with low quality evidence. A complete summary of the reoperation analysis is reported in **Supplementary material: Section 8**.

Heterogeneity was substantial within the cemented versus uncemented HA direct comparison (I^2 = 82.4%, Q = 44.22, p-value = <0.001). There was no concern of global incoherence (Q = 7.72, p = 0.6558), although comparisons between cemented and uncemented HA, screw IF and uncemented HA, and SHS vs uncemented HA were rated down in GRADE due to loop-specific incoherence. Assessment of heterogeneity and incoherence is included within **Supplementary material: Section 9**.

Hip-Related Complications

Cemented HA had favorable hip-related complication results over Screw IF (RR: 0.42, 95% CI: 0.28 - 0.64, low quality of evidence, RD: -20%, 95% CI: -28% to -12%), SHS (RR: 0.35, 95% CI: 0.22 - 0.56, very low-quality evidence, RD: -22%, 95% CI: -31% - -12%), and uncemented HA (RR: 0.65, 95% CI: 0.47 – 0.89, moderate quality evidence, RD: -5%, 95% CI: -11% - -0.3%). THA had a better complication profile than screw IF (RR: 0.55, 95% CI: 0.36 - 0.82, moderate quality evidence, RD: -18%, -26% - -10%), and SHS (RR: 0.46, 95% CI: 0.28 - 0.74, low quality evidence, RD: -20%, 95% CI: -30% - -10%), whereas uncemented HA was worse than cemented HA as described above, but better than SHS (RR: 0.54, 95% CI: 0.34 - 0.88, low quality evidence, RD: -16%, 95% CI: -26% - -6%). A complete summary of the hip-related complications analysis is reported in **Supplementary material: Section 10**.

Heterogeneity was observed in the comparison between Screw IF and THA (I^2 = 89.6%, Q = 28.86, p – value = <0.001). There was no concern of global incoherence (Q = 7.03, p = 0.6337); however, the comparisons between cemented HA and SHS, as well as THA and SHS, were rated

down in GRADE due to loop-specific incoherence. Assessment of heterogeneity and incoherence is included within **Supplementary material: Section 11**.

Short-term Function

There was low to very low-quality evidence indicating that screw IF and SHS had significantly worse short-term function outcomes than all 3 arthroplasty options (**Supplementary material: Section 5**). THA also had significantly improved short-term functional outcomes over uncemented HA (MD: -4.81, 95% CI -8.85 to -0.78, low quality evidence). Statistically significant comparisons of cemented HA versus SHS (MD: 25.36, 95% CI 11.44 to 39.28, low quality evidence), THA versus Screw IF (MD: 11.90, 95% CI 5.65 to 18.15, very low quality evidence), THA versus SHS (MD: 27.70, 95% CI 13.71 to 41.69, low quality evidence), and uncemented HA versus SHS (MD: 22.89, 95% CI 8.73 to 37.04, low quality evidence) had confidence intervals that exceeded ½ of the MID (HHS MID = 8 points), indicating an appreciable number of patients would achieve a minimally important short-term functional improvement with cemented HA, THA, or uncemented HA over SHS, and from THA over Screw IF. A complete summary of the short-term function analysis is reported in **Supplementary material: Section 12**.

There was substantial heterogeneity within the comparison of cemented and uncemented HA (I^2 = 81.2%, Q = 26.17, p-value = <0.001). There were no concerns of loop-specific or global incoherence (Q = 7.03, p = 0.6838). Assessment of heterogeneity and inconsistency is included within **Supplementary material: Section 13**.

Long-term Function

THA was significantly better than all other treatment options with regard to long-term function. THA comparisons to cemented HA (MD: 5.06, 95% CI: 2.36 to 7.77, low quality evidence), uncemented HA (MD: 5.22, 95% CI: 1.42 to 9.02, low quality evidence), screw IF (MD: 7.83, 95% CI: 2.73 to 12.93, very low-quality evidence), and SHS (MD: 9.62, 95% CI: 5.00 to 14.24, moderate quality evidence), all indicated significant improvements. All of these statistically significant improvements in function – with exception of THA vs SHS – were considered imprecise due to the confidence intervals containing ½ of the MID (HHS MID = 8 points). A complete summary of the long-term function analysis is reported in **Supplementary material: Section 14**.

Substantial heterogeneity was observed in comparisons between cemented HA and THA (I^2 = 71.9%, Q = 18.44, p-value = 0.002), as well as cemented and uncemented HA (I^2 = 80.2%, Q = 29.15, p-value = <0.001). There was no concern of loop-specific or global incoherence (Q = 0.75, p-value = 0.98). The complete assessment of heterogeneity and incoherence is provided in **Supplementary material: Section 15**.

Short-term Quality of Life

Screw IF was significantly worse than cemented HA (MD: -0.08, 95% CI: -0.13 to -0.03, very low-quality evidence), and THA MD: -0.10, 95% CI: -0.16 to -0.05), very low-quality evidence) for short-term quality of life; however, both results were rated own for imprecision due to their confidence intervals containing ½ of the MID (EQ-5D MID = 0.145). A complete summary of the short-term quality of life analysis is reported in **Supplementary material: Section 16**.

Heterogeneity was substantial in the comparison between cemented and uncemented HA, although within-design Q statistic analysis was not statistically significant (I^2 = 76.8%, Q = 2.73, p = 0.2554). There was considerable loop-specific incoherence in the cemented HA versus THA loop comparison through uncemented HA, which primarily contributed to significant global incoherence (Q = 10.85, p-value = <0.001). Assessment of heterogeneity and incoherence is included within **Supplementary material: Section 17**.

Long-term Quality of Life

All comparisons of long-term quality of life were rated of low or very low quality of evidence. The only significant difference indicated that THA had improve quality of life over uncemented HA (MD: 0.10, 95% CI: 0.01 to 0.19, low quality evidence); however, this result did not exceed ½ the MID (EQ-5D MID = 0.145). A complete summary of the long-term quality of life analysis is reported in **Supplementary material: Section 18.**

The assessment of long-term quality of life had substantial concerns of heterogeneity and incoherence throughout the network. As a result, all estimates were rated as low to very low evidence due to heterogeneity and/or incoherence. The heterogeneity and incoherence results are included within **Supplementary material: Section 19,** and the estimates that were rated down for heterogeneity and incoherence are summarized specifically within **Supplementary material: Section 5**.

Additional Analyses

Findings from the NMA in which cemented and uncemented HA were combined are reported in **Supplementary material: Section 20**. The lack of differentiation between cemented HA and uncemented HA hides important differences between these two options that were identified within the primary analysis. Results for other surgical options remained similar, as screw IF and SHS were inferior across most outcomes to both THA and HA, and THA demonstrated favorable long-term functional outcomes over HA. Additionally, results from the sensitivity analysis including only low risk of bias investigations demonstrated similar findings to the main analysis; as screw IF and SHS were consistently inferior to arthroplasty options, cemented HA was favorable in mortality outcomes (significantly better than THA - RR: 0.81, 95% CI: 0.69 to 0.96), and both cemented HA and THA had beneficial outcomes with regard to complications. The long-term functional benefits of THA were not apparent when analyzing only low risk of bias investigations (**Supplementary material: Section 21**).

A post-hoc analysis was conducted to differentiate results between cemented and uncemented THA. A limited amount of evidence was available for uncemented THA, as almost all of the included studies utilized cemented THA. Within the available evidence, uncemented THA demonstrated similar findings to cemented THA for all outcomes. A complete overview of the post-hoc analysis is provided within **Supplementary material: Section 22**.

Discussion

The results of this NMA demonstrated that internal fixation, with either cancellous screw or sliding hip screw, were consistently inferior to the arthroplasty options across most outcomes. Cemented HA demonstrated a number of favorable results over its uncemented counterpart;

notably a lower mortality and complication rates. Thus, there appears to be minimal benefit to taking an uncemented approach over a cemented HA. It is of note that the direct effect estimate and sensitivity analysis comparing cemented HA to THA demonstrated that cemented HA had a small but significant improvement in mortality over THA – however this result was not observed in the primary network estimate. This further provides support to the favorable mortality rates observed with cemented HA. THA was particularly advantageous with regard to long-function over all four other surgical options – although these statistically significant benefits in function may not always be clinically important to patients based on the MID for the HHS. Regardless, the advantageous risk profile makes cemented HA an attractive option for patients that are potentially concerns for future complications or death, and cemented THA is an advantageous option when function is a primary focus of the patient's recovery – while retaining an advantageous complication rate.

These results parallel the findings of the largest trial conducted comparing THA to HA, which suggested similar mortality and reoperation rates, while THA had small improvements in functional outcomes.[11] Other clinical decisions may drive the decision between cemented HA and THA, such as operative time, blood loss, and associated injuries/indications. Prior meta-analyses have demonstrated that HA requires a shorter operative time than THA – albeit most likely to be clinically unimportant.[10] Patient-specific factors may drive a surgeon's decision for displaced femoral neck fracture management, but cemented THA and cemented HA should be the focus of preliminary treatment decisions based on the available evidence.

Recently, displaced femoral neck fractures have been the topic of many large, high quality, randomized trials. These trials have individually compared the various arthroplasty and internal fixation treatment options, which provides a strong platform for this NMA to combine and evaluate the body of evidence as a whole.[11,12,26,27] This NMA has advanced the understanding of these differences by providing an in-depth assessment of these options, while separating the results of cemented and uncemented HA. The perceived benefit of cemented HA over uncemented HA within this NMA parallels the recently published and largest trial to date comparing cemented versus uncemented HA.[27] A cost-utility analysis conducted parallel to this RCT reported that cemented HA was cost-effective in comparison to uncemented HA, as it was cost saving while also providing benefits in quality-adjusted life years.[27,28] Although there is a large body of evidence that clearly defines the differences in outcomes between cemented and uncemented HA, the same body of evidence does not exist for cemented and uncemented THA. The available evidence within this NMA was almost exclusively from cemented THA, with minimal data available to differentiate uncemented THA. A post-hoc analysis was conducted that did not observe any differences between cemented and uncemented THA, yet the paucity of high-quality evidence in this comparison precludes any definitive conclusions to understand if the use of cement – whether it be with HA or THA – is the primary driver of favorable outcomes, or if the differences in cemented uncemented HA are not transferrable to the use of THA.

This NMA is strengthened by its thorough and systematic assessment of the different surgical options for displaced femoral neck fracture. Although a number of pairwise comparisons have been assessed between these options, this NMA provides the most comprehensive overview of the comparative evidence for all surgical options. This will prove to be an important

evaluation of the surgical landscape for displaced femoral neck fractures, which not only provides a detailed understanding of the benefits and harms of each surgical option – but also highlights areas for future research to further advance the field of hip fracture care. The use of GRADE to evaluate the quality of evidence provides a methodological strength in which few NMAs within the field of orthopaedics have fully adopted. This NMA is limited by its inclusion of trials published in English only; however, it has been proven that inclusion of studies published in languages beyond English is not likely to result in differences in findings.[29]

Displaced femoral neck fractures in the elderly population remain a major concern due to the risk of mortality and morbidity. Across all included studies, the absolute mortality rate was approximately 24% within 5 years. Perhaps one the most impactful future advancements within orthopaedic care would include strategies that create a meaningful reduction in post-hip fracture mortality – regardless of the surgical technique utilized. Mortality after hip fracture is associated with considerable morbidity and healthcare costs, as elderly patients often develop additional sequelae following a hip fracture. Future research initiatives that focus on strategies to minimize post-hip fracture mortality and morbidity could have an impact that few other investigations within orthopaedics could have.

Conclusion

This review found that cemented HA and THA generally provided the greatest benefit to harm ratios amongst the available surgical treatment options for displaced femoral neck fractures. Cemented HA had beneficial results in comparison with regard to mortality and complication rates, while THA demonstrated benefits in complication rates and function.

Notably, cemented HA demonstrated improvement in both mortality and hip-related complications over uncemented HA with moderate certainty evidence. Internal fixation options were consistently inferior to arthroplasty options across most outcomes.

References

1. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA*. 1997;7(5):407-413.

2. Papadimitropoulos EA, Coyte PC, Josse RG, Greenwood CE. Current and projected rates of hip fracture in Canada. *CMAJ Can Med Assoc J J Assoc Medicale Can*. 1997;157(10):1357-1363.

3. Bhandari M, Tornetta P, Hanson B, Swiontkowski MF. Optimal Internal Fixation for Femoral Neck Fractures: Multiple Screws or Sliding Hip Screws? *J Orthop Trauma*. 2009;23(6):403-407. doi:10.1097/BOT.0b013e318176191f

4. Daigle ME, Weinstein AM, Katz JN, Losina E. The cost-effectiveness of total joint arthroplasty: a systematic review of published literature. *Best Pract Res Clin Rheumatol*. 2012;26(5):649-658. doi:10.1016/j.berh.2012.07.013

5. Keene GS, Parker MJ, Pryor GA. Mortality and morbidity after hip fractures. *BMJ*. 1993;307(6914):1248-1250.

6. FAITH Investigators. Fixation using alternative implants for the treatment of hip fractures (FAITH): design and rationale for a multi-centre randomized trial comparing sliding hip screws

and cancellous screws on revision surgery rates and quality of life in the treatment of femoral neck fractures. *BMC Musculoskelet Disord*. 2014;15(1):219. doi:10.1186/1471-2474-15-219

7. Gjertsen JE, Baste V, Fevang JM, Furnes O, Engesæter LB. Quality of life following hip fractures: results from the Norwegian hip fracture register. *BMC Musculoskelet Disord*. 2016;17(1):265. doi:10.1186/s12891-016-1111-y

8. Khan M, Aleem IS, Poolman RW. Fixation versus primary replacement of displaced femoral neck fractures in the elderly. *Indian J Orthop*. 2011;45(1):23-26. doi:10.4103/0019-5413.73658

9. FAITH Investigators. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *Lancet Lond Engl*. 2017;389(10078):1519-1527. doi:10.1016/S0140-6736(17)30066-1

10. Ekhtiari S, Gormley J, Axelrod DE, Devji T, Bhandari M, Guyatt GH. Total Hip Arthroplasty Versus Hemiarthroplasty for Displaced Femoral Neck Fracture: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *J Bone Joint Surg Am*. 2020;102(18):1638-1645. doi:10.2106/JBJS.20.00226

11. HEALTH Investigators. Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. *N Engl J Med*. 2019;381(23):2199-2208. doi:10.1056/NEJMoa1906190

12. FAITH-2 Investigators. Fixation using alternative implants for the treatment of hip fractures (FAITH-2): design and rationale for a pilot multi-centre 2 x 2 factorial randomized

controlled trial in young femoral neck fracture patients. *Pilot Feasibility Stud*. 2019;5(101676536):70. doi:10.1186/s40814-019-0458-x

13. Tonin FS, Rotta I, Mendes AM, Pontarolo R. Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharm Pract*. 2017;15(1):943. doi:10.18549/PharmPract.2017.01.943

14. 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. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385

15. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ*. 2020;371. doi:10.1136/bmj.m3900

16. Phillips MR, Sadeghirad B, Busse JW, et al. Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study. *BMJ Open*. 2022;12(6):e056400. doi:10.1136/bmjopen-2021-056400

17. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. *Res Synth Methods*. 2011;2(3):188-203. doi:10.1002/jrsm.46

18. Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*. 2020;369:m1714. doi:10.1136/bmj.m1714

19. Carrasco-Labra A, Devji T, Qasim A, et al. Minimal important difference estimates for patient-reported outcomes: A systematic survey. *J Clin Epidemiol*. 2020;0(0). doi:10.1016/j.jclinepi.2020.11.024

20. Busse JW, Sadeghirad B, Oparin Y, et al. Management of Acute Pain From Non-Low Back, Musculoskeletal Injuries : A Systematic Review and Network Meta-analysis of Randomized Trials. *Ann Intern Med*. 2020;173(9):730-738. doi:10.7326/M19-3601

21. Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol*. 2013;13:35. doi:10.1186/1471-2288-13-35

22. Krahn U, Binder H, König J. Visualizing inconsistency in network meta-analysis by independent path decomposition. *BMC Med Res Methodol*. 2014;14:131. doi:10.1186/1471-2288-14-131

23. Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *J Clin Epidemiol*. 2012;65(3):262-267. doi:10.1016/j.jclinepi.2011.04.015

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. doi:10.1016/j.jclinepi.2017.10.005

25. 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. doi:10.1016/j.jclinepi.2017.10.005

26. FAITH Investigators. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *Lancet Lond Engl*. 2017;389(10078):1519-1527. doi:10.1016/S0140-6736(17)30066-1

27. Fernandez MA, Achten J, Parsons N, et al. Cemented or Uncemented Hemiarthroplasty for Intracapsular Hip Fracture. *N Engl J Med*. 2022;386(6):521-530. doi:10.1056/NEJMoa2108337

28. Png ME, Petrou S, Fernandez MA, et al. Cost-utility analysis of cemented hemiarthroplasty versus hydroxyapatite-coated uncemented hemiarthroplasty for the treatment of displaced intracapsular hip fractures : the World Hip Trauma Evaluation 5 (WHiTE 5) trial. *Bone Jt J*. 2022;104-B(8):922-928. doi:10.1302/0301-620X.104B8.BJJ-2022-0417.R1

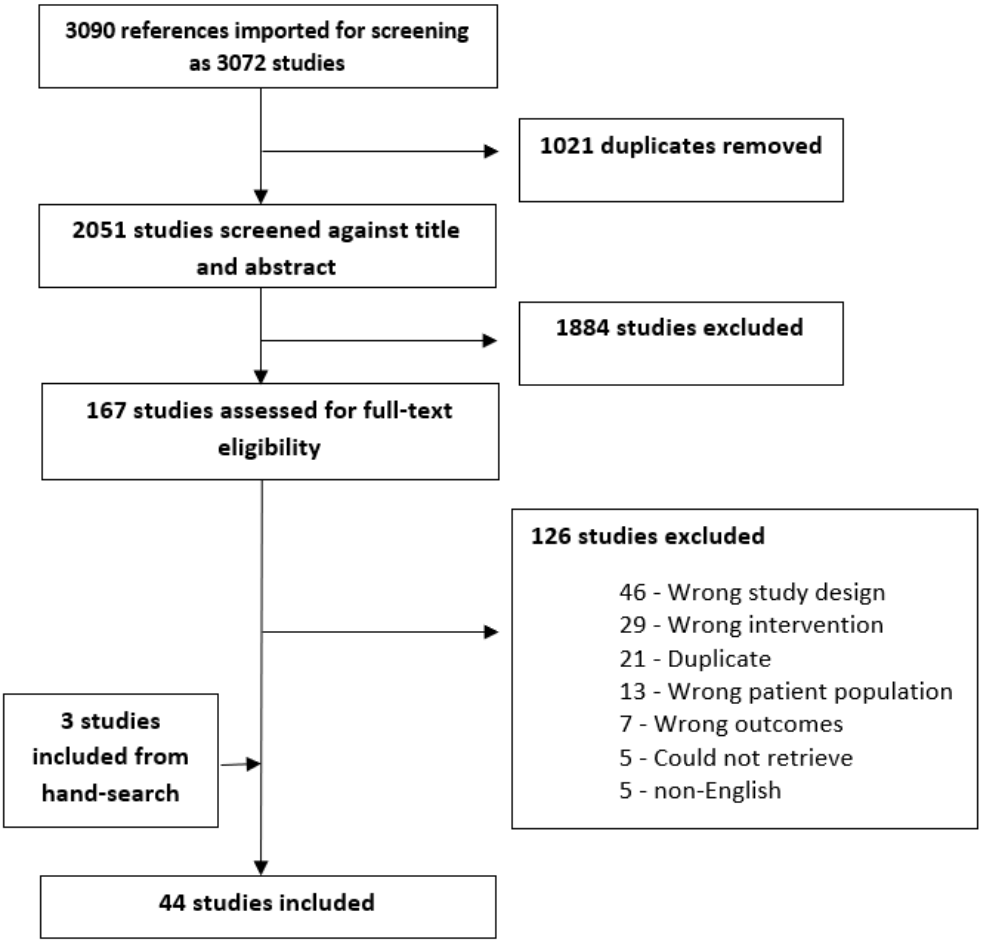
29. Nussbaumer-Streit B, Klerings I, Dobrescu AI, et al. Excluding non-English publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. *J Clin Epidemiol*. 2020;118:42-54. doi:10.1016/j.jclinepi.2019.10.011

Appendix A: Supplementary Materials

Surgical Management of Displaced Femoral Neck Fractures: A Systematic Review and Network Meta-Analysis

Supplementary Materials

Section 1: Screening Flow Diagram



Section 2: Systematic Literature Search: Medline

1. Arthroplasty, Replacement, Hip/ or Hip Prosthesis/ or total hip arthroplasty.mp. or THA.mp. or total hip replacement.mp. or THR.mp.
2. Hemiarthroplasty/ or hemiarthrop*.mp.
3. hip arthrop*.mp.
4. Fracture Fixation, Internal/
5. screw.mp.
6. internal fixation.mp.
7. ORIF.mp.
8. dynamic hip screw.mp.
9. sliding hip screw.mp.
10. surg*.mp. and Orthopedics/
11. fixat*.mp.
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. Femoral Neck Fractures/
14. (fem* neck* or prox* fem* or hip).mp.
15. 13 or 14
16. 12 and 15
17. randomized controlled trial.pt.
18. controlled clinical trial.pt.
19. randomized.ab.
20. placebo.ab.
21. clinical trials as topic.sh.
22. randomly.ab.
23. trial.ti.
24. 17 or 18 or 19 or 20 or 21 or 22 or 23
25. exp animals/ not humans.sh.
26. 24 not 25
27. 16 and 26

Section 3: Included Studies

Included Study Characteristics

Study ID	Primary Country	Interventions Compared	Sample Size (displaced)	Mean Age	Percent Male
Baker 2006	United Kingdom	THA Cemented HA	40 41	74 76	80% 78%
Blomfeldt 2005	Sweden	Uncemented HA Screw IF	30 30	84 84	7% 14%
Blomfeldt 2007	Sweden	THA Cemented HA	60 60	81 81	22% 10%
Cadossi 2013	Italy	HA THA	41 42	84 82	32% 19%
Cao 2014	China	THA Screw IF	157 128	76 77	46% 46%
Chammout 2012	Sweden	THA Screw IF	43 57	78 79	12% 28%
Davison 2001	United Kingdom	Cemented HA SHS	187 93	75 73	24% 25%
deAngelis 2011	United States	Cemented HA Uncemented HA	66 64	82 83	21% 25%
Dorr 1986	United States	THA Cemented HA Uncemented HA	39 37 13	69 72 66	41% 30% 31%
El-Abed 2005	Ireland	Uncemented HA SHS	62 60	74 72	36% 30%
Emery 1991	United Kingdom	Cemented HA Uncemented HA	27 26	78 80	11% 15%
FAITH 2017	Canada	Screw IF SHS	167 179	72[#] 72[#]	39%[#] 40%[#]
FAITH-2 2020	Canada	Screw IF SHS	31 30	39[#] 43[#]	77%[#] 23%[#]
Fernandez 2022	United Kingdom	Cemented HA Uncemented HA	610 615	85 84	31% 33%
Figved 2009	Norway	Cemented HA Uncemented HA	115 115	83 83	22% 26%
Frihagen 2013	Norway	Cemented HA Screw IF	110 112	83 83	22% 13%
HEALTH 2019*	Canada	THA HA	718 723	79 79	29% 31%
Hedbeck 2013	Sweden	Cemented HA Screw IF	30 30	84 85	17% 17%
HOPE 2019	Sweden	THA Cemented HA	60 60	85 86	25% 25%
Inngul 2015	Sweden	Cemented HA Uncemented HA	67 74	81 81	31% 28%
Iorio 2019	Italy	THA Uncemented HA	30 30	82 83	40% 43%
Johansson 2014	Sweden	THA Screw IF	50 50	84 84	20% 32%
Keating 2006	United Kingdom	Cemented HA THA	69 69	75 75	22% 25%
Linde 1986	Denmark	Screw IF SHS	47 40	76 76	70% 60%

Macaulay 2008	United States	HA THA	23 17	77 82	39% 59%
Madsen 1987	Denmark	Screw IF SHS	52 51	75 74	21% 29%
Moerman 2017	Netherlands	Cemented HA Uncemented HA	110 91	83 84	25% 33%
Mohabey 2017	India	Cemented HA Uncemented HA	20 20	70	45%
Morvin 2020	Slovenia	Cemented HA Uncemented HA	79 79	86 84	42% 39%
Mouzopolous 2008	Greece	THA HA SHS	37 34 38	73 74 75	24% 29% 32%
Parker 2002	United Kingdom	Uncemented HA Cemented HA	229 226	82 82	20% 20%
Parker 2019	United Kingdom	THA Cemented HA	52 53	77 77	15% 23%
Puolakka 2001	Finland	Cemented HA Screw IF	15 17	82 81	7% 23%
Ravikumar 2000	United Kingdom	THA Uncemented HA SHS	89 91 91	81 82 80	NR NR NR
Roden 2003	Sweden	HA Screw IF	47 53	81 81	28 30
Rogmark 2002	Sweden	HA THA	89 103	82 82	NR NR
Santini 2005	Italy	Cemented HA Uncemented HA	53 53	82 80	25% 21%
Sharma 2016	India	HA THA	40 40	73 78	28% 35%
Sonaje 2017	France	THA Cemented HA	20 20	65 66	30% 35%
Talsnes 2013	Norway	Cemented HA Uncemented HA	162 172	84 84	28% 22%
Taylor 2012	New Zealand	Cemented HA Uncemented HA	80 80	85 85	29% 66%
Tidermark 2003	Sweden	THA Screw IF	55 55	79 81	18% 21%
van den Bekerom 2010	Netherlands	Cemented HA THA	137 115	80 82	16% 22%
Vidovic 2013	Croatia	Cemented HA Uncemented HA	38 41	83 82	NR NR

THA: Total hip arthroplasty, HA: Hemiarthroplasty, SHS: Sliding hip screw, IF: Internal fixation

*Study data was provided to allow for comparison between cemented and uncemented HA

#Value for the entire study, but only displaced fractures were included in this analysis

Protocol requested extraction of BMI – however this data was infrequently and inadequately reported

Reference List of Included Studies

62 publications representing 44 unique studies

Avery PP, Baker RP, Walton MJ, et al. Total hip replacement and hemiarthroplasty in mobile, independent patients with a displaced intracapsular fracture of the femoral neck: a seven- to ten-year follow-up report of a prospective randomised controlled trial. *The Journal of bone and joint surgery British volume*. 2011;93(8):1045-1048. doi:10.1302/0301-620X.93B8.27132

Axelrod D., Comeau-Gauthier M., Bzovsky S., et al. What Predicts Health-Related Quality of Life for Patients With Displaced Femoral Neck Fractures Managed With Arthroplasty? A Secondary Analysis of the HEALTH Trial. *Journal of orthopaedic trauma*. 2020;34(Supplement 3):S29-S36. doi:10.1097/BOT.0000000000001933

Baker RP, Squires B, Gargan MF, Bannister GC. Total hip arthroplasty and hemiarthroplasty in mobile, independent patients with a displaced intracapsular fracture of the femoral neck. A randomized, controlled trial. *The Journal of bone and joint surgery American volume*. 2006;88(12):2583-2589.

Barenius B., Inngul C., Alagic Z., Enocson A. A randomized controlled trial of cemented versus cementless arthroplasty in patients with a displaced femoral neck fracture. *Bone and Joint Journal*. 2018;100B(8):1087-1093. doi:10.1302/0301-620X.100B8.BJJ-2017-1593.R1

Blankstein M., Schemitsch E.H., Bzovsky S., et al. What Factors Increase Revision Surgery Risk When Treating Displaced Femoral Neck Fractures With Arthroplasty: A Secondary Analysis of the HEALTH Trial. *Journal of orthopaedic trauma*. 2020;34(Supplement 3):S49-S54. doi:10.1097/BOT.0000000000001936

Blomfeldt R, Tornkvist H, Eriksson K, Soderqvist A, Ponzer S, Tidermark J. A randomised controlled trial comparing bipolar hemiarthroplasty with total hip replacement for displaced intracapsular fractures of the femoral neck in elderly patients. *The Journal of bone and joint surgery British volume*. 2007;89(2):160-165.

Blomfeldt R, Tornkvist H, Ponzer S, Soderqvist A, Tidermark J. Internal fixation versus hemiarthroplasty for displaced fractures of the femoral neck in elderly patients with severe cognitive impairment. *The Journal of bone and joint surgery British volume*. 2005;87(4):523-529.

Blomfeldt R, Tornkvist H, Ponzer S, Soderqvist A, Tidermark J. Comparison of internal fixation with total hip replacement for displaced femoral neck fractures. Randomized, controlled trial performed at four years. *The Journal of bone and joint surgery American volume*. 2005;87(8):1680-1688.

Cadossi M, Chiarello E, Savarino L, et al. A comparison of hemiarthroplasty with a novel polycarbonate-urethane acetabular component for displaced intracapsular fractures of the femoral neck: a randomised controlled trial in elderly patients. 2013;95-B(5):609-615. doi:10.1302/0301-620X.95B5.31083

Cao L, Wang B, Li M, et al. Closed reduction and internal fixation versus total hip arthroplasty for displaced femoral neck fracture. *Chinese journal of traumatology = Zhonghua chuang shang za zhi*. 2014;17(2):63-68.

Chammout G, Kelly-Pettersson P, Hedbeck CJ, Stark A, Mukka S, Skoldenberg O. HOPE-Trial: Hemiarthroplasty Compared with Total Hip Arthroplasty for Displaced Femoral Neck Fractures in Octogenarians: A Randomized Controlled Trial. *JB & JS open access*. 2019;4(2):e0059. doi:10.2106/JBJS.OA.18.00059

Chammout GK, Mukka SS, Carlsson T, Neander GF, Stark AWH, Skoldenberg OG. Total hip replacement versus open reduction and internal fixation of displaced femoral neck fractures: a randomized long-term follow-up study. *The Journal of bone and joint surgery American volume*. 2012;94(21):1921-1928.

Davison JN, Calder SJ, Anderson GH, et al. Treatment for displaced intracapsular fracture of the proximal femur. A prospective, randomised trial in patients aged 65 to 79 years. *The Journal of bone and joint surgery British volume*. 2001;83(2):206-212.

de Angelis J.P., Ademi A., Staff I., Lewis C.G. Cemented versus uncemented hemiarthroplasty for displaced femoral neck fractures: A prospective randomized trial with early follow-up. *Journal of Orthopaedic Trauma*. Published online 2011. doi:10.1097/BOT.0b013e318238b7a5

Dorr LD, Glousman R, Hoy AL, Vanis R, Chandler R. Treatment of femoral neck fractures with total hip replacement versus cemented and noncemented hemiarthroplasty. *The Journal of arthroplasty*. 1986;1(1):21-28.

El-Abed K, McGuinness A, Brunner J, Dallovedova P, O'Connor P, Kennedy JG. Comparison of outcomes following uncemented hemiarthroplasty and dynamic hip screw in the treatment of displaced subcapital hip fractures in patients aged greater than 70 years. *Acta orthopaedica Belgica*. 2005;71(1):48-54.

Emery RJ, Broughton NS, Desai K, Bulstrode CJ, Thomas TL. Bipolar hemiarthroplasty for subcapital fracture of the femoral neck. A prospective randomised trial of cemented Thompson and uncemented Moore stems. *The Journal of bone and joint surgery British volume*. 1991;73(2):322-324.

Fernandez MA, Achten J, Parsons N, et al. Cemented or Uncemented Hemiarthroplasty for Intracapsular Hip Fracture. *The New England journal of medicine*. 2022;386(6):521-530. doi:10.1056/NEJMoa2108337

Figved W, Opland V, Frihagen F, Jervidalo T, Madsen JE, Nordsletten L. Cemented versus uncemented hemiarthroplasty for displaced femoral neck fractures. *Clinical orthopaedics and related research*. 2009;467(9):2426-2435. doi:10.1007/s11999-008-0672-y

Fixation using Alternative Implants for the Treatment of Hip fractures (FAITH) Investigators. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *Lancet (London, England)*. 2017;389(10078):1519-1527. doi:10.1016/S0140-6736(17)30066-1

Frihagen F., Madsen J.E., Nordsletten L., Lofthus C.M., Stoen R.O. Hemiarthroplasty or internal fixation for displaced femoral neck fractures in the elderly: 6 year follow up of an RCT. *Osteoporosis International*. 2013;24(1 SUPPL. 1):S135. doi:10.1007/s00198-013-2312-y

Frihagen F, Nordsletten L, Madsen JE. Hemiarthroplasty or internal fixation for intracapsular displaced femoral neck fractures: randomised controlled trial. *BMJ (Clinical research ed)*. 2007;335(7632):1251-1254.

HEALTH Investigators, Bhandari M, Einhorn TA, et al. Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. *The New England journal of medicine*. 2019;381(23):2199-2208. doi:10.1056/NEJMoa1906190

Hedbeck CJ, Enocson A, Lapidus G, et al. Comparison of bipolar hemiarthroplasty with total hip arthroplasty for displaced femoral neck fractures: a concise four-year follow-up of a randomized trial. *The Journal of bone and joint surgery American volume*. 2011;93(5):445-450. doi:10.2106/JBJS.J.00474

Hedbeck CJ, Inngul C, Blomfeldt R, Ponzer S, Tornkvist H, Enocson A. Internal fixation versus cemented hemiarthroplasty for displaced femoral neck fractures in patients with severe cognitive dysfunction: a randomized controlled trial. *Journal of orthopaedic trauma*. 2013;27(12):690-695. doi:10.1097/BOT.0b013e318291f544

Inngul C, Blomfeldt R, Ponzer S, Enocson A. Cemented versus uncemented arthroplasty in patients with a displaced fracture of the femoral neck: a randomised controlled trial. 2015;97-B(11):1475-1480. doi:10.1302/0301-620X.97B11.36248

Iorio R, Iannotti F, Mazza D, et al. Is dual cup mobility better than hemiarthroplasty in patients with dementia and femoral neck fracture? A randomized controlled trial. *SICOT-J*. 2019;5(101675099):38. doi:10.1051/sicotj/2019035

Johansson T, Jacobsson SA, Ivarsson I, Knutsson A, Wahlstrom O. Internal fixation versus total hip arthroplasty in the treatment of displaced femoral neck fractures: a prospective randomized study of 100 hips. *Acta orthopaedica Scandinavica*. 2000;71(6):597-602.

Johansson T, Risto O, Knutsson A, Wahlstrom O. Heterotopic ossification following internal fixation or arthroplasty for displaced femoral neck fractures: a prospective randomized study. *International orthopaedics*. 2001;25(4):223-225.

Johansson T. Internal fixation compared with total hip replacement for displaced femoral neck fractures: a minimum fifteen-year follow-up study of a previously reported randomized trial. *The Journal of bone and joint surgery American volume*. 2014;96(6):e46. doi:10.2106/JBJS.K.00244

Johansson T, Bachrach-Lindstrom M, Aspenberg P, Jonsson D, Wahlstrom O. The total costs of a displaced femoral neck fracture: comparison of internal fixation and total hip replacement. A randomised study of 146 hips. *International orthopaedics*. 2006;30(1):1-6.

Keating JF, Grant A, Masson M, Scott NW, Forbes JF. Randomized comparison of reduction and fixation, bipolar hemiarthroplasty, and total hip arthroplasty. Treatment of displaced intracapsular hip fractures in healthy older patients. *J Bone Joint Surg Am*. 2006;88(2):249-260. doi:10.2106/JBJS.E.00215

Langslet E, Frihagen F, Opland V, Madsen JE, Nordsletten L, Figved W. Cemented versus uncemented hemiarthroplasty for displaced femoral neck fractures: 5-year followup of a randomized trial. *Clinical orthopaedics and related research*. 2014;472(4):1291-1299. doi:10.1007/s11999-013-3308-9

Leonardsson O, Sernbo I, Carlsson A, Akesson K, Rogmark C. Long-term follow-up of replacement compared with internal fixation for displaced femoral neck fractures: results at ten years in a randomised study of 450 patients. *The Journal of bone and joint surgery British volume*. 2010;92(3):406-412. doi:10.1302/0301-620X.92B3.23036

Linde F, Andersen E, Hvass I, Madsen F, Pallesen R. Avascular femoral head necrosis following fracture fixation. *Injury*. 1986;17(3):159-163.

Macaulay W, Nellans KW, Iorio R, et al. Total hip arthroplasty is less painful at 12 months compared with hemiarthroplasty in treatment of displaced femoral neck fracture. *HSS journal : the musculoskeletal journal of Hospital for Special Surgery*. 2008;4(1):48-54. doi:10.1007/s11420-007-9061-4

Madsen F, Linde F, Andersen E, Birke H, Hvass I, Poulsen TD. Fixation of displaced femoral neck fractures. A comparison between sliding screw plate and four cancellous bone screws. *Acta orthopaedica Scandinavica*. 1987;58(3):212-216.

Moerman S, Mathijssen NMC, Niesten DD, et al. More complications in uncemented compared to cemented hemiarthroplasty for displaced femoral neck fractures: a randomized controlled trial of 201 patients, with one year follow-up. *BMC musculoskeletal disorders*. 2017;18(1):169. doi:10.1186/s12891-017-1526-0

Mohabey AV, Warjukar PR, Ravikumar M. Functional outcome of cemented versus uncemented modular bipolar hemiarthroplasty in proximal femoral neck fractures. *Int J Orthop Sci*. 2017;3(4):609-611. doi:10.22271/ortho.2017.v3.i4i.83

Mouzopoulos G, Stamatakos M, Arabatzi H, et al. The four-year functional result after a displaced subcapital hip fracture treated with three different surgical options. *Int Orthop*. 2008;32(3):367-373. doi:10.1007/s00264-007-0321-1

Movrin I. Cemented versus uncemented hemiarthroplasty for displaced femoral neck fractures: A randomized controlled trial with two years follow-up. *Acta orthopaedica et traumatologica turcica*. 2020;54(1):83-88. doi:10.5152/j.aott.2020.01.432

Parker MJ, Khan RJK, Crawford J, Pryor GA. Hemiarthroplasty versus internal fixation for displaced intracapsular hip fractures in the elderly. A randomised trial of 455 patients. *The Journal of bone and joint surgery British volume*. 2002;84(8):1150-1155.

Parker MJ, Cawley S. Treatment of the displaced intracapsular fracture for the “fitter” elderly patients: A randomised trial of total hip arthroplasty versus hemiarthroplasty for 105 patients. *Injury*. 2019;50(11):2009-2013. doi:10.1016/j.injury.2019.09.018

Puolakka TJ, Laine HJ, Tarvainen T, Aho H. Thompson hemiarthroplasty is superior to Ullevaal screws in treating displaced femoral neck fractures in patients over 75 years. A prospective randomized study with two-year follow-up. *Annales chirurgiae et gynaecologiae*. 2001;90(3):225-228.

Ravikumar KJ, Marsh G. Internal fixation versus hemiarthroplasty versus total hip arthroplasty for displaced subcapital fractures of femur--13 year results of a prospective randomised study. *Injury*. 2000;31(10):793-797.

Roden M, Schon M, Fredin H. Treatment of displaced femoral neck fractures: a randomized minimum 5-year follow-up study of screws and bipolar hemiprostheses in 100 patients. *Acta orthopaedica Scandinavica*. 2003;74(1):42-44.

Rogmark C., Carlsson A., Johnell O., Sernbo I. A prospective randomised trial of internal fixation versus arthroplasty for displaced fractures of the neck of the femur. *Journal of Bone and Joint Surgery - Series B*. 2002;84(2):183-188. doi:10.1302/0301-620X.84B2.11923

Santini S, Rebeccato A, Bolgan I, Turi G. Hip fractures in elderly patients treated with bipolar hemiarthroplasty: comparison between cemented and cementless implants. 2005;6(2):80-87. doi:10.1007/s10195-005-0086-5

Sharma V, Awasthi B, Kumar K, Kohli N, Katoch P. Outcome Analysis of Hemiarthroplasty vs. Total Hip Replacement in Displaced Femoral Neck Fractures in the Elderly. *Journal of clinical and diagnostic research : JCDR*. 2016;10(5):RC11-3. doi:10.7860/JCDR/2016/18638.7877

Slobogean GP, Sprague S, Bzovsky S, et al. Fixation Using Alternative Implants for the Treatment of Hip Fractures (FAITH-2): The Clinical Outcomes of a Multicenter 2 x 2 Factorial Randomized Controlled Pilot Trial in Young Femoral Neck Fracture Patients. *Journal of orthopaedic trauma*. 2020;34(10):524-532. doi:10.1097/BOT.0000000000001773

Slobogean GP, Sprague S, Bzovsky S, et al. Fixation using Alternative Implants for the Treatment of Hip Fractures (FAITH-2): The Exploratory Health-Related Quality of Life and Patient-Reported Functional Outcomes of a Multi-Centre 2 x 2 Factorial Randomized Controlled Pilot Trial in Young Femoral. *Injury*. 2021;(0226040, gon). doi:10.1016/j.injury.2021.02.030

Sonaje J, Meena P, Bansiwal R, Bobade S. Comparison of functional outcome of bipolar hip arthroplasty and total hip replacement in displaced femoral neck fractures in elderly in a developing country: a 2-year prospective study. Published online 2017:1-6. doi:10.1007/s00590-017-2057-y

Sprague S, Bhandari M, Heetveld M, et al. Factors associated with health-related quality of life, hip function, and health utility after operative management of femoral neck fractures. 2018;100B(3):361-369. doi:10.1302/0301-620X.100B3.BJJ-2017-0853.R1

Stoen R.O., Lofthus C.M., Nordsletten L., Madsen J.E., Frihagen F. Randomized trial of hemiarthroplasty versus internal fixation for femoral neck fractures: No differences at 6 years hip. *Clinical Orthopaedics and Related Research*. 2014;472(1):360-367. doi:10.1007/s11999-013-3245-7

Talsnes O, Hjelmstedt F, Pripp AH, Reikeras O, Dahl OE. No difference in mortality between cemented and uncemented hemiprosthesis for elderly patients with cervical hip fracture. A prospective randomized study on 334 patients over 75 years. *Archives of orthopaedic and trauma surgery*. 2013;133(6):805-809. doi:10.1007/s00402-013-1726-5

Taylor F, Wright M, Zhu M. Hemiarthroplasty of the hip with and without cement: a randomized clinical trial. *The Journal of bone and joint surgery American volume*. 2012;94(7):577-583. doi:10.2106/JBJS.K.00006

Tidermark J, Ponzer S, Svensson O, Soderqvist A, Tornkvist H. Internal fixation compared with total hip replacement for displaced femoral neck fractures in the elderly. A randomised, controlled trial. *The Journal of bone and joint surgery British volume*. 2003;85(3):380-388.

Tol M, van den Bekerom M, Sierevelt I, Hilverdink E, Raaymakers E, Goslings J. Hemiarthroplasty or total hip arthroplasty for the treatment of a displaced intracapsular fracture in active elderly patients: 12-year follow-up of randomised trial. 2017;99-B(2):250-254. doi:10.1302/0301-620X.99B2.BJJ-2016-0479.R1

van den Bekerom MPJ, Hilverdink EF, Sierevelt IN, et al. A comparison of hemiarthroplasty with total hip replacement for displaced intracapsular fracture of the femoral neck: a randomised controlled multicentre trial in patients aged 70 years and over. *The Journal of bone and joint surgery British volume*. 2010;92(10):1422-1428. doi:10.1302/0301-620X.92B10.24899

Vidovic D, Matejcic A, Punda M, et al. Periprosthetic bone loss following hemiarthroplasty: a comparison between cemented and cementless hip prosthesis. *Injury*. 2013;44 Suppl 3(0226040, gon):S62-6. doi:10.1016/S0020-1383(13)70201-8

Vidovic D, Punda M, Darabos N, Bekavac-Beslin M, Bakota B, Matejcic A. Regional bone loss following femoral neck fracture: A comparison between cemented and cementless hemiarthroplasty. *Injury*. 2015;46 Suppl 6(0226040, gon):S52-6. doi:10.1016/j.injury.2015.10.069

Waaler Bjornelv GM, Frihagen F, Madsen JE, Nordsletten L, Aas E. Hemiarthroplasty compared to internal fixation with percutaneous cannulated screws as treatment of displaced femoral neck fractures in the elderly: cost-utility analysis performed alongside a randomized, controlled trial. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. 2012;23(6):1711-1719. doi:10.1007/s00198-011-1772-1

Section 4: Risk of Bias Assessments for All Included Trials

Study ID	Intervention 1	Intervention 2	Intervention 3	Sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting
Baker 2006	THA	Cemented HA	NA	High	High	High	High	Low	High
Blomfeldt 2005	Uncemented HA	Screw IF	NA	High	Low	High	High	High	Low
Blomfeldt 2007	THA	Cemented HA	NA	High	High	High	Low	Low	Low
Cadossi 2013	THA	HA	NA	High	Low	High	High	High	Low
Cao 2014	THA	Screw IF	NA	High	Low	High	High	High	High
Chammout 2012	THA	Screw IF	NA	High	High	High	High	Low	Low
Chammout 2019	THA	Cemented HA	NA	Low	Low	Low	Low	Low	Low
Davison 2001	Cemented HA	SHS	NA	Low	Low	High	High	Low	Low
deAngelis 2011	Cemented HA	Uncemented HA	NA	Low	Low	High	High	Low	Low
Dorr 1986	THA	Cemented HA	Uncemented HA	High	High	High	High	High	High
El-Abed 2005	Uncemented HA	SHS	NA	High	High	High	High	High	Low
Emery 1991	Cemented HA	Uncemented HA	NA	High	High	High	High	Low	Low
FAITH 2017	Screw IF	SHS	NA	Low	Low	Low	Low	Low	Low
FAITH-2 2020	Screw IF	SHS	NA	Low	Low	Low	Low	Low	Low
Fernandez 2022	Cemented HA	Uncemented HA	NA	Low	Low	Low	Low	Low	Low
Figved 2009	Cemented HA	Uncemented HA	NA	Low	Low	High	Low	Low	Low
Frihagen 2013	Cemented HA	Screw IF	NA	Low	Low	High	High	Low	Low
HEALTH 2019	THA	Uncemented HA	Cemented HA	Low	Low	Low	Low	Low	Low
Hedbeck 2013	Cemented HA	Screw IF	NA	High	Low	High	High	High	Low
Inngul 2015	Cemented HA	Uncemented HA	NA	Low	Low	High	Low	High	Low
Iorio 2019	THA	Uncemented HA	NA	High	High	High	High	Low	Low
Johansson 2014	THA	Screw IF	NA	Low	Low	High	High	Low	Low
Keating 2006	Cemented HA	THA	NA	Low	Low	Low	Low	Low	Low

Linde 1986	Screw IF	SHS	NA	High	High	High	High	High	High
Macaulay 2008	THA	HA	NA	Low	Low	High	High	Low	Low
Madsen 1987	Screw IF	SHS	NA	High	High	High	High	Low	High
Moerman 2017	Cemented HA	Uncemented HA	NA	Low	Low	Low	High	High	Low
Mohabey 2017	Cemented HA	Uncemented HA	NA	High	High	High	High	High	High
Morvin 2020	Cemented HA	Uncemented HA	NA	High	Low	Low	High	High	Low
Mouzopolous 2008	THA	HA	SHS	High	High	High	High	Low	Low
Parker 2002	Uncemented HA	Cemented HA	NA	High	Low	High	High	Low	Low
Parker 2019	THA	Cemented HA	NA	High	Low	High	Low	Low	Low
Puolakka 2001	Cemented HA	Screw IF	NA	High	Low	High	High	High	High
Ravikumar 2000	THA	Uncemented HA	SHS	High	High	High	High	High	Low
Roden 2003	Screw IF	HA	NA	High	Low	High	High	Low	Low
Rogmark 2002	THA	HA	NA	Low	Low	High	High	Low	Low
Santini 2005	Cemented HA	Uncemented HA	NA	High	High	High	High	Low	Low
Sharma 2016	THA	HA	NA	High	High	High	High	High	Low
Sonaje 2017	THA	Cemented HA	NA	High	High	High	High	High	Low
Talsnes 2013	Cemented HA	Uncemented HA	NA	Low	Low	High	Low	High	Low
Taylor 2012	Cemented HA	Uncemented HA	NA	Low	Low	Low	Low	Low	Low
Tidermark 2003	THA	Screw IF	NA	High	Low	High	Low	Low	Low
van den Bekerom 2010	Cemented HA	THA	NA	Low	Low	High	High	High	Low
Vidovic 2013	Cemented HA	Uncemented HA	NA	High	Low	High	High	High	Low

IF: Internal Fixation, HA: Hemi-Arthroplasty, SHS: Sliding Hip Screw, THA: Total Hip Arthroplasty

Section 5: GRADE Assessment for all Comparisons

Comparison	Direct	Direct GRADE	Indirect	Indirect GRADE	Network	Network GRADE	Direct Weight	I^2
Mortality								
Cemented HA vs Screw IF	0.99 (0.77, 1.27)	Low[1,3]	0.82 (0.60, 1.11)	Low	0.92 (0.75, 1.11)	Low	0.60	0%
Cemented HA vs SHS	1.08 (0.65, 1.78)	Low[3,5]	1.03 (0.77, 1.37)	Very Low	1.04 (0.81, 1.34)	Low	0.26	NA
Cemented HA vs THA	**0.80 (0.67, 0.95)**	Moderate[1]	1.07 (0.82, 1.39)	Low	0.87 (0.75, 1.00)	Moderate[8]	0.71	0%
Cemented vs Uncemented HA	0.91 (0.83, 1.01)	Moderate[3]	0.76 (0.55, 1.07)	Moderate	**0.90 (0.82, 0.99)**	Moderate	0.92	0%
Screw IF vs SHS	0.80 (0.43, 1.52)	Low[3,5]	1.24 (0.89, 1.73)	Very Low	1.13 (0.85, 1.52)	Low	0.21	NA
Screw IF vs THA	1.21 (0.83, 1.75)	Low[1,3]	0.84 (0.64, 1.09)	Low	0.95 (0.76, 1.17)	Low	0.34	0%
Screw IF vs Uncemented HA	1.08 (0.59, 1.97)	Very Low[1,3,5]	0.97 (0.78, 1.21)	Low	0.98 (0.80, 1.21)	Low	0.12	NA
SHS vs THA	0.93 (0.62, 1.39)	Very Low[1,3,5]	0.77 (0.55, 1.10)	Low	0.84 (0.64, 1.09)	Low	0.43	0%
SHS vs Uncemented HA	0.76 (0.52, 1.11)	Very Low[1,3,5]	0.97 (0.68, 1.36)	Low	0.87 (0.67, 1.12)	Low	0.45	34.4%
THA vs Uncemented HA	0.95 (0.71, 1.26)	Moderate[3]	1.08 (0.89, 1.32)	Moderate	1.04 (0.88, 1.22)	Moderate	0.32	0%
Reoperation								
Cemented HA vs Screw IF	**0.18 (0.05, 0.61)**	Moderate[1]	**0.30 (0.13, 0.70)**	Moderate	**0.25 (0.13, 0.51)**	Low[8]	0.32	0%
Cemented HA vs SHS	**0.06 (0.01, 0.36)**	Moderate[5]	**0.30 (0.13, 0.71)**	Moderate	**0.22 (0.10, 0.49)**	Low[8]	0.17	NA
Cemented HA vs THA	1.02 (0.48, 2.17)	Moderate[3]	1.69 (0.62, 4.58)	Moderate	1.23 (0.67, 2.24)	Moderate	0.64	0%
Cemented vs Uncemented HA	1.19 (0.64, 2.22)	Low[2,3]	**0.24 (0.08, 0.76)**	Moderate	0.83 (0.48, 1.43)	Low[7]	0.77	82.4%
Screw IF vs SHS	0.76 (0.29, 2.03)	Moderate[3]	1.07 (0.35, 3.29)	Moderate	0.88 (0.42, 1.85)	Moderate	0.57	61.1%
Screw IF vs THA	**4.83 (1.94, 12.03)**	Moderate[1]	**4.86 (1.82, 12.98)**	Moderate	**4.84 (2.48, 9.45)**	Low[8]	0.54	57.7%
Screw IF vs Uncemented HA	2.50 (0.41, 15.32)	Very Low[1,3,5]	**3.45 (1.52, 7.83)**	Moderate	**3.26 (1.54, 6.89)**	Low[7]	0.17	NA
SHS vs THA	**6.51 (1.60, 26.42)**	Low[1,5]	**5.06 (1.94, 13.15)**	Moderate	**5.48 (2.49, 12.07)**	Low[8]	0.32	0%
SHS vs Uncemented HA	1.40 (0.45, 4.35)	Very Low[1,3,5]	**8.90 (3.02, 26.21)**	Moderate	**3.69 (1.69, 8.06)**	Low[7]	0.48	0%
THA vs Uncemented HA	0.50 (0.18, 1.38)	Moderate[3]	0.85 (0.34, 2.11)	Low	0.67 (0.34, 1.32)	Moderate	0.44	36.1%
Hip-Related Complications								
Cemented HA vs Screw IF	**0.32 (0.15, 0.69)**	Moderate[1]	**0.47 (0.28, 0.79)**	Very Low	**0.42 (0.28, 0.64)**	Low[8]	0.31	9.3%
Cemented HA vs SHS	**0.12 (0.04, 0.37)**	Moderate[5]	**0.45 (0.26, 0.75)**	Moderate	**0.35 (0.22, 0.56)**	Very Low[7,8]	0.19	NA
Cemented HA vs THA	0.89 (0.56, 1.41)	Moderate[3]	0.60 (0.33, 1.10)	Low	0.77 (0.53, 1.11)	Moderate	0.64	0%
Cemented vs Uncemented HA	0.72 (0.51, 1.02)	Moderate[3]	**0.42 (0.20, 0.85)**	Moderate	**0.65 (0.47, 0.89)**	Moderate	0.80	41.0%
Screw IF vs SHS	0.91 (0.54, 1.53)	Moderate[3]	0.72 (0.35, 1.47)	Very Low	0.84 (0.55, 1.28)	Moderate	0.65	44.1%
Screw IF vs THA	1.51 (0.86, 2.64)	Very Low[1,2,3]	**2.30 (1.26, 4.22)**	Moderate	**1.83 (1.21, 2.77)**	Moderate	0.54	89.6%
Screw IF vs Uncemented HA	1.29 (0.40, 4.15)	Very Low[1,3,5]	1.59 (0.97, 2.59)	Low	1.54 (0.98, 2.42)	Low	0.15	NA
SHS vs THA	1.83 (0.70, 4.83)	Very Low[1,3,5]	**2.33 (1.32, 4.10)**	Moderate	**2.19 (1.35, 3.57)**	Low[7]	0.25	NA
SHS vs Uncemented HA	1.40 (0.69, 2.84)	Very Low[1,3,5]	**2.31 (1.21, 4.42)**	Low	**1.84 (1.14, 2.97)**	Low	0.45	0%
THA vs Uncemented HA	0.82 (0.43, 1.55)	Moderate[3]	0.85 (0.50, 1.47)	Low	0.84 (0.56, 1.27)	Moderate	0.42	47.3%
Short-term Function								
Cemented HA vs Screw IF	**8.10 (0.38, 15.82)**	Low[1,5]	**11.96 (2.07, 21.85)**	Low	**9.56 (3.47, 15.65)**	Very Low[8]	0.62	NA

Comparison	Direct	Direct GRADE	Indirect	Indirect GRADE	Network	Network GRADE	Direct Weight	I^2
Cemented HA vs SHS	NA	NA	**25.36 (11.44, 39.28)**	Low	**25.36 (11.44, 39.28)**	Low	0	NA
Cemented HA vs THA	-2.65 (-6.26, 0.96)	Moderate[3]	-0.46 (-9.29, 8.38)	Low	-2.34 (-5.68, 1.00)	Moderate	0.86	0%
Cemented vs Uncemented HA	2.66 (-0.15, 5.46)	Low[2,3]	-2.66 (-17.55, 12,24)	Low	2.47 (-0.28, 5.23)	Low	0.97	81.2%
Screw IF vs SHS	**15.80 (3.28, 28.32)**	Moderate[5]	NA	NA	**15.80 (3.28, 28.32)**	Low[8]	1.00	NA
Screw IF vs THA	**-14.00 (-23.26, -4.74)**	Low[1,5]	**-10.14 (-18.61, -1.68)**	Low	**-11.90 (-18.15, -5.65)**	Very Low[8]	0.46	NA
Screw IF vs Uncemented HA	NA	NA	**-7.09 (-13.69, -0.48)**	Low	**-7.09 (-13.69, -0.48)**	Very Low[8]	0	NA
SHS vs THA	NA	NA	**-27.70 (-41.69, -13.71)**	Low	**-27.70 (-41.69, -13.71)**	Low	0	NA
SHS vs Uncemented HA	NA	NA	**-22.89 (-37.04, -8.73)**	Low	**-22.89 (-37.04, -8.73)**	Low	0	NA
THA vs Uncemented HA	1.50 (-5.18, 8.18)	Low[3,5]	**6.72 (1.65, 11.78)**	Low	**4.81 (0.78, 8.85)**	Low	0.37	NA
Long-term Function								
Cemented HA vs Screw IF	3.30 (-4.59, 11.19)	Very Low[1,3,5]	2.40 (-4.15, 8.95)	Very Low	2.76 (-2.28, 7.80)	Very Low	0.41	NA
Cemented HA vs SHS	2.70 (-4.95, 10.35)	Moderate[3]	5.64 (-0.21, 11.49)	Low	4.56 (-0.09, 9.20)	Moderate	0.37	NA
Cemented HA vs THA	**-5.07 (-8.01, -2.12)**	Moderate[2]	**-5.05 (-11.84, 1.74)**	Low	**-5.06 (-7.77, -2.36)**	Low[8]	0.84	71.9%
Cemented vs Uncemented HA	0.46 (-2.61, 3.53)	Low[2,3]	-5.84 (-19.51, 7.83)	Low	-0.16 (-2.84, 3.16)	Low	0.95	80.2%
Screw IF vs SHS	2.69 (-3.76, 9.15)	Moderate[3]	0.43 (-7.52, 8.37)	Very Low	1.79 (-3.22, 6.80)	Moderate	0.60	0%
Screw IF vs THA	-9.00 (-18.49, 0.49)	Very Low[1,3,5]	**-7.35 (-13.40, -1.31)**	Very Low	**-7.83 (-12.93, -2.73)**	Very Low	0.29	NA
Screw IF vs Uncemented HA	NA	NA	-2.61 (-8.40, 3.19)	Very Low	-2.61 (-8.40, 3.19)	Very Low	0	NA
SHS vs THA	**-10.10 (-16.98, -3.22)**	Low[1,5]	**-9.23 (-15.46, -3.00)**	Moderate	**-9.62 (-14.24, -5.00)**	Moderate	0.45	NA
SHS vs Uncemented HA	NA	NA	-4.40 (-9.84, 1.04)	Low	-4.40 (-9.84, 1.04)	Low	0	NA
THA vs Uncemented HA	2.61 (-4.24, 9.46)	Low[3,5]	**6.38 (1.81, 10.96)**	Low	**5.22 (1.42, 9.02)**	Low	0.31	NA
Short-term Quality of Life								
Cemented HA vs Screw IF	**0.09 (0.01, 0.16)**	Low[1,5]	0.07 (0.00, 0.15)	Low	**0.08 (0.03, 0.13)**	Very Low[8]	0.51	0%
Cemented HA vs SHS	NA	NA	0.07 (-0.02, 0.16)	Low	0.07 (-0.02, 0.16)	Low	0	NA
Cemented HA vs THA	-0.03 (-0.07, 0.01)	Moderate[3]	0.02 (-0.07, 0.10)	Low	-0.02 (-0.06, 0.01)	Low[7]	0.81	0%
Cemented vs Uncemented HA	0.02 (-0.01, 0.05)	Moderate[3]	0.03 (-0.08, 0.15)	Low	0.02 (-0.01, 0.05)	Moderate	0.92	76.8%
Screw IF vs SHS	-0.01 (-0.08, 0.06)	Low[3,5]	NA	NA	-0.01 (-0.08, 0.06)	Low	1.00	NA
Screw IF vs THA	**-0.12 (-0.21, -0.03)**	Low[1,5]	**-0.09 (-0.17, -0.01)**	Low	**-0.10 (-0.16, -0.05)**	Very Low[8]	0.44	NA
Screw IF vs Uncemented HA	0.00 (-0.13, 0.13)	Very Low[1,3,5]	**-0.07 (-0.14, -0.01)**	Low	-0.06 (-0.12, 0.00)	Low	0.20	NA
SHS vs THA	NA	NA	-0.09 (-0.18, 0.00)	Low	-0.09 (-0.18, 0.00)	Very Low[8]	0	NA
SHS vs Uncemented HA	NA	NA	-0.05 (-0.14, 0.04)	Low	-0.05 (-0.14, 0.04)	Very Low[8]	0	NA
THA vs Uncemented HA	0.00 (-0.06, 0.06)	Low[3,5]	**0.09 (0.03, 0.15)**	Moderate	0.04 (0.00, 0.09)	Moderate	0.51	NA
Long-term Quality of Life								
Cemented HA vs Screw IF	0.12 (0.00, 0.25)	Very Low[1,2,5]	-0.11 (-0.23, 0.01)	Very Low	0.01 (-0.08, 0.09)	Very Low[7]	0.50	0%
Cemented HA vs SHS	NA	NA	-0.06 (-0.18, 0.06)	Very Low	-0.06 (-0.18, 0.06)	Very Low	0	NA
Cemented HA vs THA	-0.07 (-0.15, 0.01)	Low[2,3]	-0.04 (-0.21, 0.14)	Low	-0.06 (-0.13, 0.01)	Low	0.83	0%
Cemented vs Uncemented HA	-0.02 (-0.10, 0.06)	Low[2,3]	**0.23 (0.07, 0.39)**	Low	0.04 (-0.04, 0.11)	Very Low[7]	0.78	83.9%
Screw IF vs SHS	0.01 (-0.11, 0.12)	Low[2,3]	**-0.27 (-0.47, -0.08)**	Very Low	-0.06 (-0.16, 0.03)	Very Low[7]	0.74	0%
Screw IF vs THA	-0.11 (-0.28, 0.06)	Very Low[1,2,5]	-0.05 (-0.17, 0.07)	Very Low	-0.07 (-0.17, 0.03)	Very Low	0.33	NA

Comparison	Direct	Direct GRADE	Indirect	Indirect GRADE	Network	Network GRADE	Direct Weight	I^2
Screw IF vs Uncemented HA	0.15 (-0.03, 0.33)	Very Low[1,2,3,5]	-0.01 (-0.12, 0.09)	Very Low	0.03 (-0.06, 0.12)	Very Low	0.26	NA
SHS vs THA	NA	NA	0.00 (-0.13, 0.12)	Very Low	0.00 (-0.13, 0.12)	Very Low	0	NA
SHS vs Uncemented HA	**0.24 (0.08, 0.40)**	Very Low[1,2,5]	-0.04 (-0.19, 0.12)	Very Low	0.09 (-0.02, 0.21)	Very Low[7]	0.47	NA
THA vs Uncemented HA	0.01 (-0.15, 0.17)	Low[3,5]	**0.14 (0.03, 0.25)**	Low	**0.10 (0.01, 0.19)**	Low	0.33	NA

GRADE Assessment: Reasons for downgrading direct evidence:

1. Rated down due to risk of bias
2. Rated down due to inconsistency
3. Rated down for imprecision
4. Rated down due to indirectness
5. Rated down due to publication bias

Reasons for downgrading indirect evidence:

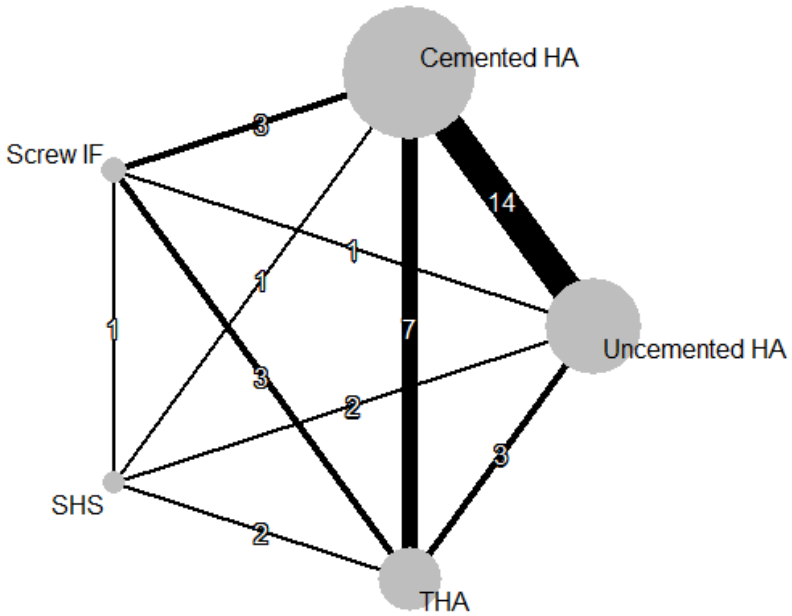
6. Rated down for intransitivity

Reasons for downgrading network evidence:

7. Rated down due to incoherence
8. Rated down due to imprecision (Not rated down if direct/indirect estimate was already rated down for imprecision)

Section 6: Mortality Results

Mortality network diagram



Number of studies: 33
Number of treatments: 5
Number of designs: 12

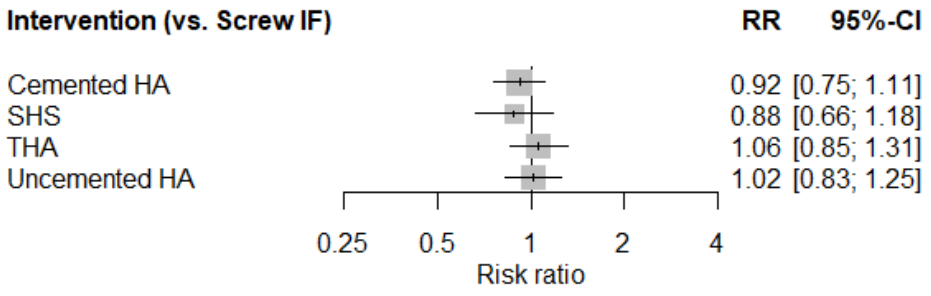
Mortality league table

Cemented HA	0.99 (0.77, 1.27)	1.08 (0.65, 1.78)	**0.80 (0.67, 0.95)**	0.91 (0.83, 1.01)
0.92 (0.75, 1.11)	**Screw IF**	0.80 (0.43, 1.52)	1.21 (0.83, 1.75)	1.08 (0.59, 1.97)
1.04 (0.81, 1.34)	1.13 (0.85, 1.52)	**SHS**	0.93 (0.62, 1.39)	0.76 (0.52, 1.11)
0.87 (0.75, 1.00)	0.95 (0.76, 1.17)	0.84 (0.64, 1.09)	**THA**	0.95 (0.71, 1.26)
0.90 (0.82, 0.99)	0.98 (0.80, 1.21)	0.87 (0.67, 1.12)	1.04 (0.88, 1.22)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

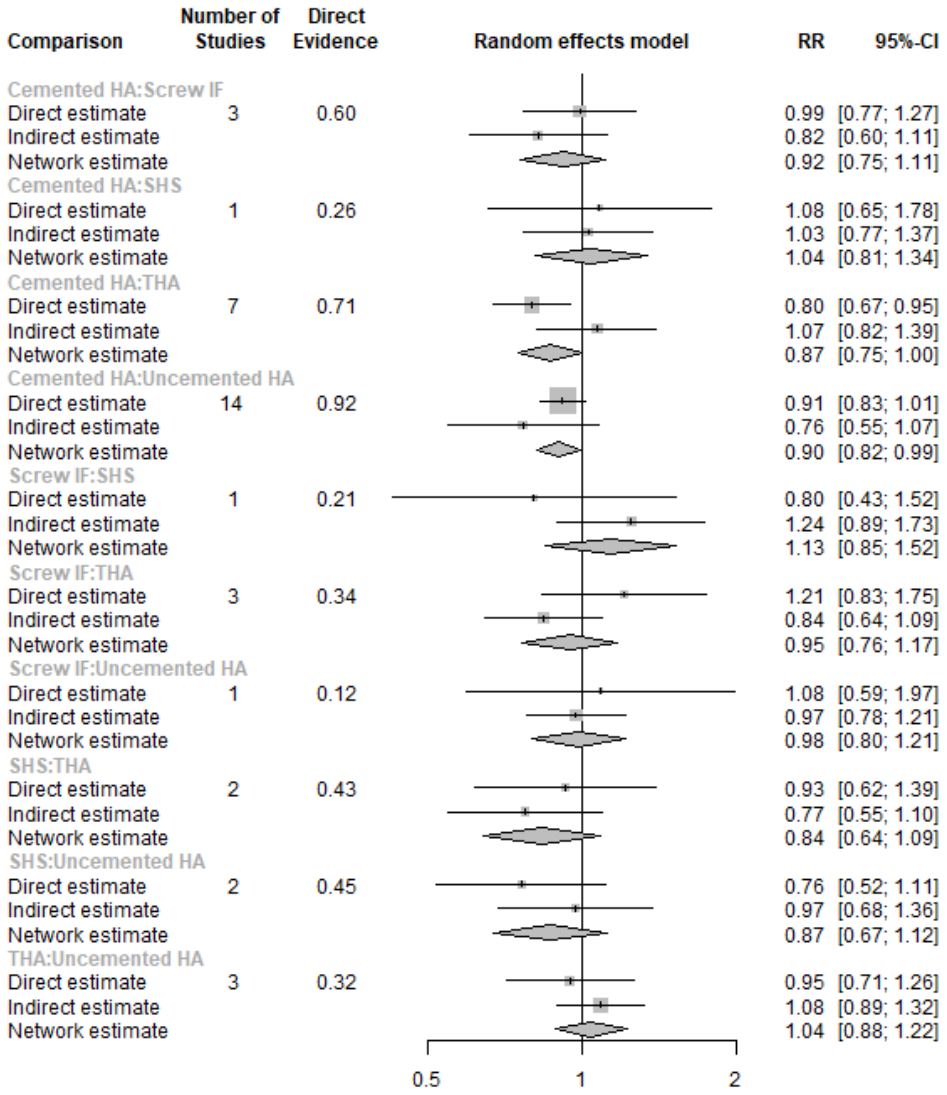
Bold: Statistical Significance

Mortality forest plot



Section 7: Mortality Heterogeneity and Incoherence Assessment

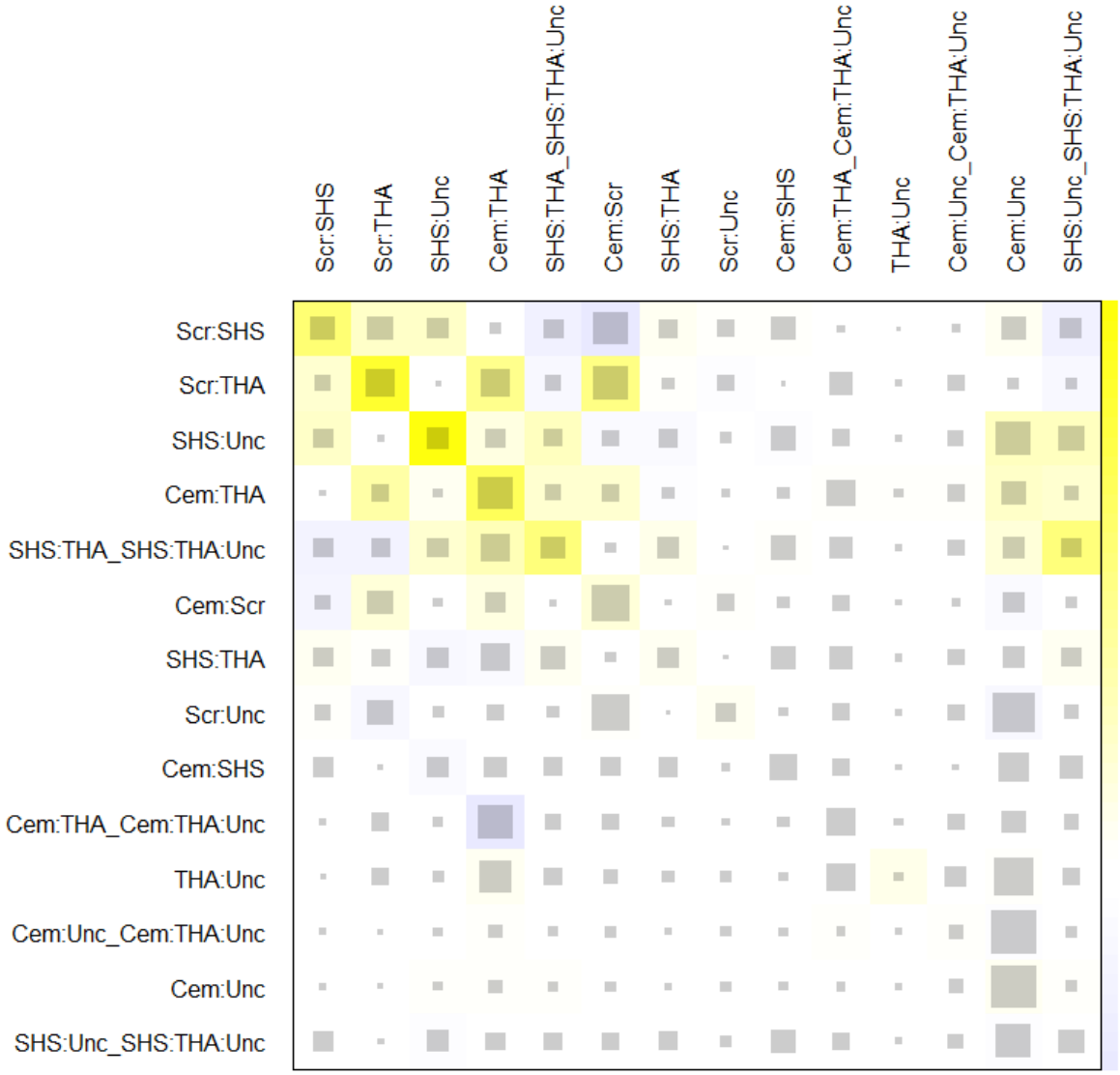
Mortality node-splitting assessment of incoherence



Within-design heterogeneity

Design	Q	p-value
Cemented HA: Screw IF	0.07	0.9658
Cemented HA: THA	4.01	0.5480
Cemented HA: Uncemented HA	5.08	0.9553
Screw IF: THA	0.26	0.8761

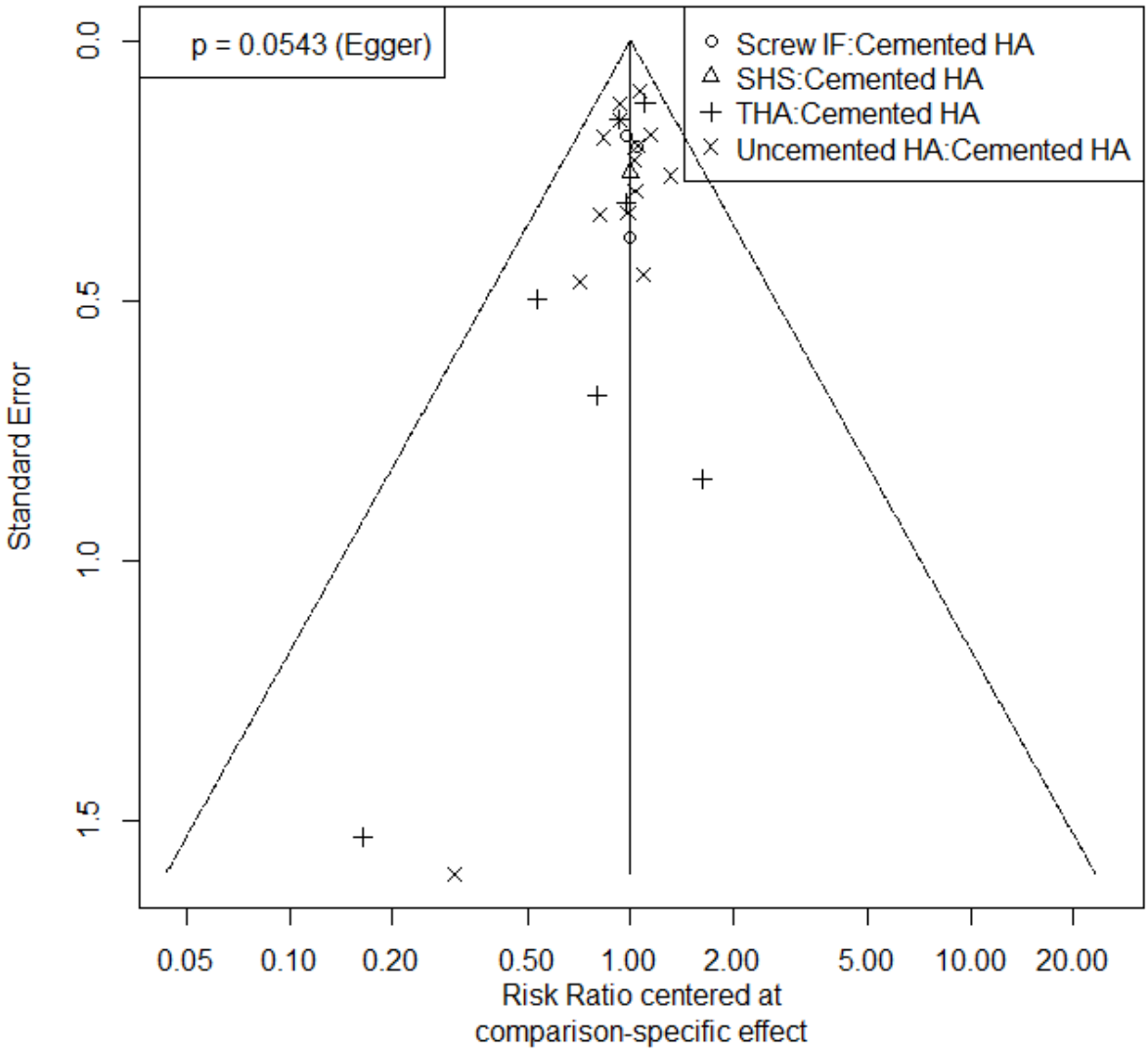
Mortality incoherence heat plot



Design-by-treatment interaction random effects model for incoherence

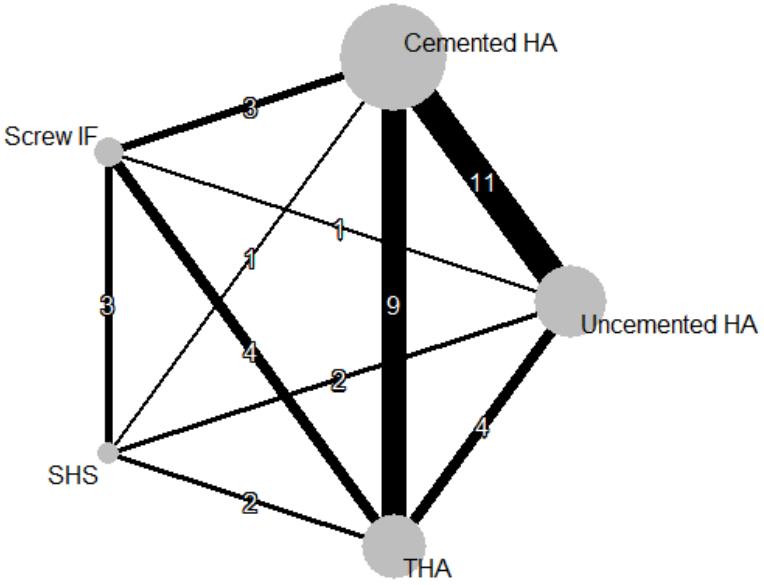
	Q	p-value	Tau^2
Between designs	7.94	0.6345	0

Mortality funnel plot



Section 8: Reoperation Results

Reoperation network diagram



Number of studies = 34
Number of treatments = 5
Number of designs = 12

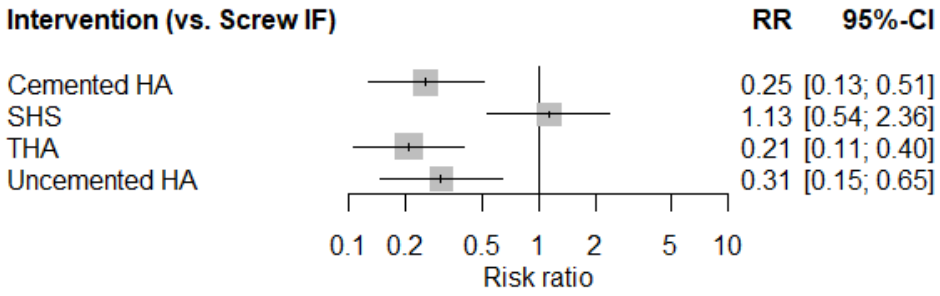
Reoperation league table

Cemented HA	**0.18 (0.05, 0.61)**	**0.06 (0.01, 0.36)**	1.02 (0.48, 2.17)	1.19 (0.64, 2.22)
0.25 (0.13, 0.51)	**Screw IF**	0.76 (0.29, 2.03)	**4.83 (1.94, 12.03)**	2.50 (0.41, 15.32)
0.22 (0.10, 0.49)	0.88 (0.42, 1.85)	**SHS**	**6.51 (1.60, 26.42)**	1.40 (0.45, 4.35)
1.23 (0.67, 2.24)	**4.84 (2.48, 9.45)**	**5.48 (2.49, 12.07)**	**THA**	0.50 (0.18, 1.38)
0.83 (0.48, 1.43)	**3.26 (1.54, 6.89)**	**3.69 (1.69, 8.06)**	0.67 (0.34, 1.32)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

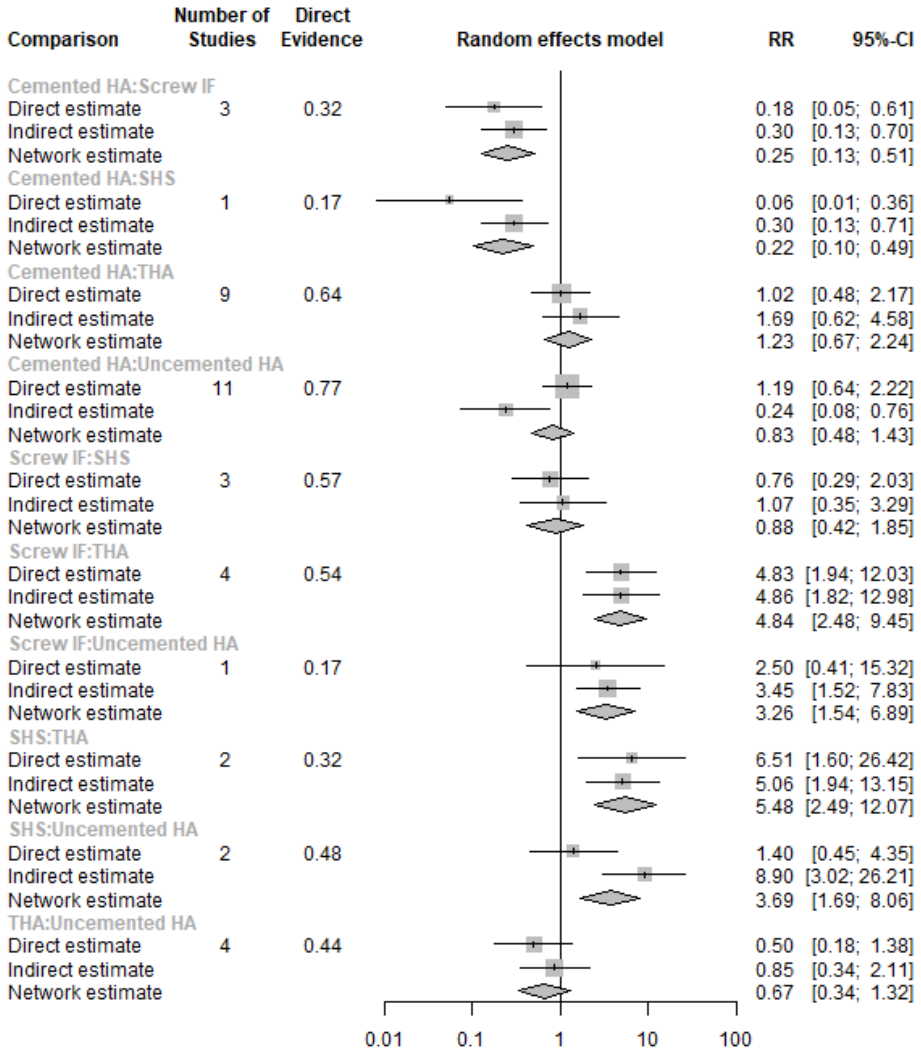
Bold: Statistical Significance

Reoperation forest plot



Section 9: Reoperation Heterogeneity and Incoherence Assessment

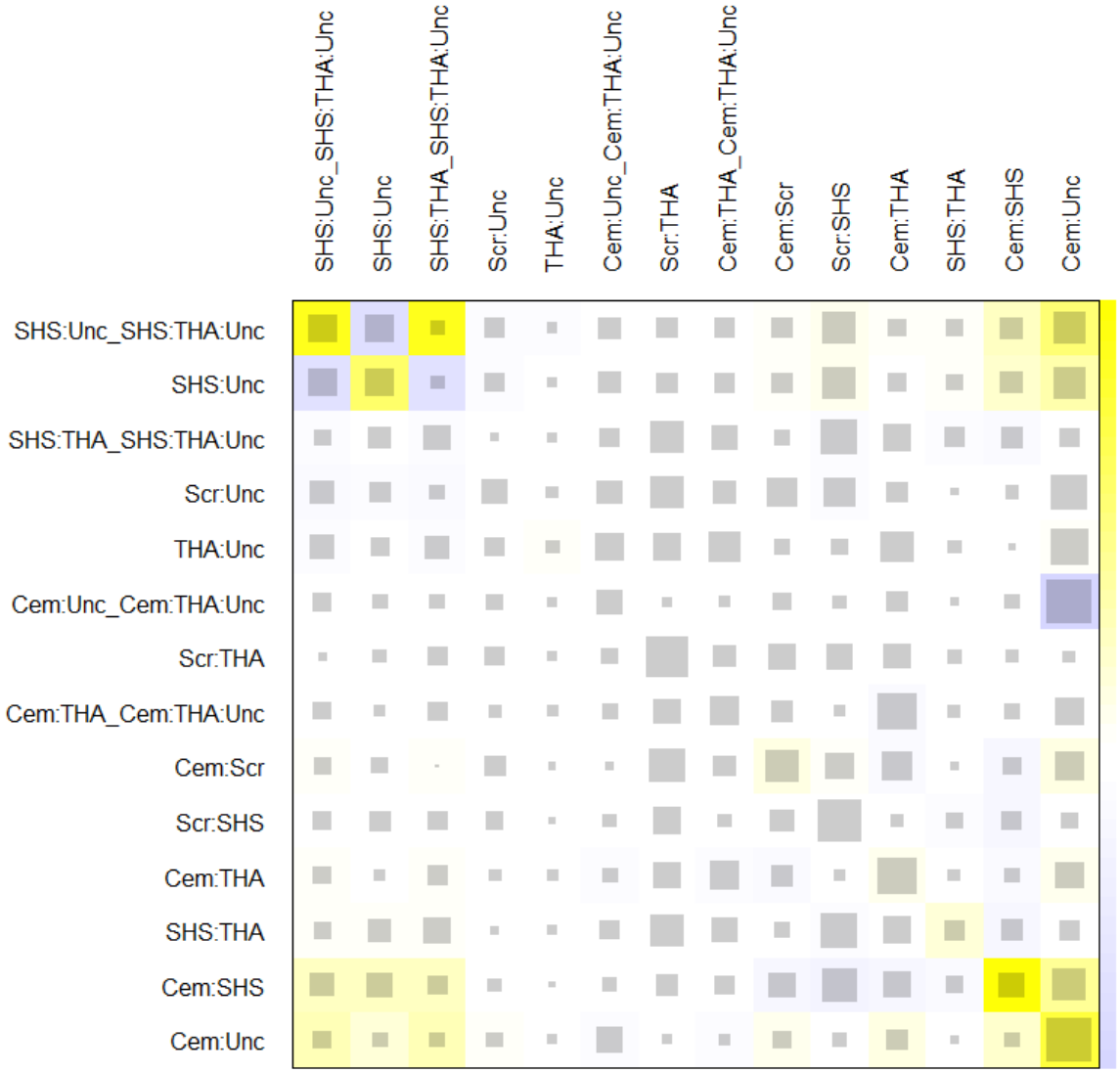
Reoperation node-splitting assessment of incoherence



Within-design heterogeneity

Design	Q	p-value
Cemented HA: Screw IF	0.81	0.6662
Cemented HA: THA	5.50	0.4818
Cemented HA: Uncemented HA	44.22	< 0.0001
Screw IF: SHS	5.14	0.0767
Screw IF: THA	7.10	0.0688
Cemented HA: THA: Uncemented HA	0.32	0.8540

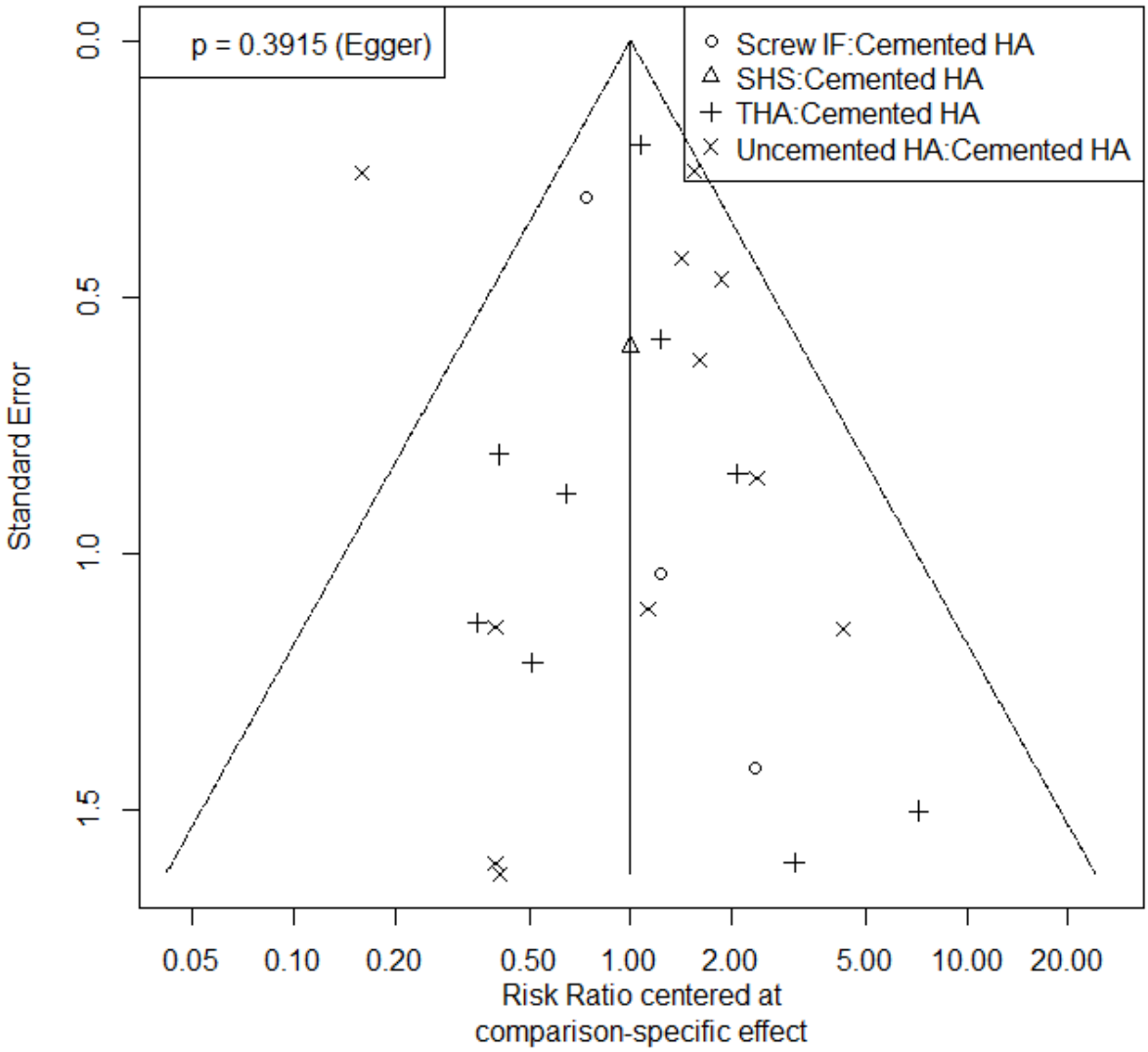
Reoperation incoherence heat plot



Design-by-treatment interaction random effects model for incoherence

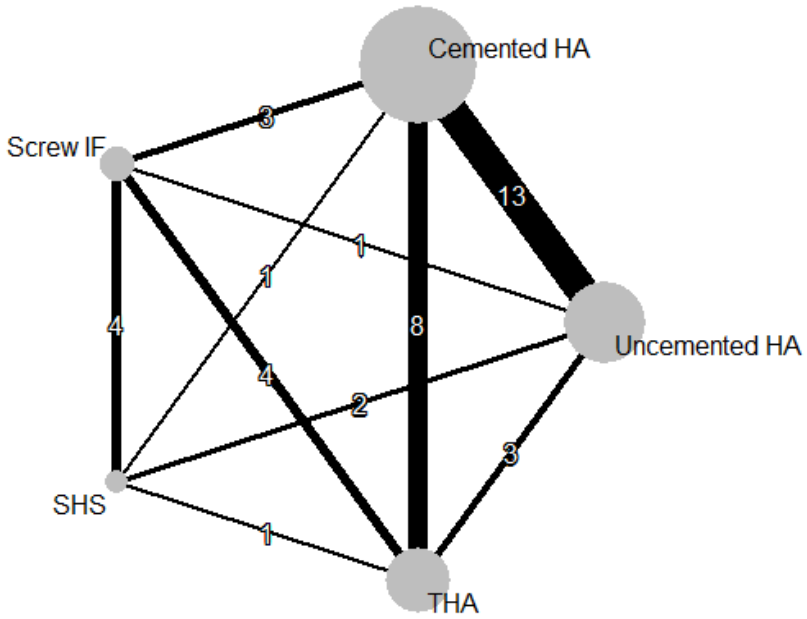
	Q	p-value	Tau^2
Between designs	7.72	0.6558	0.6111

Reoperation funnel plot



Section 10: Complications Results

Complications network diagram



Number of studies = 36
Number of treatments = 5
Number of designs = 11

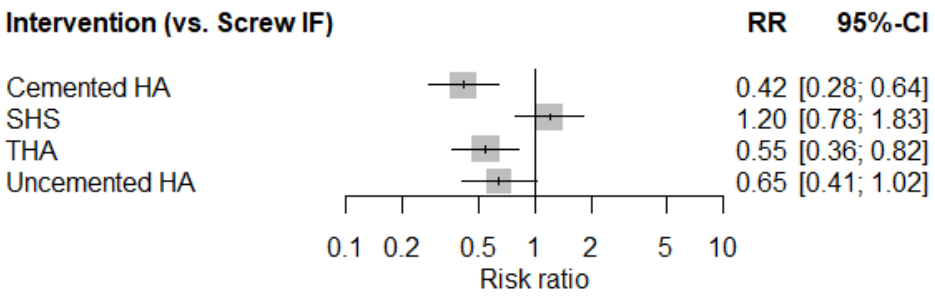
Complications league table

Cemented HA	**0.32 (0.15, 0.69)**	**0.12 (0.04, 0.37)**	0.89 (0.56, 1.41)	0.72 (0.51, 1.02)
0.42 (0.28, 0.64)	**Screw IF**	0.91 (0.54, 1.53)	1.51 (0.86, 2.64)	1.29 (0.40, 4.15)
0.35 (0.22, 0.56)	0.84 (0.55, 1.28)	**SHS**	1.83 (0.70, 4.83)	1.40 (0.69, 2.84)
0.77 (0.53, 1.11)	**1.83 (1.21, 2.77)**	**2.19 (1.35, 3.57)**	**THA**	0.82 (0.43, 1.55)
0.65 (0.47, 0.89)	1.54 (0.98, 2.42)	**1.84 (1.14, 2.97)**	0.84 (0.56, 1.27)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

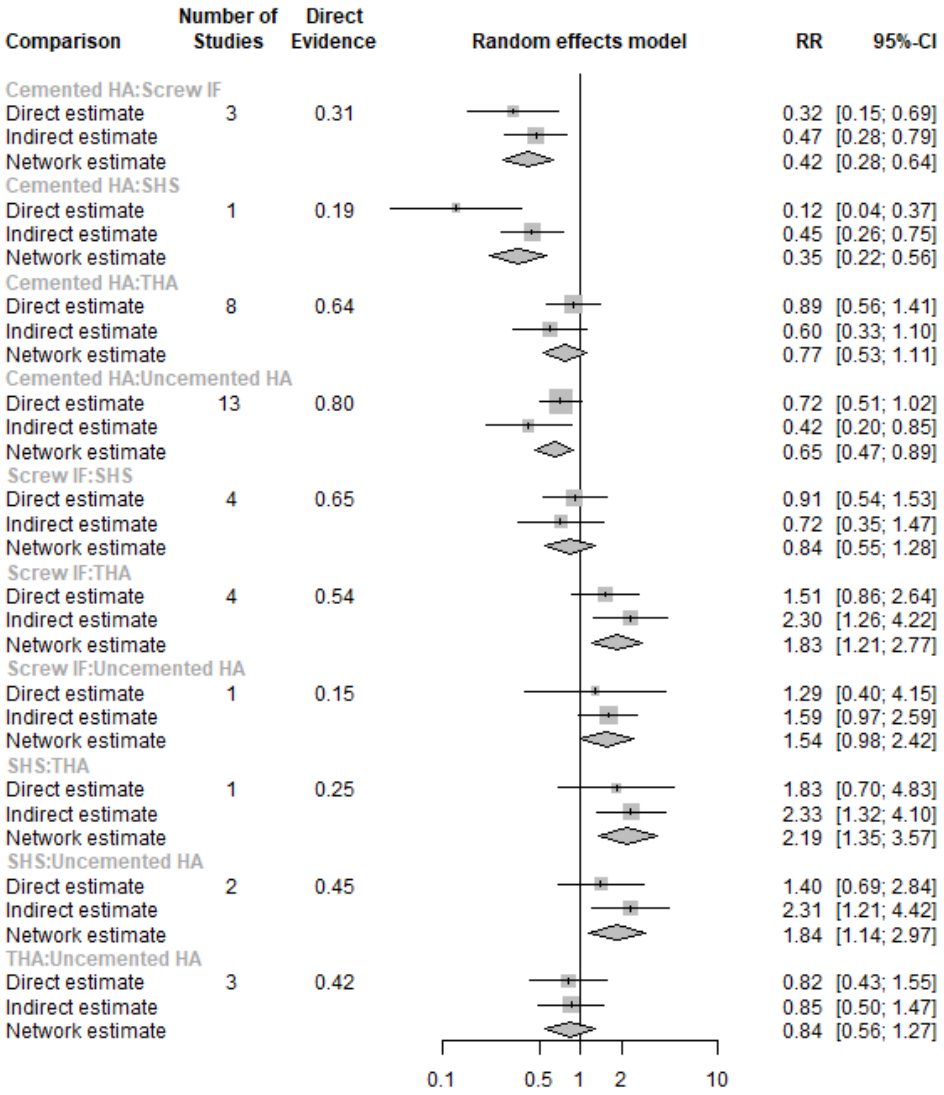
Bold: Statistical Significance

Complications forest plot



Section 11: Complications Heterogeneity and Incoherence Assessment

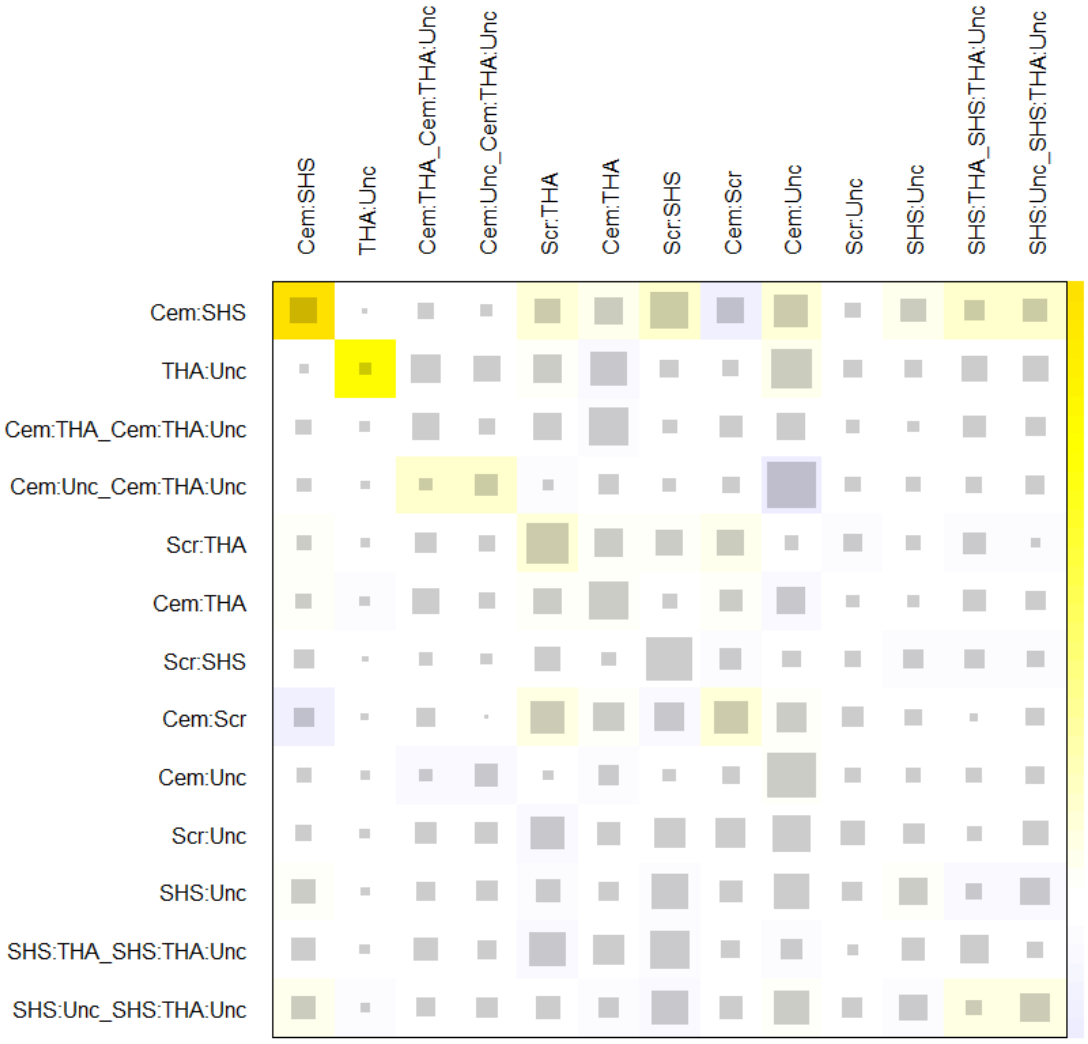
Complications node-splitting assessment of incoherence



Within-design heterogeneity

Design	Q	p-value
Cemented HA: Screw IF	2.21	0.3318
Cemented HA: THA	4.50	0.6087
Cemented HA: Uncemented HA	18.70	0.0667
Screw IF: SHS	5.37	0.1469
Screw IF: THA	28.86	< 0.0001

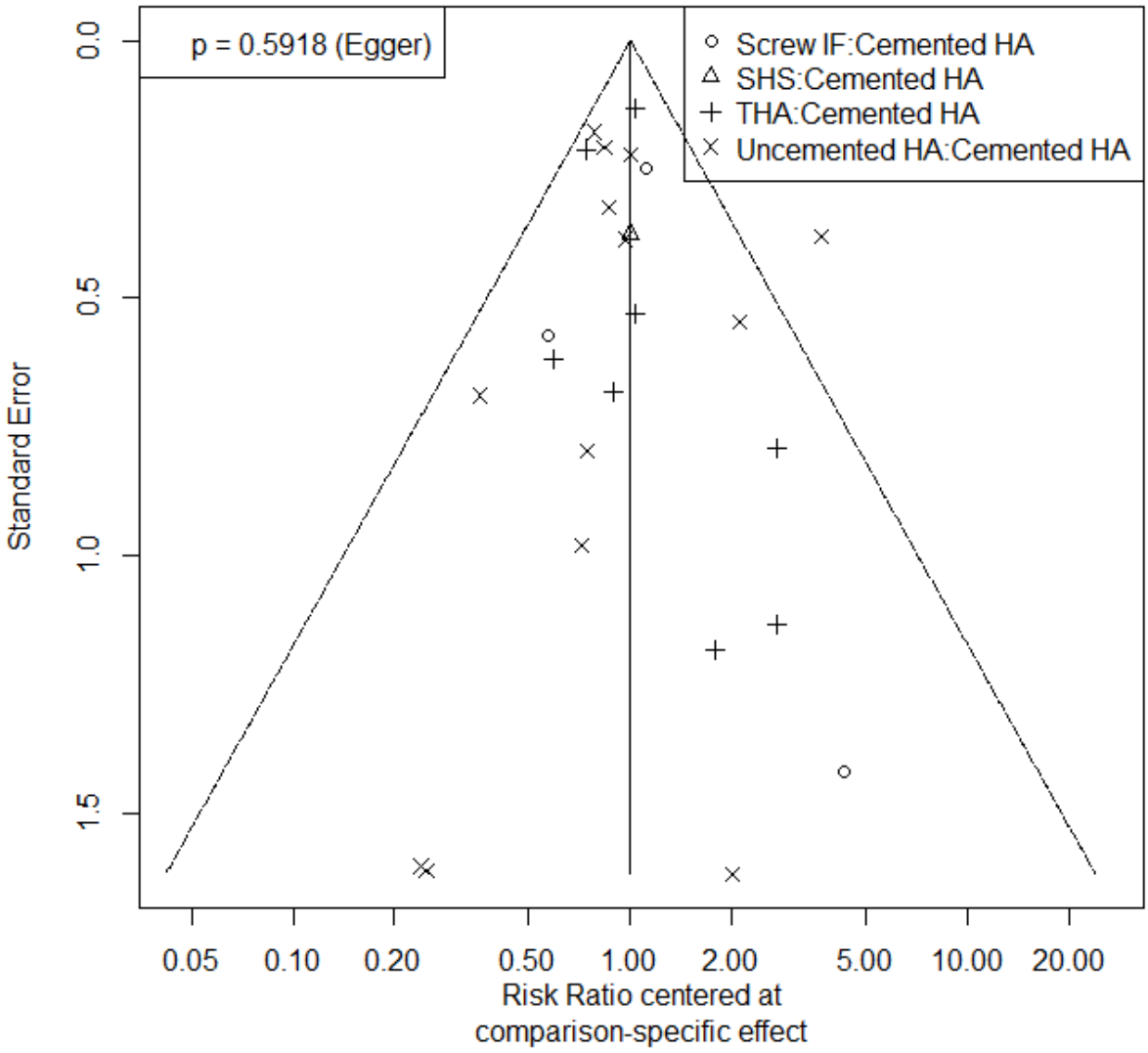
Complications incoherence heat plot



Design-by-treatment interaction random effects model for incoherence

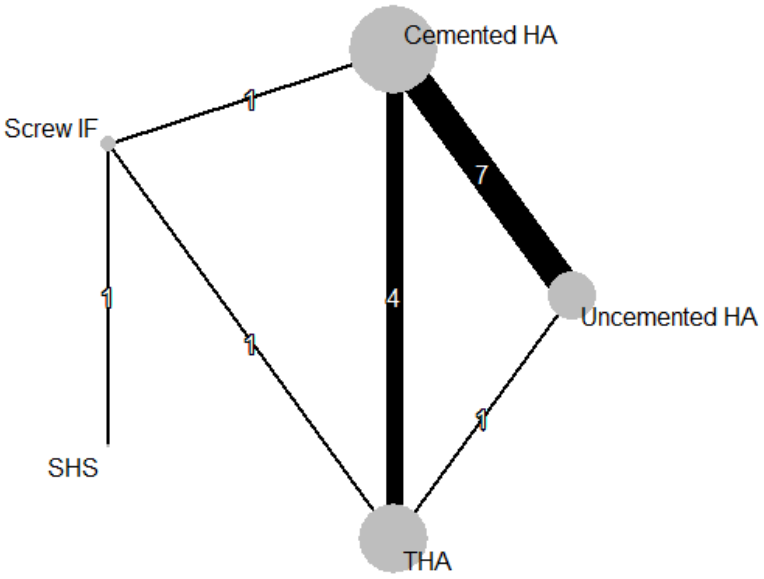
	Q	p-value	Tau^2
Between designs	7.03	0.6337	0.2408

Complications funnel plot



Section 12: Short-term Function Results

Short-term function network diagram



Number of studies = 13
Number of treatments = 5
Number of designs = 6

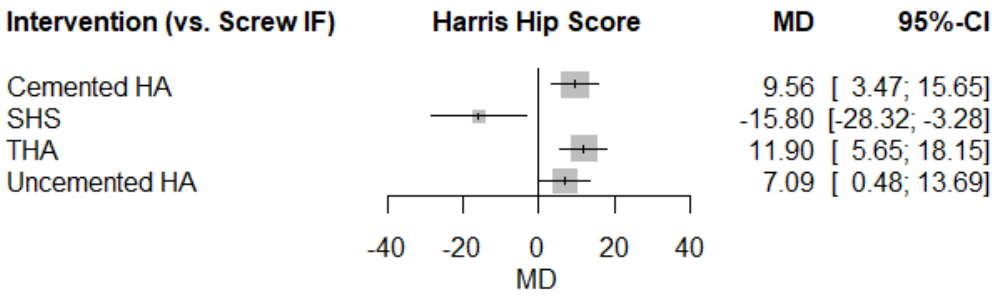
Short-term function league table

Cemented HA	**8.10 (0.38, 15.82)**	.	-2.65 (-6.26, 0.96)	2.66 (-0.15, 5.46)
9.56 (3.47, 15.65)	**Screw IF**	**15.80 (3.28, 28.32)**	**-14.00 (-23.26, -4.74)**	.
25.36 (11.44, 39.28)	**15.80 (3.28, 28.32)**	**SHS**	.	.
-2.34 (-5.68, 1.00)	**-11.90 (-18.15, -5.65)**	**-27.70 (-41.69, -13.71)**	**THA**	1.50 (-5.18, 8.18)
2.47 (-0.28, 5.23)	**-7.09 (-13.69, -0.48)**	**-22.89 (-37.04, -8.73)**	**4.81 (0.78, 8.85)**	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

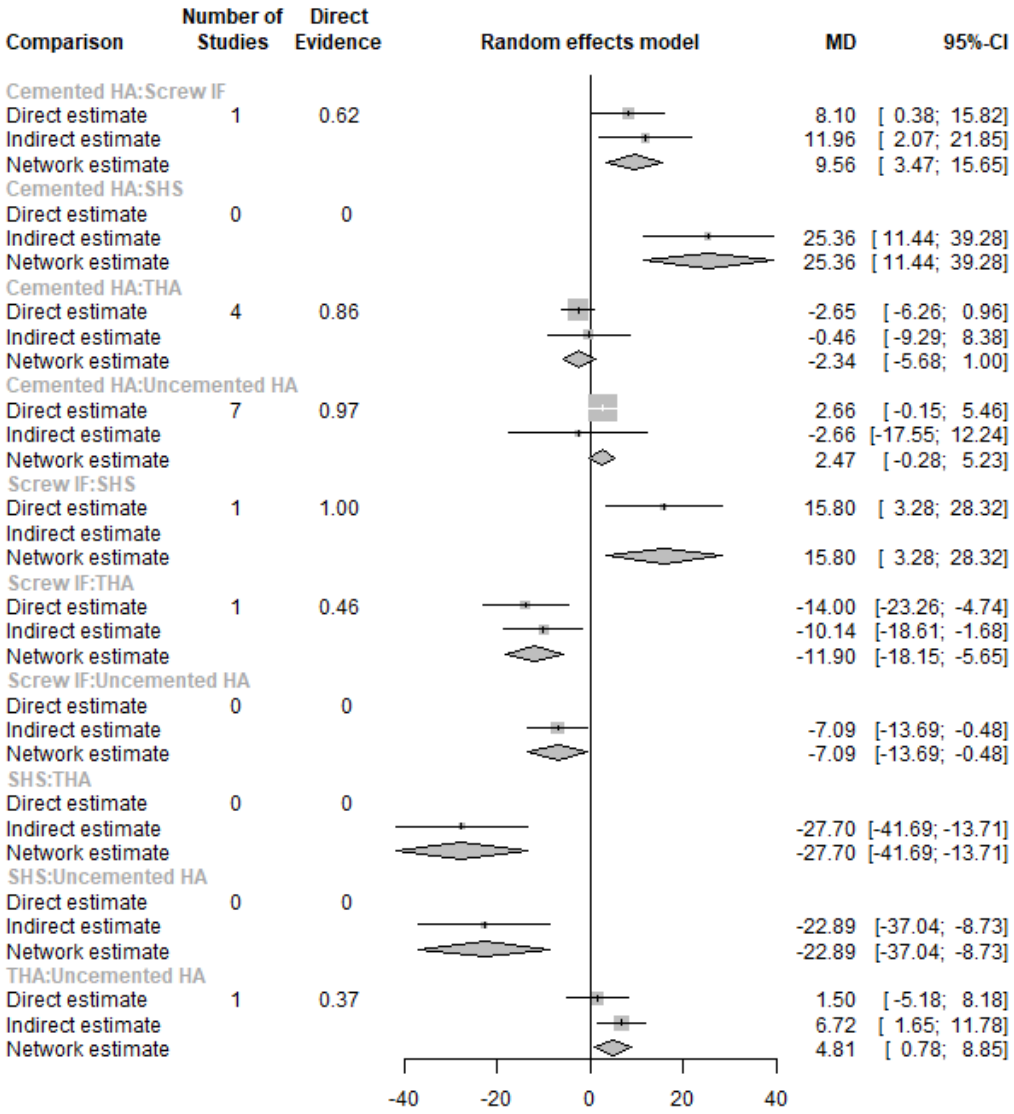
Bold: Statistical Significance

Short-term function forest plot



Section 13: Short-term Function Heterogeneity and Incoherence Assessment

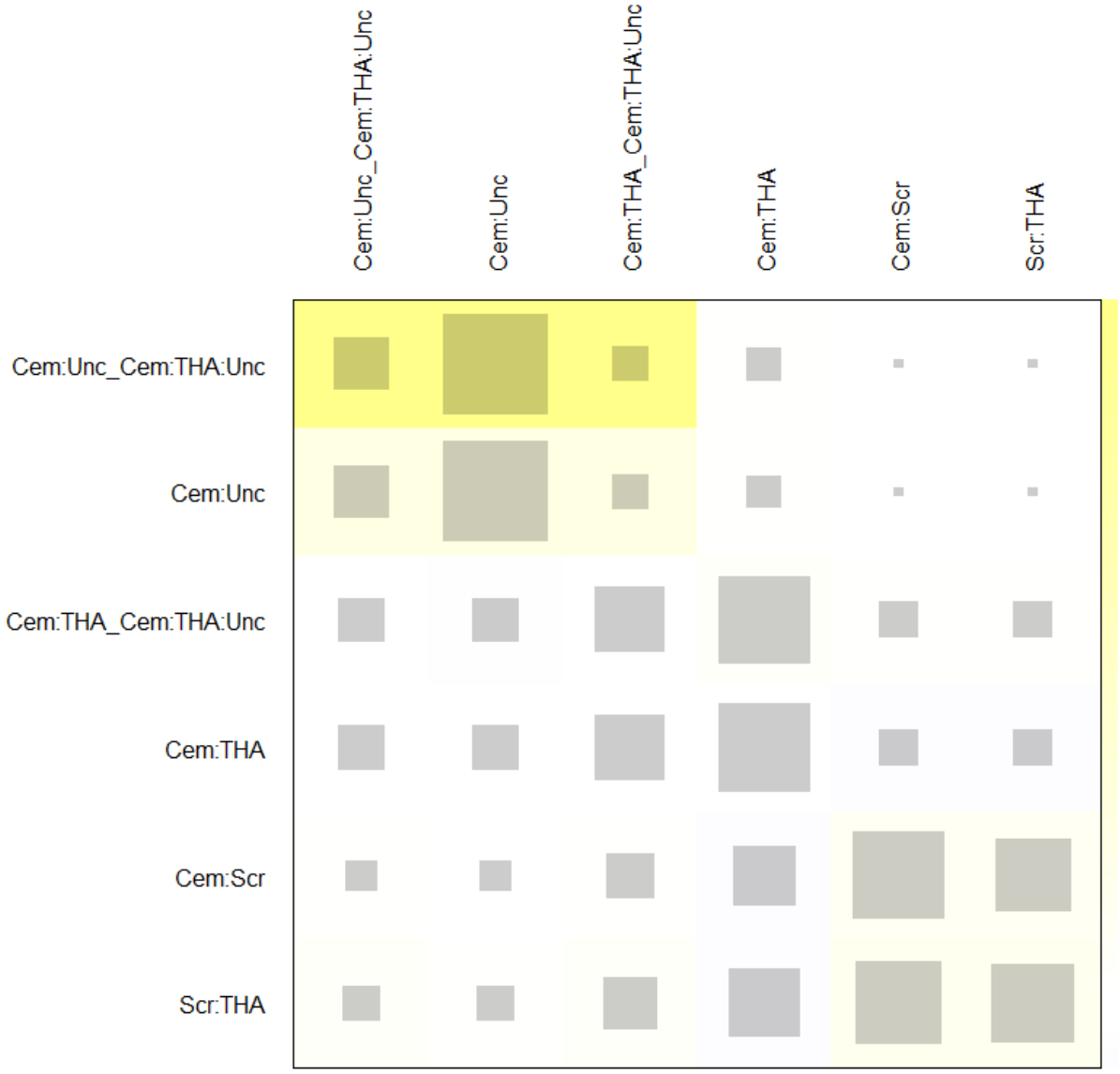
Short-term function node-splitting assessment of incoherence



Within-design heterogeneity

Design	Q	p-value
Cemented HA: THA	1.83	0.4015
Cemented HA: Uncemented HA	26.17	< 0.0001

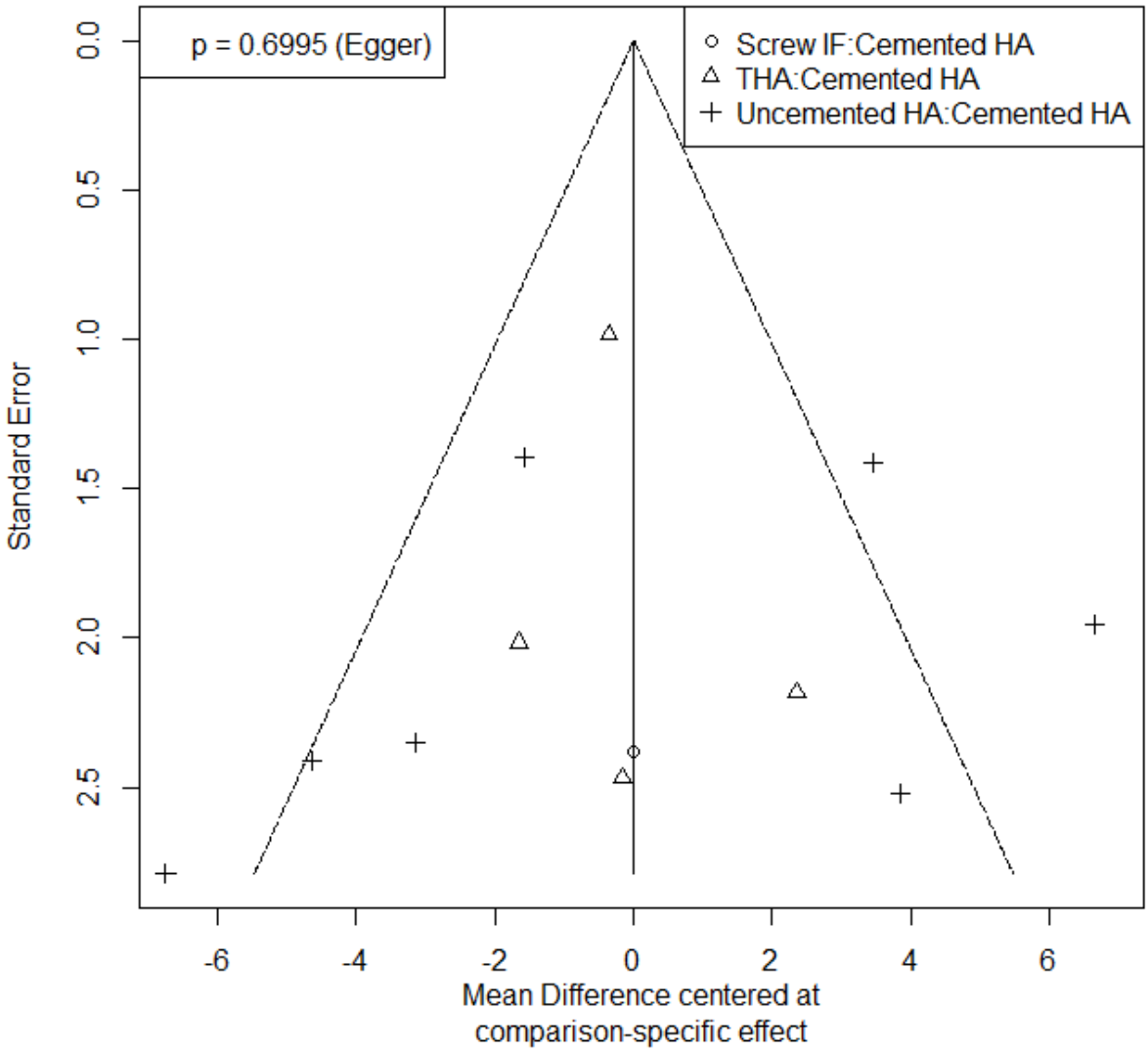
Short-term function incoherence heat plot



Design-by-treatment interaction random effects model for incoherence

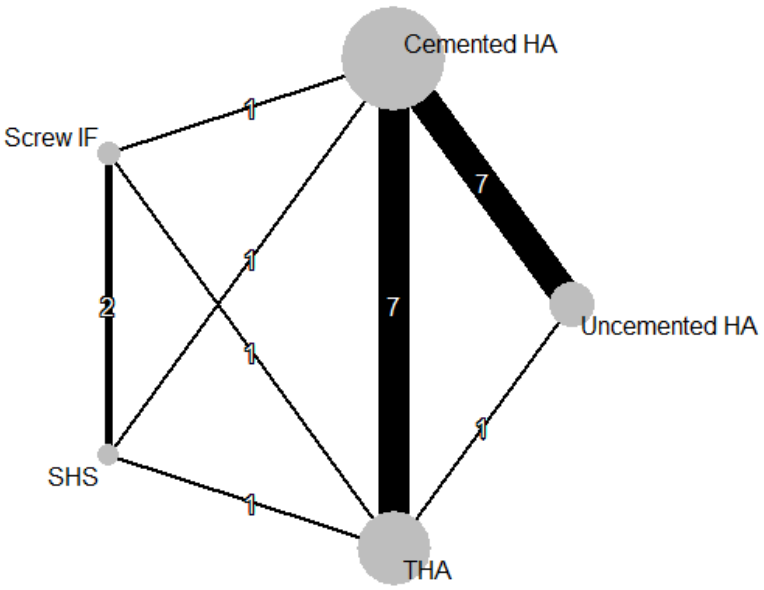
	Q	p-value	Tau^2
Between designs	7.03	0.6838	13.8839

Short-term function funnel plot



Section 14: Long-term Function Results

Long-term function network diagram



Number of studies = 19
Number of treatments = 5
Number of designs = 8

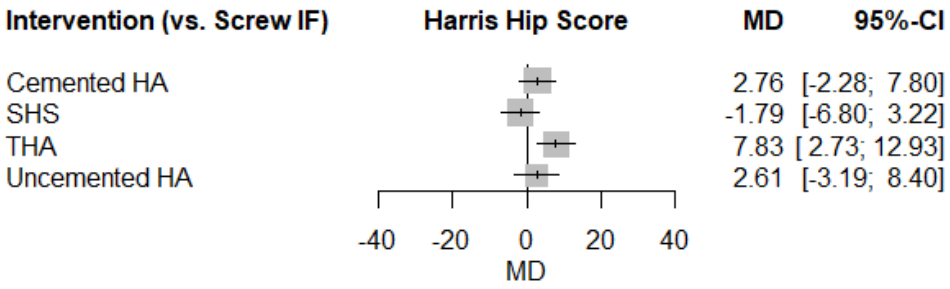
Long-term function league table

Cemented HA	3.30 (-4.59, 11.19)	2.70 (-4.95, 10.35)	**-5.07 (-8.01, -2.12)**	0.46 (-2.61, 3.53)
2.76 (-2.28, 7.80)	**Screw IF**	2.69 (-3.76, 9.15)	-9.00 (-18.49, 0.49)	.
4.56 (-0.09, 9.20)	1.79 (-3.22, 6.80)	**SHS**	**-10.10 (-16.98, -3.22)**	.
-5.06 (-7.77, -2.36)	**-7.83 (-12.93, -2.73)**	**-9.62 (-14.24, -5.00)**	**THA**	2.61 (-4.24, 9.46)
0.16 (-2.84, 3.16)	-2.61 (-8.40, 3.19)	-4.40 (-9.84, 1.04)	**5.22 (1.42, 9.02)**	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

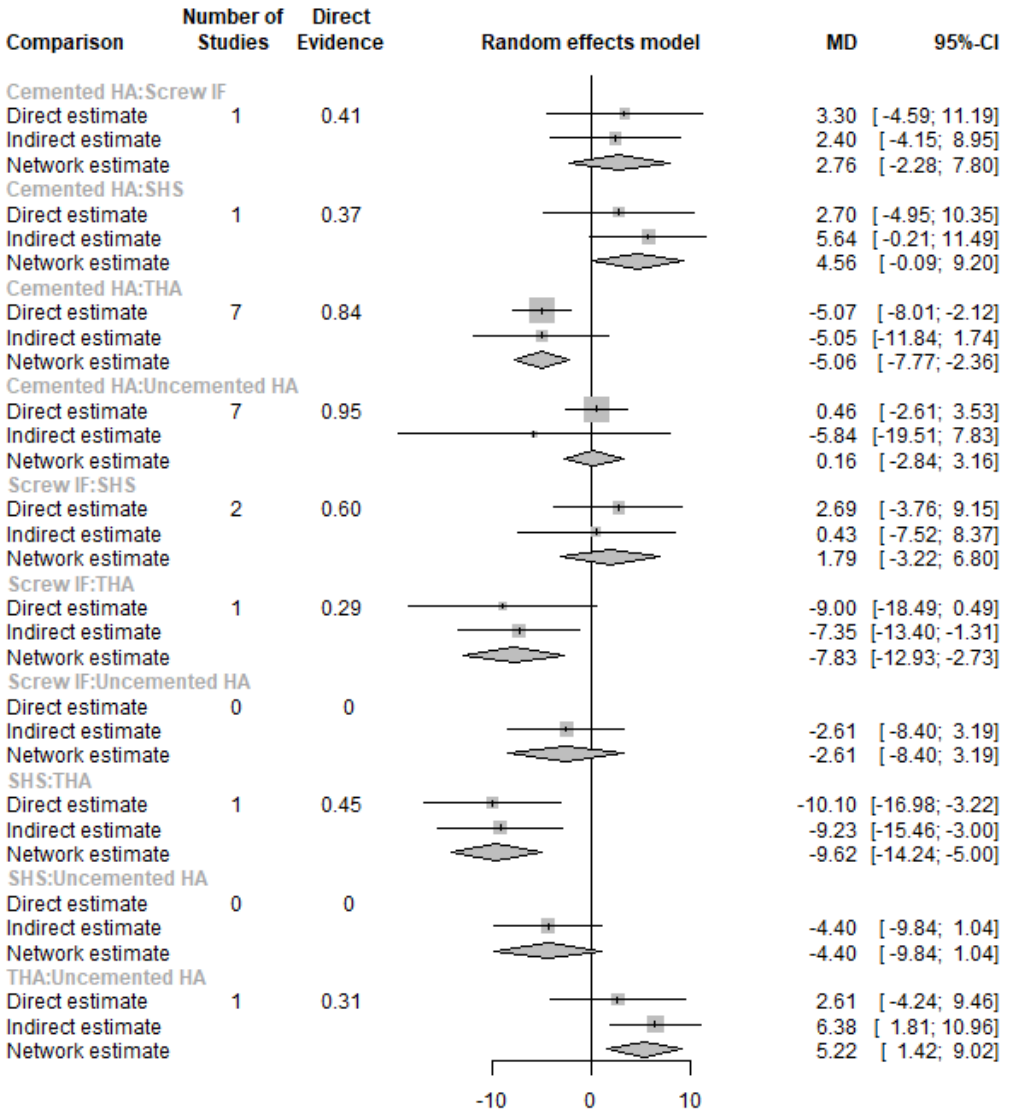
Bold: Statistical Significance

Long-term function forest plot



Section 15: Long-term Function Heterogeneity and Incoherence Assessment

Long-term function node-splitting assessment of incoherence



Within-design heterogeneity

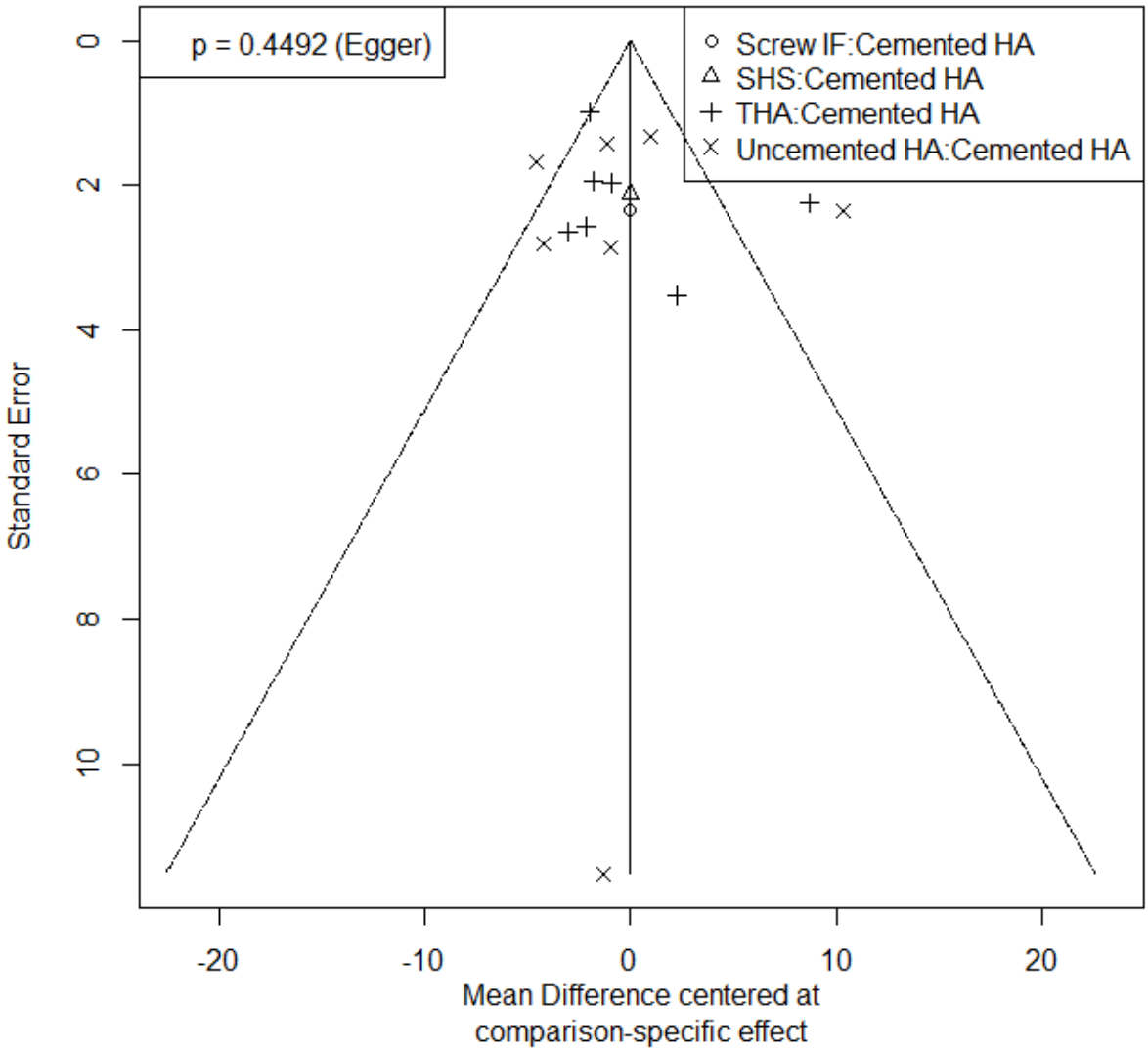
Design	Q	p-value
Cemented HA: THA	18.44	0.0024
Cemented HA: Uncemented HA	29.15	< 0.0001
Screw IF: SHS	0.55	0.4587

Long-term function incoherence heat plot

Design-by-treatment interaction random effects model for incoherence

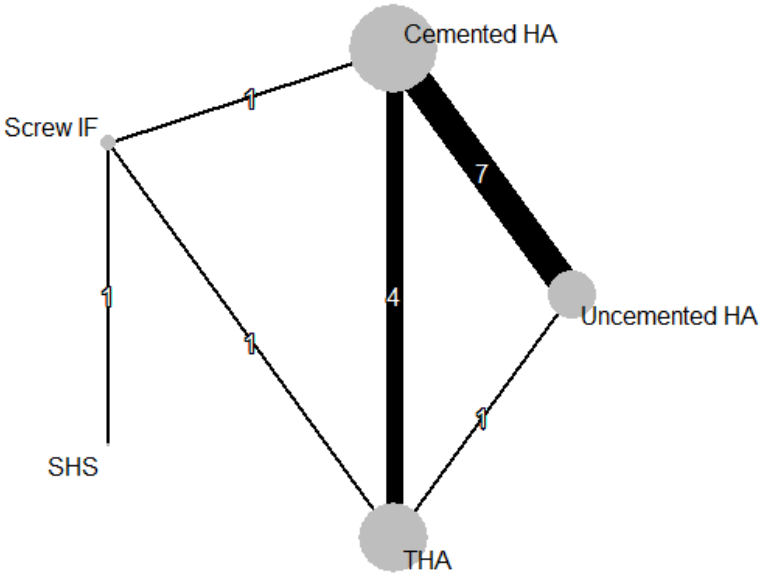
	Q	p-value	Tau^2
Between designs	0.75	0.9803	19.6669

Long-term function funnel plot



Section 16: Short-term Quality of Life Results

Short-term quality of life network diagram



Number of studies = 12
Number of treatments = 5
Number of designs = 7

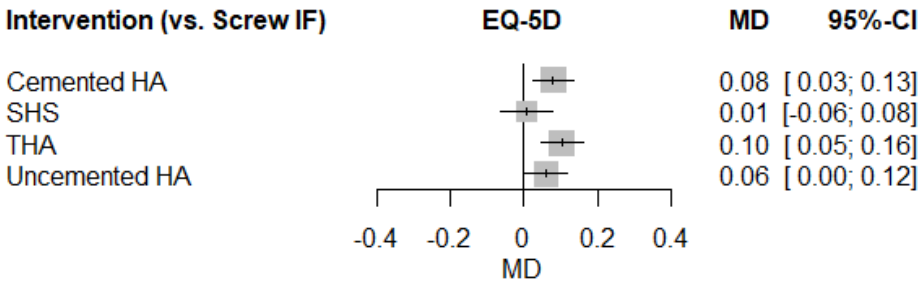
Short-term quality of life league table

Cemented HA	**0.09 (0.01, 0.16)**	.	-0.03 (-0.07, 0.01)	0.02 (-0.01, 0.05)
0.08 (0.03, 0.13)	**Screw IF**	-0.01 (-0.08, 0.06)	**-0.12 (-0.21, -0.03)**	0.00 (-0.13, 0.13)
0.07 (-0.02, 0.16)	-0.01 (-0.08, 0.06)	**SHS**	.	.
-0.02 (-0.06, 0.01)	**-0.10 (-0.16, -0.05)**	-0.09 (-0.18, 0.00)	**THA**	0.00 (-0.06, 0.06)
0.02 (-0.01, 0.05)	-0.06 (-0.12, 0.00)	-0.05 (-0.14, 0.04)	0.04 (0.00, 0.09)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

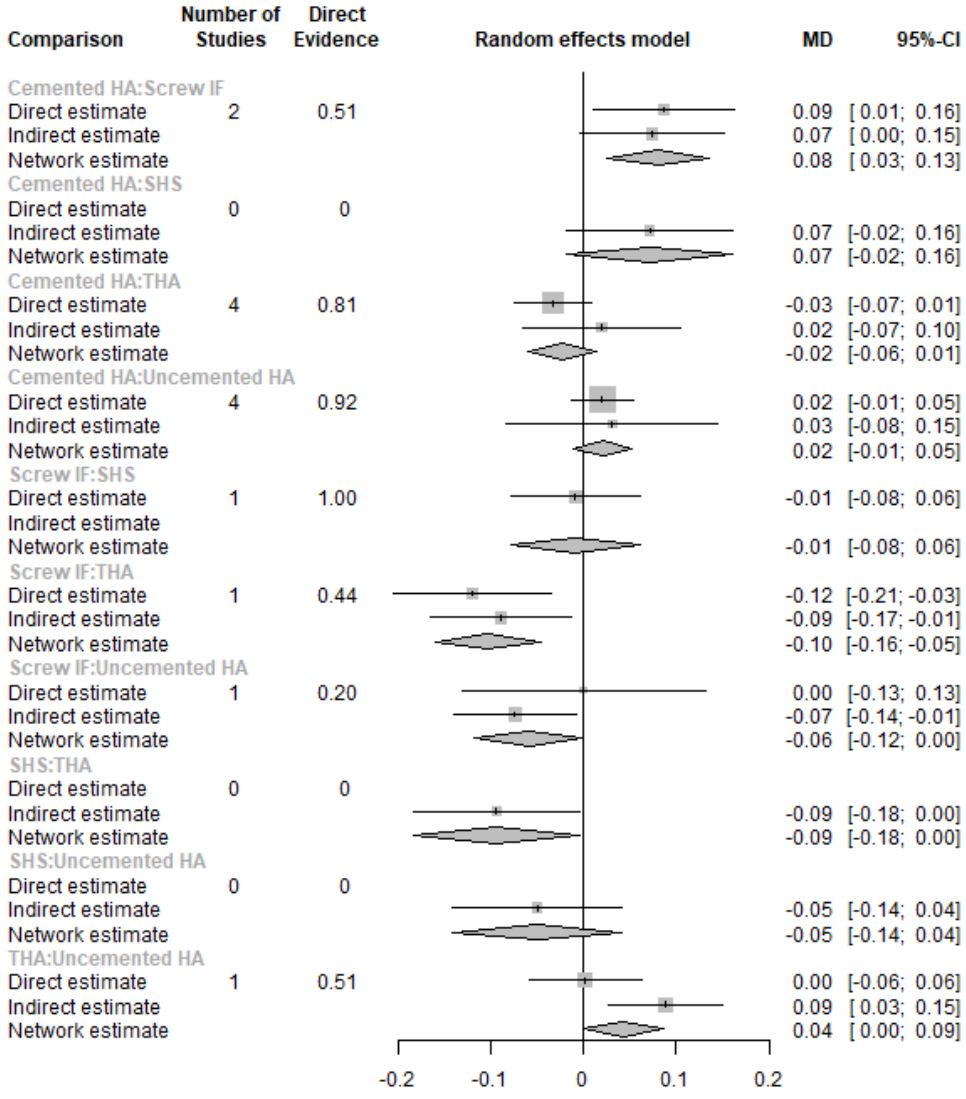
Bold: Statistical Significance

Short-term quality of life forest plot



Section 17: Short-term Quality of Life Heterogeneity and Incoherence Assessment

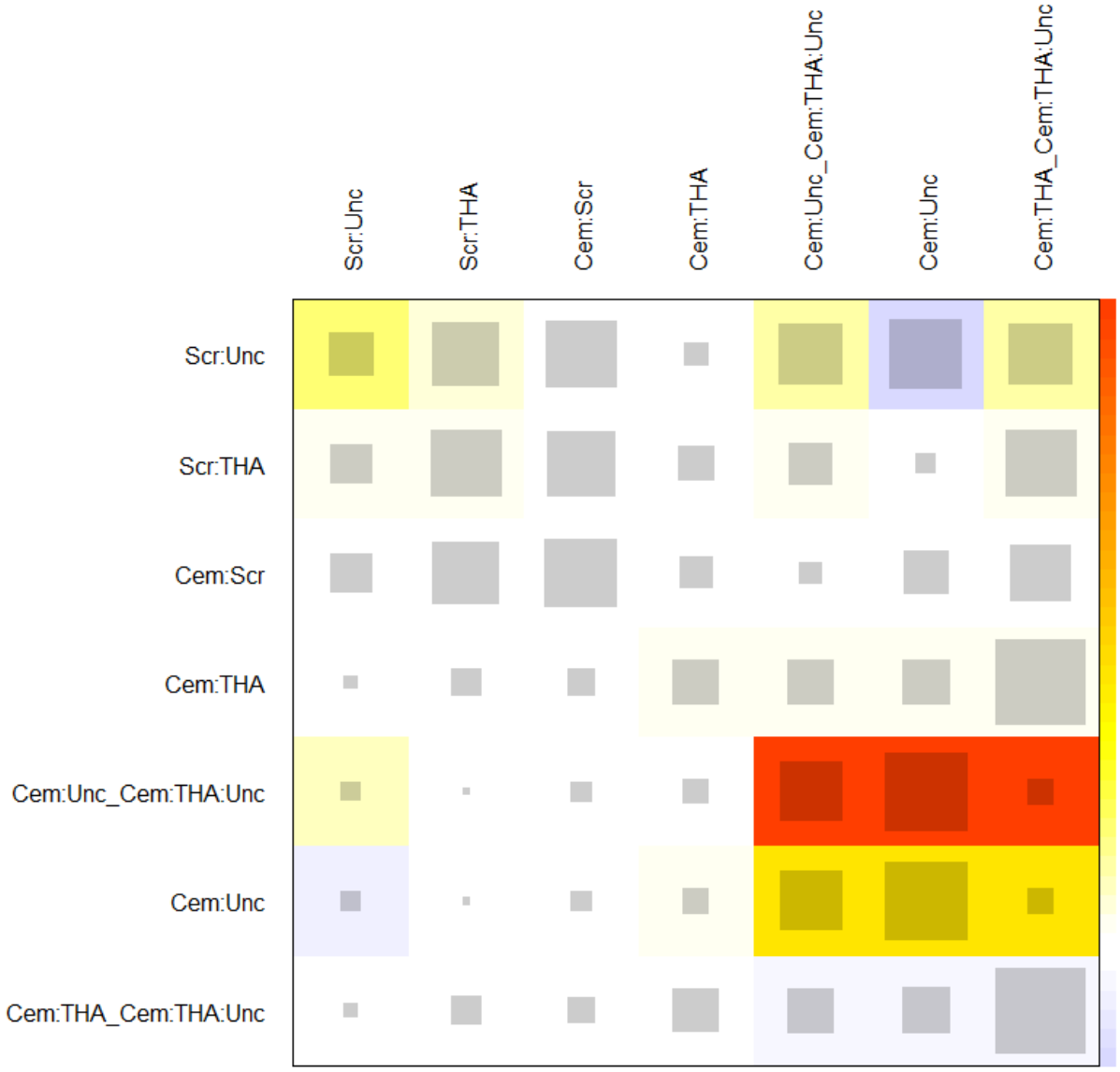
Short-term quality of life node-splitting assessment of incoherence



Within-design heterogeneity

Design	Q	p-value
Cemented HA: Screw IF	0.08	0.7842
Cemented HA: THA	2.47	0.2914
Cemented HA: Uncemented HA	2.73	0.2554

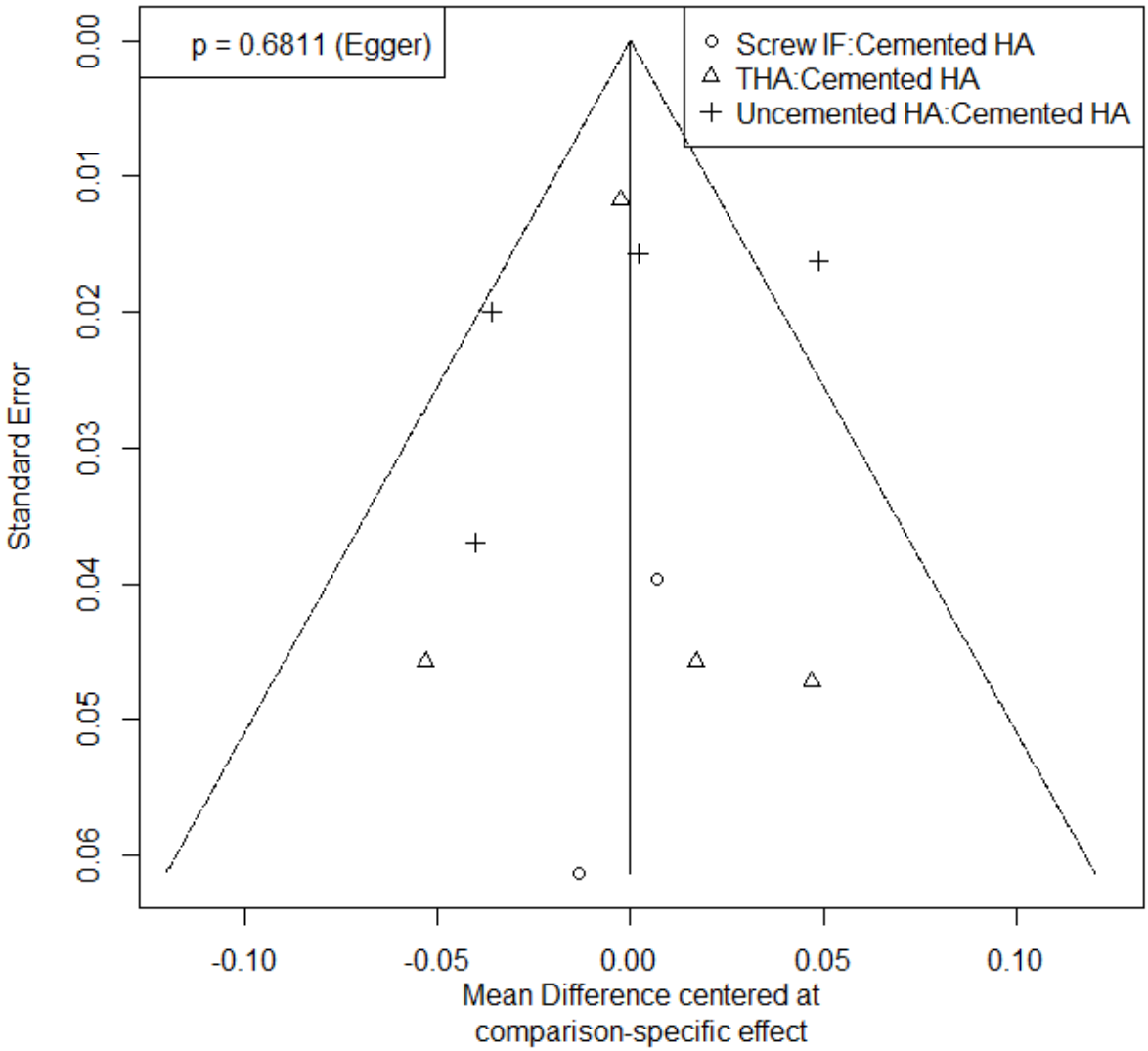
Short-term quality of life incoherence heat plot



Design-by-treatment interaction random effects model for incoherence

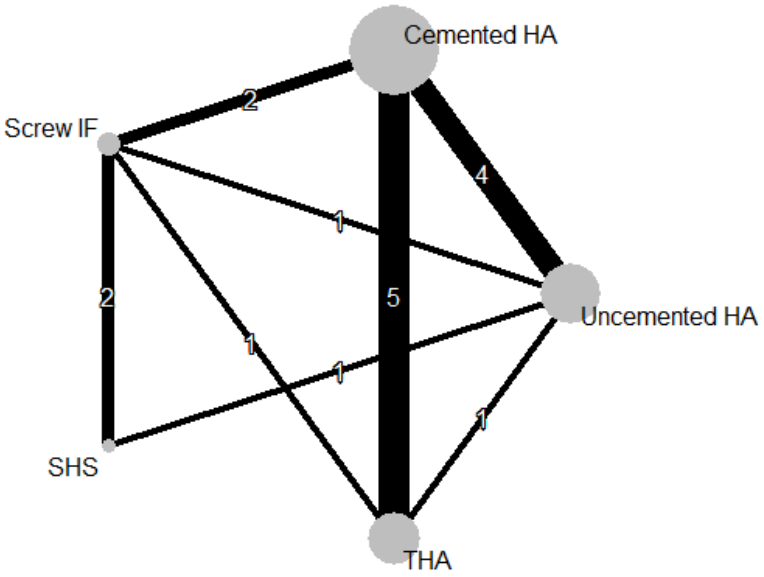
	Q	p-value	Tau^2
Between designs	10.85	0.0283	< 0.0001

Short-term quality of life funnel plot



Section 18: Long-term Quality of Life Results

Long-term quality of life network diagram



Number of studies = 15
Number of treatments = 5
Number of designs = 8

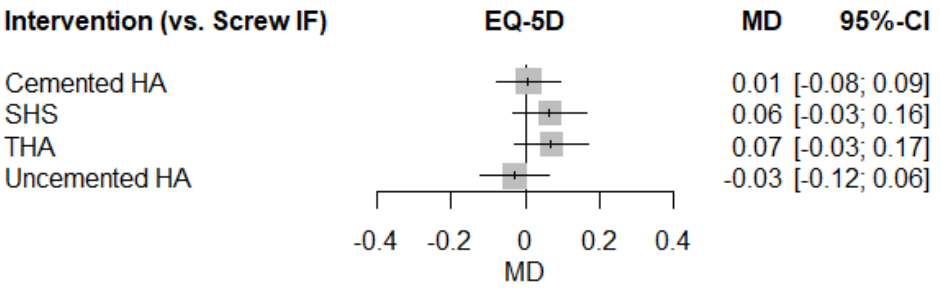
Long-term quality of life league table

Cemented HA	0.12 (0.00, 0.25)	.	-0.07 (-0.15, 0.01)	-0.02 (-0.10, 0.06)
0.01 (-0.08, 0.09)	**Screw IF**	0.01 (-0.11, 0.12)	-0.11 (-0.28, 0.06)	0.15 (-0.03, 0.33)
-0.06 (-0.18, 0.06)	-0.06 (-0.16, 0.03)	**SHS**	.	**0.24 (0.08, 0.40)**
-0.06 (-0.13, 0.01)	-0.07 (-0.17, 0.03)	-0.00 (-0.13, 0.12)	**THA**	0.01 (-0.15, 0.17)
0.04 (-0.04, 0.11)	0.03 (-0.06, 0.12)	0.09 (-0.02, 0.21)	**0.10 (0.01, 0.19)**	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

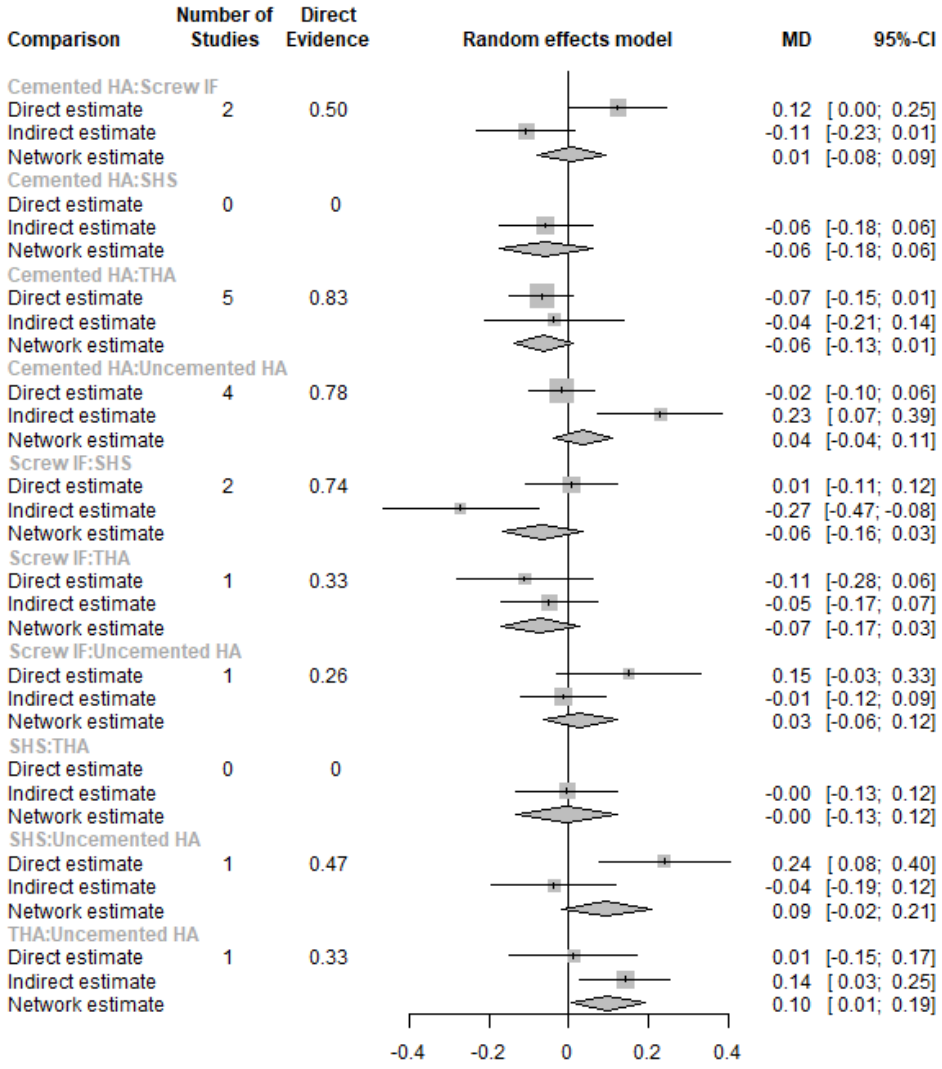
Bold: Statistical Significance

Long-term quality of life forest plot



Section 19: Long-term Quality of Life Heterogeneity and Incoherence Assessment

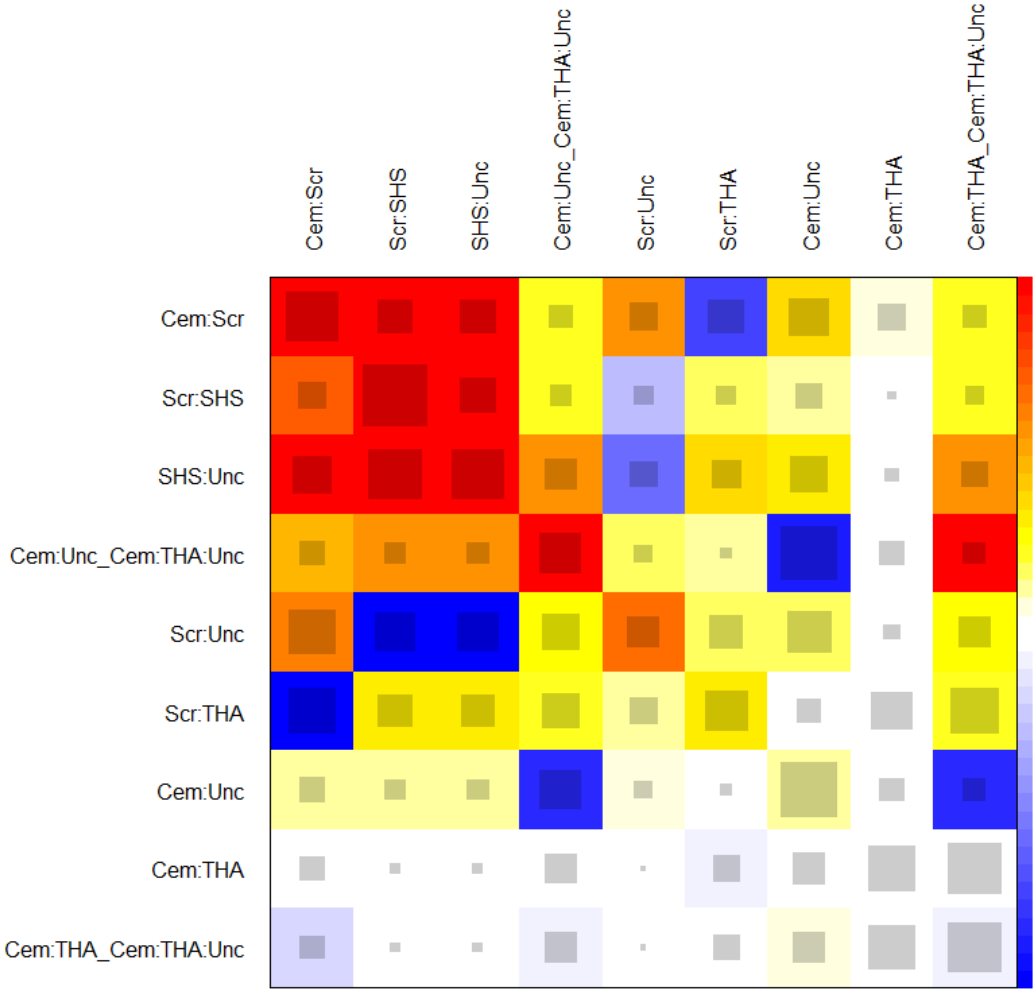
Long-term quality of life node-splitting assessment of incoherence



Within-design heterogeneity

Design	Q	p-value
Cemented HA: Screw IF	56.42	< 0.0001
Cemented HA: THA	108.42	< 0.0001
Cemented HA: Uncemented HA	108.81	< 0.0001
Screw IF: SHS	36.83	< 0.0001
Screw IF: THA	101.17	< 0.0001
Screw IF: Uncemented HA	101.80	< 0.0001
SHS: Uncemented HA	36.83	< 0.0001
Cemented HA: THA: Uncemented HA	79.62	< 0.0001

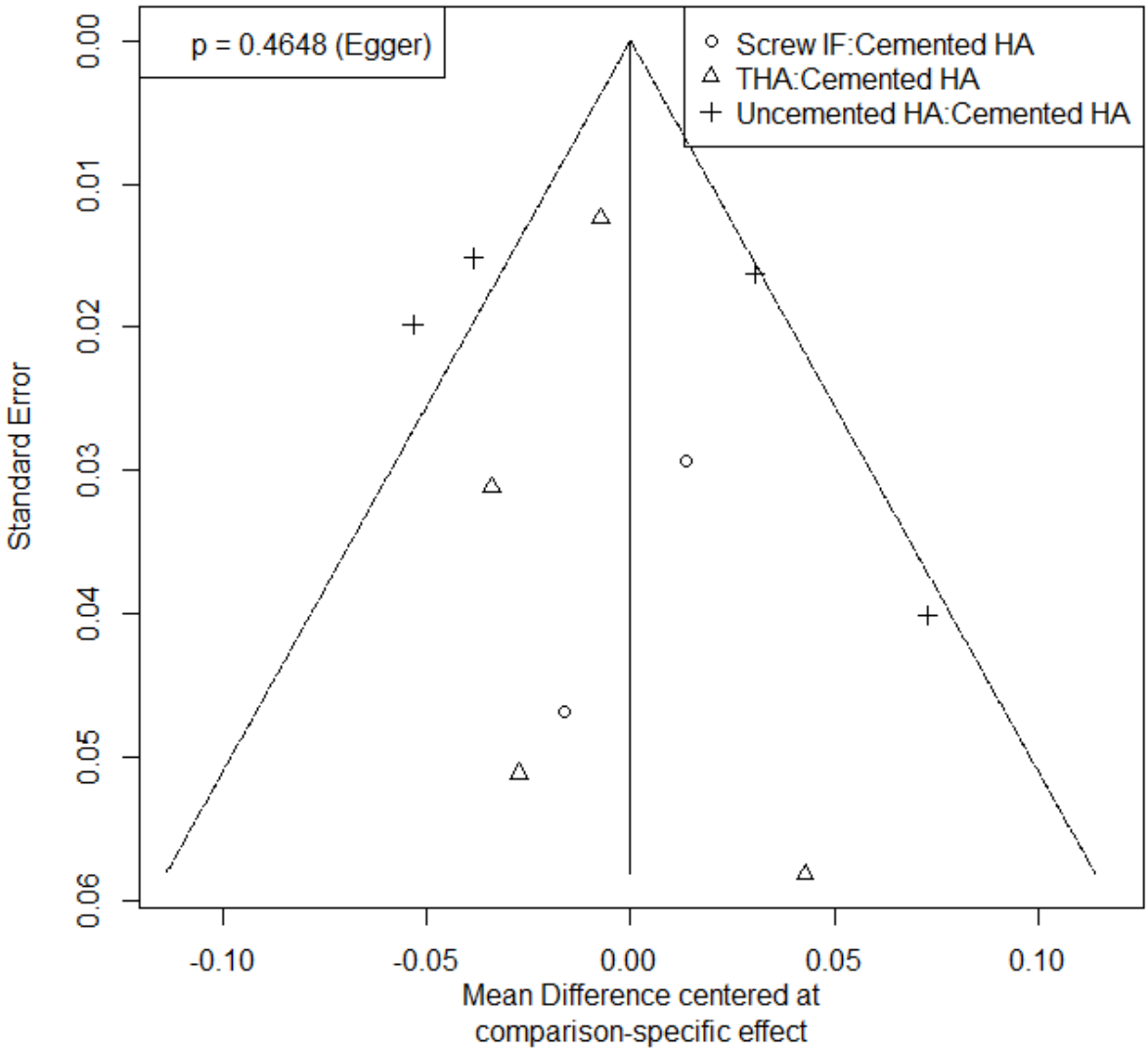
Long-term quality of life incoherence heat plot



Design-by-treatment interaction random effects model for incoherence

	Q	p-value	Tau^2
Between designs	63.76	<0.0001	0.0005

Long-term quality of life funnel plot

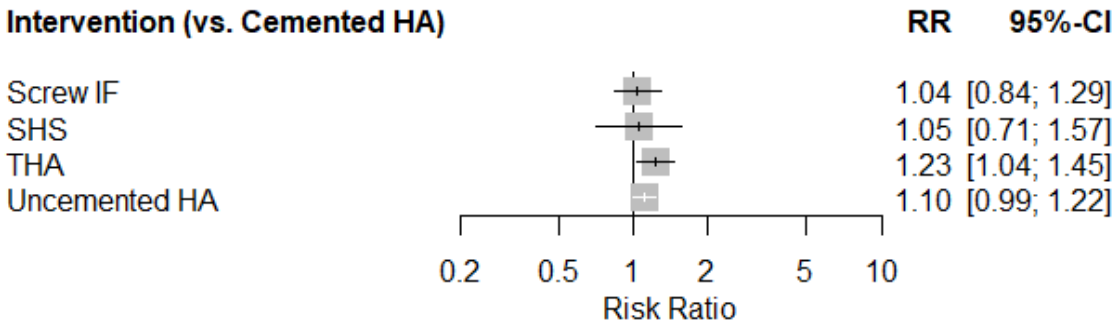


Section 20: Summary of Secondary Analysis with Combined HA groups

Comparison	Network Estimate	# Of Direct Trials	I^2
Mortality			
Screw IF vs HA	1.08 (0.89, 1.31)	5	0%
SHS vs HA	0.92 (0.72, 1.17)	4	0%
THA vs HA	1.12 (0.98, 1.29)	12	0%
Screw IF vs SHS	1.17 (0.88, 1.56)	1	NA
Screw IF vs THA	0.96 (0.78, 1.19)	3	0%
SHS vs THA	0.82 (0.63, 1.06)	2	0%
Reoperation			
Screw IF vs HA	**3.47 (2.25, 5.37)**	5	0%
SHS vs HA	**3.33 (2.07, 5.35)**	4	82.4%
THA vs HA	0.84 (0.56, 1.27)	14	33.4%
Screw IF vs SHS	1.04 (0.64, 1.71)	3	61.1%
Screw IF vs THA	**4.11 (2.59, 6.53)**	4	57.7%
SHS vs THA	**3.94 (2.28, 6.80)**	2	0%
Complications			
Screw IF vs HA	**2.11 (1.44, 3.11)**	5	30.0%
SHS vs HA	**2.40 (1.52, 3.78)**	3	88.3%
THA vs HA	1.18 (0.83, 1.66)	14	0%
Screw IF vs SHS	0.88 (0.57, 1.36)	4	44.1%
Screw IF vs THA	**1.80 (1.19, 2.71)**	4	89.6%
SHS vs THA	**2.04 (1.24, 3.36)**	1	NA
Short-term Function			
Screw IF vs HA	**-9.53 (-13.42, -5.64)**	1	NA
SHS vs HA	**-25.33 (-36.91, -13.75)**	0	NA
THA vs HA	1.32 (-0.07, 2.71)	6	0%
Screw IF vs SHS	15.80 (4.90, 26.70)	1	NA
Screw IF vs THA	**-10.85 (-14.84, -6.87)**	1	NA
SHS vs THA	**-26.65 (-38.26, -15.04)**	0	NA
Long-term Function			
Screw IF vs HA	-3.20 (-8.35, 1.94)	1	NA
SHS vs HA	**-5.12 (-9.61, -0.64)**	2	36.5%
THA vs HA	**4.17 (1.71, 6.64)**	11	80.6%
Screw IF vs SHS	1.92 (-3.26, 7.10)	2	0%
Screw IF vs THA	**-7.38 (-12.66, -2.09)**	1	NA
SHS vs THA	**-9.30 (-14.00, -4.60)**	1	NA
Short-term QoL			
Screw IF vs HA	**-0.08 (-0.13, -0.04)**	3	0%
SHS vs HA	**-0.07 (-0.13, -0.01)**	0	NA
THA vs HA	**0.02 (0.004, 0.04)**	5	0%
Screw IF vs SHS	-0.01, -0.06, 0.04)	1	NA
Screw IF vs THA	**-0.10 (-0.15, -0.06)**	1	NA
SHS vs THA	**-0.09 (-0.16, -0.03)**	0	NA
Long-term QoL			
Screw IF vs HA	0.01 (-0.09, 0.10)	3	92.3%
SHS vs HA	0.08 (-0.05, 0.21)	1	NA
THA vs HA	0.07 (-0.01, 0.15)	6	0%
Screw IF vs SHS	-0.07 (-0.19, 0.05)	2	0%
Screw IF vs THA	-0.06 (-0.18, 0.05)	1	NA
SHS vs THA	0.01 (-0.14, 0.16)	0	NA

Section 21: Sensitivity Analysis of Low Risk of Bias Studies

Mortality

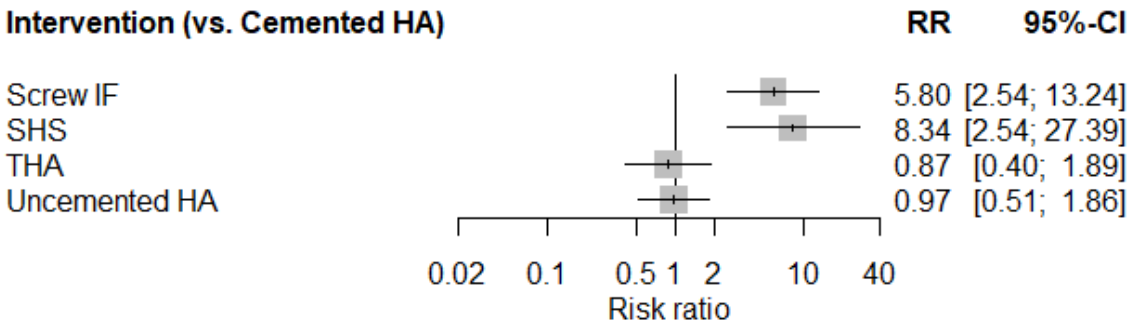


Cemented HA	0.99 (0.77, 1.27)	1.08 (0.65, 1.78)	**0.79 (0.66, 0.94)**	0.91 (0.82, 1.01)
0.96 (0.78, 1.19)	**Screw IF**	0.80 (0.43, 1.52)	1.08 (0.54, 2.16)	1.08 (0.59, 1.97)
0.95 (0.64, 1.42)	0.99 (0.65, 1.49)	**SHS**	.	.
0.81 (0.69, 0.96)	0.85 (0.65, 1.10)	0.86 (0.56, 1.32)	**THA**	1.01 (0.70, 1.45)
0.91 (0.82, 1.01)	0.94 (0.75, 1.19)	0.96 (0.63, 1.45)	1.12 (0.93, 1.35)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Reoperation

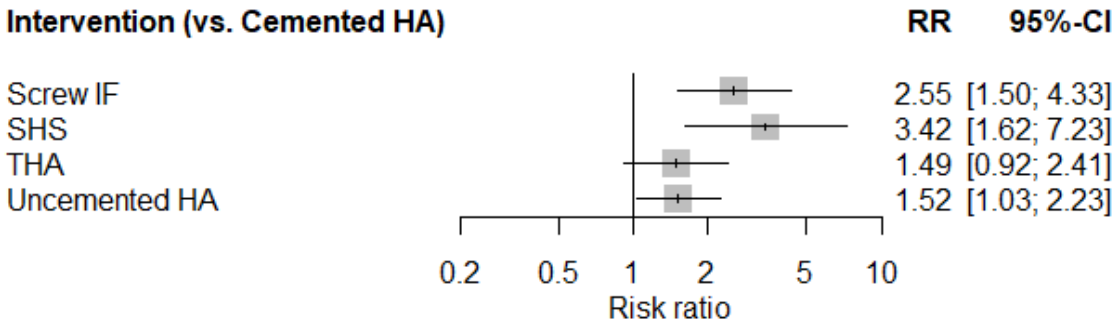


Cemented HA	**0.17 (0.05, 0.63)**	**0.06 (0.01, 0.40)**	1.06 (0.42, 2.71)	1.15 (0.58, 2.29)
0.17 (0.08, 0.39)	**Screw IF**	0.93 (0.28, 3.08)	**7.11 (2.22, 22.83)**	2.50 (0.37, 16.75)
0.12 (0.04, 0.39)	0.70 (0.24, 1.99)	**SHS**	.	.
1.15 (0.53, 2.51)	**6.69 (2.80, 16.03)**	**9.62 (2.68, 34.59)**	**THA**	0.80 (0.15, 4.19)
1.03 (0.54, 1.97)	**5.97 (2.32, 15.42)**	**8.59 (2.34, 31.47)**	0.89 (0.36, 2.24)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Complications

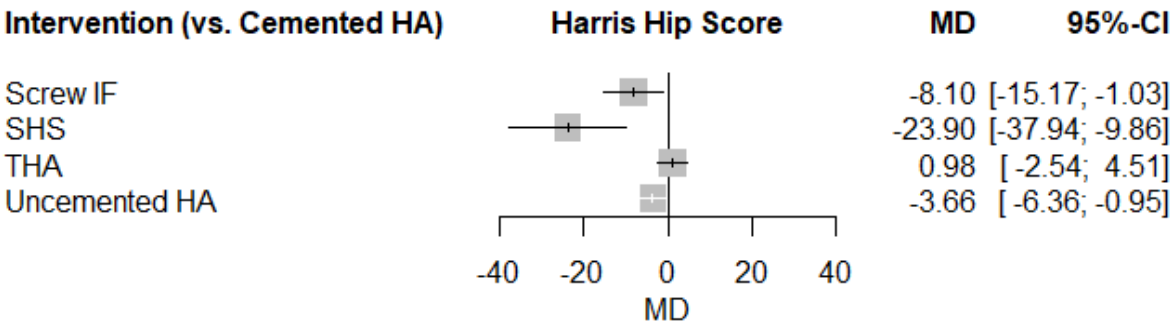


Cemented HA	**0.32 (0.14, 0.74)**	**0.12 (0.04, 0.41)**	0.88 (0.50, 1.55)	0.68 (0.46, 1.02)
0.39 (0.23, 0.67)	**Screw IF**	1.05 (0.49, 2.27)	1.11 (0.51, 2.43)	1.29 (0.36, 4.62)
0.29 (0.14, 0.62)	0.74 (0.38, 1.45)	**SHS**	.	.
0.67 (0.42, 1.09)	1.71 (0.97, 3.02)	**2.30 (1.02, 5.17)**	**THA**	1.07 (0.39, 2.92)
0.66 (0.45, 0.97)	1.68 (0.92, 3.06)	2.25 (1.00, 5.08)	0.98 (0.56, 1.72)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Short-term function

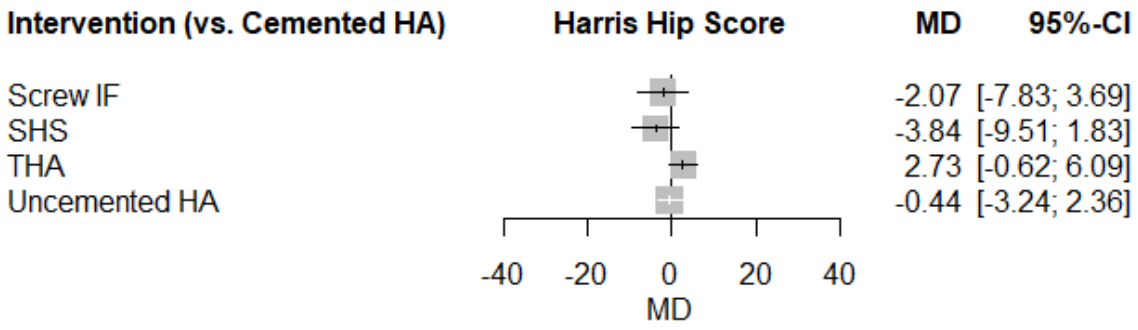


Cemented HA	**8.10 (1.03, 15.17)**	.	-1.95 (-5.64, 1.74)	**3.76 (1.01, 6.51)**
8.10 (1.03, 15.17)	**Screw IF**	**15.80 (3.67, 27.93)**	.	.
23.90 (9.86, 37.94)	**15.80 (3.67, 27.93)**	**SHS**	.	.
-0.98 (-4.51, 2.54)	**-9.08 (-16.98, -1.18)**	**-24.88 (-39.36, -10.41)**	**THA**	1.50 (-4.41, 7.41)
3.66 (0.95, 6.36)	-4.44 (-12.02, 3.13)	**-20.24 (-34.54, -5.94)**	**4.64 (0.59, 8.69)**	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Long-term function

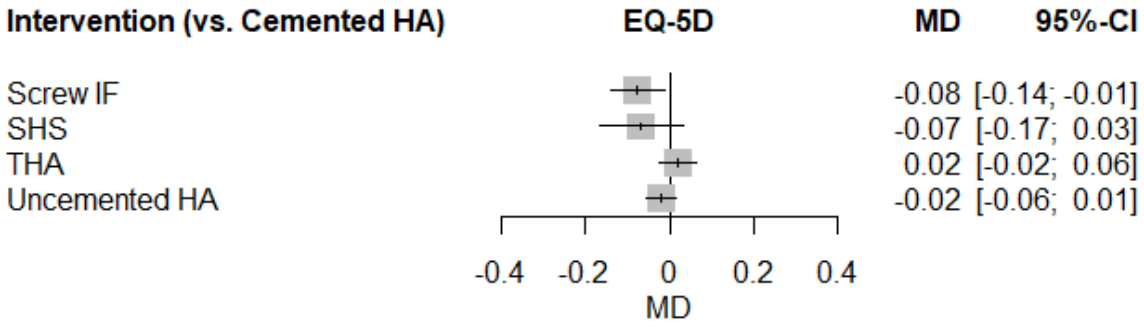


Cemented HA	3.30 (-4.04, 10.64)	2.70 (-4.37, 9.77)	-2.89 (-6.38, 0.59)	0.45 (-2.41, 3.31)
2.07 (-3.69, 7.83)	**Screw IF**	2.60 (-3.42, 8.61)	.	.
3.84 (-1.83, 9.51)	1.77 (-3.41, 6.95)	**SHS**	.	.
-2.73 (-6.09, 0.62)	-4.81 (-11.47, 1.86)	-6.58 (-13.17, 0.01)	**THA**	2.61 (-3.59, 8.82)
0.44 (-2.36, 3.24)	-1.63 (-8.03, 4.77)	-3.40 (-9.73, 2.92)	3.18 (-0.84, 7.19)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Short-term quality of life

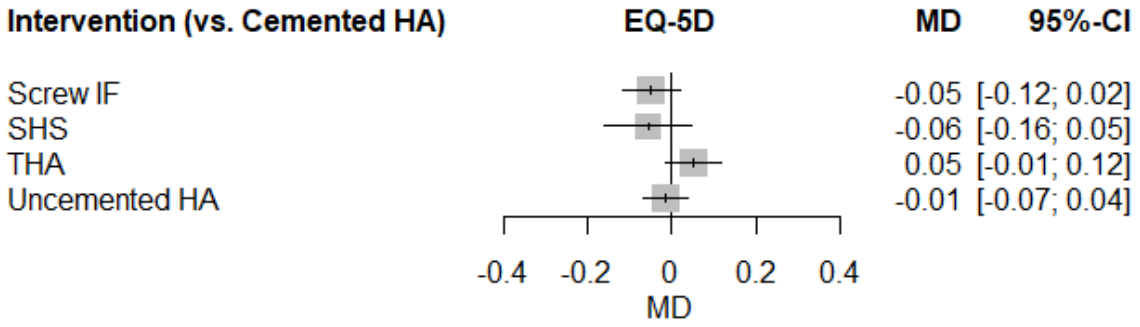


Cemented HA	0.08 (-0.02, 0.18)	.	-0.03 (-0.08, 0.02)	0.02 (-0.02, 0.06)
0.08 (0.01, 0.14)	**Screw IF**	-0.01 (-0.08, 0.07)	**-0.12 (-0.21, -0.03)**	0.00 (-0.13, 0.13)
0.07 (-0.03, 0.17)	-0.01 (-0.08, 0.07)	**SHS**	.	.
-0.02 (-0.06, 0.02)	**-0.10 (-0.16, -0.03)**	-0.09 (-0.19, 0.01)	**THA**	0.00 (-0.07, 0.07)
0.02 (-0.01, 0.06)	-0.06 (-0.12, 0.01)	-0.05 (-0.15, 0.05)	0.04 (-0.01, 0.09)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Long-term quality of life



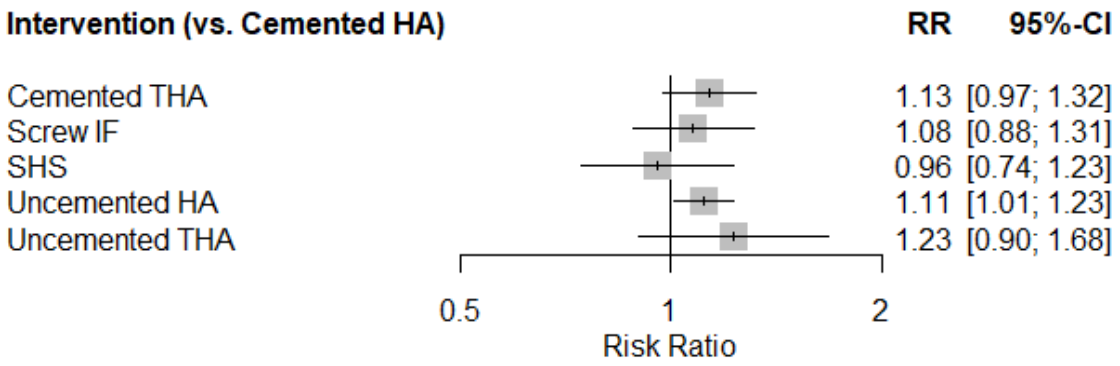
Cemented HA	**0.12 (0.03, 0.21)**	.	-0.07 (-0.14, 0.01)	-0.01 (-0.07, 0.04)
0.05 (-0.02, 0.12)	**Screw IF**	0.01 (-0.07, 0.09)	-0.11 (-0.23, 0.01)	
0.06 (-0.05, 0.16)	0.01 (-0.07, 0.09)	**SHS**	.	. **0.15 (0.01, 0.29)**
-0.05 (-0.12, 0.01)	**-0.10 (-0.18, -0.02)**	-0.11 (-0.22, 0.00)	**THA**	0.01 (-0.10, 0.12)
0.01 (-0.04, 0.07)	-0.04 (-0.11, 0.04)	-0.04 (-0.15, 0.07)	0.06 (-0.01, 0.14)	**Uncemented HA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Section 22: Secondary Analysis of Cemented and Uncemented THA

Mortality

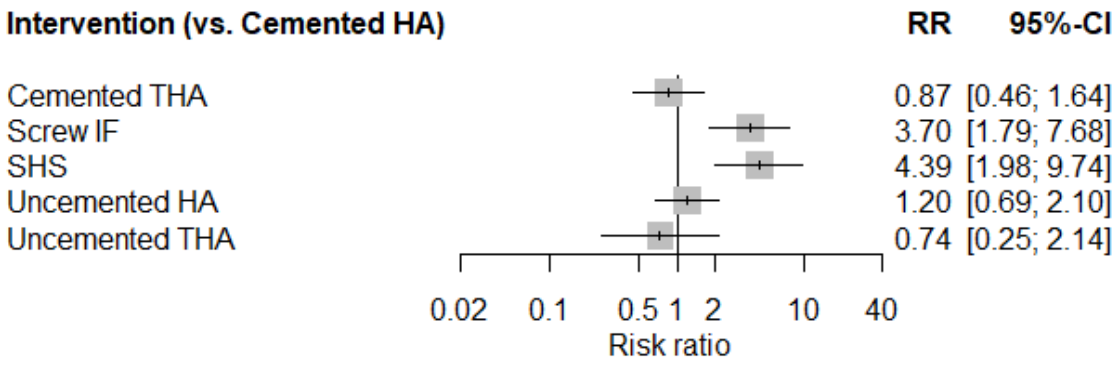


Cemented HA	**0.81 (0.68, 0.97)**	0.99 (0.77, 1.27)	1.08 (0.65, 1.78)	0.91 (0.83, 1.01)	0.81 (0.56, 1.17)
0.88 (0.76, 1.03)	**Cemented THA**	0.79 (0.51, 1.23)	1.08 (0.72, 1.62)	0.89 (0.65, 1.22)	0.85 (0.59, 1.23)
0.93 (0.76, 1.13)	1.05 (0.84, 1.33)	**Screw IF**	0.80 (0.43, 1.52)	1.08 (0.59, 1.97)	.
1.04 (0.81, 1.35)	1.19 (0.91, 1.55)	1.12 (0.84, 1.51)	**SHS**	0.76 (0.52, 1.11)	.
0.90 (0.82, 0.99)	1.02 (0.86, 1.21)	0.97 (0.78, 1.20)	0.86 (0.67, 1.11)	**Uncemented HA**	0.96 (0.65, 1.44)
0.81 (0.59, 1.11)	0.92 (0.67, 1.27)	0.87 (0.61, 1.26)	0.78 (0.53, 1.15)	0.90 (0.66, 1.24)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Reoperation

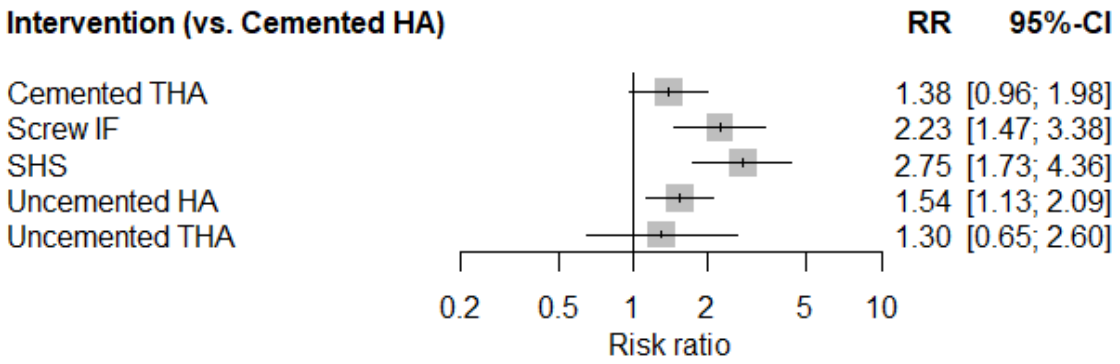


Cemented HA	0.98 (0.46, 2.10)	**0.18 (0.05, 0.62)**	**0.06 (0.01, 0.37)**	1.19 (0.63, 2.24)	1.31 (0.26, 6.59)
1.15 (0.61, 2.17)	**Cemented THA**	**0.23 (0.06, 0.91)**	**0.15 (0.04, 0.63)**	0.56 (0.19, 1.65)	1.64 (0.33, 8.19)
0.27 (0.13, 0.56)	**0.23 (0.11, 0.51)**	**Screw IF**	0.76 (0.28, 2.06)	2.50 (0.40, 15.67)	3.76 (0.76, 18.52)
0.23 (0.10, 0.50)	**0.20 (0.09, 0.46)**	0.84 (0.39, 1.80)	**SHS**	1.40 (0.44, 4.43)	.
0.83 (0.48, 1.45)	0.72 (0.35, 1.48)	**3.08 (1.43, 6.64)**	**3.66 (1.66, 8.07)**	**Uncemented HA**	1.89 (0.43, 8.29)
1.36 (0.47, 3.94)	1.18 (0.39, 3.54)	**5.03 (1.73, 14.59)**	**5.97 (1.83, 19.48)**	1.63 (0.56, 4.78)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Hip-related Complications

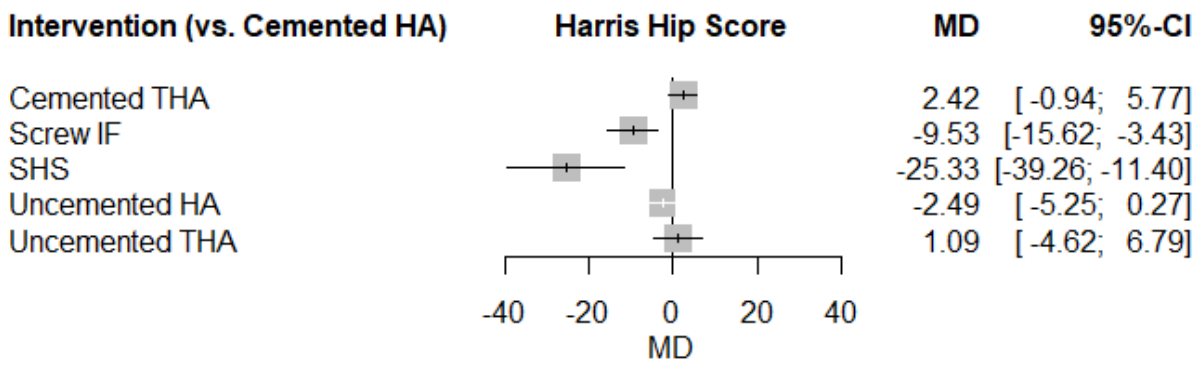


Cemented HA	0.90 (0.57, 1.40)	**0.32 (0.16, 0.68)**	**0.12 (0.04, 0.36)**	0.72 (0.51, 1.01)	0.81 (0.35, 1.87)
0.73 (0.50,1.04)	**Cemented THA**	0.86 (0.48, 1.53)	0.55 (0.21, 1.39)	0.90 (0.48, 1.70)	0.92 (0.40, 2.12)
0.45 (0.30,0.68)	**0.62 (0.41,0.94)**	**Screw IF**	0.91 (0.55, 1.52)	1.29 (0.41, 4.05)	.
0.36 (0.23,0.58)	**0.50 (0.31,0.81)**	0.81 (0.54,1.22)	**SHS**	1.40 (0.70, 2.78)	.
0.65 (0.48,0.89)	0.90 (0.60,1.35)	1.45 (0.93,2.26)	1.79 (1.12,2.85)	**Uncemented HA**	1.07 (0.47, 2.44)
0.77 (0.38,1.54)	1.06 (0.52,2.16)	1.71 (0.80,3.68)	2.11 (0.96,4.66)	1.18 (0.59,2.38)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Short-term Function

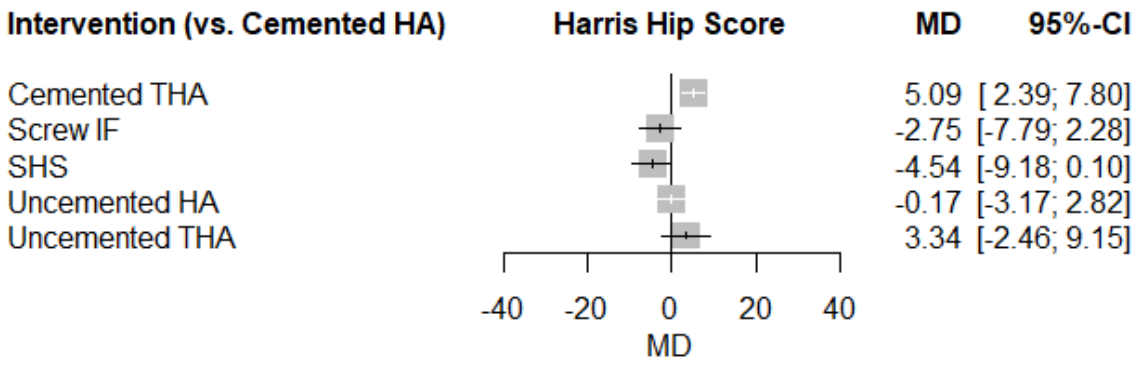


Cemented HA	-2.71 (-6.34, 0.91)	**8.10 (0.37, 15.83)**	.	2.66 (-0.15, 5.46)	-2.12 (-8.76, 4.53)
-2.42 (-5.77, 0.94)	**Cemented THA**	**14.00 (4.73, 23.27)**	.	1.70 (-5.05, 8.45)	0.38 (-6.25, 7.02)
9.53 (3.43, 15.62)	**11.95 (5.69, 18.20)**	Screw IF	**15.80 (3.27, 28.32)**	.	.
25.33 (11.40, 39.26)	**27.74 (13.74, 41.75)**	**15.80 (3.27, 28.32)**	**SHS**	.	.
2.49 (-0.27, 5.25)	**4.91 (0.85, 8.97)**	**-7.04 (-13.65, -0.42)**	**-22.84 (-37.00, -8.67)**	**Uncemented HA**	-1.32 (-8.18, 5.55)
-1.09 (-6.79, 4.62)	1.33 (-4.60, 7.26)	**-10.62 (-18.75, -2.48)**	**-26.41 (-41.35, -11.48)**	-3.58 (-9.47, 2.31)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Long-term Function

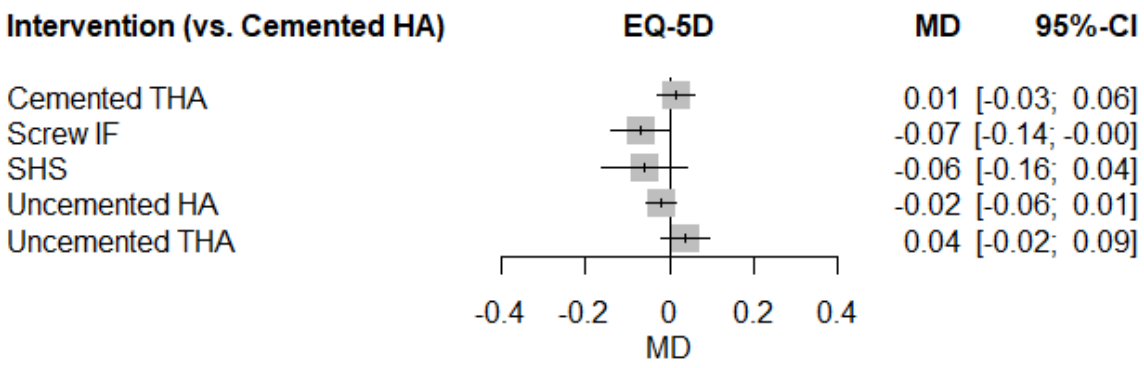


Cemented HA	**-5.09 (-8.04, -2.15)**	3.30 (-4.58, 11.18)	2.70 (-4.94,10.34)	0.46 (-2.61, 3.53)	-2.91 (-9.76, 3.94)
-5.09 (-7.80, -2.39)	**Cemented THA**	9.00 (-0.48, 18.48)	**10.10 (3.23,16.97)**	2.70 (-4.20, 9.61)	0.30 (-6.53, 7.13)
2.75 (-2.28, 7.79)	**7.84 (2.75, 12.94)**	**Screw IF**	2.69 (-3.76, 9.14)	.	.
4.54 (-0.10, 9.18)	**9.63 (5.02, 14.25)**	1.79 (-3.21, 6.79)	**SHS**	.	.
0.17 (-2.82, 3.17)	**5.27 (1.46, 9.07)**	-2.58 (-8.37, 3.21)	-4.37 (-9.81, 1.07)	**Uncemented HA**	-2.40 (-9.39, 4.58)
-3.34 (-9.15, 2.46)	1.75 (-4.20, 7.70)	-6.10 (-13.61, 1.42)	**-7.88 (-15.12, -0.65)**	-3.52 (-9.55, 2.52)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Short-term Quality of Life

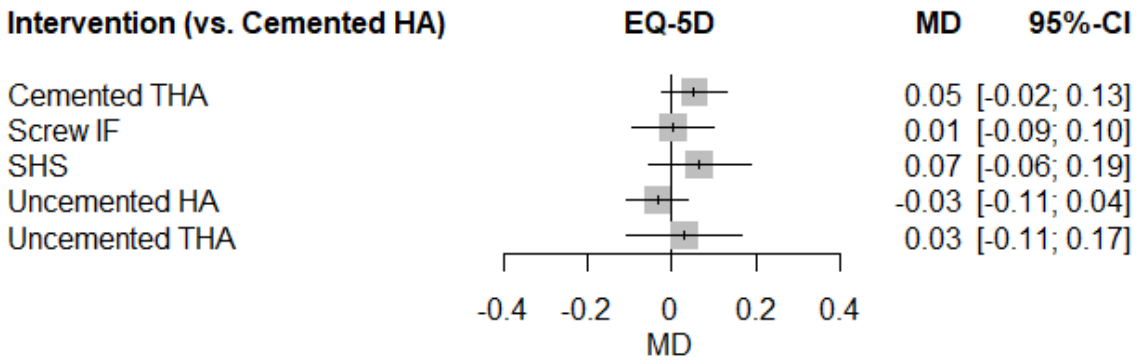


Cemented HA	-0.03 (-0.07, 0.02)	**0.09 (0.01, 0.17)**	.	0.02 (-0.02, 0.06)	-0.05 (-0.12, 0.01)
-0.01 (-0.06, 0.03)	**Cemented THA**	.	.	-0.01 (-0.08, 0.06)	-0.03 (-0.10, 0.03)
0.07 (0.00, 0.14)	0.08 (0.01, 0.16)	**Screw IF**	-0.01 (-0.08, 0.07)	0.00 (-0.13, 0.13)	.
0.06 (-0.04, 0.16)	0.08 (-0.03, 0.18)	-0.01 (-0.08, 0.07)	**SHS**	.	.
0.02 (-0.01, 0.06)	0.04 (-0.01, 0.08)	-0.05 (-0.12, 0.02)	-0.04 (-0.14, 0.06)	**Uncemented HA**	-0.02 (-0.09, 0.04)
-0.04 (-0.09, 0.02)	-0.02 (-0.08, 0.04)	**-0.10 (-0.19, -0.02)**	-0.10 (-0.21, 0.02)	-0.06 (-0.12, 0.00)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Long-term Quality of Life



Cemented HA	-0.07 (-0.15, 0.01)	0.12 (0.00, 0.25)	.	-0.02 (-0.10, 0.07)	-0.06 (-0.22, 0.10)
-0.05 (-0.13, 0.02)	**Cemented THA**	.	.	0.01 (-0.15, 0.17)	-0.00 (-0.16, 0.16)
-0.01 (-0.10, 0.09)	0.05 (-0.07, 0.17)	**Screw IF**	0.01 (-0.11, 0.12)	0.15 (-0.03, 0.33)	.
-0.07 (-0.19, 0.06)	-0.01 (-0.15, 0.13)	-0.06 (-0.16, 0.04)	**SHS**	**0.24 (0.07, 0.41)**	.
0.03 (-0.04, 0.11)	0.09 (-0.01, 0.19)	0.04 (-0.06, 0.14)	0.10 (-0.02, 0.22)	**Uncemented HA**	-0.01 (-0.17, 0.15)
-0.03 (-0.17, 0.11)	0.02 (-0.12, 0.16)	-0.03 (-0.19, 0.14)	0.04 (-0.14, 0.21)	-0.07 (-0.21, 0.08)	**Uncemented THA**

Upper grey: Direct estimates, Lower white: Network estimates

Bold: Statistical Significance

Part B: Surgical Management of Displaced Femoral Neck Fractures: A Protocol for a Systematic Review and Network Meta-Analysis

Mark R Phillips, BSc, PhD(c)[1]; Jason W. Busse, DC PhD[1,2,3,4]; Lehana Thabane PhD[1,5,6]; Raveendhara R Bannuru MD PhD[7]; Mohit Bhandari MD PhD[1,8]

12. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
13. The Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Ontario, Canada
14. Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada
15. The Michael G. DeGroote Centre for Medicinal Cannabis Research, McMaster University, Hamilton, Ontario, Canada
16. Biostatistics Unit, St Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada
17. Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa
18. Center for Treatment Comparison and Integrative Analysis, Tufts Medical Center, Boston, MA, USA
19. Department of Surgery, McMaster University, Hamilton, Ontario, Canada

Abstract

The search for an optimal femoral neck fracture treatment option to limit mortality and reoperation, while maximizing patient function and quality of life has been taking place within orthopaedic research for many years. This study will use a network meta-analysis (NMA) approach to compare the relative effectiveness of total hip arthroplasty (THA), hemiarthroplasty (HA), cancellous screws (CS) and dynamic/sliding hip screws (DHS) for the surgical management of displaced femoral neck fractures.

Introduction

The incidence of hip fractures worldwide was 1.3 million in 1990 with projections to increase to 2.6 million cases in the year 2025, and 4.5 million by 2050.[1] Hip fractures have a significant impact on patients; mortality rates range between 5-10% at one month after hip fracture, and may be as high as 30% at one year.[2] Reoperation or revision surgery is associated with increased risk of mortality after hip fracture surgery.[3] For this reason, a large body of evidence has emerged to identify optimal treatment methods to limit the risk of requiring an additional surgery.[4] Hip fractures also pose a significant impact on patient daily function and quality of life. A recent cohort study of over 10,000 hip fracture patients found that 71% of patients had trouble walking four months after hip fracture surgery, and 58% had difficulty ambulating one year after surgery.[5] The search for an optimal treatment option to limit mortality and reoperation, while maximizing patient function and quality of life has been an important focus in orthopaedic research.

There are multiple treatment options for hip fractures – specifically fractures of the femoral neck – that are broadly categorized as arthroplasty or internal fixation.[3] Arthroplasty options include total hip arthroplasty (TKA) or hemiarthroplasty (HA).[6] The primary options for internal fixation are the use of cancellous screws or a dynamic/sliding hip screw.[4] While there have been numerous randomized trials and pairwise meta-analyses comparing the different hip fracture treatment options to one another, there is opportunity for network meta-analysis (NMA) to provide a detailed analysis of the current evidence for all potential displaced femoral neck fracture treatment options.

Methods

This protocol reports relevant information to the standards of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols checklist extension for network meta-analysis[7]. This study is registered on PROSPERO (303952).

Eligibility Criteria

This systematic review will include randomized trials that enroll adult patients (aged ≥18 years) with a femoral neck fracture, and randomly assign them to two or more of the following interventions: (1) total hip arthroplasty (THA), (2) dynamic hip screw/sliding hip screw (DHS), (3) hemiarthroplasty (HA), or (4) cancellous screws (CS).

Information Sources and Article Selection

We will conduct a literature search of MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials (CENTRAL) to identify all relevant studies published in English. Additionally, the reference lists of relevant studies will be hand-searched to identify additional articles for

inclusion. Articles retrieved through the systematic search will be screened independently and in duplicate using Covidence software (covidence.org). Titles and abstracts of retrieved articles will be screened for eligibility, followed by the full-texts of potentially eligible studies. Articles included by at least one reviewer at the title and abstract will be reviewed for eligibility at the full-text stage. Any disagreements between reviewers will be resolved by a consensus meeting. If consensus is not met, a third reviewer will be consulted.

Data Collection

All data will be collected using a standardized pilot-tested data extraction form. Data items extracted will include: year of publication, location of research, study follow-up timepoints, mean age, Body Mass Index (BMI), percent female, number of treatment arms, type of implant, and cemented/ uncemented implant. The outcomes of interest include:

1. unplanned secondary procedures (re-operation)
2. mortality
3. hip-related complications (instability, dislocation, avascular necrosis, non/ mal-union).
4. Function (both the short (≤6 months) and long-term (>1 year))
5. Quality of life (both the short (≤6 months) and long-term (>1 year))

Network Geometry

Network geometry will be described, including the number of unique treatments, trials assessing each treatment, comparisons between each treatment option, and number of patients informing each comparison. A network plot will be provided to visualize the network, which will weight nodes by the number of patients and connections weighted by the number of studies comparing the connected interventions. The NMA will assess a 4-node network with the

treatment nodes: THA, HA, cancellous screws, and DHS. A secondary NMA will be conducted to assess the cemented THA, uncemented THA, cemented HA, and uncemented HA, separately, in order to determine the differences in treatment effect for cemented and uncemented implants. If studies did not specify or standardize the use of cement with HA or THA options, they will not be included within this secondary network analysis.

Methods of Analysis

This study will compare the available treatment options for femoral neck fractures by combining the direct and indirect estimates to develop a summary network estimate. Analyses for each outcome will be conducted using a Frequentist random-effect model. Analyses will be conducted using the netmeta package in R software (v3.6.2). Analyses will be evaluated and reported using a minimally contextualized framework.[8]

Summary Measures

All continuous outcomes will be analyzed as weighted mean differences (WMD) with 95% confidence intervals (CIs) of change scores from baseline. If only baseline and follow-up scores are reported, we will calculate the change from baseline and the associated SD assuming a within group correlation of 0.5. If insufficient information is reported to calculate the change from baseline SD, the SD from baseline will be used. If there are no SDs or other measures of variance reported that could be used to calculate an SD, an SD will be imputed from a trial of similar sample size that has the lowest risk of bias.

When studies report continuous outcomes for the same construct using different measurement instruments, we will convert the outcome scores to the most commonly utilized measurement instrument for that outcome. These estimates will be compared to relevant

anchor-based minimally important difference (MID) estimates for the utilized measures. Interpretation of estimates against an MID will cautiously consider the risk of inappropriate conclusions that estimates below an MID are unimportant. To address this, the RD of achieving the MID will be modelled from the pooled WMD.[9]

All dichotomous outcomes (unplanned secondary procedures (re-operation), mortality, and hip-related complications) will be reported as Relative Risk (RR) with 95% CIs, as well as Risk Differences (RD) with 95% CIs. The baseline risk (BR) will be based on the absolute risk of event within large, well-conducted observational studies that appropriately represent the patient demographics of the included studies.[10] If unavailable, the baseline risk will be derived from the absolute risk of event within the THA arm of included studies.

Assessment of Heterogeneity and Inconsistency

Heterogeneity will be assessed using the chi-squared test and I^2 statistic. We have determined a number of potential hypotheses to explain potential variability between studies. First, trials that evaluate older patient populations may see worse results across outcomes than studies with generally younger participants. Finally, studies with greater risk of bias may demonstrate larger effects than studies with low risk of bias. Subgroup analyses will be conducted to explore these hypotheses, which is explained in more detail in the next section of this protocol.

Inconsistency within the network will be assessed using the node splitting method. I^2 values for each split node will be provided. Additionally, inconsistency will be assessed by assessment of the Cochran's Q statistic. Specific areas within the network causing inconsistency will be visualized using an inconsistency heat plot.[11]

Additional Analyses

Network subgroup analyses will be conducted for age (60 or older versus younger than 60), and risk of bias (high vs low risk of bias). The age subgroup will be determined by the mean age of the study population if data is not specifically available for age categories within studies.

Risk of Bias and Certainty of Evidence Assessment

Risk of bias within individual studies will be assessed using a modified version of the Cochrane Risk of Bias assessment tool, as proposed by Akl et al.[12,13] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach and GRADE NMA extension will be utilized to assess the overall certainty of evidence across assessed outcomes.[14,15] We will rate down the certainty for incoherence between the indirect and direct estimates, and we will use the direct or indirect estimate of effect instead of the network estimate if supported by higher certainty of evidence. We will not rate down the certainty rating of the network estimate twice if both intransitivity and incoherence are present. We will assess imprecision by using the network estimate; if the 95% CI excluded the null effect for dichotomous outcome or ½ the MID for continuous outcomes, then we will not rate down for imprecision unless the comparison is informed by fewer than 300 observations for continuous outcomes or 300 events for binary outcomes. When 10 or more trials are included for a comparison, publication bias will be visually assessed for asymmetry using a funnel plot and Egger's test for publication bias.[16] Risk of bias and certainty of evidence will both be assessed by two independent reviewers, with discrepancies resolved through a third reviewer if necessary.

Discussion

This study aims to explore the relative effectiveness of all available surgical options for displaced femoral neck fracture. The use of NMA is of particular benefit in clinical scenarios where there is uncertainty regarding the optimal treatment option with multiple potential options. Two large randomized trials have recently been conducted comparing arthroplasty and internal fixation options, however the clear differences between all possible surgical options remains unclear.[4,17] By assessing multiple outcomes, the proposed study will clearly distinguish the advantages – and disadvantages – for each of these surgical options, as well as potential subgroups that may have divergent effects from the displaced femoral neck fracture population as a whole.

References

1. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA*. 1997;7(5):407-413.

2. Keene GS, Parker MJ, Pryor GA. Mortality and morbidity after hip fractures. *BMJ*. 1993;307(6914):1248-1250.

3. FAITH Investigators TF. Fixation using alternative implants for the treatment of hip fractures (FAITH): design and rationale for a multi-centre randomized trial comparing sliding hip screws and cancellous screws on revision surgery rates and quality of life in the treatment of femoral neck fractures. *BMC Musculoskelet Disord*. 2014;15(1):219. doi:10.1186/1471-2474-15-219

4. Fixation using Alternative Implants for the Treatment of Hip fractures (FAITH) Investigators. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *Lancet Lond Engl*. 2017;389(10078):1519-1527. doi:10.1016/S0140-6736(17)30066-1

5. Gjertsen JE, Baste V, Fevang JM, Furnes O, Engesæter LB. Quality of life following hip fractures: results from the Norwegian hip fracture register. *BMC Musculoskelet Disord*. 2016;17(1):265. doi:10.1186/s12891-016-1111-y

6. Khan M, Aleem IS, Poolman RW. Fixation versus primary replacement of displaced femoral neck fractures in the elderly. *Indian J Orthop*. 2011;45(1):23-26. doi:10.4103/0019-5413.73658

7. 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. *Ann Intern Med*. 2015;162(11):777-784. doi:10.7326/M14-2385

8. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ*. 2020;371. doi:10.1136/bmj.m3900

9. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. *Res Synth Methods*. 2011;2(3):188-203. doi:10.1002/jrsm.46

10. Newcombe RG, Bender R. Implementing GRADE: calculating the risk difference from the baseline risk and the relative risk. *Evid Based Med*. 2014;19(1):6-8. doi:10.1136/eb-2013-101340

11. Neupane B, Richer D, Bonner AJ, Kibret T, Beyene J. Network Meta-Analysis Using R: A Review of Currently Available Automated Packages. *PLOS ONE*. 2014;9(12):e115065. doi:10.1371/journal.pone.0115065

12. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343. doi:10.1136/bmj.d5928

13. Akl EA, Sun X, Busse JW, et al. Specific instructions for estimating unclearly reported blinding status in randomized trials were reliable and valid. *J Clin Epidemiol*. 2012;65(3):262-267. doi:10.1016/j.jclinepi.2011.04.015

14. Puhan MA, Schünemann 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. doi:10.1136/bmj.g5630

15. 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. doi:10.1016/j.jclinepi.2017.10.005

16. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634. doi:10.1136/bmj.315.7109.629

17. HEALTH Investigators, Bhandari M, Einhorn TA, et al. Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. *N Engl J Med*. 2019;381(23):2199-2208. doi:10.1056/NEJMoa1906190

Chapter 6: Discussion and Opportunities for Future Research

This thesis has provided an important contribution to the future of NMA quality, advancing the uptake of NMA results into clinical practice. The NMA conducted in displaced femoral neck fractures demonstrates the applicability of this tool, yet also highlights some important future directions. From a clinical perspective, the NMA on displaced femoral neck fractures has provided a strong framework for understanding the implications of surgical options.

This NMA has demonstrated the generally poor prognosis for elderly patients with a displaced femoral neck fracture. The absolute mortality across all surgical options was roughly 24% over 5 years post-surgery, illustrating the high risk of death in the years immediately following an elderly displaced femoral neck fracture. Small differences in mortality existed across different surgical options, which suggests that the surgical procedure chosen may not be a driving factor in these high mortality rates. Thus, future areas of clinical research should focus on other potential ways to mitigate mortality following hip fracture.

There is great potential for future research to expand our ability to prevent mortality after an elderly hip fracture. Some promising research that has put focus on these topics include research into post-operative monitoring[1,2], hip fracture prevention[3], accelerated surgery versus standard care[4], and holistic post-operative fracture care tailored to elderly fracture patients.[5] Additionally, a better understanding of patient prognosis could provide meaningful improvements to the post-hip fracture care pathway. Providing an increased understanding of the mortality risk factors in patients may improve clinician's ability to identify at-risk patients, and implement a more involved post-fracture care plan tailored to the specific mortality risks

identified. Secondary analyses have been conducted following major hip fracture trials – such as FAITH[6] and HEALTH[7] – to understand the prognostic factors associated with poor patient outcomes.[8] This research has highlighted that factors such as age, body mass index, American Society of Anesthesiologists (ASA) class, the use of ambulatory aid prior to hip fracture, and kidney disease, are all associated with an increased risk of mortality.[8] All of these fields of research may be important in improving the current mortality rate following elderly hip fracture – as this thesis has illustrated that the mortality rate after elderly hip fracture is high, regardless of surgical treatment option.

There are also numerous implications and future directions that arise from this thesis from a health research methodology perspective. As highlighted in the introduction of this thesis, a lack of standardization of methodology exists for NMAs. Recent insights have shown that some methods decisions, such as frequentist versus Bayesian analyses, may not have meaningful impact on the results observed.[9] Thus, a focus on maximizing interpretability of results is of primary importance, opposed to discerning the difference between analysis approaches. Whether a frequentist or Bayesian NMA is conducted, authors should ensure the results are reported to the highest standard – incorporating both the novel methods for interpreting NMA findings from minimally or partially contextualized approaches[10,11], while presenting those findings within our proposed presentation tool for maximal interpretability.

Standardization takes time and buy-in from future NMA authors, which requires knowledge translation and dissemination initiatives. A parallel can be drawn to the use of forest plots to present pairwise meta-analysis results; which organically grew to be the known and consistent standard for presenting results. Dissemination through a grass-roots uptake of the

NMA presentation tool, with high-impact NMAs continuing to provide examples of the utilization of this tool, will be invaluable to its adoption. To facilitate this, a number of recent NMAs in highly visible journals have already utilized and displayed this presentation method across various clinical areas.[12,13] A true test of the improved interpretability from this presentation method will be the adoption in future NMAs, as buy-in from NMA authors will be justification of the benefits this tool provides. With continued use, iterative improvements or unforeseen challenges may be addressed to ensure this tool serves its intended purpose. A passive uptake is, however, not considered as a sufficient knowledge translation plan. Future NMA workshops led within our research team plan to discuss and promote the implementation of this tool, and appropriate NMA methodology in general. Additionally, subsequent methodology initiatives will continue to improve this presentation format in response to feedback within the rapidly changing NMA methodology landscape. It is anticipated that the presentation format will not be a rigid "one-size-fits-all" approach to presenting NMA results. Instead, it provides a framework, backed by user feedback, to inform the optimal approach to presenting NMA results. This framework may continue to evolve through iterative methodological investigations and initiatives to further enhance the presentation of NMA results.

Even within this thesis, minor adjustments and considerations were needed to best implement the presentation tool within the femoral neck fracture NMA. The minimally contextualized approach to categorizing treatments had developed three categories within the initial NMA on acute pain management; which was used in the development of the presentation format. The femoral neck fracture NMA, however, required four categories to adequately differentiate the treatments from one another. This highlights a very important consideration –

of which the minimally contextualized approach to categorization manuscript by Brignardello-Petersen et al also suggests – that this is a high-level framework to guide authors in presenting results.[10,11] It is not a rigid tool that must be utilized in the exact manner presented within the acute pain NMA example. Instead, authors may utilize their best judgement within this tool to develop the most logical presentation format for their specific NMA scenario.

The totality of this thesis provides important context and background to the landscape of NMA methodology, an improvement to current NMA presentation methods for multiple outcomes, and an example NMA that utilizes this presentation tool to inform an important clinical area. The NMA methods landscape will continue to evolve, and the components of this thesis provide the groundwork to solidify and advance the implementation of NMA results into practice.

References

1. McGillion MH, Parlow J, Borges FK, et al. Post-discharge after surgery Virtual Care with Remote Automated Monitoring-1 (PVC-RAM-1) technology versus standard care: randomised controlled trial. *BMJ*. 2021;374:n2209. doi:10.1136/bmj.n2209

2. McGillion MH, Dvirnik N, Yang S, et al. Continuous Noninvasive Remote Automated Blood Pressure Monitoring With Novel Wearable Technology: A Preliminary Validation Study. *JMIR MHealth UHealth*. 2022;10(2):e24916. doi:10.2196/24916

3. Mackey DC, Lachance CC, Wang PT, et al. The Flooring for Injury Prevention (FLIP) Study of compliant flooring for the prevention of fall-related injuries in long-term care: A randomized trial. *PLoS Med*. 2019;16(6):e1002843. doi:10.1371/journal.pmed.1002843

4. HIP ATTACK Investigators. Accelerated surgery versus standard care in hip fracture (HIP ATTACK): an international, randomised, controlled trial. *Lancet Lond Engl*. 2020;395(10225):698-708. doi:10.1016/S0140-6736(20)30058-1

5. Blauth M, Joeris A, Rometsch E, et al. Geriatric fracture centre vs usual care after proximal femur fracture in older patients: what are the benefits? Results of a large international prospective multicentre study. *BMJ Open*. 2021;11(5):e039960. doi:10.1136/bmjopen-2020-039960

6. FAITH Investigators. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. *Lancet Lond Engl*. 2017;389(10078):1519-1527. doi:10.1016/S0140-6736(17)30066-1

7. HEALTH Investigators, Bhandari M, Einhorn TA, et al. Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. *N Engl J Med*. 2019;381(23):2199-2208. doi:10.1056/NEJMoa1906190

8. Bzovsky S, Comeau-Gauthier M, Schemitsch EH, et al. Factors Associated With Mortality After Surgical Management of Femoral Neck Fractures. *J Orthop Trauma*. 2020;34 Suppl 3:S15-S21. doi:10.1097/BOT.0000000000001937

9. Sadeghirad B, Foroutan F, Zoratti MJ, et al. Theory and practice of Bayesian and frequentist frameworks for network meta-analysis. *BMJ Evid-Based Med*. Published online June 27, 2022:bmjebm-2022-111928. doi:10.1136/bmjebm-2022-111928

10. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ*. 2020;371:m3900. doi:10.1136/bmj.m3900

11. Brignardello-Petersen R, Izcovich A, Rochwerg B, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ*. 2020;371:m3907. doi:10.1136/bmj.m3907

12. Busse JW, Sadeghirad B, Oparin Y, et al. Management of Acute Pain From Non-Low Back, Musculoskeletal Injuries : A Systematic Review and Network Meta-analysis of Randomized Trials. *Ann Intern Med*. 2020;173(9):730-738. doi:10.7326/M19-3601

13. Nanji K, Sarohia GS, Kennedy K, et al. The 12- and 24-Month Effects of Intravitreal Ranibizumab, Aflibercept, and Bevacizumab on Intraocular Pressure: A Network Meta-Analysis. *Ophthalmology*. 2022;129(5):498-508. doi:10.1016/j.ophtha.2021.11.024