Evaluating the quality of evidence from a network meta-analysis

Julian Higgins¹
with Cinzia Del Giovane², Anna Chaimani³, Deborah Caldwell¹, Georgia Salanti³

¹School of Social and Community Medicine, University of Bristol, UK
²Statistics Unit, Department of Clinical and Diagnostic Medicine and Public Health, University of Modena and Reggio Emilia, Modena, Italy
³Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

Overview

- Evidence network
  - e.g. topical antibiotics for chronically discharging ears
- GRADE
- Network meta-analysis
- Application of GRADE to network meta-analysis
- Particular attention to
  - transitivity of effects (indirect comparison)
  - contributions of different bits of evidence
Evidence network: topical antibiotics for chronically discharging ears

Outcome: persistent discharge from the ear after 1 week

Quality of this evidence?

Ou coe of this evidence?
GRADE in a nutshell

- **High**: Further research is very unlikely to change our confidence in the estimate of effect.
- **Moderate**: Further research likely to have impact on confidence in the estimate of effect and may change the estimate.
- **Low**: Further research is very likely to have an important impact on our confidence in the estimate of effect, and is likely to change the estimate.
- **Very low**: Any estimate of effect is very uncertain.

Randomized trials start here

5 reasons to **downgrade**
What might decrease quality of evidence

1. Risk of bias
2. Indirectness of evidence
3. Inconsistency of results
4. Imprecision
5. Publication bias

• For each category, 3 scoring options:
  • No concerns
  • Serious −1 level
  • Very serious −2 levels

Network meta-analysis: simultaneous analysis of all the evidence
Network meta-analysis: simultaneous analysis of all the evidence

Comparison

Indirect comparison
Assumes transitivity

Network meta-analysis
assumes
Coherence
among sources of evidence

Network meta-analysis: treatment rankings
Contributions to the analysis

<table>
<thead>
<tr>
<th>Direct comparisons in the network</th>
<th>AB</th>
<th>AD</th>
<th>BC</th>
<th>BD</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct &amp; Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>12.0</td>
<td>39.0</td>
<td>10.0</td>
<td>26.0</td>
<td>10.0</td>
</tr>
<tr>
<td>AD</td>
<td>12.3</td>
<td>72.3</td>
<td>3.1</td>
<td>9.1</td>
<td>3.2</td>
</tr>
<tr>
<td>BC</td>
<td>1.9</td>
<td>1.9</td>
<td>66.6</td>
<td>13.9</td>
<td>15.8</td>
</tr>
<tr>
<td>BD</td>
<td>6.8</td>
<td>6.8</td>
<td>17.6</td>
<td>51.2</td>
<td>17.6</td>
</tr>
<tr>
<td>CD</td>
<td>4.1</td>
<td>4.1</td>
<td>68.0</td>
<td>30.9</td>
<td>25.9</td>
</tr>
<tr>
<td>Indirect only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>8.9</td>
<td>32.8</td>
<td>25.5</td>
<td>16.7</td>
<td>16.1</td>
</tr>
<tr>
<td>Entire network</td>
<td>8.0</td>
<td>26.6</td>
<td>26.3</td>
<td>26.0</td>
<td>15.1</td>
</tr>
<tr>
<td>Included studies</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

1. Evaluating risk of bias

**Pairwise estimates**

Risk of Bias (RoB) judgments for each direct estimate (as in GRADE)

- moderate
- low
- high

For each pairwise comparison, integrate the RoB assessments and the respective contributions

**Ranking**

Integrate the RoB assessments and the contribution of each direct evidence to the network as a whole
2. Evaluating indirectness

Consider populations, treatments and outcomes (as in GRADE)

Examine similarity of effects modifiers across sources of direct evidence
- transitivity assumption

- consider contributions of direct evidence to each pairwise network estimates
- consider whole network

3. Evaluating inconsistency

Inconsistency
- Heterogeneity
- Incoherence

Disagreement (between-study variance) within direct evidence
- Comparison-specific $\tau^2$
- Assumed-common $\tau^2$
- Empirical evidence on $\tau^2$

Disagreement between direct and indirect evidence
- Loop coherence
- Node splitting
- Design-by-treatment test
- Comparison of model fit

Statistical consequences of between-study differences in population, treatments, outcomes and biases
4. Evaluating imprecision

- Examine the width of the confidence intervals (as in GRADE)
- Exclude clinically relevant effect sizes?

5. Evaluating publication bias

- Use GRADE criteria
- Non-statistical consideration of likelihood of non-publication of evidence
- Could consider a symmetry in contour-enhanced funnel plots for each pair-wise comparison
- Could consider comparison-adjusted funnel plot for the network
- Take into account the contributions of each direct piece of evidence
Summary of confidence in effect estimates

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Nature of the evidence</th>
<th>Confidence</th>
<th>Downgrading due to</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB: Quinolone antibiotic vs no treatment</td>
<td>Mixed</td>
<td>Low</td>
<td>Study limitations¹; Indirectness²</td>
</tr>
<tr>
<td>AC: Non-quinolone antibiotic vs no treatment</td>
<td>Indirect only</td>
<td>Low</td>
<td>Study limitations¹; Indirectness²</td>
</tr>
<tr>
<td>AD: Antiseptic vs no treatment</td>
<td>Mixed</td>
<td>Very low</td>
<td>Study limitations¹; Indirectness²; Imprecision³</td>
</tr>
<tr>
<td>BC: Non-quinolone antibiotic vs quinolone antibiotic</td>
<td>Mixed</td>
<td>Very low</td>
<td>Study limitations¹; Indirectness²; Imprecision³</td>
</tr>
<tr>
<td>BD: Antiseptic vs quinolone antibiotic</td>
<td>Mixed</td>
<td>Moderate</td>
<td>Inconsistency⁴</td>
</tr>
<tr>
<td>CD: Antiseptic vs non-quinolone antibiotic</td>
<td>Mixed</td>
<td>Very low</td>
<td>Study limitations¹; Indirectness²; Imprecision³</td>
</tr>
</tbody>
</table>

| Ranking of treatments                  | Low                     | Study limitations¹; Indirectness² |

¹Dominated by evidence at high or moderate risk of bias.
²No convincing evidence for the plausibility of the transitivity assumption.
³Predictive intervals include effects with different interpretations (also no convincing evidence for plausibility of transitivity assumption).
⁴Confidence intervals include values favouring either treatment.
⁵60% of the information is from studies at moderate risk of bias.
⁶Moderate level of heterogeneity, and some evidence of incoherence in the network.

Concluding remarks

- Confidence in findings of network meta-analysis should consider both pair-wise estimates and any ranking of treatments
- Key issues in addition to pair-wise meta-analysis are
  - transitivity assumption required for indirect comparisons
    - extension of ‘directness’ in GRADE
  - coherence between direct and indirect evidence
    - extension to ‘inconsistency’ in GRADE
- Our suggestions are workable but subjective
- We encourage use of the contributions of direct evidence to the network estimates and the network as whole
References


Example:

Contributions of direct evidence: