Judging the usefulness of a network meta-analysis

Neil Hawkins, LSHTM
In three parts

• Brief introduction to network meta-analysis
• A review of the underlying consistency assumption
• “Tools” for evaluating reliability of the consistency assumption
An independent meta-analysis of trials ... showed that Cipralex was one of two anti-depressants judged to have achieved the best possible balance between efficacy and acceptability.”
"Cipralex was among the top two antidepressant drugs for both efficacy... and acceptability...(fig 2)"

Figure 1. Acute remission rates by treatment step. Adapted from STAR*D, Rush et al.

Figure 2. Probability of being among the top 4 drugs for both efficacy (at least 50% reduction in depression score, or considered much or very much improved on clinical global impression) and acceptability (all cause withdrawal) at mean 8 weeks' treatment. Adapted from Bandolier, 2009

References

Prescribing information can be found overleaf.
These claims refer to Cipriani et al. 2009

Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis

Andrea Cipriani, Toshiaki A Furukawa, Georgia Salanti, John R Geddes, Julian PT Higgins, Rachel Churchill, Norio Watanabe, Atsu Nakagawa, Ichiro M Omori, Hugh McGinue, Michele Tansella, Corrado Barbori

Summary

Background Conventional meta-analyses have shown inconsistent results for efficacy of second-generation antidepressants. We therefore did a multiple-treatments meta-analysis, which accounts for both direct and indirect comparisons, to assess the effects of 12 new-generation antidepressants on major depression.

Methods We systematically reviewed 117 randomised controlled trials (25 928 participants) from 1991 up to Nov 30, 2007, which compared any of the following antidepressants at therapeutic dose range for the acute treatment of unipolar major depression in adults: bupropion, citalopram, duloxetine, escitalopram, fluoxetine, fluvoxamine, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, and venlafaxine. The main outcomes were the proportion of patients who responded to or dropped out of the allocated treatment. Analysis was done on an intention-to-treat basis.

Findings Mirtazapine, escitalopram, venlafaxine, and sertraline were significantly more efficacious than duloxetine (odds ratios [OR] 1·39, 1·33, 1·30, and 1·27, respectively), fluoxetine (1·37, 1·32, 1·28, and 1·25, respectively), fluvoxamine (1·41, 1·35, 1·30, and 1·27, respectively), paroxetine (1·35, 1·30, 1·27, and 1·22, respectively), and reboxetine (2·03, 1·95, 1·89, and 1·85, respectively). Reboxetine was significantly less efficacious than all the other antidepressants tested. Escitalopram and sertraline showed the best profile of acceptability, leading to significantly fewer discontinuations than did duloxetine, fluvoxamine, paroxetine, reboxetine, and venlafaxine.

Interpretation Clinically important differences exist between commonly prescribed antidepressants for both efficacy and acceptability in favour of escitalopram and sertraline. Sertraline might be the best choice when starting treatment for moderate to severe major depression in adults because it has the most favourable balance between benefits, acceptability, and acquisition cost.

Funding None.
The analysis is readily interpretable.

Figure 4: Ranking for efficacy (solid line) and acceptability (dotted line).
Ranking indicates the probability to be the best treatment, the second best, the third best, and so on, among the 12 antidepressants.
But do these claims represent knowledge?
A taxonomy of comparisons

**Direct Comparison (head to head RCT)**

**‘Naïve’ or ‘Unadjusted’ Indirect Comparison:**
Absolute effect estimates from individual trial arms

**‘Adjusted’ Indirect Comparison:**
Relative effect estimates between treatments

**Mixed Treatment Comparison/Network Meta-Analysis:**
‘Adjusted’ indirect comparison extended to more complex networks of trial evidence (i.e. head to head and indirect evidence)
Simultaneous comparison of multiple treatments: combining direct and indirect evidence

Deborah M Caldwell, A E Ades, J P T Higgins

How can policy makers decide which of five treatments is the best? Standard meta-analysis provides little help but evidence based decisions are possible

Several possible treatments are often available to treat patients with the same condition. Decisions about optimal care, and the clinical practice guidelines that inform these decisions, rely on evidence based evaluation of the different treatment options. Systematic reviews and meta-analyses of randomised controlled trials are the main sources of evidence. However, most systematic reviews focus on pair-wise, direct comparisons of treatments (often with the comparator being a placebo or control group), which can make it difficult to determine the best treatment. In the absence of a collection of large, high quality, randomised trials comparing all eligible treatments (which is invariably the situation), we have to rely on indirect comparisons of multiple treatments. For example, an indirect estimate of the benefit of A over B can be obtained by comparing trials of A v C with trials of B v C, even though indirect comparisons produce relatively imprecise estimates. We describe comparisons of three or more treatments, based on pair-wise or multi-arm comparative studies, as a multiple treatment comparison evidence structure.

The need to combine direct and indirect evidence

Angioplasty balloon device used to unblock and widen arteries
The network of trial evidence for thrombolysis and angioplasty after myocardial infarction

Number of trials

Alteplase

Streptokinase

Reteplase

PTCA*

Strep + Acc. Alteplase

Tenecteplase

*Percutaneous transluminal coronary angioplasty

All rights reserved
The trial evidence summarised as a set of pairwise comparisons

Mortality at 35 Days
Mean Odds Ratio (95% CI)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Compared to:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Streptokinase</td>
</tr>
<tr>
<td>Streptokinase</td>
<td></td>
</tr>
<tr>
<td>Alteplase</td>
<td>0.89 (0.54 to 1.14)</td>
</tr>
<tr>
<td>Acc. Alteplase</td>
<td>0.86 (0.78 to 0.94)</td>
</tr>
<tr>
<td>Streptokinase + Alteplase</td>
<td>0.96 (0.87 to 1.05)</td>
</tr>
<tr>
<td>Reteplase</td>
<td>0.95 (0.79 to 1.12)</td>
</tr>
<tr>
<td>Tenecteplase</td>
<td>-</td>
</tr>
<tr>
<td>PTCA</td>
<td>0.49 (0.20 to 0.91)</td>
</tr>
</tbody>
</table>
Network meta-analyses provide estimates of treatment effects compared to a common reference

### Mortality at 35 Days
**Mean Odds Ratio (95% CI)**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streptokinase</td>
<td>1.04 (0.91 to 1.35)</td>
</tr>
<tr>
<td>Alteplase</td>
<td>1</td>
</tr>
<tr>
<td>Acc. Alteplase</td>
<td>0.88 (0.70 to 1.19)</td>
</tr>
<tr>
<td>Streptokinase+Alteplase</td>
<td>1.02 (0.78 to 1.51)</td>
</tr>
<tr>
<td>Reteplase</td>
<td>0.92 (0.70 to 1.24)</td>
</tr>
<tr>
<td>Tenecteplase</td>
<td>0.90 (0.61 to 1.35)</td>
</tr>
<tr>
<td>PTCA</td>
<td>0.65 (0.49 to 0.86)</td>
</tr>
</tbody>
</table>

Direct Comparison: 0.63 (0.25 to 1.29)
(Bayesian) network meta-analysis provide a readily interpretable summary of joint uncertainty

<table>
<thead>
<tr>
<th>Table 3  Percentage mortality at 35 days and the probability that each treatment is best (lowest mortality) in multiple treatment comparison analysis*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effect model</strong></td>
</tr>
<tr>
<td>35 day Mortality %</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Streptokinase</td>
</tr>
<tr>
<td>Alteplase</td>
</tr>
<tr>
<td>Accelerated alteplase</td>
</tr>
<tr>
<td>Streptokinase + alteplase</td>
</tr>
<tr>
<td>Retepase</td>
</tr>
<tr>
<td>Tenecteplase</td>
</tr>
<tr>
<td>Percutaneous transluminal coronary angioplasty</td>
</tr>
</tbody>
</table>

*Absolute mortality is based on the average mortality with streptokinase in the 19 randomised controlled trials that included it (see bmj.com for further details).
The simplest form of network meta-analysis is the adjusted indirect comparison.

Indirect Comparison of PTCA vs Alteplase

$$\text{OR}_{\text{PTCA vs Alteplase}} = \frac{0.49}{0.89} = 0.55$$
Estimation of uncertainty in an adjusted indirect comparison

\[ OR_{AB} = \frac{OR_{AC}}{OR_{BC}} \]

\[ \log OR_{AB} = \log OR_{AC} - \log OR_{BC} \]

\[
\text{var}(\log OR_{AB}) = \text{var}(\log OR_{AC}) + \text{var}(\log OR_{BC})
\]

\[
\text{sd}(\log OR_{AB}) = \sqrt{\text{sd}(\log OR_{AC})^2 + \text{sd}(\log OR_{BC})^2}
\]
The final result for the PTCA vs Alteplase AIC

- OR: 0.49 (0.20 to 0.91)
- OR: 0.55 (0.24 to 1.29)
- OR: 0.89 (0.54 to 1.14)
Uncertainty in indirect estimates

• 95% Confidence intervals are estimated by adding the variance for the contributing indirect comparisons

• Only represents uncertainty arising from the sampling error in the contributing trials

• Does not represent uncertainty in the fundamental assumptions

• Absolute ‘Best Case’ estimate of uncertainty
Network meta-analysis can be viewed as extension of the adjusted indirect comparison to more complex networks

- Treatment effects are estimated that best ‘fit’ the network of trial comparisons

1. $d_{\text{Streptokinase}}$, $d_{\text{Reteplase}}$, $d_{\text{PTCA}}$, … are estimates of the Log Odds Ratio (LOR) of Streprokinase, Reteplase and PTCA compared to a reference comparator (e.g. Alteplase). These are the “basic” parameters

2. $\text{LOR}_{\text{Streptokinase vs. Alteplase}} = d_{\text{Streptokinase}}$

3. $\text{LOR}_{\text{Reteplase vs. Alteplase}} = d_{\text{Reteplase}}$

4. $\text{LOR}_{\text{Streptokinase vs. PTCA}} = d_{\text{Streptokinase}} - d_{\text{PTCA}}$

5. …
The basic assumption underlying network meta-analysis is that:

\[ \hat{\partial}_{AB} = \hat{\partial}_{AC} - \hat{\partial}_{BC} \]

Referred to as:

• Consistency
  – Indirect and direct estimates are consistent

• Exchangeability
  – If treatments were exchanged between trials estimates would be the same (allowing for random variation)

• Similarity
  – The trials are similar and comparable

• Transitivity

\[ \hat{\partial}_{AB} = \hat{\partial}_{AC} - \hat{\partial}_{BC} \quad \hat{\partial}_{AC} = \hat{\partial}_{AB} - \hat{\partial}_{CB} \]
Consider a single trial

A (Response = 30%)

B (20%)

C (10%)
By definition consistent on the relative risk scale

\[ RR_{AvsB} = \frac{RR_{AvsC}}{RR_{BvsC}} \]

\[ RR_{AvsB} = \frac{3}{2} = 1.5 \]
And also on the odds ratio scale...

\[ \text{OR}_{A\text{vs}B} = \frac{\text{OR}_{A\text{vs}C}}{\text{OR}_{B\text{vs}C}} \]

\[ \text{OR}_{A\text{vs}B} = \frac{3.86}{2.25} = 1.71 \]
And on the risk difference (RD) scale...

\[ RD_{AvsB} = RD_{AvsC} - RD_{BvsC} \]
\[ RD_{AvsB} = 20\% - 10\% = +10\% \]
Whereas multiple trials may be consistent

\[ RR_{A\text{vs}B} = \frac{3}{2} = 1.5 \]
Or may be inconsistent...

What determines whether networks of multiple trials will be consistent?
Prognostic factors alter response in individual treatment arms

Severe Patients

A (30%) → C (10%)  RR: 3

Mild Patients

A (60%) → C (20%)  RR: 3

But do not alter the relative treatment effect (on a given scale)
Predictive factors alter response in individual treatment arms

And alter the relative treatment effect (on a given scale)
A completely homogeneous set of trials...

\[ RR_{A vs B} = \frac{3}{2} = 1.5 \]

Will behave like a single multi-arm trial and be consistent...
A heterogeneous set of trials

\[ RR_{AvsB} = \frac{3}{2} = 1.5 \]

Will still be consistent if they differ in terms of prognostic factors
However, a heterogeneous set of trials

Severe Patients

A (30%) \[\text{RR}: 1.5\] B (20%)

\[RR_{\text{A vs B}} = \frac{1.5}{2} = 0.75 \neq 1.5\]

Mild Patients

A (30%) \[\text{RR}: 1.5\] C (20%)

B (20%) \[\text{RR}: 2\] C (10%)

Will be inconsistent if they differ in terms of predictive factors
A heterogeneous set of trials

May be consistent on one treatment effect scale
A heterogeneous set of trials

\[ RD_{A vs B} = 40\% - 10\% = 30\% \neq 10\% \]

But be inconsistent on a different treatment effect scale
Consistency is a “model” applied to a connected network of trial data

- It is an “assumption”, a convenience*, not a natural law
- Network meta-analyses are confounded by variation in predictive factors (treatment effect modifiers)
- Network meta-analyses are not confounded by variation in prognostic factors
- Naïve indirect comparisons are confounded by variation in prognostic factors and predictive factors
- Factors may be prognostic on one scale but not another
- The reliability of the model is a function of the degree of heterogeneity

**"essentially, all models are wrong, but some are useful"  George Box**
Assessing Heterogeneity

- Assess Extent
  - Study / Patient Characteristics
    - Observed Characteristics
  - Subgroup / Regression
    - Observed Characteristics
- Assess Impact
  - (In)consistency
    - Unobserved Characteristics
- Account for
  - Regression Stratification
    - Matching
    - Hierarchy
  - Observed Characteristics
An example network meta-analysis: treatments for advanced NSCLC

• Comparators
  – Placebo, Docetaxel, Erlotinib, Gefitinib, Pemetrexed
• Continuous Endpoint
  – Hazard Ratio: Overall Survival
• 4,672 patients in 6 studies
• NMA conducted on multiplicative hazard ratio scale: HRAB=HRAC / HRBC

The trials form a connected network
Which includes “loops”

Compared With

<table>
<thead>
<tr>
<th>Drug</th>
<th>Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gefitinib</td>
<td>r: 1.02 (0.91 to 1.14)</td>
</tr>
<tr>
<td>Remetrexed</td>
<td>r: 0.99 (0.82 to 1.2)</td>
</tr>
<tr>
<td>Docetaxel</td>
<td>r: 0.89 (0.79 to 1.01)</td>
</tr>
<tr>
<td>Erlotinib</td>
<td>r: 0.48 (0.24 to 0.96)</td>
</tr>
</tbody>
</table>

Hazard ratios: mean (95% CI)

All rights reserved
The results can be expressed against a common reference comparator

However, due to correlation we cannot directly derive all possible pairwise comparisons from this
There is, however, heterogeneity in study characteristics

<table>
<thead>
<tr>
<th>Author and date</th>
<th>Trial name</th>
<th>Trial design</th>
<th>Jadad score</th>
<th>Treatment</th>
<th>Number randomized</th>
<th>Mean treatment duration (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shepherd 2005</td>
<td>BR21</td>
<td>Double-blind</td>
<td>3</td>
<td>Erlotinib</td>
<td>488</td>
<td>Not stated</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>Hanna 2004</td>
<td>JMEI</td>
<td>Open-label</td>
<td>2</td>
<td>Pemetrexed</td>
<td>283</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Docetaxel</td>
<td>288</td>
<td>4</td>
</tr>
<tr>
<td>Shepherd 2000</td>
<td>TAX 317</td>
<td>Not stated</td>
<td>2</td>
<td>Docetaxel</td>
<td>55</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Thatcher 2005</td>
<td>ISEL</td>
<td>Double blind</td>
<td>4</td>
<td>Gefitinib</td>
<td>1129</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Placebo</td>
<td>563</td>
<td>2.7</td>
</tr>
<tr>
<td>Douillard 2007</td>
<td>INTEREST</td>
<td>Open-label</td>
<td>1</td>
<td>Gefitinib</td>
<td>723</td>
<td>4.4</td>
</tr>
<tr>
<td>(conference presentation)</td>
<td></td>
<td></td>
<td></td>
<td>Docetaxel</td>
<td>710</td>
<td>3.0</td>
</tr>
<tr>
<td>Cufer 2006</td>
<td>SIGN</td>
<td>Open-label</td>
<td>2</td>
<td>Gefitinib</td>
<td>68</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Docetaxel</td>
<td>73</td>
<td>2.8</td>
</tr>
</tbody>
</table>
(Some) methods for assessing inconsistency

- Node splitting
- Comparison with an inconsistency model
- Treatment by design model
In Node Splitting, the inconsistency parameter $\omega_{jk}$ represents the discrepancy between the direct and indirect estimates of treatment effect for a comparison between treatment $j$ vs. treatment $k$.

The formula for the inconsistency parameter is:

$$\omega_{jk} = d_{jk}^{dir} - d_{jk}^{ind}$$

- The inconsistency parameter can be tested against the null hypothesis $\omega_{jk} = 0$.
- Can be tested against null: inconsistency $= 0$.
The results of the node splitting analysis for the NSCLC NMA

<table>
<thead>
<tr>
<th>Comparison</th>
<th>FE Pairwise</th>
<th>RE Pairwise (Q:0 P:1 I^2:NaN)</th>
<th>Bayesian Network Meta–Analysis</th>
<th>Direct Network Meta–Analysis</th>
<th>Indirect (p=0.104)</th>
<th>Indirect (p=0.0974)</th>
<th>Indirect (p=0.0979)</th>
<th>Median (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Docetaxel vs. Placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NA (NA to NA)</td>
</tr>
<tr>
<td></td>
<td>0.48 (0.24 to 0.96)</td>
<td>0.48 (0.24 to 0.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.85 (0.72 to 1.00)</td>
</tr>
<tr>
<td></td>
<td>0.48 (0.24 to 0.97)</td>
<td>0.88 (0.74 to 1.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gefitinib vs. Placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.89 (0.79 to 1.01)</td>
</tr>
<tr>
<td></td>
<td>0.89 (0.79 to 1.01)</td>
<td>0.88 (0.78 to 0.99)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.89 (0.79 to 1.01)</td>
</tr>
<tr>
<td></td>
<td>0.49 (0.25 to 0.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gefitinib vs. Docetaxel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.02 (0.91 to 1.14)</td>
</tr>
<tr>
<td></td>
<td>1.02 (0.91 to 1.14)</td>
<td>1.03 (0.92 to 1.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.02 (0.91 to 1.14)</td>
</tr>
<tr>
<td></td>
<td>1.85 (0.92 to 3.68)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NA (NA to NA)</td>
</tr>
</tbody>
</table>
Evidence Synthesis for Decision Making 4: Inconsistency in Networks of Evidence Based on Randomized Controlled Trials

Sofia Dias, PhD, Nicky J. Welton, PhD, Alex J. Sutton, PhD, Deborah M. Caldwell, PhD, Guobing Lu, MSc, A. E. Ades, PhD

Inconsistency can be thought of as a conflict between “direct” evidence on a comparison between treatments B and C and “indirect” evidence gained from AC and AB trials. Like heterogeneity, inconsistency is caused by effect modifiers and specifically by an imbalance in the distribution of effect modifiers in the direct and indirect evidence. Defining inconsistency as a property of loops of evidence, the relation between inconsistency and heterogeneity and the difficulties created by multiarm trials are described. We set out an approach to assessing consistency in 3-treatment triangular networks and in larger circuit structures, its extension to certain special structures in which independent tests for inconsistencies can be created, and describe methods suitable for more complex networks. Sample WinBUGS code is given in an appendix. Steps that can be taken to minimize the risk of drawing incorrect conclusions from indirect comparisons and network meta-analysis are the same steps that will minimize heterogeneity in pairwise meta-analysis. Empirical indicators that can provide reassurance and the question of how to respond to inconsistency are also discussed. Key words: Network meta-analysis; inconsistency; indirect evidence, Bayesian. (Med Decis Making 2013;33:641–656)
Consistency Model

Consistent treatment effect

\[
\eta_{jk} = d^J - d^K
\]

Estimated Treatment effect for treatment \( j \) vs. treatment \( k \)
Inconsistency Model

Estimated Treatment effect for treatment \( j \) vs. treatment \( k \)

Independent treatment effect for each comparison

\[ \eta_{jk} = d_{jk} \]
Comparison of posterior residual deviance

<table>
<thead>
<tr>
<th>Analysis</th>
<th>DIC (lower = better fit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency Model</td>
<td>-13.89</td>
</tr>
<tr>
<td>Inconsistency Model</td>
<td>-14.75</td>
</tr>
</tbody>
</table>
CD study is inconsistent (RED) and of low precision (thin line). CD study will be the outlier
CD study is inconsistent (RED) and of high precision (thin line). AB study (low precision) will be the outlier.

Inconsistency is a property of “loops”, not individual studies or comparisons.
Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies

J. P. T. Higgins, D Jackson, J. K. Barrett, G Lu, A. E. Ades and I. R. White

Meta-analyses that simultaneously compare multiple treatments (usually referred to as network meta-analyses or mixed treatment comparisons) are becoming increasingly common. An important component of a network meta-analysis is an assessment of the extent to which different sources of evidence are compatible, both substantively and statistically. A simple indirect comparison may be confounded if the studies involving one of the treatments of interest are fundamentally different from the studies involving the other treatment of interest. Here, we discuss methods for addressing inconsistency of evidence from comparative studies of different treatments. We define and review basic concepts of heterogeneity and inconsistency, and attempt
Definitions of inconsistency

• Consistency
  \[ \delta^{AB} = \delta^{AC} - \delta^{BC} \]

• Heterogeneity (Variation within a comparison)
  Treatment specific variance: \( \delta_{iJK} \sim N(\delta^{JK}, \tau_{JK}^2) \)
  Common variance: \( \delta_{iJK} \sim N(\delta^{JK}, \tau^2) \)

• Design inconsistency
  – Treatment effects vary by study design
    (design=comparator set)
Design by Treatment Interaction Model

Main (consistent) treatment effect

Estimated Treatment effect for treatment A vs. Treatment J from study i with design d

Between design variation (aka inconsistency) Fixed Effect

Within trial between design variation (aka heterogeneity) Random Effect
An Example

<table>
<thead>
<tr>
<th>Trial Design</th>
<th>Treatment</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Ref</td>
<td>$\delta^{AB}$</td>
<td>$\delta^{AC}$</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>Ref</td>
<td>$\delta^{AB} + \omega^{AB}_2$</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>Ref</td>
<td>-</td>
<td>$\delta^{AC} + \omega^{AC}_3$</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>Ref</td>
<td>$\delta^{AB}$</td>
<td>$\delta^{AB} + \omega^{AC}_4$</td>
<td></td>
</tr>
</tbody>
</table>

Diagram of A, B, and C relationships.
Results of the treatment by design analysis for the NSCLC NMA

The gefitinib effect (vs Placebo) as estimated from the gefitinib versus docetaxel trial was 0.55 (95% CrI 0.27 to 1.12) times the effect as estimated from the other trials.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>DIC (lower = better fit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency Model</td>
<td>-13.89</td>
</tr>
<tr>
<td>Treatment by Design</td>
<td>-14.80</td>
</tr>
</tbody>
</table>
## Quick Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>Addresses Question</th>
<th>Provides Global Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node splitting</td>
<td>Does each link agree with the rest of network</td>
<td>No</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>How does imposing consistency affect fit (globally and per study)</td>
<td>Yes</td>
</tr>
<tr>
<td>Treatment by design</td>
<td>What is the difference between treatment effects estimated from different study designs?</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Inconsistency cannot be observed in “star shaped” networks

Inconsistency can be (and is likely) to be observed in “well connected” networks

Which should we trust most?

We can use tools such as the ISPOR-AMCP-NPC checklist

**Evidence base**
- Attempt to include all relevant RCTs?
- 1 network?
- No poor quality RCTs?
- No differences in effect modifiers between direct comparisons?

**Analysis**
- Naive comparisons avoided?
- Consistency assessed?
- With consistency, was direct & indirect evidence included?
- Account for inconsistency/ Minimize bias?
- Valid rationale for FE/RE model?
- Rationale for heterogeneity assumptions in RE model discussed?
- Subgroup or meta-regression analysis?

**Reporting quality & transparency**
- Network & source data presented?
- Direct & indirect results reported
- Are all contrasts presented with uncertainty?
- Ranking of treatments presented?
- Results by subgroup or levels of effect-modifiers presented?

**Interpretation**
- Conclusions fair & balanced?

**Conflict of interest**
- Conflict of interest? If yes, steps taken to address these?

Fig. 1 - Overview of domains related to assessment of the credibility of a network meta-analysis. FE, fixed effects; RCTs, randomized controlled trials; RE, random effects.

All rights reserved
But ultimately, the assessment of credibility is a judgment.
And credibility is

- is weakened by known differences between trials in factors that (might) act as treatment effect modifiers
- is strengthened by consistency between direct and indirect evidence (if “loops” exists)
- is strengthened by analyses that adjust or account for observed treatment effect modifiers or inconsistency
In Cipriani 2009 et al.

“Analysis indicated statistical incoherence in three out of 70 comparisons of direct with indirect evidence for response rate ... and three out of 63 comparisons for dropout rate ... These numbers are compatible with chance because about six significant findings would be expected out of 133 statistical tests.”

“Overall, heterogeneity was moderate, although for most comparisons the 95% CI included values that showed very high or no heterogeneity, reflecting the small number of included studies for each pair-wise comparison. In the meta-analyses of direct comparisons, we found I2 values higher than 75% for the comparisons citalopram and reboxetine (I2=85.0%), and escitalopram and fluoxetine (I2=82.7%).”
Final thought: exchangeability is implicit in clinical decision-making