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 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:
  - $I^2$
  - $\tau^2$
  - Cochran Q test
  - Forest plots

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 receive 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.

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.


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:
  \[
  \log(OR_{AC}) = \log(OR_{AB}) - \log(OR_{BC})
  \\
  Var(\log(OR_{AC})) = Var(\log(OR_{AB})) + Var(\log(OR_{BC}))
  \]
- This method is valid only assuming normality on the log scale.

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 RR, RD, SMD, 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.

---

Frequentist NMA models

• The first frequentist NMA model is a random-effects inconsistency model proposed by Lumley et al.\textsuperscript{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.\textsuperscript{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.\textsuperscript{3,9}

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.

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

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 \(\omega\).\(^7\)
  - \(\omega\) quantifies the overall inconsistency across the network.
  - A value of \(\omega >0.25\) suggests substantial inconsistency and recommends against NMA.\(^{11}\)

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.

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:

```
Simple indirect comparison
A  C
B

Star
A  E
C

Network w/ close loop
A  C  E
B  D
```
## Network geometry and model choice

<table>
<thead>
<tr>
<th>Methods</th>
<th>Simple indirect comparison</th>
<th>Star network</th>
<th>Network with at least one closed loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative assessment</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted Indirect comparison, random-effects meta regression, logistic regression</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lumley’s mixed-effects linear regression</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Mixed effects and hierarchical models</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**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. 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 for each one.

• 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)**.\(^\text{14}\)
  
  ▶ 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.

---

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.

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

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.
• 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


