Mixed Treatment Comparisons (MTC)  
– Concepts and Problems –  

Ralf Bender, Sibylle Sturtz  
Institute for Quality and Efficiency in Health Care (IQWiG), Cologne

Disclosures

No relationships to disclose
Agenda

- Introduction
- Concepts and methods
  - Simple adjusted indirect comparison
  - Mixed treatment comparison (MTC)
  - Basic assumptions
- Problems
  - Network size
  - Inconsistency
- Importance of MTC
- Conclusions

Introduction

Idea

- Indirect comparison: IC
- Effect of intervention C relative to B: \( d_{BC} = d_{AC} - d_{AB} \)
- Mixed treatment comparison (MTC) meta-analysis
  (Also called: Multiple treatment meta-analysis, Network meta-analysis)

\[ Lu & Ades , JASA 2006 \]
Introduction

Reasons for MTC

- For many clinical indications there are often several possible interventions
- Combined analysis of all relevant data is to be preferred for health care decisions

Problems solved by MTC

- Direct comparisons between active interventions A and B may not always be available
- Even if direct evidence is available, there may be only a few studies
- Direct evidence from separate pairwise comparisons cannot determine which of several interventions is most effective

Example: Thrombolysis

Boland et al., HTA 2003

6 treatments for acute myocardial infarction:

1. Streptokinase (SK)
2. Tissue plasminogen activator (t-PA)
3. Accelerated alteplase (Acct-PA)
4. Tenecteplase (TNK)
5. Retepase (r-PA)
6. SK+t-PA

14 studies, 15 possible pairwise comparisons
Introduction

Thrombolysis: Results from Boland et al. (*HTA*, 2003)

- Pairwise comparisons performed by applying usual meta-analyses

  - "... SK is as effective as t-PA ..."
  - "... TNK is as effective as Acct-PA ..."
  - "... r-PA is at least as effective as SK ..."

- "... SK is as effective as, or inferior to Acct-PA ..."
  - "... r-PA is as effective as Acct-PA or not ...

- "... two further questions on indirect comparisons arise, whether TNK is superior to SK or not and whether r-PA is as effective as TNK or not ...

Results Thrombolysis (MTC FE model)

<table>
<thead>
<tr>
<th></th>
<th>SK</th>
<th>t-PA</th>
<th>Acct-PA</th>
<th>SK+T-PA</th>
<th>r-PA</th>
<th>TNK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SK</strong></td>
<td>**0.87</td>
<td>0.91</td>
<td>1.04</td>
<td>0.94</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>t-PA</strong></td>
<td>1.00</td>
<td>0.87</td>
<td>0.86</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Acct-PA</strong></td>
<td>0.96</td>
<td>0.97</td>
<td>1.11</td>
<td><strong>1.12</strong></td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td><strong>SK + t-PA</strong></td>
<td>0.96</td>
<td>0.97</td>
<td>1.11</td>
<td><strong>1.12</strong></td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td><strong>r-PA</strong></td>
<td>0.95</td>
<td>0.97</td>
<td>1.11</td>
<td><strong>1.12</strong></td>
<td>1.02</td>
<td>1.01</td>
</tr>
<tr>
<td><strong>TNK</strong></td>
<td>0.87</td>
<td>0.88</td>
<td>1.01</td>
<td>0.91</td>
<td>0.87</td>
<td>0.87</td>
</tr>
</tbody>
</table>

- Consistent
- Possibly with narrower credibility intervals
- Additional information (indirect comparisons)
## Introduction

### Probability of treatment $x$ being best

<table>
<thead>
<tr>
<th>Treatment $x$</th>
<th>MTC FE model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35 day mortality (%)</td>
</tr>
<tr>
<td>(1) SK</td>
<td>6,5</td>
</tr>
<tr>
<td>(2) t-PA</td>
<td>6,4</td>
</tr>
<tr>
<td>(3) Acct-PA</td>
<td>5,6</td>
</tr>
<tr>
<td>(4) SK + t-PA</td>
<td>6,2</td>
</tr>
<tr>
<td>(5) r-PA</td>
<td>5,8</td>
</tr>
<tr>
<td>(6) TNK</td>
<td>5,6</td>
</tr>
</tbody>
</table>

### Indirect comparisons are of increasing popularity

_Schöttker et al., DIMDI 2009_
Naive Approaches

Unadjusted indirect comparisons

- Meta-analysis using data of single arms:
  - MA using all data for intervention A,
  - MA using all data for intervention B,
  - MA using all data for intervention C, ...
- Breaks randomisation
- Should never be used!
- A correct analysis has to be based on the estimated effects of each RCT

Bucher Approach

Simple adjusted indirect comparison

- No direct evidence available
- One mutual comparator
- On log-odds scale:
  \[ \hat{d}_{BC}^{\text{indirect}} = \hat{d}_{AC}^{\text{direct}} - \hat{d}_{AB}^{\text{direct}} \]
  \[ \text{Var}(\hat{d}_{BC}^{\text{indirect}}) = \text{Var}(\hat{d}_{AC}^{\text{direct}}) + \text{Var}(\hat{d}_{AB}^{\text{direct}}) \]
- Assumes independence of pairwise comparisons
- Extensions available for several direct comparisons linked by common comparators (‘ladder’ design)
- Not applicable for more complex networks

*Bucher et al., JCE 1997
Wells et al., CADTH 2009*
Frequentist Network Approaches
White et al., RSM 2012

Frequentist network meta-analysis

- Combination of direct and indirect evidence in a complete network
- At least one closed loop required
- Original approach by Lumley (2002) only for 2-arm trials
- Extended by White et al. (2012) to situation of multi-arm trials
- Computations: Any software for linear mixed models can be used (SAS, R, Stata etc.)

Bayesian Network Approaches
Lu & Ades, JASA 2006

Mixed treatment comparison (MTC) meta-analysis

- Also called: Multiple treatment meta-analysis, network meta-analysis
- Combination of direct and indirect evidence in a complete network
- Applicable in all kinds of (connected) networks
- Can be applied to multi-armed trials
- Study level covariates can be incorporated
- Most flexible approach
- Bayesian approach requires specification of prior distributions
Bayesian Network Approaches

Lu & Ades, JASA 2006

Models for MTC: Basic & Functional Parameters

Take A as reference treatment

Treatment effects of B,C,D relative to A:
Basic parameters with priors
\[ d_{Ak} \sim N(0,10000) \quad k=B,C,D \]

Remaining contrasts: Functional parameters:

\[ d_{BC} = d_{AC} - d_{AB} \]
\[ d_{BD} = d_{AD} - d_{AB} \]
\[ d_{CD} = d_{AD} - d_{AC} \]

Fixed effects (FE) model

For intervention k in study j:
\[ r_{jk} \sim \text{Bin}(p_{jk}, n_{jk}) \]
\[ b = A,B,C \quad \text{if} \quad k = b \]
\[ \mu_{jb} + d_{bk} \quad \text{if} \quad k \text{ 'after' } b \]

Functional p.
\[ d_{bk} = d_{Ak} - d_{Ab} \]

Basic p.
\[ d_{Ak} \sim N(0,100000) \]
\[ \mu_{jb} \sim N(0,10000) \]
Random effects (RE) model

For intervention $k$ in study $j$:

$$\logit(p_{jk}) = \begin{cases} 
\mu_{jb} & \text{if } k = b \\
\mu_{jb} + \delta_{bk} & \text{if } k \text{ 'after' } b 
\end{cases}$$

$\mu_{jb}$: study specific effect of $b$

$\delta_{bk}, \sigma^2$: random effects

Functional p.

$\delta_{jbk} \sim N(d_{bk}, \sigma^2) \sim N(d_{Ak} - d_{Ab}, \sigma^2)$

Basic p.

$d_{Ak} \sim N(0, 10000)$

$\mu_{jb} \sim N(0, 10000)$

Bayesian Network Approaches

Computations: WinBUGS

- NICE Decision Support Unit:

- Multi-Parameter Evidence Synthesis Research Group
  http://www.bris.ac.uk/social-community-medicine/projects/mpes/mtc/

  WinBUGS Code for MTC meta-analyses:
  - FE model
  - RE model for 2- and 3-arm trials
  - RE model for multi-arm trials
Basic Assumptions

Basic assumptions for IC and MTC

- **Similarity assumption:**
  Trials are similar concerning moderators of the relevant treatment effect

- **Homogeneity assumption:**
  Trials are sufficiently homogeneous to be quantitatively combined

→ Same assumptions as for usual pairwise MA

+ **Consistency assumption:**
  Direct and indirect evidence estimate the same effect

Adequate MTC

- **Definition of relevant interventions**
  - Primary interventions
  - Comparators
  - Connecting interventions

- **Information retrieval**
  - Systematic literature search
  - Network may be never "complete"

- **Assessment of assumptions**
  - Similarity: PICO for the whole network
  - Homogeneity: Pairwise meta-analyses (forest plot, I², Q)
  - Consistency: In the framework of MTC meta-analysis
Basic Assumptions

Assessing similarity

- Similarity assumption:
  Comparability of studies regarding possible effect modifiers across all interventions
- PICO for the whole network
- Subjective evaluation of study characteristics
- Subgroup analyses
- Meta-regression

*Song et al., BMJ 2009*

Basic Assumptions

Assessing homogeneity

- Homogeneity assumption:
  Sufficient homogeneity of effect estimates across all studies comparing interventions
- Assessment of homogeneity in each pairwise meta-analysis
- Forest plots
- Tests for heterogeneity: Cochrans’ Q
- Measures for heterogeneity: $\tau^2$, $I^2$
Basic Assumptions

Assessing consistency

- Consistency assumption: Comparability of effect estimates from direct and indirect evidence
- Assessment of consistency within MTC meta-analysis
- Tests for inconsistency
- Models containing inconsistency parameters
- Graphical approaches
- No clear standard yet!

Problems: Example

Direct comparison

Possible networks
Example
Possible networks

MTC incl. placebo arms

Placebo

A

B

Example
Possible networks

MTC incl. drug classes

Placebo

A

B
drug class C
drug class D

F

E
Example
Possible networks

Placebo

MTC incl. all drugs sep.

drug class C

drug class D

Example 1
Possible networks

OR [95% CI]

Direct comparison  1.33 [0.93; 1.91]
MTC incl. placebo arms  1.10 [0.89; 1.35]
MTC incl. drug classes  1.14 [0.96; 1.35]
MTC incl. all drugs sep.  1.12 [0.93; 1.33]

Random effects MTC
Example 2
Possible networks

OR [95% CI]

Direct comparison  $\sim$  0.77 [0.46; 1.28]
MTC incl. placebo arms  $\sim$  1.02 [0.74; 1.37]
MTC incl. drug classes  $\sim$  1.07 [0.90; 1.28]
MTC incl. all drugs sep.  $\sim$  1.09 [0.89; 1.32]

Random effects MTC

Definition of relevant interventions

Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency
Sibylle Sturtz$^{a,†}$ and Ralf Bender$^{a,b}$

Impact of network size:
Larger networks are based upon more evidence but have more potential for heterogeneity and inconsistency
Network size

Issues regarding network size

- Results of different networks may or may not differ
- CI width in general smaller in large networks – but not always!
- Claim of inclusions of "all relevant evidence" represents an unworkably vague phrase (Cooper et al., Value Health 2011)
- Literature search for a "complete" evidence base may be a never ending story
- Network size connected with consistency

For 6 interventions: 15 possible pairs
For 12 interventions: 66 possible pairs!
Inconsistency

Methods used to investigate inconsistency

- Test for inconsistency
  \((Bucher et al., JCE 1997; Caldwell et al., JCE 2010)\)

- Graphical approach:
  Leverage vs. Bayesian deviance residuals
  \((Dias et al., Stat. Med. 2010)\)

Example

Test for consistency based on Bucher’s approach:

\[
z_{BC} = \frac{\hat{\omega}_{BC}}{\sqrt{\text{Var}(\hat{\omega}_{BC})}} \sim N(0, 1)
\]

with \(\hat{\omega}_{BC} = \hat{d}_{BC}^{\text{direct}} - \hat{d}_{BC}^{\text{indirect}}\)

and \(\text{Var}(\hat{\omega}_{BC}) = \text{Var}_{BC}^{\text{direct}} + \text{Var}_{BC}^{\text{indirect}}\)

\[
= \text{Var}_{BC}^{\text{direct}} + \text{Var}_{AB}^{\text{direct}} + \text{Var}_{AC}^{\text{direct}}
\]

<table>
<thead>
<tr>
<th>Direct</th>
<th>Placebo</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.99</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>[1.65; 2.39]</td>
<td>[0.93; 1.91]</td>
</tr>
</tbody>
</table>
Leverage Plot
Curves of form $x^2+y=c$, c=1,2,3,4

Example
Inconsistency

<table>
<thead>
<tr>
<th>All studies (MTC)</th>
<th>Placebo</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.93 [1.61; 2.29]</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2.11 [1.82; 2.44]</td>
<td>1.10 [0.89; 1.35]</td>
</tr>
</tbody>
</table>

- Leverage Plot
- Curves of form $x^2+y=c$, c=1,2,3,4

Example
Inconsistency

<table>
<thead>
<tr>
<th>Direct</th>
<th>Placebo</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.99</td>
<td>p=0.466</td>
</tr>
<tr>
<td></td>
<td>[1.65; 2.39]</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2.04</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>[1.74; 2.38]</td>
<td>[0.93; 1.91]</td>
</tr>
</tbody>
</table>

Example
Inconsistency

<table>
<thead>
<tr>
<th>Without No. 4+32</th>
<th>Placebo</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.84 [1.58; 2.13]</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2.19 [1.93; 2.50]</td>
<td>1.19 [1.00; 1.42]</td>
</tr>
</tbody>
</table>

- Deviance: no evidence of inconsistency
- Test for consistency n.s.
- Results for MTC and MA comparable
- A vs B: MTC smaller CI

Example
Inconsistency

<table>
<thead>
<tr>
<th>Direct</th>
<th>Placebo</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.86</td>
<td>p=0.653</td>
</tr>
<tr>
<td></td>
<td>[1.59; 2.19]</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2.13</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>[1.87; 2.44]</td>
<td>[0.93; 1.91]</td>
</tr>
</tbody>
</table>
Important Questions:
- Which network is the most appropriate?
- Is lumping of drug classes reasonable?
- Which amount of inconsistency is relevant?
- Which method is appropriate to evaluate inconsistency?
  - Graphical approach seems to be useful
  - Test for inconsistency is not reliable (for $\alpha=0.05$)

- Many things are still unclear!
- More experience is required!

Problematic Issue:
- Only consistent networks should be used in practice
- All relevant evidence should be used

Challenge

Network meta-analyses should only be performed if

- Similarity assumption is sufficiently plausible
- Pairwise meta-analyses do not show relevant heterogeneity
- Network does not show relevant inconsistency

*Song et al., BMJ 2011*
Challenge

Network meta-analyses should only be performed if

- Similarity assumption is sufficiently plausible
- Pairwise meta-analyses do not show relevant heterogeneity
- Network does not show relevant inconsistency

"This review shows that the underlying assumptions are not routinely explored or reported when undertaking indirect comparisons."

Example: ACEI/ARB in Diabetes

Reno-protective effects of renin-angiotensin system blockade in type 2 diabetic patients: a systematic review and network meta-analysis

P. Vejakama · A. Thakkinstian · D. Lertprattananon · A. Ingsathit · C. Ngarmukos · J. Attia

A network meta-analysis was performed to compare indirectly all treatment effects.

Typical problems:

- Heterogeneity in meta-analyses explored – but without consequence
- No assessment of consistency in network meta-analysis
Example: 3 types of stents

Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis

Christoph Stettler, senior research fellow,1 Sabine Allemann, research fellow,2 Simon Wandel, research fellow,3 Adrian Krauth, professor of cardiology,4 Marie-Da Costa Moraes, professor of cardiology,4 Albert Schöning, professor of medicine,5 Matthias E Pflaumer, professor of cardiology,6 Gregg W Stone, professor of medicine,7 Martin B Leon, professor of medicine,8 José Ramón de la Rosa, professor of cardiology,9 Anjan Iaques Goy, professor of interventional cardiology,10 Seung-Jung Park, professor of cardiology,11 Merelابلv, associate professor of cardiology,12 Maximon J Sutton, head of department,13 Henrik Kolbke, associate professor of cardiology,14 Christian Strobing, professor of cardiology,15 Maurin Menchelli, interventional cardiologist16 Paul Vermeersch, interventional cardiologist17 Mauri Tirkkonen, training fellow in cardiology18 Paolo Cervinka, cardiologist19 Marco De Carlo, visit director20 Andreas Efrati, associate professor of cardiology21 Tania Cherchi, interventional cardiologist22 Paulo Orlandi, interventional cardiologist23 Martin Schuel, professor of cardiology24 Peter Biers, head of division25 Bernhard Meier, professor of cardiology26 Stephan Windcker, head of massive cardiology27 Peter Kirk, head of division28

Stettler et al., BMJ 2008

3 web appendixes (17 pages in total):
- Comprehensive description of applied models
- Methods to assess goodness-of-fit, heterogeneity, inconsistency
- Description of intermediate data

Stellenwert von Ergebnissen aus indirekten Vergleichen
Gemeinsame Stellungnahme von IQWiG, GMDS und IBS-DR
Autoren: Ralf Bender, Carsten Schwenke, Claudia Schmoor, Dieter Haushcke

Joint statement of IQWiG, GMDS and IBS-DR (07.03.2012):
Network meta-analyses lead to lower certainty of results compared to meta-analyses of direct head-to-head studies
http://www.gmds.de/pdf/publikationen/stellungnahmen/120202_IQWiG_GMDS_IBS_DR.pdf
(In-)direct comparisons

„Direct randomized comparisons of treatments are usually more trustworthy than indirect comparisons …“

Ioannidis, CMAJ 2009

Always?

Reviews and Overviews

Why Olanzapine Beats Risperidone, Risperidone Beats Quetiapine, and Quetiapine Beats Olanzapine: An Exploratory Analysis of Head-to-Head Comparison Studies of Second-Generation Antipsychotics

Heres et al., Am. J. Psychiatry 2006

Sometimes an infinite stair

(Impossible figure, by M.C. Escher)

Conclusions

- MTC represents an important and promising method for health technology assessment
- In practice, choice of an appropriate network required
- Only consistent networks should be used
- A network may never be "complete"
- Exclusion of a few studies may be reasonable
- Clear standards for identification of inconsistency and dealing with inconsistency are currently lacking