Lecture 7

Detecting and exploring inconsistency

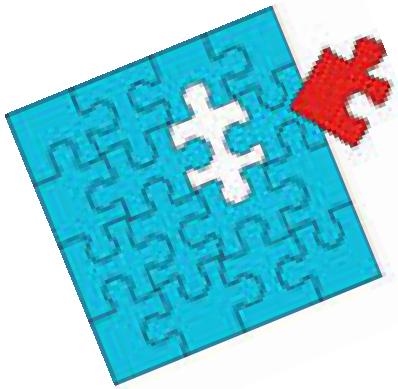
Orestis Efthimiou

Network meta-analysis

A project-based course using R

Kea island, April 2018

Outline



- Definition of inconsistency
- Transitivity vs. consistency
- Local methods for inconsistency
- A note on multi-arm studies
- Global methods for inconsistency
- Which method to use?
- What if we find inconsistency?
- What if we don’t find inconsistency?

Inconsistency can be viewed as a special form of heterogeneity

"Dispertion in the relative treatment effects evaluated in ***different settings***"

Different settings : different studies

'classic' notion of heterogeneity

Inconsistency can be viewed as a special form of heterogeneity

"Dispertion in the relative treatment effects evaluated in ***different settings***"

Different settings : different sources of evidence

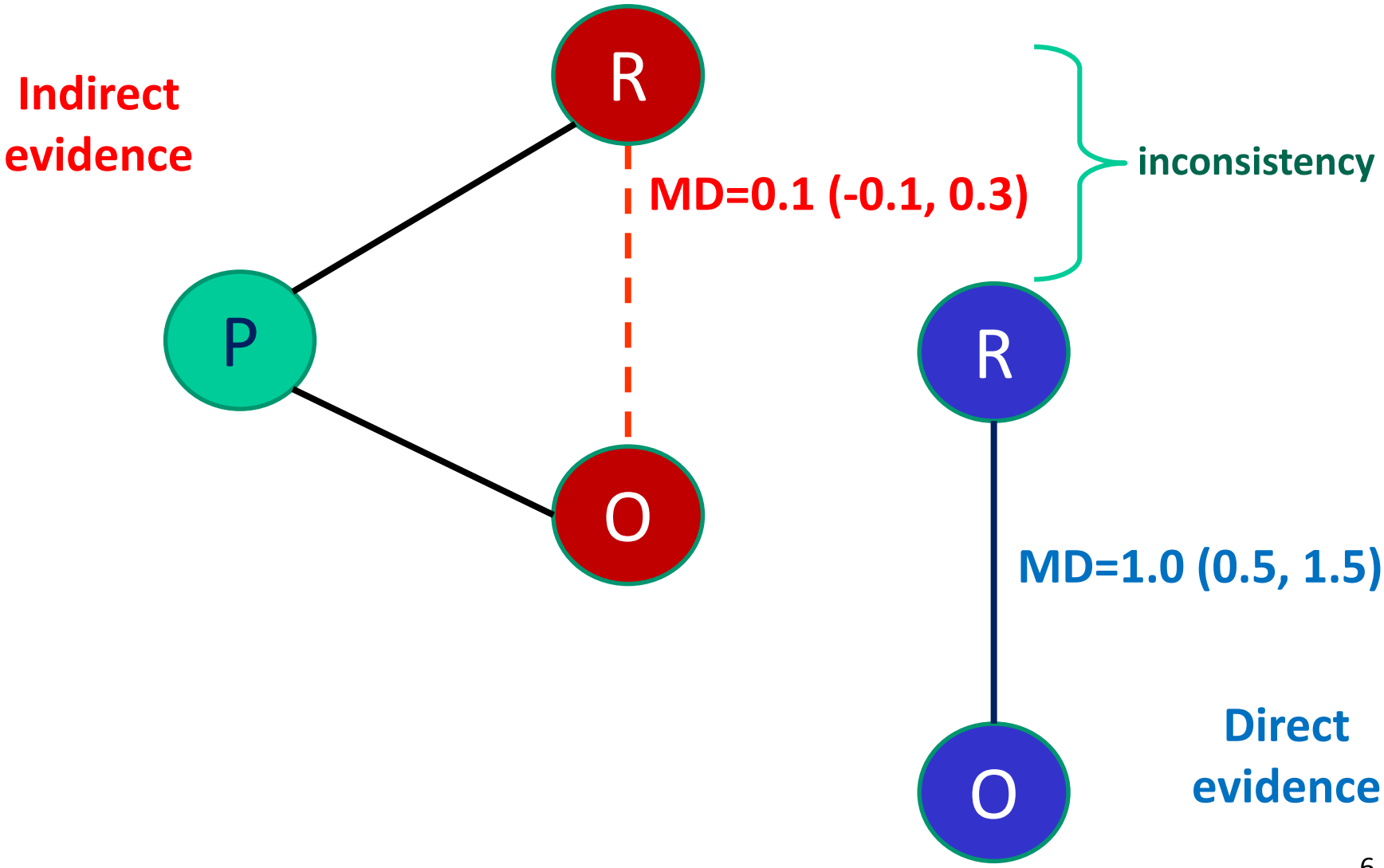
inconsistency

Inconsistency can be viewed as a special form of heterogeneity

Heterogeneity: *variability within a comparison*

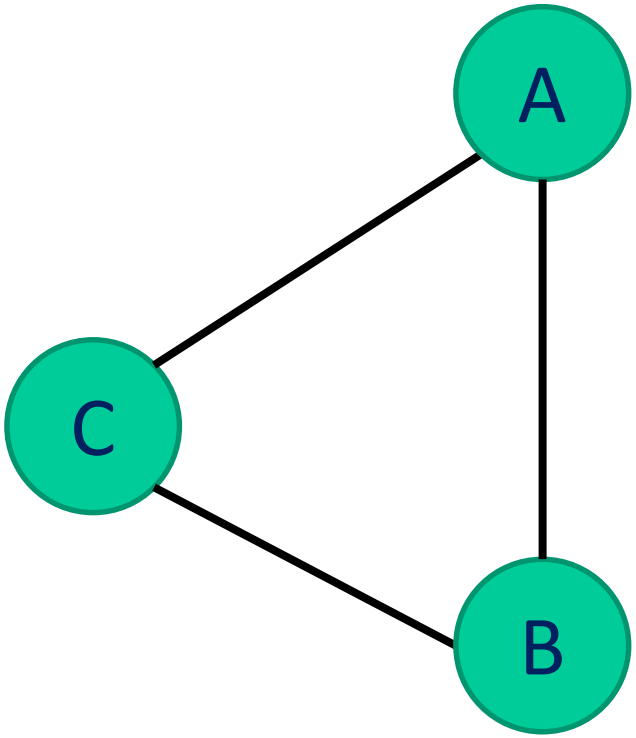
Inconsistency: *variability across comparisons*

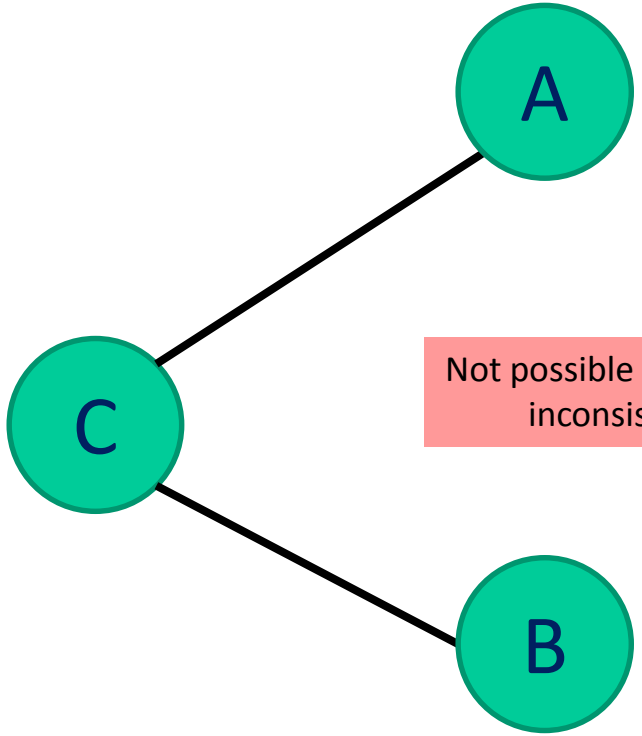
Inconsistency



Inconsistency

- It's important to examine the validity of the consistency equations
- A breach in the consistency equations might indicate a breach in the **transitivity assumption**
- It's only possible to examine inconsistency if we have **closed loops** in the network plot





Not possible to measure inconsistency

The concept of consistency in a network

- **Transitivity in a network implies:**

✓ Every treatment in the network has a ‘fixed’ definition irrespectively of the comparator

✓ Treatments are ‘jointly randomizable’.

✓ Patients could in principle receive all treatments.

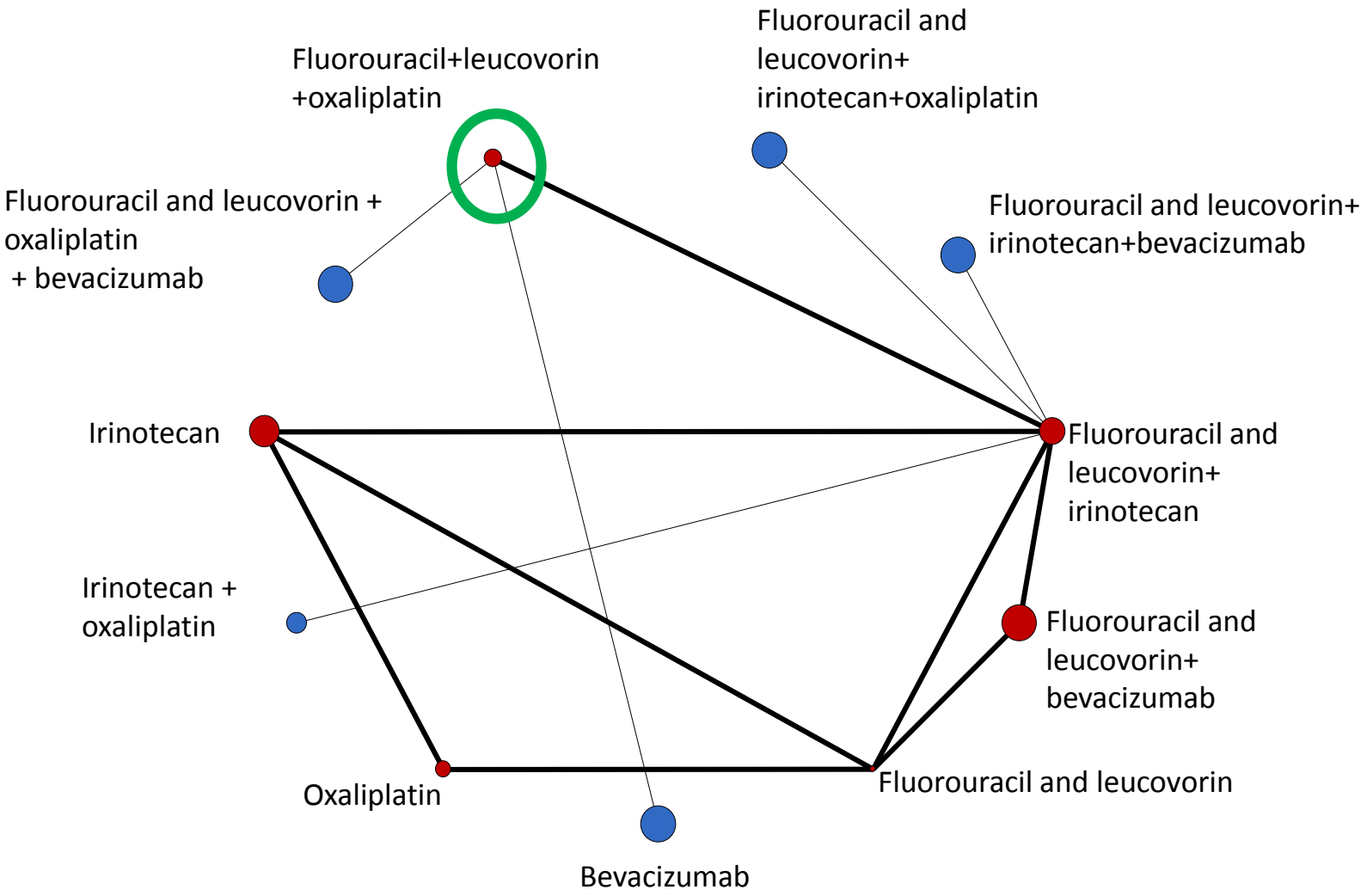
✓ All sets of trials grouped by comparison are similar with respect to the distribution of effect modifiers

The concept of consistency in a network

A network is **consistent** when the various pieces of evidence (direct and many indirect sources) are in agreement

If the transitivity assumption does not hold this **may** lead to **inconsistencies** (**but not always!**)

Consistency and transitivity in a network



Lancet Oncol 2007 Golfinopoulos V, Salanti G et al

Regarding the node "fluorouracil+leucovorin +oxaliplatin", we are...

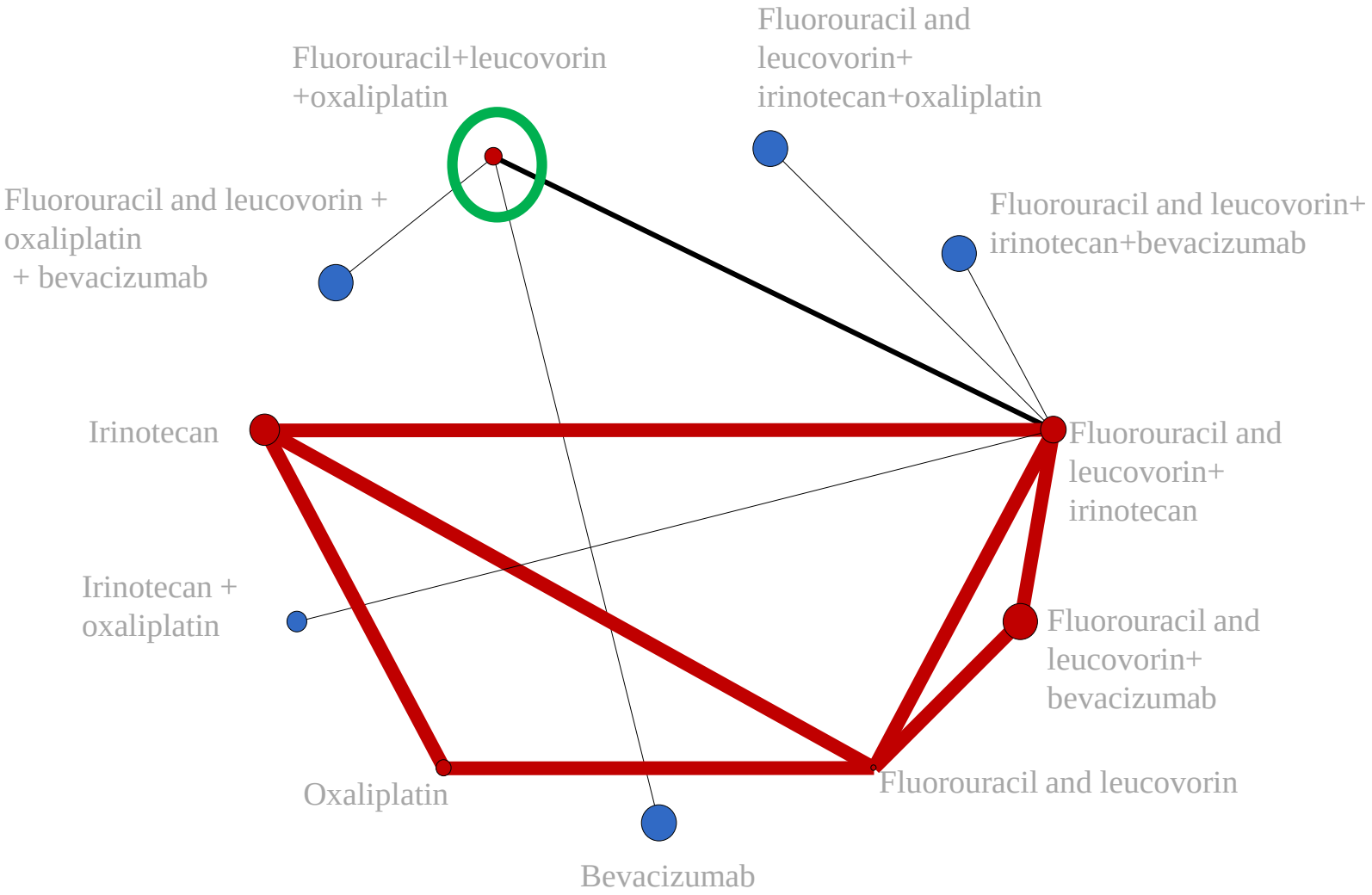
able to assess inconsistency

able to assess inconsistency but not transitivity

able to assess inconsistency and transitivity

not able to assess inconsistency

Consistency and transitivity in a network



Lancet Oncol 2007 Golfinopoulos V, Salanti G et al

Transitivity and consistency in NMA

In the outset

The treatments we compare are in principle **jointly randomizable**

They have the same indication, I can imagine a mega-trial with all treatments being compared etc

When we find the studies

The groups of studies that compare them do not differ with respect to the **distribution of effect modifiers**

You can test this assumption if you have enough studies per comparison

When we extract the outcomes

Direct and indirect treatment effects are in statistical agreement

Various statistical tests

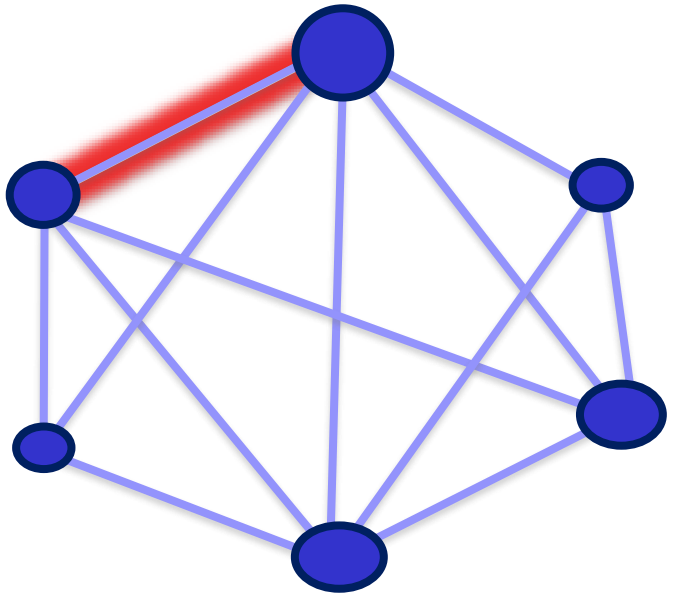
Cipriani A et al. *Conceptual and Technical Challenges in Network Meta-analysis* **Annals of Internal Medicine** 2013

How to estimate the inconsistency of a network?

- Either “**locally**”
- Or **“globally”**

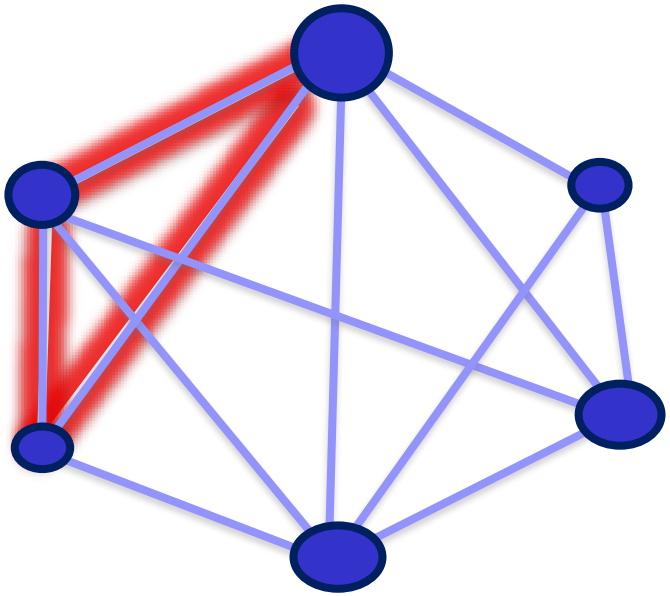
Appropriate methods to evaluate inconsistency

- **Local tests**
  - focus on a comparison



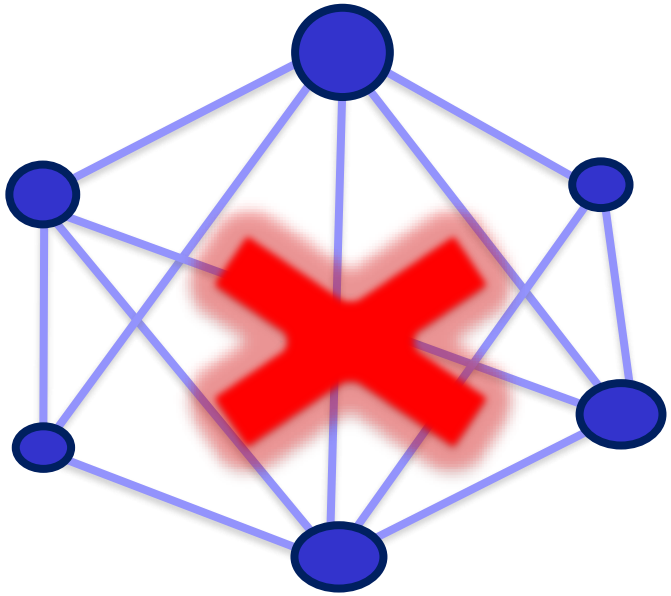
Appropriate methods to evaluate inconsistency

- **Local tests**
  - focus on a comparison
  - focus on a loop



Appropriate methods to evaluate inconsistency

- Global tests



Statistical approaches for evaluating inconsistency in a network

Local methods

1. Loop-specific approach
2. SIDE ("node-splitting")
3. Back-calculation

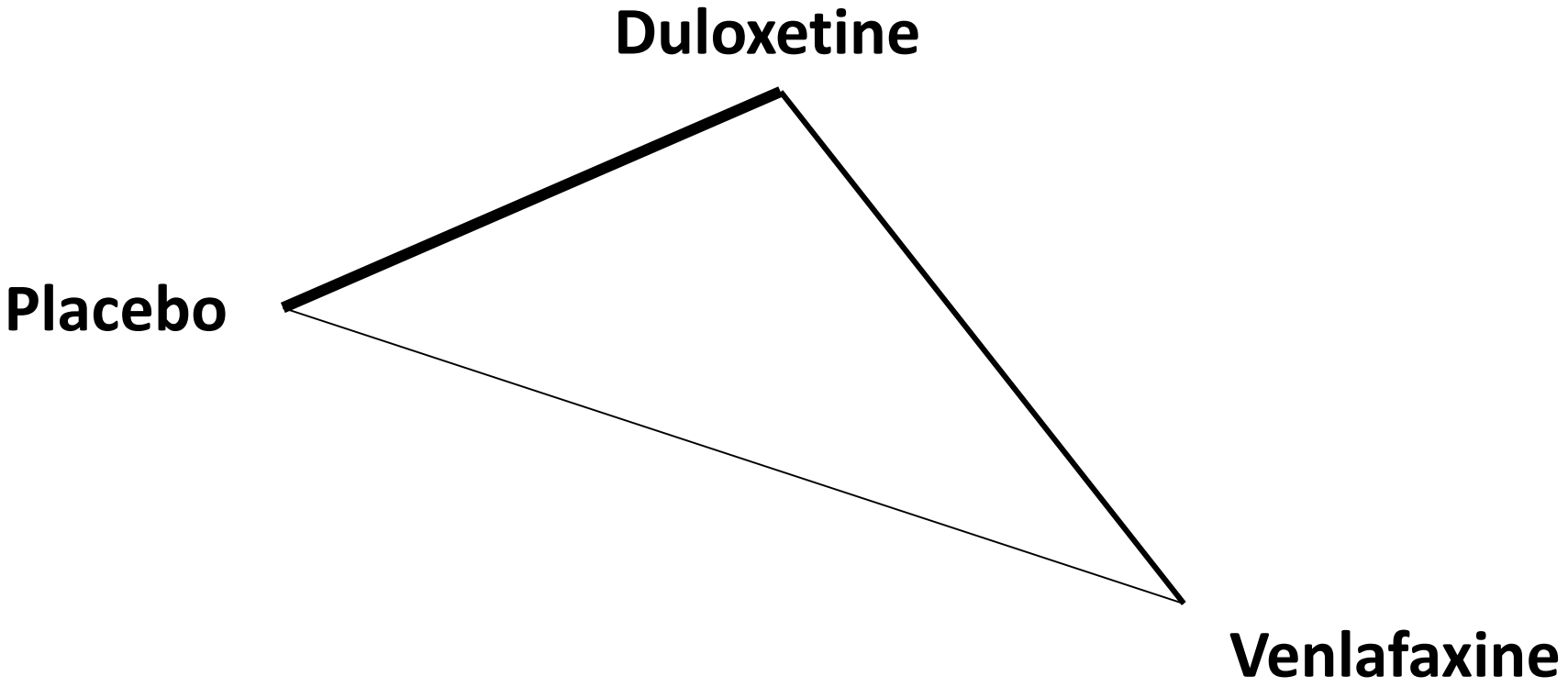
Global methods

4. The Lu & Ades model
5. The design-by-treatment inconsistency model

**other methods have also been proposed*

LOCAL TESTS FOR INCONSISTENCY

1. Loop-specific approach



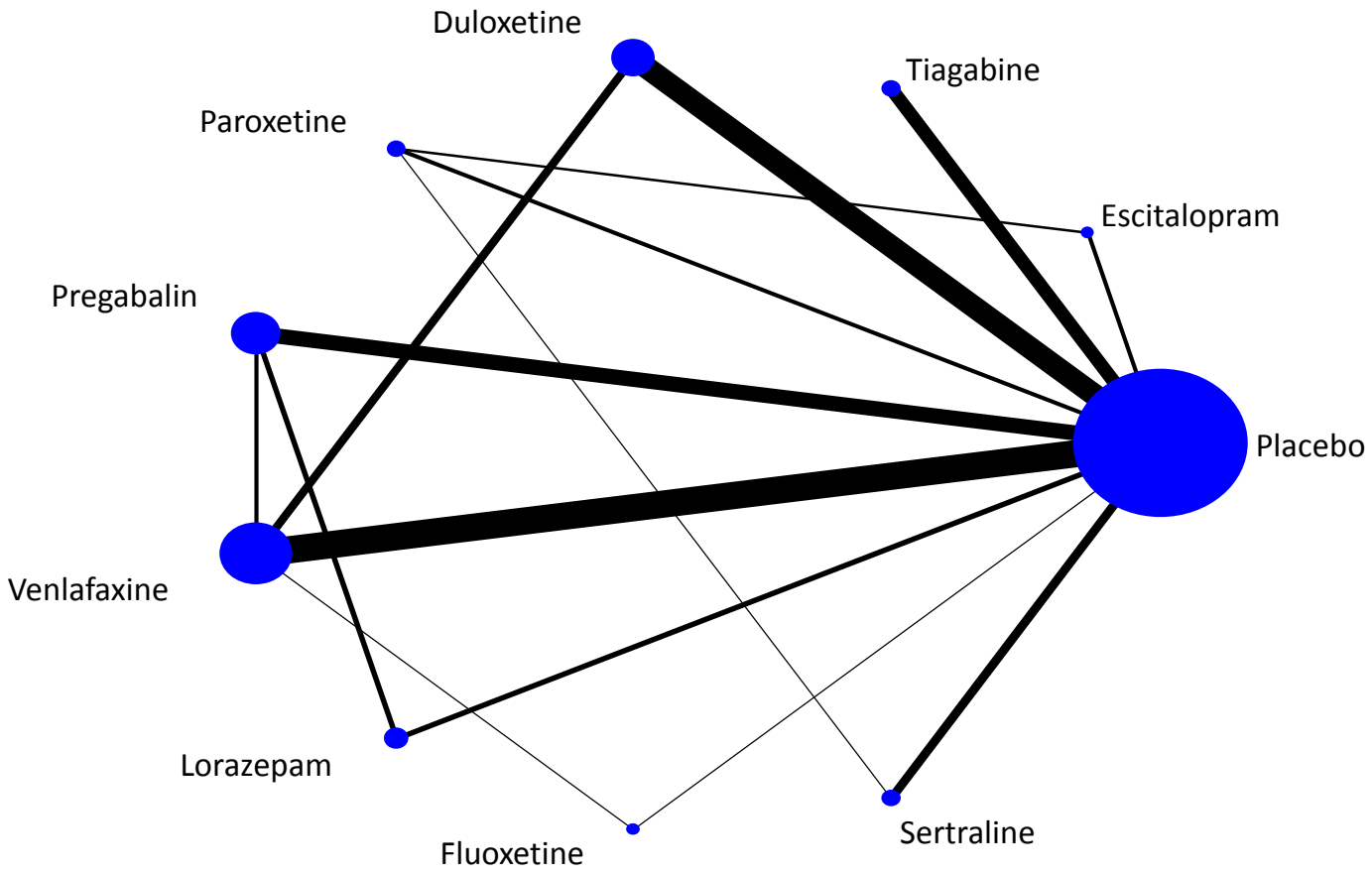
We calculate the inconsistency factor (IF) for the loop

$$IF_{Pla\text{-}Dulo\text{-}Ven} = SMD^{ind}_{Dulo\ vs\ Ven} - SMD^{dir}_{Dulo\ vs\ Ven}$$

Calculate 95% CI for IF

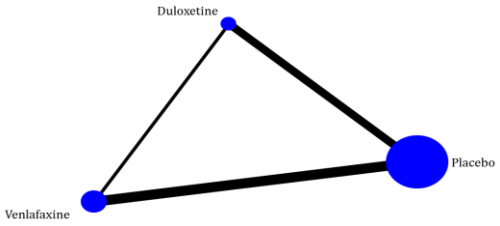
1. Loop-specific approach

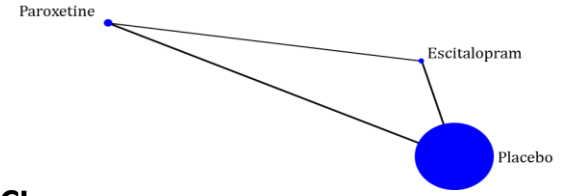
Extend this idea to all **loops** in the network



1. Loop-specific approach

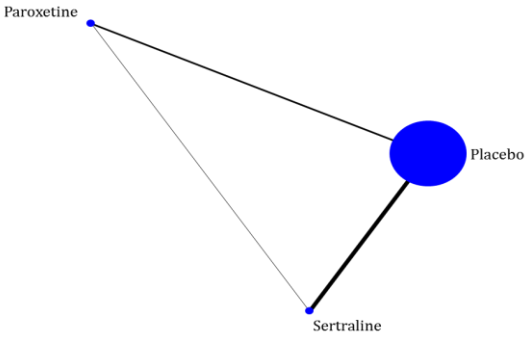
Extend this idea to all **loops** in the network

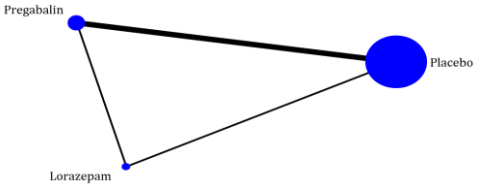


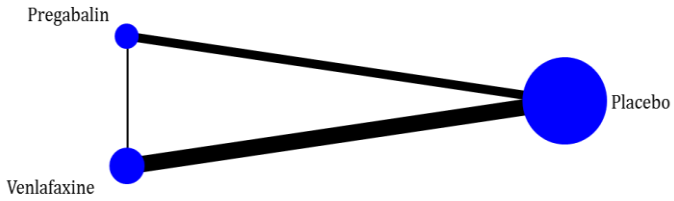


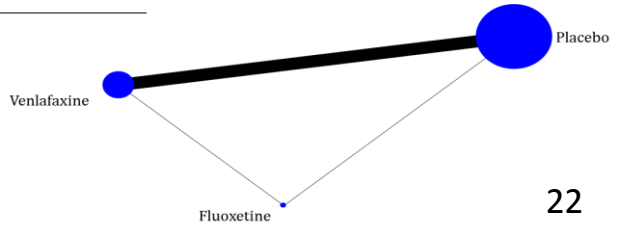
Loop	ROR	95%CI (truncated)
PLA-VEN-FLUO	2.508	(1.00,11.43)
PLA-ESC-PAR	2.028	(1.00,5.58)
PLA-PAR-SER	1.461	(1.00,5.16)
PLA-DULO-VEN	1.178	(1.00,1.80)
PLA-PREG-VEN	1.163	(1.00,2.19)
PLA-PREG-LOR	1.058	(1.00,2.33)

1 2 4 8 15









1. Loop-specific approach

Good because

- ✓ It is simple and easy to apply (`ifplot` command in Stata)
- ✓ Can indicate loops with large inconsistency

BUT it is problematic as

- ✓ There are multiple, correlated tests which cannot be combined to infer about the consistency of the entire network
- ✓ We don't contrast direct vs ***all*** indirect evidence
- ✓ Does not account for **multi-arm studies**

a detour: multi-arm studies

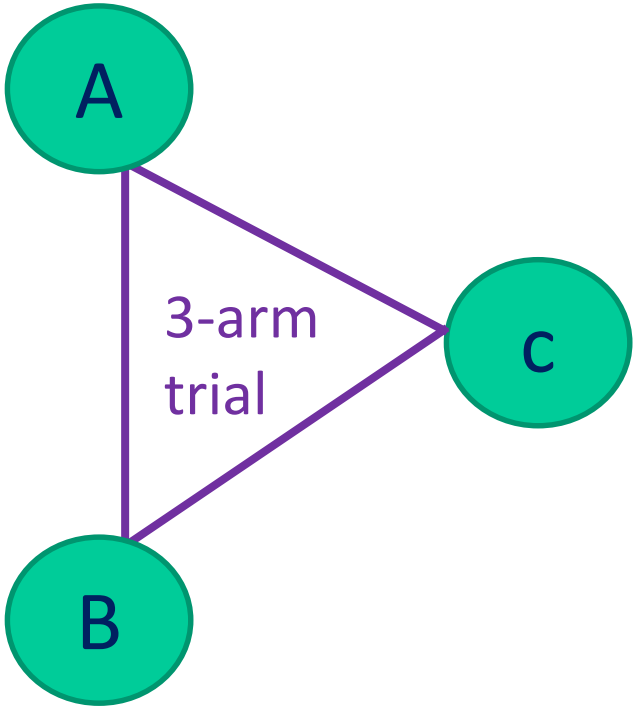
- ✓ Studies comparing more than two treatments
- ✓ Treatment effects in multi-arm studies are correlated
- ✓ E.g. in an ABC study, the relative treatment effects AB and AC are not independent, because they share the same arm, A

- ✓ By definition multi-arm studies always form loops
- ✓ But these loops are **by definition consistent**
- ✓ This **complicates** things regarding inconsistency

a detour: multi-arm studies

- Example: three-arm trial ABC
- Three different treatment effects can be estimated: AB, AC, BC
- $\mu_{AB} = d_A - d_B$
- $\mu_{AC} = d_A - d_C$
- $\mu_{BC} = d_B - d_C$

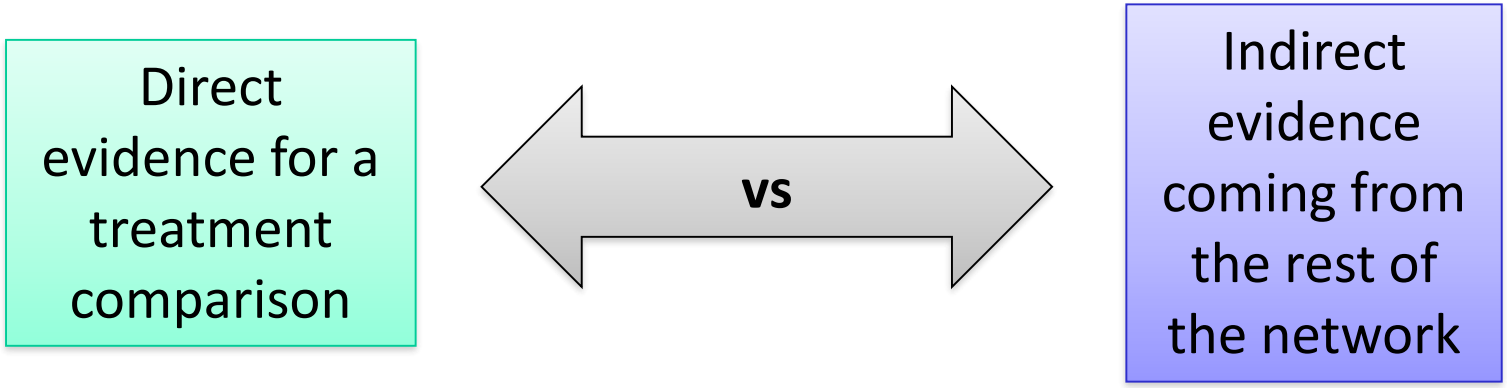
$\mu_{BC} = \mu_{AC} - \mu_{AB}$ ⟹ Consistent by definition



2. SIDE (*Separating Indirect from Direct Evidence*)

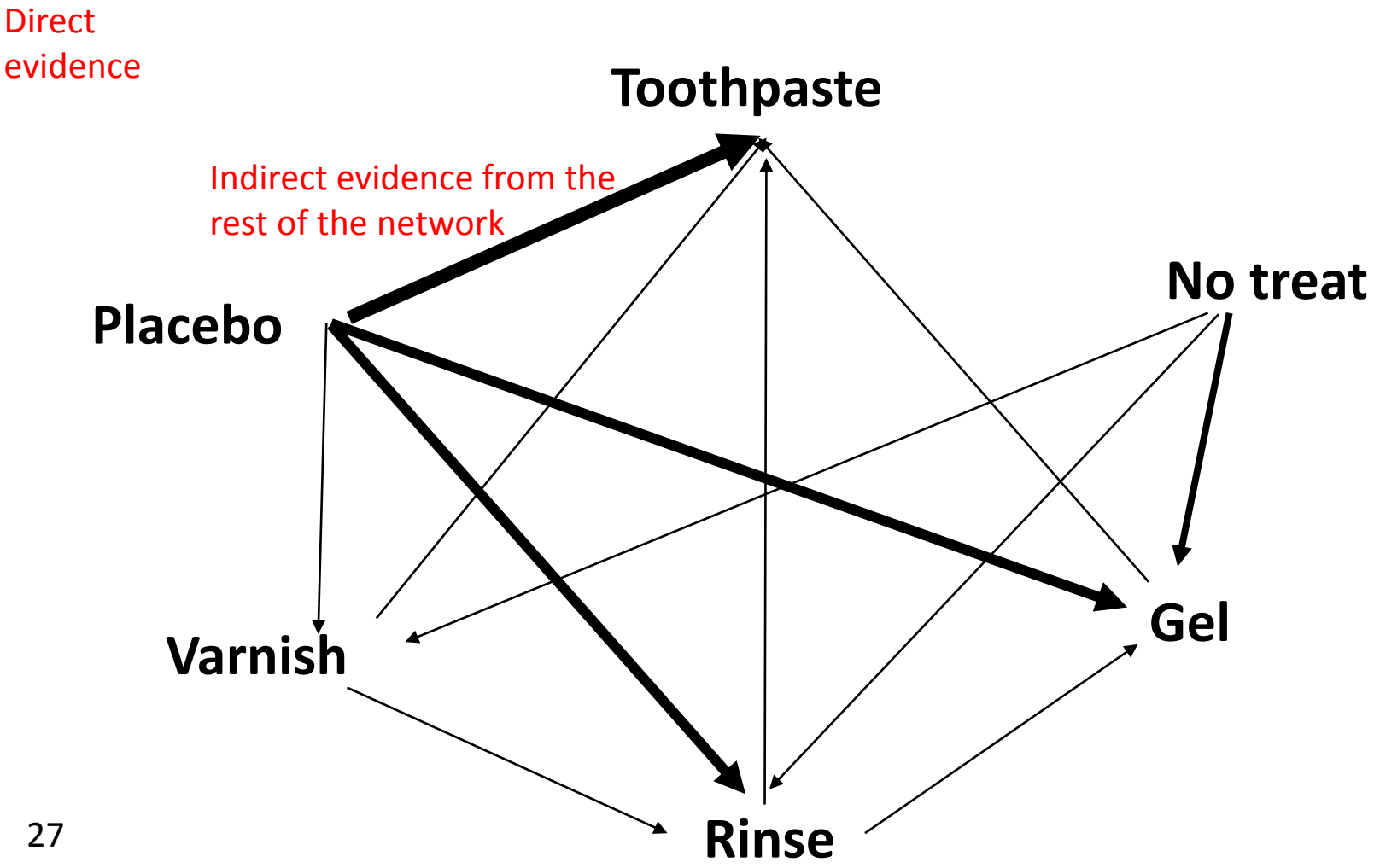
or "node-splitting" or "side-splitting"

- For every pairwise comparison available in the network, contrast **direct** and **indirect** evidence from the entire network:



2. SIDE (*Separating Indirect from Direct Evidence*)

or "node-splitting" or "side-splitting"



2. SIDE (*Separating Indirect from Direct Evidence*)

or "node-splitting" or "side-splitting"

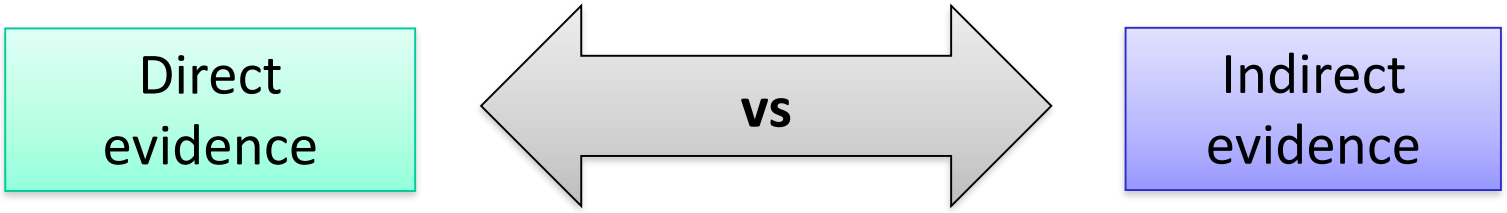
- This procedure can be repeated for **all** available comparisons

- Good because it synthesizes indirect information coming from all the network
- Resource intensive, especially for large networks
- Does not **properly** account for multi-arm studies

- Currently available only in Stata (`network sidesplit` command)

3. Back calculation method

- This is an approach **very similar to SIDE**
- Using this method, for each treatment comparison we partition the evidence in the network in two independent parts: **direct** and **indirect**



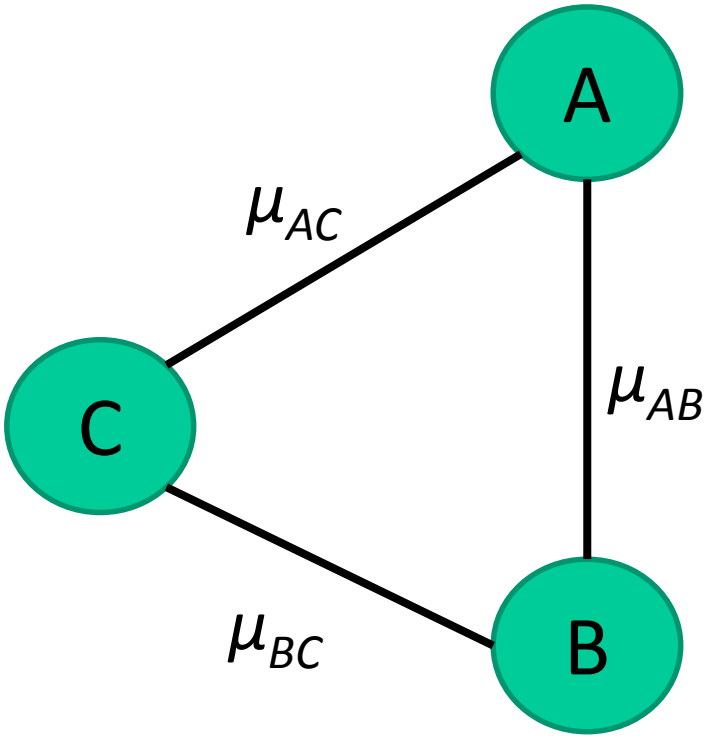
3. Back calculation method

- Bypasses the problems that previous methods face in the presence of multi-arm studies
- Easy to use, very quick in R (`netsplit` command)

GLOBAL TESTS FOR INCONSISTENCY

Modelling inconsistency

The consistency model

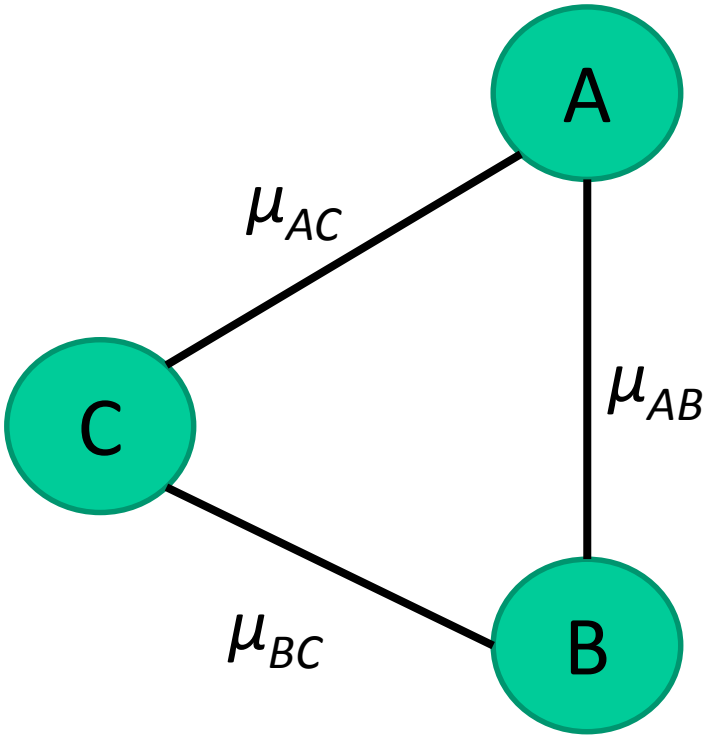


Consistency equation

$$\mu_{BC} = \mu_{AC} - \mu_{AB}$$

Modelling inconsistency

The inconsistency model



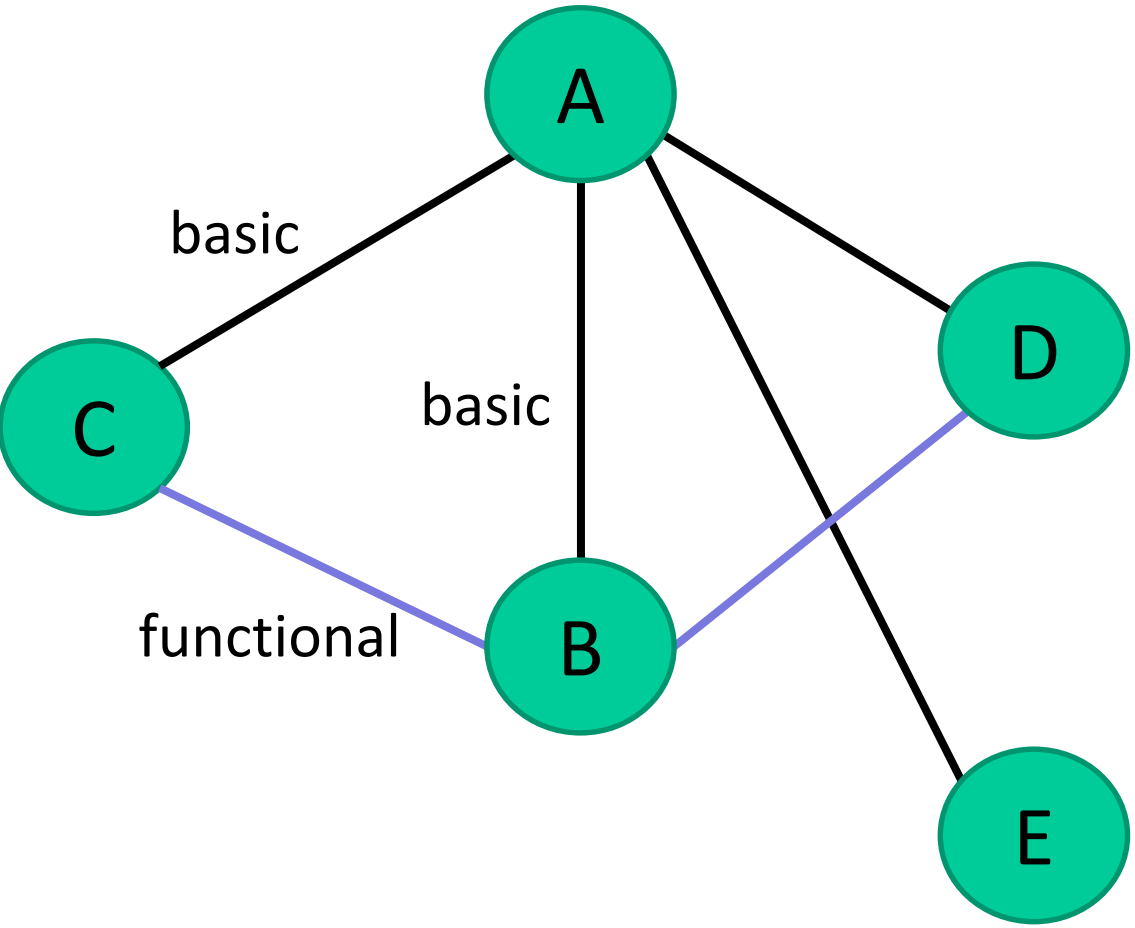
"Inconsistency" equation

$$\mu_{BC} = \mu_{AC} - \mu_{AB} + w_{ABC}$$

inconsistency factor

4. Lu & Ades inconsistency model

Parameterisation

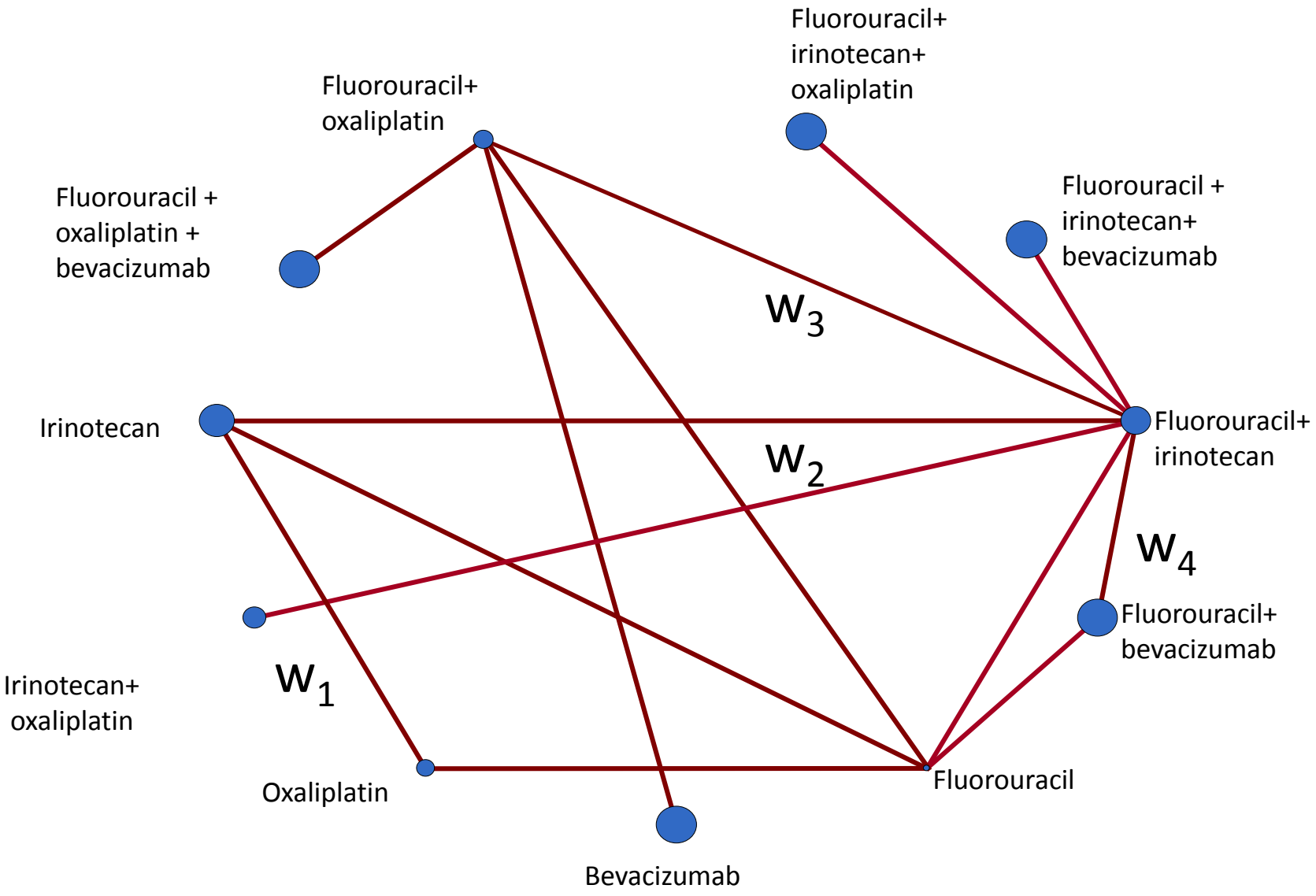


Consistency equations

$$\mu_{BC} = \mu_{AC} - \mu_{AB} + w_{ABC}$$
$$\mu_{BD} = \mu_{AD} - \mu_{AB} + w_{ABD}$$
$$\mu_{DC} = \mu_{AC} - \mu_{AD}$$
$$\mu_{BE} = \mu_{AE} - \mu_{AB}$$
$$\mu_{DE} = \mu_{AE} - \mu_{AD}$$
$$\mu_{CE} = \mu_{AE} - \mu_{AC}$$

inconsistency factors

Example: Survival with 11 chemotherapy regimens in colorectal cancer



Example: Survival with 11 chemotherapy regimens in colorectal cancer

$w1 = -0.08$, $w2 = -0.07$, $w3 = -0.06$, $w4 = -0.03$ (on logHR scale)

No loop is remarkably inconsistent

No important changes in the estimated HRs

The assumption of consistency is reasonably supported from the model

Issues with the Lu and Ades model

- In the presence of **multi-arm trials**, the Lu and Ades inconsistency model is not uniquely defined
- Different parameterizations (choice of reference treatment) may lead to different results
- **In the presence of multi-arm studies the Lu and Ades model should not be used**

5. Inconsistency as 'design-by-treatment interaction'

- A more general approach to inconsistency
- It assumes alternative types of inconsistency
  - ✓ inconsistency within loops made up of different trials
  - ✓ inconsistency between two-arm and three-arm trials
  - ✓ and beyond...
- Such a model has been termed a **design-by-treatment interaction model**

5. Inconsistency as 'design-by-treatment interaction'

- Incorporates, but goes beyond "loop inconsistency"

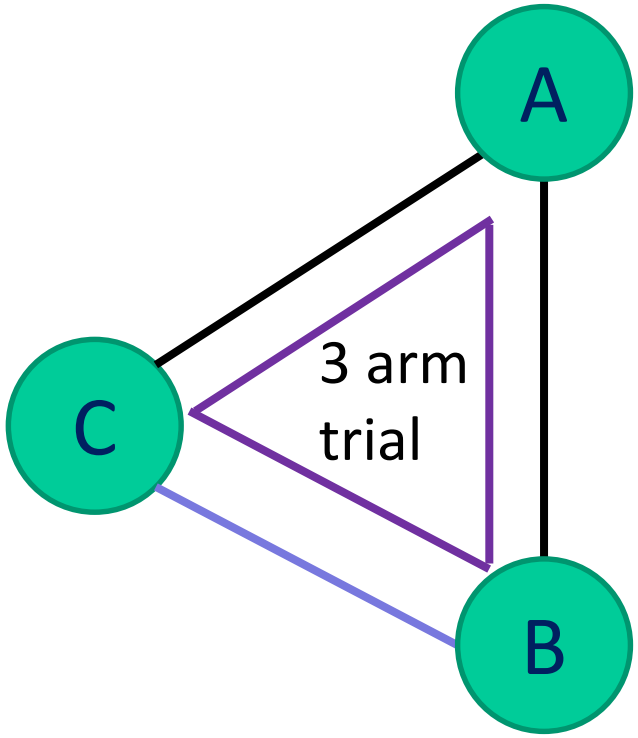
Design: the set of treatments compared in a study. ABC is a different design than AB, AC or BC

Design inconsistency: when the relative effects of A vs B is different when measured in studies of different designs.

- μ_{AB} *is different when estimated in AB and ABC studies*

More degrees of freedom (inconsistency factors) are needed

5. Inconsistency as 'design-by-treatment interaction'



inconsistency model

$\mu_{BC} = \mu_{AC} - \mu_{AB} + w_{BC}$

μ_{AB} from AB = μ_{AB} from ABC + w_{AB}

μ_{AC} from AC = μ_{AC} from ABC + w_{AC}

- *For the design inconsistency model we need three inconsistency factors*
- *The Lu and Ades model would only include one*

5. Inconsistency as 'design-by-treatment interaction'

Key assumption of the design-by-treatment model:

μ_{AB} is different when estimated in AB or ABC studies

- ✧ Is this a plausible assumption?
- ✧ It might be, when the study design is a proxy for important differences

For example, maybe

- ✧ Multi-arm studies are more recent?
- ✧ Two-arm studies mainly sponsored by pharma?
- ✧ Multi-arm studies are of better methodological quality?
- ✧ Two-arm studies more susceptible to publication bias?
- ✧ Other (possibly unknown) factors?

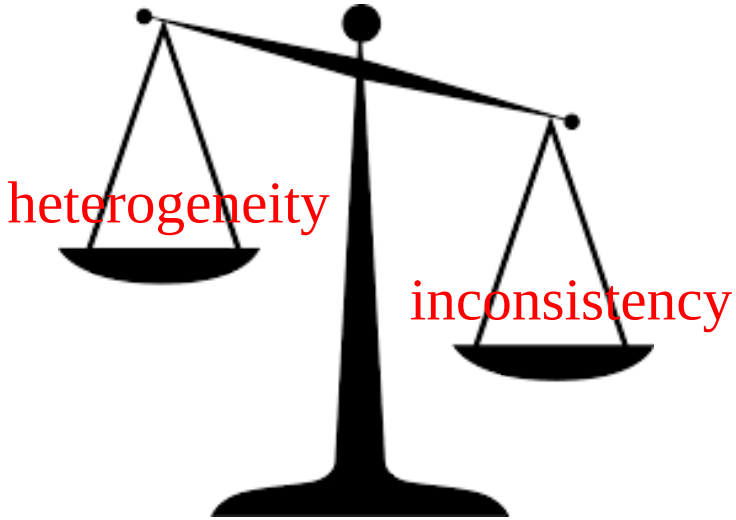
5. Inconsistency as 'design-by-treatment interaction'

- The hypothesis of consistency in the network can be checked using a **global test**
- Unlike the Lu and Ades model, this global test does not have problems in the presence of **multi-arm studies**

- The notion of "design inconsistency" seems to be **arbitrary**. Why is an ABC trial expected to give different results than AB, ABD, or ABCD studies?
- The model seems to be driven by mathematical, rather than epidemiological considerations
- The inconsistency factors depend on parameterization, so they are not interpretable on their own

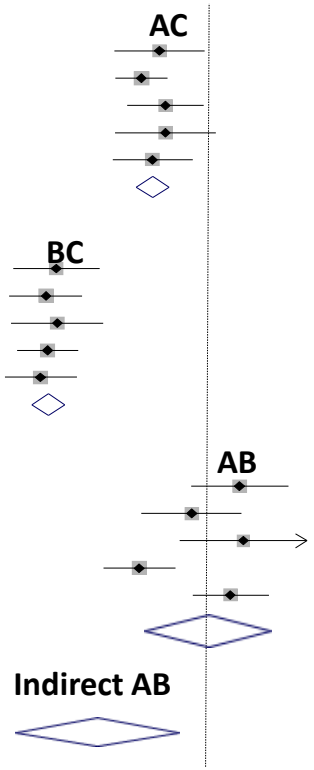
Many methods for detecting inconsistency, which one to use?

- One thing to keep in mind is that all approaches have low power and trade-off with heterogeneity



High heterogeneity may mask inconsistency

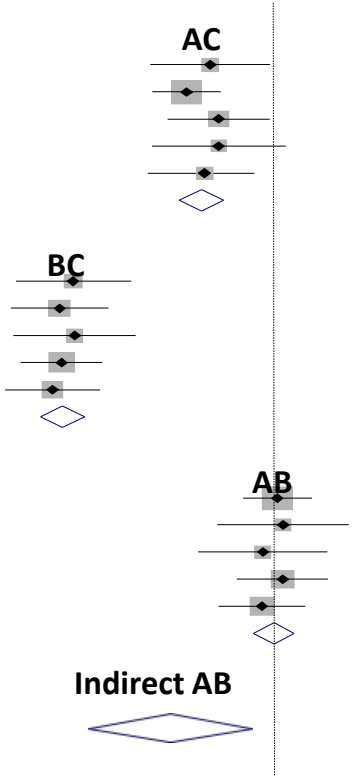
High heterogeneity in AB



CIs form indirect and direct evidence on AB overlap

no evidence of inconsistency

Low heterogeneity in AB



Indirect and direct evidence on AB disagree

statistical inconsistency

Many methods for detecting inconsistency, which one to use?

- Song et al. performed **simulations** to evaluate the statistical properties of various methods for detecting inconsistency

- They found that the methods have **low power** in detecting inconsistency, especially in the presence of large heterogeneity

- We recommend researchers to apply **both global and local methods** in order to get a better sense of possible inconsistencies in the network as a whole and also to identify problematic studies or comparisons

Song F et al. (2012) Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. BMC

Inappropriate methods for assessing inconsistency

Inconsistency cannot be assessed by:

- ✗Comparing direct estimates with NMA results
- ✗Comparing what other meta-analyses found to the NMA results
- ✗Comparing prior beliefs with NMA results

What if we don't find any inconsistency?

- The absence of significant inconsistency **does not mean** there is consistency
  - ✓ Issues of power and the trade-off with heterogeneity may limit the usefulness of the consistency tests

- Being unable to test for inconsistency **does not mean** there is transitivity
  - ✓ Because inconsistency can only be assessed in loops

What if we don’t find any inconsistency?

- **Conceptual evaluation** of the transitivity assumption should always take place
  - ✓ Look at the distribution of effect modifiers across studies
  - ✓ Studies should always be checked for systematic differences in populations, interventions, comparators and outcomes

What if we find inconsistency?

Check your data for extraction errors!

Might consider

- splitting intervention nodes in the network
- presenting a variety of separate direct, indirect and mixed comparisons
- Not performing a NMA at all
- Subgroup analyses or network meta-regression

Be careful! Selective inclusion of evidence pieces might lead to bias

Which of the following is correct?

Intransitivity will always lead to inconsistency

Intransitivity will always lead to inconsistency if there are closed loops

Intransitivity may lead to inconsistency, if there are closed loops

Beware of difference in terminology

Unfortunately NMA terminology in the literature has not been yet completely harmonized

- Coherence, similarity or exchangeability are also used as a term for (what we described as) transitivity
- Consistency is sometimes used to describe both the assumption and the statistical disagreement between direct and indirect evidence
- Side splitting is sometimes used to describe the back-calculation, although the two methods are not identical
- And also mixed treatment comparisons (MTC), multiple treatments meta-analysis (MTM) are used instead of network meta-analysis

A very nice series of 7 methodological papers for evidence synthesis and NMA was published in MDM

Evidence Synthesis for Decision Making 1: Introduction

Sofia Dias, PhD, Nicky J. Welton, PhD, Alex J. Sutton, PhD, A. E. Ades, PhD

We introduce the series of 7 tutorial papers on evidence synthesis methods for decision making, based on the Technical Support Documents in Evidence Synthesis prepared for the National Institute for Health and Clinical Excellence (NICE) Decision Support Unit. Although oriented to NICE's Technology Appraisal process, which examines new pharmaceutical products in a cost-effectiveness framework, the methods presented throughout the tutorials are equally relevant to clinical guideline development and to comparisons between medical devices, or public health interventions. Detailed guidance is given on how to use the other tutorials in the series, which propose a single evidence synthesis framework that covers fixed and random effects models, pairwise meta-analysis, indirect comparisons, and network meta-analysis, and where outcomes expressed in several different reporting formats can be analyzed without recourse to normal approximations. We describe the principles of evidence synthesis required by the 2008 revision of the NICE Guide to the Methods of Technology Appraisal and explain how the approach proposed in these tutorials was designed to conform to those requirements. We finish with some suggestions on how to present the evidence, the synthesis methods, and the results. ***Key words:*** *cost-effectiveness analysis; Bayesian meta-analysis; systematic reviews.* ***(Med Decis Making 2013;33:597–606)***

This paper is the first of 7 tutorial papers on evidence synthesis methods in decision making, which are based on the Technical Support Documents (TSDs) in Evidence Synthesis prepared for worked examples. Although aimed at those making, reviewing, and appraising submissions to NICE, the TSDs do not attempt to "prescribe" the form that analyses must take or the methods that must be

Also, check out our recent review on the methodology of NMA

Tutorial

Research Synthesis Methods

Received 21 November 2014, Revised 30 September 2015, Accepted 06 November 2015 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1195

GetReal in network meta-analysis: a review of the methodology

Orestis Efthimiou,[a*] Thomas P. A. Debray,[b,c] Gert van Valkenhoef,[d] Sven Trelle,[e,f] Klea Panayidou,[e] Karel G. M. Moons,[b,c] Johannes B. Reitsma,[b,c] Aijing Shang[g] and Georgia Salanti[a,†] on behalf of GetReal Methods Review Group[‡]

Pairwise meta-analysis is an established statistical tool for synthesizing evidence from multiple trials, but it is informative only about the relative efficacy of two specific interventions. The usefulness of pairwise meta-analysis is thus limited in real-life medical practice, where many competing interventions may be available for a certain condition and studies informing some of the pairwise comparisons may be lacking. This commonly encountered scenario has led to the development of network meta-analysis (NMA). In the last decade, several applications, methodological developments, and empirical studies in NMA have been published, and the area is thriving as its relevance to public health is increasingly recognized. This article presents a review of the relevant literature on NMA methodology aiming to pinpoint the developments that have appeared in the field.

Literature

- Efthimiou O, Debray TPA, van Valkenhoef G, et al (2016) GetReal in network meta-analysis: a review of the methodology. Res Synth Methods. doi: 10.1002/jrsm.1195
- Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* 2012; 3: 80-97.
- Cooper NJ, Sutton AJ et al. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine* 2009; 28: 1861–1881.
- Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine* 2010; 29: 932-944.
- Lu G, Ades AE. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *Journal of American Statistical Association* 2006; 101: 447-459.
- Higgins JPT, Jackson D, Barrett J et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Research Synthesis Methods* 2012; 3: 98-110.