

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

Quantitative Synthesis
Chapter 5. Network Meta-Analysis

Prepared for:
The Agency for Healthcare Research and Quality (AHRQ)
Training Modules for Systematic Reviews Methods Guide
www.ahrq.gov

Learning objective

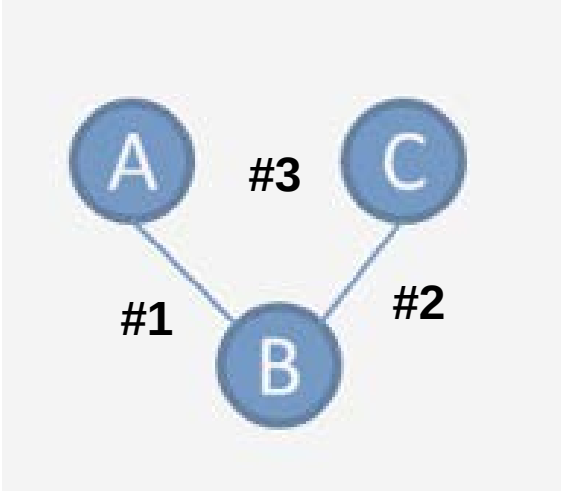
- Understand the network meta-analysis approach and when it can be implemented.

Network meta-analysis

- Decision-makers want head-to-head estimates of different interventions that are being chosen, but head-to-head trials are rare.
  - Most trials compare active agent to placebo.
  - This leaves patients and clinicians largely unable to directly compare across treatments
- **Network meta-analysis (NMA)** can be used to compare one intervention to another, when they have not been analyzed in the same trial.
  - This approach involves comparing agents indirectly.

Network meta-analysis schematic

- **Network meta-analysis** can be explained thus:
  - If we know that intervention A is better than B by a certain amount (#1 in figure),
  - And we know how B compares with C (#2 in figure),
  - We can *indirectly infer* the effect magnitude of A versus C (#3 in figure).
- NMA terminology:
  - A direct comparison is referred to as a *closed loop* (e.g., #1 and #2)
  - When there is no closed loop, direct comparison is not possible and indirect comparison is required (e.g., #3)



Assumptions

- 3 assumptions are required for NMA to be valid:
  - 1. Homogeneity of direct evidence
  - 2. Transitivity, similarity, or exchangeability
  - 3. Consistency between direct and indirect evidence

1. Homogeneity of direct evidence

- Homogeneity of treatment effects across trials decreases confidence in pooled results in any meta-analysis, including NMA.[1]
- In NMA direct evidence (within each pairwise comparison) should be sufficiently homogeneous.
- This homogeneity of direct evidence can be assessed using standard methods for evaluating homogeneity:
  - I^2
  - $\boldsymbol{\tau^2}$
  - Cochran *Q* test
  - Forest plots

1. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. 2011;64(12):1294-302. PMID: 21803546. http://dx.doi.org/10.1016/j.jclinepi.2011.03.017

Transitivity

- Participants enrolled in trials of different comparisons in a network must be sufficiently similar.[2]
  - ► This is defined with reference to distribution of effect modifiers.
- Patients should be similar such that it's plausible that they were equally likely to received any of the treatments in the network.
- Active and placebo controlled interventions must be sufficiently similar across trials as well, to attribute treatment effects to intervention.
- Transitivity cannot be empirically tested but should be conceptually considered.

2. Mills EJ, Ioannidis JP, Thorlund K, et al. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 2012;308(12):1246-53. PMID: 23011714. http://dx.doi.org/10.1001/2012.jama.11228

Consistency (indirect and direct)

- Consistency refers to agreement between direct and indirect estimates for the same treatment comparison.[3,4]
  - In closed loops, direct and indirect estimates of treatment effect should be similar to ensure consistency (previously known as coherence).
- Inconsistency may be due to several factors, including: differences in patients, treatments, settings, and timing.
- Statistical models exist that assume consistency (consistency models) and which allow for direct/indirect estimate inconsistency (inconsistency models).
- Consistency can and should be empirically evaluated.
- Notable inconsistency suggests NMA should not be performed.

3. Higgins JPT, Jackson D, Barrett JK, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods. 2012;3(2):98-110. http://dx.doi.org/10.1002/jrsm.1044 172.
4. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3(2):80-97. http://dx.doi.org/10.1002/jrsm.1037

Statistical approaches

- Simple indirect comparisons apply when there is no closed loop in the evidence network.
- The ***simple indirect comparison*** approach is to qualitatively compare the point estimates and overlap of confidence intervals.
  - The treatments likely have comparable effectiveness if their direct effects relative to common comparator (e.g., placebo) have same direction and magnitude, with substantial confidence interval overlap.
  - Such results must have interpreted cautiously because confidence interval overlap is not a reliable substitute for formal hypothesis testing.

Simple indirect comparison

- At least 3 statistical methods exist to conduct simple indirect comparison:
  - 1. Adjusted indirect comparison method of Bucher *et al*.[5]
  - 2. Logistic regression
  - 3. Random effects meta-regression
- With only 2 sets of trials (e.g., A vs B; B vs C), ***Bucher's method*** is sufficient:

$$\log(OR_{AC}) = \log(OR_{AB}) - \log(ORBC)$$

$$Var(\log(OR_{AC})) = Var(\log(OR_{AB})) + \mathrm{Var}(\log(OR_{BC}))$$

- This method is valid only assuming normality on the log scale.

5. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683-91. PMID: 9250266

Simple indirect comparison

- ***Logistic regression*** works with arm-level binary outcomes and generates OR as the outcome.[6]
- ***Meta-regression*** uses contrast-level data and can be extended to RRs, RDs, SMDs, and other effect measures.[6]
- Given assumptions (i.e., no differences in prognostic factors between included studies), all 3 methods yield unbiased estimates of direct effects.
- Meta-regression and adjusted indirect comparisons are more convenient approaches for comparing trials with 2 treatment arms; random effects should be used for both.

6. Glenny A, Altman D, Song F, et al. Indirect comparisons of competing interventions. Health Technol Assess 2005; PMID: 16014203

Frequentist NMA models

- The first frequentist NMA model is a random-effects inconsistency model proposed by Lumley *et al*.[7]
  - Included studies cannot have >2 arms.
  - Incorporates sampling variability, heterogeneity, and inconsistency.
- Further developments in the frequentist framework have enabled >2 treatment arms per study and utilized new methods of addressing inconsistency.
  - A general NMA formulation has been proposed by Salanti *et al*., defining inconsistency in a standard way.[8]
  - A treatment-by-design interaction has been proposed as an alternative inconsistency definition by White *et al*. and Higgins *et al*., within a meta-regression framework.[3,9]

7. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002;21(16):2313-24. PMID: 12210616. http://dx.doi.org/10.1002/sim.1201
8. Salanti G, Higgins JP, Ades A, et al. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008;17(3):279-301. PMID: 17925316. http://dx.doi.org/10.1177/0962280207080643
9. White IR, Barrett JK, Jackson D, et al. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods. 2012;3(2):111-25. http://dx.doi.org/10.1002/jrsm.1045

Bayesian NMA models

- A Bayesian NMA approach has been introduced that represents treatment effects as ***basic parameters*** and ***functional parameters***.[10]
  - Basic parameters are effect parameters that are directly compared to baseline treatment.
  - Functional parameters are represented as functions of basic parameters.
- In this framework, evidence inconsistency is defined as a function of a functional parameter with at least 2 basic parameters.
- This model has been extended to include study-level covariates (explaining heterogeneity), repeated measurements, or to appraise novelty effects.

10. Lu G, Ades A. Assessing Evidence Inconsistency in Mixed Treatment Comparisons J Am Stat Assoc. 2006;101(20):447-59 https://doi.org/10.1198/016214505000001302

Bayesian NMA models

- A vague (flat or uniform) prior is commonly chosen for the treatment effect and heterogeneity parameters in Bayesian NMA.[11]
  - A vague prior distribution for heterogeneity may not be appropriate when the number of studies is small.
  - An informative prior for heterogeneity can be obtained from the empirically observed distribution of heterogeneity in various settings.
- In NMA, frequentist and Bayesian approaches often yield similar results due to the common practice of using non-informative priors in Bayesian models.[12,13]

11. Higgins J, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med. 1996;15(24):2733-49. PMID: 8981683. http://dx.doi.org/10.1002/(SICI)10970258(19961230)15:243.0.CO;2-0
12. Greco T, Landoni G, Biondi-Zoccai G, et al. A Bayesian network meta-analysis for binary outcome: how to do it. Stat Methods Med Res. 2013. http://dx.doi.org/10.1177/096228021350018 5 188.
13. Hong H, Carlin BP, Shamliyan TA, et al. Comparing Bayesian and Frequentist Approaches for Multiple Outcome Mixed Treatment Comparisons. Med Decis Making. 2013;33(5):702-14. http://dx.doi.org/10.1177/0272989X13481110

Arm-based, contrast-based models

- In NMA, ***arm-based models*** differ from ***contrast-based models***.
  - This distinction refers to how outcomes are reported in the included studies.
  - Arm-level data are raw data per study arm.
  - Contrast-level data show the difference in outcomes between study arms, whether absolute (e.g., RD) or relative (e.g., RR, OR).
- ***Contrast-based models*** resemble traditional meta-analysis approaches for direct-effect pooling.
  - These models preserve randomization and alleviate concern about between-arm differences.
  - They use effect sizes relative to a common comparison group.

Arm-based models in NMA

- Although contrast-based models are the dominant NMA framework, ***arm-based models*** are an important NMA variant.
  - These combine observed absolute effect size in individual arms across studies, thereby producing a pooled rate/mean outcome per arm.
  - Estimates can then be compared across pooled arms to yield treatment effects.
- Because they break randomization, arm-based models lack the protections against bias afforded by randomization. Comparative estimates are thus at increased risk of bias.
  - Thus, non-randomized studies may be included in this framework.
  - The validity of the arm-based framework is an active area of debate.

Assessing consistency

- NMA generates results for all pairwise comparisons, but only closed loops enable assessment of consistency of evidence.
  - Network must have one comparison with direct evidence.
- 2 broad types of methods exist for evaluating consistency:
  - ***Overall consistency measures*** for the entire network.
  - ***Loop-based approaches*** in which direct and indirect estimates are compared.

Single measures for consistency

- These approaches use a single measure to represent consistency within a whole network.
- The approach of Lumley assumes that each treatment comparison has a different inconsistency factor, which follows a common random-effects distribution.[7]
- The variance of the differences is called ***incoherence***, and is represented as ω.[7]
  - ω quantifies the overall inconsistency across the network.
  - A value of ω >0.25 suggests substantial inconsistency and recommends against NMA.[11]

11. van der Valk R, Webers CAB, Lumley T, et al. A network meta-analysis combined direct and indirect comparisons between glaucoma drugs to rank effectiveness in lowering intraocular pressure. J Clin Epidemiol. 2009;62(12):1279-83. http://dx.doi.org/10.1016/j.jclinepi.2008.04.012

Single measures for consistency

- ***Global Wald test:*** Tests an inconsistency factors that follows a X^2 distribution under a null consistency assumption.[9]
  - P<0.10 supports rejecting the null of consistency.

Loop-based approaches

- This approach involves comparing direct and indirect estimates for each comparison in a network.
- This framework preferred over a global approach:
  - Global approaches conceal important sources of inconsistency, despite being easily interpreted.
- Various approaches exist for comparing direct and indirect estimates:
  - Z-test
  - Side-splitting

Loop-based approaches

- ***Z-test:*** Used to compare difference of pooled effect sizes for direct and indirect comparisons. [5]
  - Advantages include simplicity, ease of application, ability to identify specific loops with large inconsistency.
  - Limitations include the need for multiple correlated tests.
- ***Side-splitting:*** A *node* is a treatment and a *side/edge* is a comparison.[12]
  - Compares the pooled estimate from direct evidence only to the pooled estimate including both direct and indirect evidence.

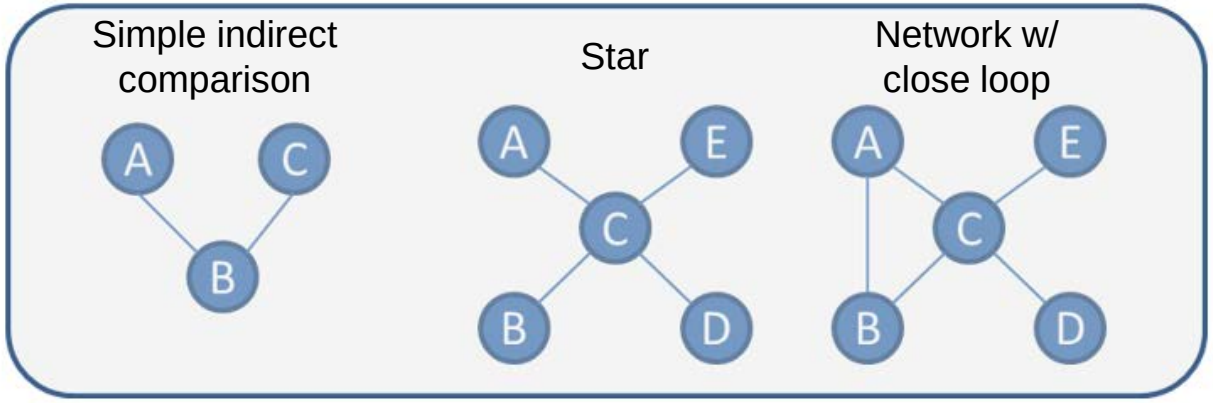
12. Dias S, Welton NJ, Caldwell DM, et al. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29(7-8):932-44. http://dx.doi.org/10.1002/sim.3767

Evaluating consistency

- Lack of statistical significance of inconsistency does not prove consistency within a network.
- Statistical tests for inconsistency are commonly under-powered due to limited number of studies with direct comparisons.
- When important inconsistency is identified within a network, options include:
  - Abandon NMA and perform only traditional meta-analysis.
  - Present results from inconsistency models and acknowledge interpretation issues that exist.
  - Split the network to eliminate inconsistent nodes.
  - Attempt to explain inconsistency with network meta-regression.
  - Use only direct estimates for NMA nodes that demonstrate inconsistency.

Choice of NMA method

- There has been no systematic evaluation of the comparative performance of NMA methods.
- Investigators should test relevant assumptions where possible, regardless of method chosen.
- Investigators should refrain from combining multiple sources of evidence in an inconsistent network.
- ***Network geometry*** also affects model choice:



Network geometry and model choice

Methods	Simple indirect comparison	Star network	Network with at least one closed loop
Qualitative assessment	X		
Adjusted Indirect comparison, random-effects meta regression, logistic regression	X	X	
Lumley's mixed-effects linear regression			X
Mixed effects and hierarchical models		X	X

X: Appropriate method for the network geometry.

Rating strength of evidence

- Original EPC and GRADE evidence was simple and involved rating down all evidence from indirect comparisons.
- More recently, GRADE published a new approach based on evaluating strength of evidence separately rather than for the entire network collectively.[13]
  - The rationale is that the strength of evidence likely varies by the specific comparison.
  - The approach involves presenting the 3 estimates for each comparison (direct, indirect, and network estimates), then rating strength of evidence separately fore ach one.

13. 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. PMID: 26085374. http://dx.doi.org/10.1136/bmj.h3326

Ranking probabilities

- NMA results are commonly presented as probability of being most effective and as rankings of treatments.
- Results are also presented as the ***Surface Under the Cumulative Ranking Curve (SUCRA)***.[14]
  - SUCRA is a simple transformation of the mean rank that is used to provide a treatment hierarchy accounting for location and variance of all relative treatment effects.
  - SUCRA=1: treatment certain to be the best; SUCRA=0: treatment certain to be the worst.
- These presentations must be interpreted with caution.

14. 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(2):163-71. http://dx.doi.org/10.1016/j.jclinepi.2010.03.016

Ranking interpretation concerns

- 3 concerns should be considered, whether results are presented as probabilities, rankings, or SUCRA:
  - Such estimates are usually very imprecise.
  - When rankings suggest superiority of one agent above others, the absolute difference might be trivial.[15]
    - Converting relative effect measure to an absolute effect aids in interpretation of clinical significance.
  - Rankings conceal that each comparison has its own risks of bias, limitations, and strength of evidence.

15. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA. 2014;312(2):171-9. PMID: 25005654. http://dx.doi.org/10.1001/jama.2014.555916

Reporting NMA results

- Published NMA results reflect a high degree of heterogeneity and numerous deficiencies.
- NMAs commonly reflect unclear understanding of assumptions and inappropriate search strategies, methods, and comparison of direct/indirect evidence.[16-18]
- The PRISMA statement has been extended to cover NMAs.[19]

16. Bafeta A, Trinquart L, Seror R, et al. Reporting of results from network meta-analyses: methodological systematic review. BMJ. 2014;348; PMID: 24618053. http://dx.doi.org/10.1136/bmj.g1741. 215.
17.Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ. 2009;338:b1147.216
18. Hutton B, Salanti G, Chaimani A, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. PloS One. 2014;9(3):e92508. http://dx.doi.org/10.1371/journal.pone.0092508
19. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. PMID: 19621072 http://dx.doi.org/10.1371/journal.pmed.1000097

Reporting NMA results

- The following information should be presented in NMA reporting:
  - Rationale for conducting NMA
  - Mode of inference (e.g., Bayesian, frequentist)
    - Choice of priors for Bayesian analysis
  - Model choice (random vs. fixed effects; consistency vs. inconsistency model; etc)
  - Software and syntax used
  - Graphical presentation of the network structure and geometry
  - Pairwise effect sizes
  - Assessment of consistency between direct and indirect estimates.

Recommendations

- Always base a network meta-analysis (NMA) on a rigorous systematic review.
- **For a NMA, three assumptions must be met:**
  - Homogeneity of direct evidence
  - Transitivity, similarity, or exchangeability
  - Consistency (between direct and indirect evidence)
- Investigators may choose a frequentist or Bayesian mode of inference based on the research team's expertise, complexity of evidence network, and the research question.
- **Evaluating inconsistency** is a major and mandatory component of network meta-analysis.
  - Conducting a global test should not be the only method used to evaluate inconsistency. A loop-based approach can identify the comparisons that cause inconsistency.
- **Cautiously use inference based on the rankings and probabilities of treatments being most effective.**
  - Rankings and probabilities can be misleading and should be interpreted based on the magnitude of pairwise effect sizes. Despite such rankings, differences across interventions may not be clinically important.

Author

- This presentation was prepared by Jonathan Snowden, Ph.D.
- Special thanks to Ethan Balk for his comments on this presentation.
- The presentation is based on the chapter entitled “Network Meta-Analysis (Mixed Treatment Comparisons/Indirect Comparisons)” in the Methods Guide for Comparative Effectiveness Reviews (available at: https://doi.org/10.23970/AHRQEPCMETHGUIDE3

References

1. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. 2011;64(12):1294-302. PMID: 21803546. http://dx.doi.org/10.1016/j.jclinepi.2011.03.017
2. Mills EJ, Ioannidis JP, Thorlund K, et al. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 2012;308(12):1246-53. PMID: 23011714. http://dx.doi.org/10.1001/2012.jama.11228
3. Higgins JPT, Jackson D, Barrett JK, et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods. 2012;3(2):98-110. http://dx.doi.org/10.1002/jrsm.1044 172.
4. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3(2):80-97. http://dx.doi.org/10.1002/jrsm.1037
5. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683-91. PMID: 9250266
6. Glenny A, Altman D, Song F, et al. Indirect comparisons of competing interventions. Health Technol Assess 2005; PMID: 16014203
7. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med. 2002;21(16):2313-24. PMID: 12210616. http://dx.doi.org/10.1002/sim.1201
8. Salanti G, Higgins JP, Ades A, et al. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008;17(3):279-301. PMID: 17925316. http://dx.doi.org/10.1177/0962280207080643
9. White IR, Barrett JK, Jackson D, et al. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods. 2012;3(2):111-25. http://dx.doi.org/10.1002/jrsm.1045
10. Lu G, Ades A. Assessing Evidence Inconsistency in Mixed Treatment Comparisons J Am Stat Assoc. 2006;101(20):447-59 https://doi.org/10.1198/016214505000000130 2
11. van der Valk R, Webers CAB, Lumley T, et al. A network meta-analysis combined direct and indirect comparisons between glaucoma drugs to rank effectiveness in lowering intraocular pressure. J Clin Epidemiol. 2009;62(12):1279-83. http://dx.doi.org/10.1016/j.jclinepi.2008.04.012
12. Dias S, Welton NJ, Caldwell DM, et al. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29(7-8):932-44. http://dx.doi.org/10.1002/sim.3767
13. 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. PMID: 26085374. http://dx.doi.org/10.1136/bmj.h3326
14. 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(2):163-71. http://dx.doi.org/10.1016/j.jclinepi.2010.03.016
15. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA. 2014;312(2):171-9. PMID: 25005654. http://dx.doi.org/10.1001/jama.2014.555916
16. Bafeta A, Trinquart L, Seror R, et al. Reporting of results from network meta-analyses: methodological systematic review. BMJ. 2014;348; PMID: 24618053. http://dx.doi.org/10.1136/bmj.g1741.215.
17.Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ. 2009;338:b1147.216
18. Hutton B, Salanti G, Chaimani A, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. PloS One. 2014;9(3):e92508. http://dx.doi.org/10.1371/journal.pone.0092508
19. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. PMID: 19621072 http://dx.doi.org/10.1371/journal.pmed.1000097