Network Meta-Analysis for Comparative Effectiveness Research

Joseph C. Cappelleri, Ph.D., M.P.H.
Pfizer Inc
New London, CT

e-mail:joseph.c.cappelleri@pfizer.com

Seminar presented at the Meeting of the New York Chapter of the American Statistical Association, November 22, 2010
The latest research shows that we really should do something with all this research.
Outline

• Introduce comparative effectiveness research (CER)
• Explain the concepts in network meta-analysis
• Describe the assumptions of network meta-analysis
• Illustrate its application
Comparative Effectiveness Research and the 2009 Stimulus Bill

- 2009 American Recovery and Reinvestment Act, or “Stimulus Bill,” provided $1.1 billion to support comparative effectiveness research:
  - $300 million to Agency for Healthcare Research and Quality
  - $300 million to the National Institutes of Health
  - $400 million to the Office of the Secretary of Health and Human Services (HHS)

- To evaluate the relative effectiveness of different health care services and treatment options

- To encourage the development and use of clinical registries, clinical data networks, and other forms of electronic data to generate outcomes data

- $1.5 million to support an Institute of Medicine study – to make recommendations to HHS Secretary to establish national priorities on comparative effectiveness research
What is Comparative Effectiveness Research?

♦ A type of systematic review
  – Synthesizes available scientific evidence on a specific topic

♦ Expands the scope of a typical systematic review
  – Goes beyond the effectiveness of a single intervention
  – Compares the relative benefits and harms among a range of available treatments or interventions for a given condition

♦ Parallels decisions facing clinicians, patients, and policy makers who must choose among a variety of alternatives in making diagnostic, treatment, and health-care delivery decisions
  – Three key elements: relevance, timeliness, transparency
Motivation for Comparative Effectiveness Research

♦ Decision makers are often faced with more than one viable treatment option

♦ Consider a 55-year-old woman whose bone scan shows greatly decreased bone density
  – Should she take drugs, increase vitamin D and calcium intake, focus on weight-bearing exercises, or watchfully wait?
  – Drugs are effective but limited information on their long-term effects
  – Some women will develop kidney stones after calcium intake
  – No precise formulation on effective exercise prescription
Despite mounting evidence from 18 trials spanning and evaluating 24 regimens, evidence is available only on a few direct comparisons.

Source: Ioannidis 2006
Four Principal Steps in Comparative Effectiveness Reviews

- Step 1: Formulate the problem
- Step 2: Define the studies and search strategies
- Step 3: Evaluate applicability of studies
- Step 4: Assess benefits and harms of treatments
Examples of Comparative Effectiveness Research

♦ Oral Medications for Type 2 Diabetes Mellitus
  – Annals of Internal Medicine 2007; 147:386-399
  – Includes indirect treatment comparisons

♦ Diagnosis and Treatment of Erectile Dysfunction

♦ Treatment of Overactive Bladder in Women
Quantitative Synthesis: Meta-Analysis

♦ Integral part of comparative effectiveness research and reviews

♦ Should be performed to address pre-specified questions, following PRISMA guidelines (Liberati et al. 2009)

♦ Clinical and methodological diversity, as well as statistical heterogeneity, should be considered before pooling studies to calculate summary effect
How Does CER Meta-Analysis Differ from Traditional Meta-Analysis?

- CER meta-analysis is more expansive
  - Standard meta-analysis is subsumed within CER meta-analysis

- CER meta-analysis involves all relevant treatments (even if not directly compared), not just one particular treatment or class of treatments

- CER meta-analysis considers wider net of evidence, not just from a particular type of study design and not just efficacy

- CER meta-analysis places even greater emphasis on heterogeneity, Bayesian methods, and updating results
Network Meta-Analysis

♦ Network meta-analysis is a key part of CER
  – Needed when there’s little or no evidence from head-to-head (direct) comparisons
  – Interventions of interest with a common comparator

♦ Network meta-analysis enables us to combine trials involving different sets of treatments, using a network of evidence, within a single analysis

♦ This integrated and unified analysis incorporates all direct and indirect comparative evidence about treatments
The main drawback is that meta-analysis focuses on comparing only two alternatives at a time, leading to a plethora of analyses to interpret with no quantitatively rigorous methods for integrating them (Source: Schmid 2010)
Two Specific Types of Network Meta-Analysis

- Indirect comparison – when only two (or one pair of) treatments are being compared indirectly

- Mixed treatment comparisons – a generalization of indirect comparisons with more than two (or multiple pairs of) treatments being compared indirectly
  - At least one pair of treatments is compared both directly and indirectly

- Extensions of standard pairwise meta-analysis of randomized control trials
  - Fixed-effect and random-effect network meta-analysis

- Relies on statistical methods that maintain benefits of randomization within each trial
Examples of Evidence Networks

- **Indirect Treatment Comparison**
- **Mixed Treatment Comparisons**

**Closed loops in network: combination of direct and indirect evidence**
Indirect Comparison

Solid Line: Direct Comparison
Dashed Line: Indirect Comparison

Source: Adapted from Jansen et al. 2008
Mixed Treatment Comparisons

Source: Adapted from Jansen et al. 2008
## Indirect Comparisons of Multiple Treatments

<table>
<thead>
<tr>
<th>Trial</th>
<th>Arms</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A, B</td>
<td>Want to compare A vs. B</td>
</tr>
<tr>
<td>2</td>
<td>A, B</td>
<td>Direct evidence from trials 1, 2 and 7</td>
</tr>
<tr>
<td>3</td>
<td>B, C</td>
<td>Indirect evidence from trials 3, 4, 5, 6 and 7</td>
</tr>
<tr>
<td>4</td>
<td>B, C</td>
<td>Combining all “A” arms and comparing with all “B” arms destroys randomization</td>
</tr>
<tr>
<td>5</td>
<td>A, C</td>
<td>Use indirect evidence of A vs. C and B vs. C comparisons as additional evidence to preserve randomization and within-study comparison</td>
</tr>
<tr>
<td>6</td>
<td>A, C</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>A, B, C</td>
<td></td>
</tr>
</tbody>
</table>

Source: Schmid 2010
How Is An Indirect Comparison Made?
Frequentist Approach

Calculate effect of A vs. C and B vs. C separately

\[ T_{AB} = T_{AC} - T_{BC} \]

with its standard error being the square root of sum of variances

\[ \text{square root of } \text{Var}(T_{AC}) + \text{Var}(T_{BC}) \]

(Bucher et al. 1997)
Four treatments (A, B, C, D) with treatment A as reference.

Relative treatment effects (e.g., log odds ratios) of B, C, D relative to A are the **basic** parameters

\[ d_{AB}, d_{AC}, d_{AD} \]

Remaining contrasts are **functional** parameters

\[
\begin{align*}
    d_{BC} &= d_{AC} - d_{AB} \\
    d_{BD} &= d_{AD} - d_{AB} \\
    d_{CD} &= d_{AD} - d_{AC}
\end{align*}
\]

Basic parameters determine functional parameters.

Functional parameters inform indirectly from basic parameters.
Multiple Treatments Model

Again, same four treatments with treatment A as reference
Consider a binary outcome
Likelihood: \( r_{ik} \sim \text{Bin}(\pi_{ik}, n_{ik}) \) for treatment k in study i

Model: Each trial compares treatments b and k indirectly through A

\[
\text{Logit}(\pi_{ik}) = \eta_{ib} \quad \text{if } k = b
\]
\[
= \eta_{ib} + \delta_{i(b,k)} \quad \text{if } k \neq b
\]
\[
\delta_{i(b,k)} \sim \mathcal{N}(d_{bk} = d_{Ak} - d_{Ab}, \sigma^2) \quad \text{(random-effects model)}
\]

\( b = A, B, C \quad (d_{AA} = 0) \)

\( k = B, C, D \)

Note how functional and basic parameters inform each other
Note: Add priors on \( d_{Ak} \) and \( \sigma^2 \) for Bayesian analysis

Source: Lu and Ades 2004
What are the Basic Assumptions of Network Meta-Analysis?

- Homogeneity assumption for standard meta-analysis
- Similarity assumption for indirect comparison
- Consistency assumption for the combination of direct and indirect evidence
What Can Go Wrong If Assumptions Are Not Met?
Meta-analysis of Risperidone versus Haloperidol for Schizophrenia: Outcome is No Clinical Improvement (Frequentist Approach with Random-Effects Model)

Adjusted indirect comparison:

\[ \text{lnOR}'_{RH} = \text{lnOR}_{RP} - \text{lnOR}_{HP} \]

\[ \text{lnOR}'_{RH} = -0.909 - (-1.707) = 0.798 \]

Standard error \( SE(\text{lnOR}'_{RH}) = \text{square root of } \{[SE(\text{lnOR}_{RP})]^2 + [SE(\text{lnOR}_{HP})]^2] \}

\[ SE(\text{lnOR}'_{RH}) = \text{square root of } [(0.218)^2 + (0.318)^2] = 0.386 \]

Results suggest that risperidone was less efficacious than haloperidol: Odds Ratio = \( \exp(0.798) = 2.22; \) 95% CI = \( \exp(0.798 \pm 1.96 \times 0.386) = 1.04 \) to 4.72

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Number of trials</th>
<th>Log odds ratio (SE)</th>
<th>Odds ratio (95% CI)</th>
<th>I²%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo controlled trials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risperidone vs placebo</td>
<td>3</td>
<td>-0.909 (0.218)</td>
<td>0.40 (0.26, 0.62)</td>
<td>37%</td>
</tr>
<tr>
<td>Haloperidol vs placebo</td>
<td>9</td>
<td>-1.707 (0.318)</td>
<td>0.18 (0.10, 0.34)</td>
<td>11%</td>
</tr>
<tr>
<td>Risperidone vs haloperidol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct comparison</td>
<td>10</td>
<td>-0.262 (0.142)</td>
<td>0.77 (0.58, 1.02)</td>
<td>14%</td>
</tr>
<tr>
<td>Adjusted indirect comparison</td>
<td>3/9</td>
<td>0.798 (0.386)</td>
<td>2.22 (1.04, 4.72)</td>
<td></td>
</tr>
<tr>
<td>Combination of direct and indirect estimate</td>
<td>10+(3/9)</td>
<td>0.207 (0.527)</td>
<td>1.23 (0.44, 3.45)</td>
<td>85%</td>
</tr>
</tbody>
</table>

NB: Random-effects model was used in meta-analyses of trials and for the combination of the direct and indirect estimates. Odds ratio = \( \exp(\text{log odds ratio}) \).

Cl: confidence interval; SE: standard error

Source: Song 2009

But in the 10 head-to-head comparisons, risperidone tended to be more efficacious than haloperidol: Odds ratio = 0.77, 95% CI = 0.58 to 1.02. This example of inconsistent evidence between indirect and direct estimates calls into question combining them.
Results of Different Methods of Comparing Risperidone and Haloperidol for Schizophrenia
(Outcome: Not Clinically Improved)

Informal indirect comparison: Odds ratio of haloperidol vs. placebo suggested a greater treatment effect than the odds ratio of risperidone vs. placebo, despite the overlapping confidence intervals.

Formal indirect comparison: Favors haloperidol over risperidone.

Direct comparison: Favors risperidone over haloperidol.

Combination of direct and indirect: Validity doubtful given their inconsistent evidence.

Source: Song 2009
Homogeneity Assumption for Standard Meta-Analysis

- Results from multiple trials can be pooled in meta-analyses before an indirect comparison is conducted.

- In standard meta-analysis, it is assumed that different trials estimate the same single effect (fixed-effects model) or different effects are distributed around a typical value (random-effects model).
  - The underlying assumption is the trials are sufficiently homogeneous to be quantitatively combined.
  - Heterogeneity can be tested using chi-square test and I-square (the proportion of total variation in results that is due to heterogeneity rather than chance).
Indirect Comparison Between Treatments A and B: Example from Three Trials

Note: In contrast to direct within-trial comparison, indirect comparison means a between-study comparison of different interventions. Here the indirect comparison of A and B is adjusted according to the results of their direct comparison with a common intervention, C.
Similarity Assumption for Indirect Comparison

♦ This assumption requires that the patients included should sufficiently similar in the sets of randomized-controlled studies
  – If so, the relative effect estimated by trials of A vs. C is generalizable to patients in trials of B vs. C (and vice versa)

♦ In addition to clinical similarity, methodological similarity (e.g., quality, definition of outcomes) is required for valid estimates
  – If there is imbalance in the distribution of effect modifiers (treatment-by-covariate interactions) between trials, then estimates become biased

♦ Indirect assessment of risperidone versus haloperidol for schizophrenia
  – Patient characteristics, dose of drug, and treatment duration were similar between the two sets of placebo-controlled trials
  – But clinical improvement was defined differently
    • Placebo-controlled trials of risperidone: 20% or more greater reduction in total score on the Positive and Negative Syndrome Scale or Brief Psychiatric Rating Scale
    • Placebo-controlled trials of haloperidol: rated by clinicians using the Clinical Global Impression or other scales
When both direct and indirect evidence are available, an assumption of consistency is required to combine the direct and indirect estimates.

Inconsistent results between them (say, as measured by I-square) may give invalid and misleading results.

Example: risperidone versus haloperidol for schizophrenia (large I-square = 85%)

When results are inconsistent, it is important to investigate possible causes of discrepancy.

Methods have been proposed to evaluate consistency (Salanti et al. 2008; Dias et al. 2010)
Suppose we have AB, AC, BC direct evidence

Indirect estimate:

\[ \hat{d}_{BC}^{\text{indirect}} = \hat{d}_{AC}^{\text{direct}} - \hat{d}_{AB}^{\text{direct}} \]

Measure of inconsistency:

\[ \hat{\omega}_{BC} = \hat{d}_{BC}^{\text{indirect}} - \hat{d}_{BC}^{\text{direct}} \]

Approximate test (normal distribution):

\[ z_{BC} = \frac{\hat{\omega}_{BC}}{\sqrt{V_{\hat{\omega}_{BC}}}} \]

with variance:

\[ V_{\hat{\omega}_{BC}} = V_{\hat{d}_{BC}^{\text{direct}}} + V_{\hat{d}_{AC}^{\text{direct}}} + V_{\hat{d}_{AB}^{\text{direct}}} \]
Bayesian Network Meta-Analysis

- Bayesian methods have been developed to conduct network meta-analysis of multiple treatments and to combine direct and indirect evidence
  - Well-suited for mixed treatment comparisons

- Allows probability statements that one drug is better (e.g., more efficacious, safer) than another

- Provides probability calculation that a particular drug is best (rank-order the interventions)

- Lu and Ades 2004; Jansen et al. 2008; Sutton et al. 2008
Bayesian Approach

- Posterior distribution is a weighted average of currently observed data and prior information.

- Output is a probability distribution.

- As more data become available, the influence of the prior gets reduced.

- Inferences are intuitive and relevant to decision making.
Frequentist vs. Bayesian Output

Frequentist output:
Point estimate (with 95% CI)

“*No significant difference between Tx A and Tx B*”

Bayesian output:
Probability distribution

“75% Probability that Tx A is better than Tx B”

Probability that Tx B is better

Probability that Tx A is better

Tx A – Tx B

μ

Tx A – Tx B
Health Assessment Questionnaire (HAQ) in Rheumatoid Arthritis

- Results of Bayesian MTC (with a random-effects model) showed that, based on reduced HAQ scores, the probability that
  - Placebo is best = 1%
  - Methotrexate is best = 1%
  - Anti-TNFα is best = 8%
  - Anti-TNFα + methotrexate = 90%

- There was a 95% chance that anti-TNFα + methotrexate resulted in more favorable (reduced) HAQ scores than placebo
  - The probability is 95% that this reduction in HAQ scores is between 0.10 and 1.06 points
  - Range on HAQ: 0 to 3 points

Source: Jansen et al. 2008
### Multiple Treatments for Acute Myocardial Infarction

One overview covered 14 randomized controlled trials with two or three-way comparisons of six thrombolytic treatments.

The other overview featured 22 randomized controlled trials in which primary percutaneous transluminal coronary angioplasty (PCTA) was compared with thrombolytic treatment (streptokinase, alteplase, or accelerated alteplase). Because this meta-analysis collapsed the three thrombolytic treatments as a single comparator, the approach was criticized as the relevant comparator should have been the best thrombolytic drug, not the average one.

#### Evidence structure for comparison of multiple treatments used in two meta-analyses: number of randomised controlled trials directly comparing seven treatments for acute myocardial infarction. P's denote the treatments compared

<table>
<thead>
<tr>
<th>No of Trials</th>
<th>Streptokinase</th>
<th>Alleplase</th>
<th>Accelerated</th>
<th>Streptokinase</th>
<th>Alleplase</th>
<th>Releplase</th>
<th>Tenecteplase</th>
<th>PCTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boland et al</td>
<td>8</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>P</td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td>Keeley et al</td>
<td>8</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
</tbody>
</table>

PCTA = primary percutaneous transluminal coronary angioplasty.

Source: Caldwell et al. 2005.
The two sets of overviews with 36 randomized controlled trials were integrated across the seven treatments.

Estimates for all the 21 possible pair-wise comparisons (10 based on direct data and 11 based on indirect data) rendered empirical evidence of which treatment is most likely to have the lowest mortality (the winner was PCTA, with at least a 95% probability of being best) (Caldwell et al. 2005).

Mixed treatment comparison integrated and connected all available data so that relevant treatments can be compared, including those that would not have been otherwise, and direct evidence can be made more precise (by combining direct comparisons for which data were available with their corresponding indirect estimates).

For example, although the direct evidence showed PCTA not to be statistically better than alteplase [fixed effects odds ratio = 0.81, 95% confidence interval (CI) = 0.64 to 1.02], the mixed treatment comparison clearly did (odd ratio = 0.89, 95% CI = 0.61 to 0.89) as it capitalized on information from the indirect comparisons as well as the available direct comparison.
Thrombolytic Treatments for Acute Myocardial Infarction

Reconsider the overview of 14 randomized controlled trials with two or three-way comparisons of six thrombolytic treatments.

In terms of reducing 35-day mortality, a Bayesian analysis revealed that the probability that tenecteplase is best is 43%; accelerated alteplase, 40%; reteplase, 15%; streptokinase plus alteplase, 1%; streptokinase, 0%; and alteplase, 0%
International Society for Pharmacoeconomics and Outcomes Research (ISPOR)

- ISPOR Task Force on Indirect Treatment Comparisons
- ISPOR conference presentations
- Two articles prepared for the journal Value in Health
  - Guidance for decision-makers
  - Guidance on conducting
Network Meta-Analysis Requires Utmost Care

In the same spirit as
the Lesson Learned from Choosing the Right Study Design

An Early Clinical Trial (n = 2)


Coffee Example

In the late 18th century, King Gustav III of Sweden decided that coffee was poison and ordered a clinical trial.

1. Study:
   1) The king condemned a convicted murderer to drink coffee every day
   2) Control: another murderer was condemned to drink tea daily
   3) Outcome - death
   4) Two physicians were appointed to determine the outcome

2. Results:
   1) The two doctors died first
   2) The king was murdered
   3) Both convicts enjoyed long life until the tea drinker died at age 83.
      (No age was given for the coffee drinker)
3. Discussion:
– One should not rely on such a small sample size
– Perhaps the end point was too hard
– The outcome of the trial had no effect on the decision makers
– Coffee was forbidden in Sweden in 1794 and again in 1822

4. Conclusion:
– None possible
– External events and other biases may have confounded the result
  (Kings should not mess with clinical trials)
Placed network meta-analysis in context of CER

Highlighted concepts on network analysis

Described its basic assumptions

Provided example on schizophrenia with frequentist indirect comparison

Illustrated applications with Bayesian mixed treatment comparisons
References


♦ Caldwell et al. 2005. BMJ. 331:897-900

♦ Dias et al. 2010. Statistics in Medicine. 29:932-944

♦ Ioannidis. 2006. Lancet. 368:1470-1472

♦ Jansen et al. 2008. Value in Health. 11:956-964


♦ Salanti et al. 2008. Statistical Methods in Medical Research. 17:279-301


♦ Song. 2009 (February). What is indirect comparison. Part of the “What is...?” series