

Methods to detect and account for inconsistency

Georgia Salanti

Institute of Social and Preventive Medicine
University of Bern

Acknowledgements for some slides: Julian Higgins

Who is tallest?

Anna

Boq

Cora



Inconsistency can be viewed as a special form of heterogeneity

“Dispertion in the relative treatment effects evaluated in *different settings*”

When different settings means different studies= **heterogeneity**

Inconsistency can be viewed as a special form of heterogeneity

“Dispertion in the relative treatment effects evaluated in *different settings*”

When different settings means different sources of evidence= **inconsistency**

Inconsistency can be viewed as a special form of heterogeneity

Heterogeneity: variability beyond chance within a comparison

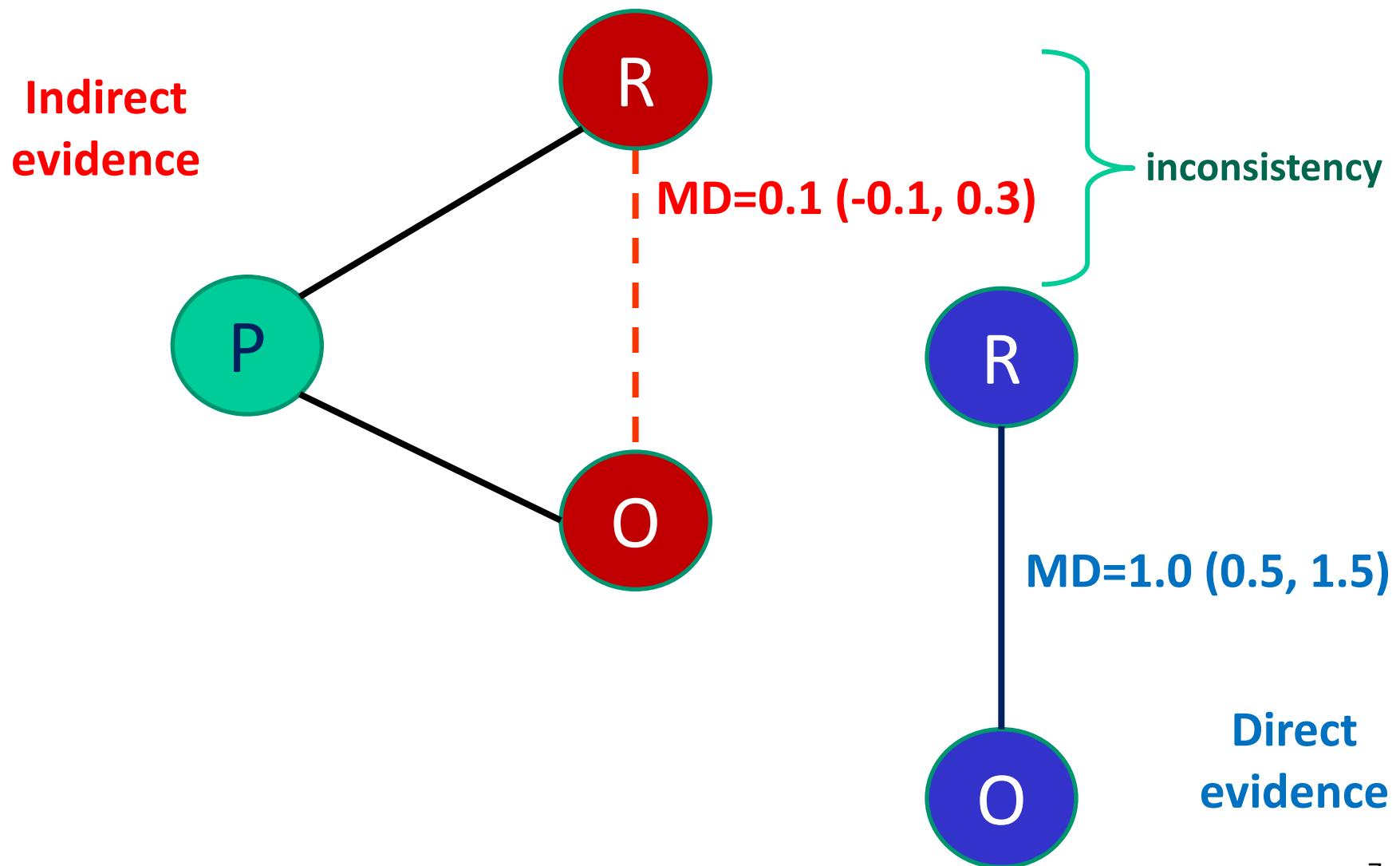
Inconsistency: variability beyond chance across comparisons

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!**)

Inconsistency



Estimating inconsistency

- In a ABC loop of evidence:

$$IF = |\mu_{BC}^I - \mu_{BC}^D| = |\mu_{AC}^I - \mu_{AC}^D| = |\mu_{AB}^I - \mu_{AB}^D|$$

$$\text{var(IF)} = \text{var}(\mu_{BC}^I) + \text{var}(\mu_{BC}^D)$$

$$95\% \text{CI}: IF \pm 1.96 \sqrt{\text{var(IF)}}$$

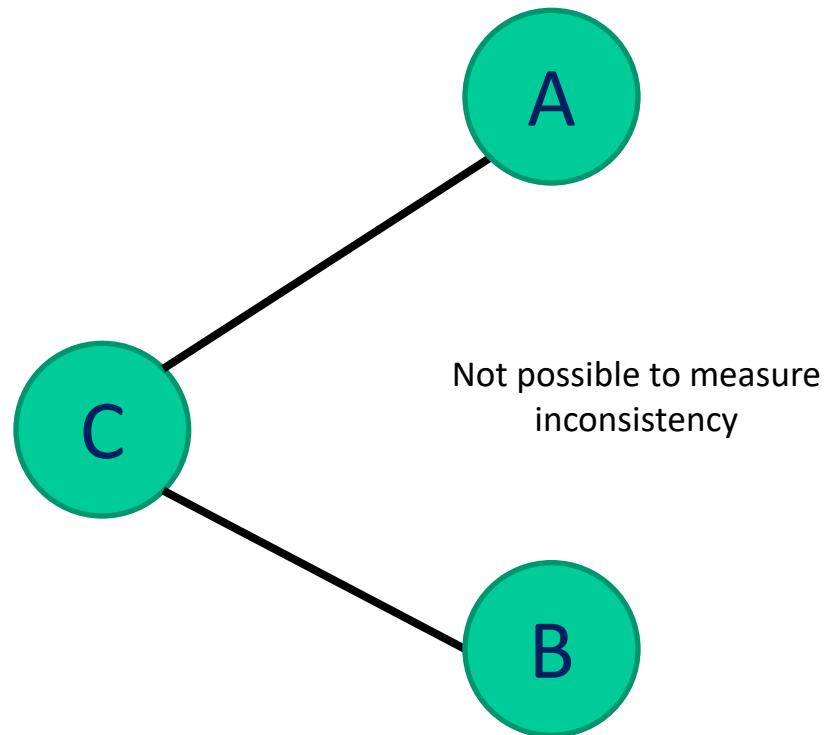
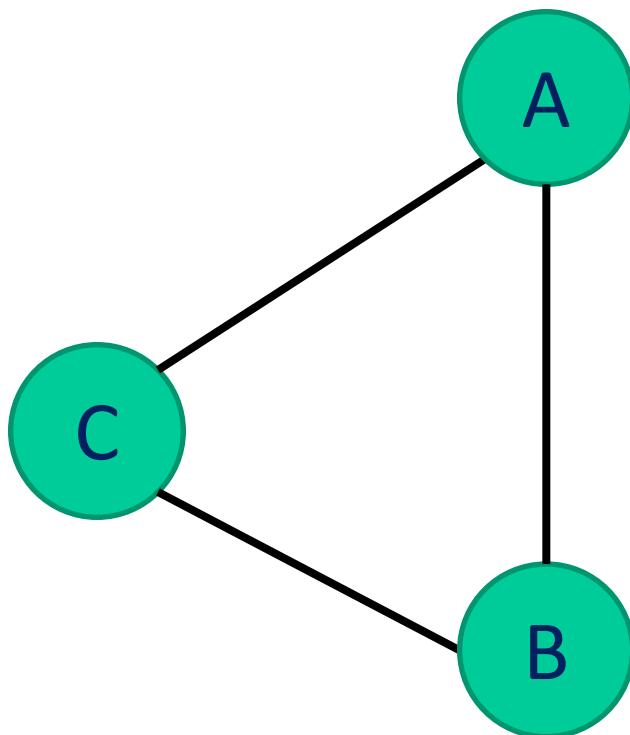
- If the 95% CI excludes zero, then there is statistically significant inconsistency
- A test for $H_0: IF=0$

$$z = \frac{IF}{\sqrt{\text{var(IF)}}} \sim N(0,1)$$



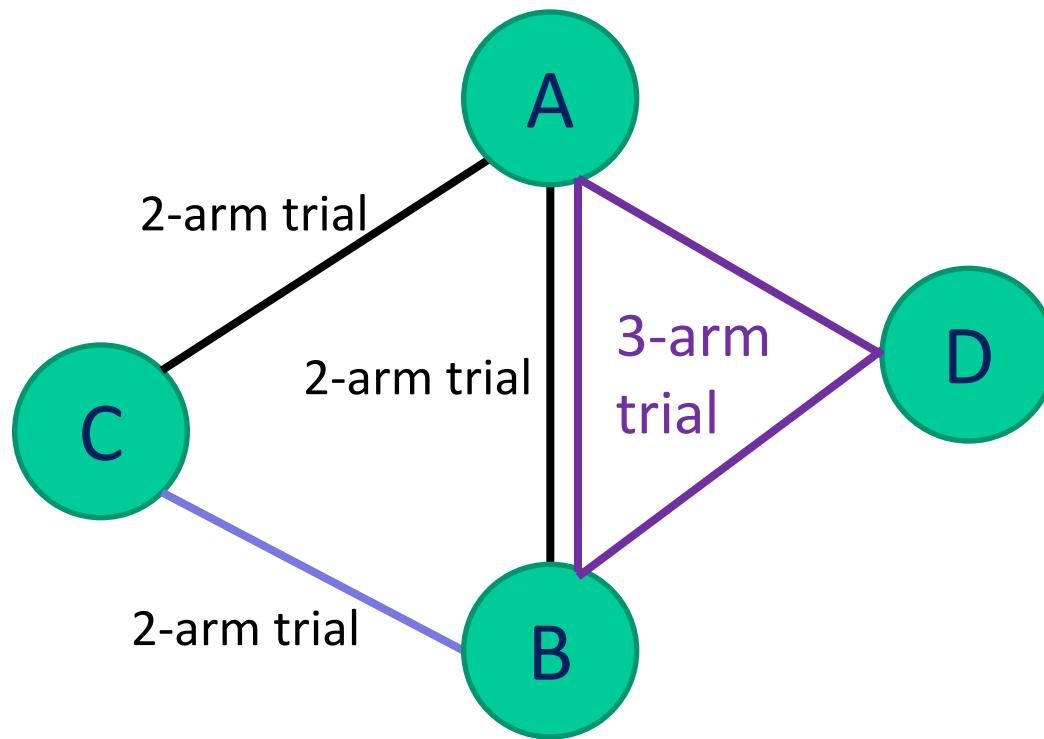
Inconsistency

- 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



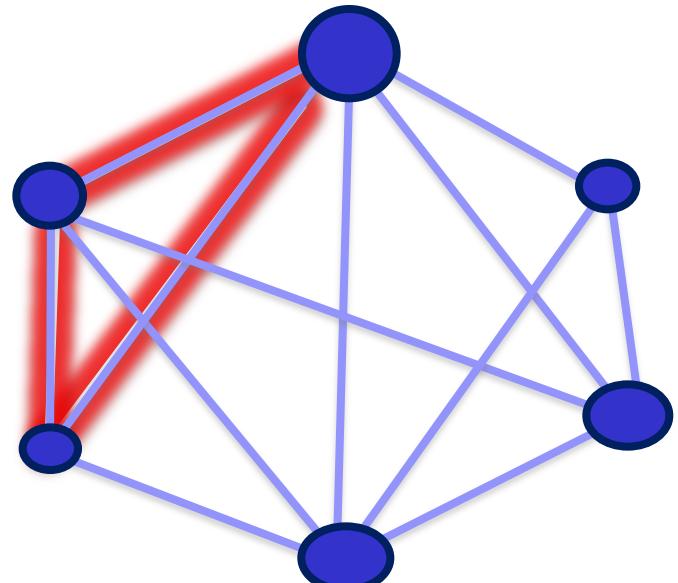
Inconsistency in network meta-analysis

- It's important to examine the validity of the consistency equations if possible
- It's only possible if we have **closed loops** in the network plot
 - and these do not comprise only multi-arm studies

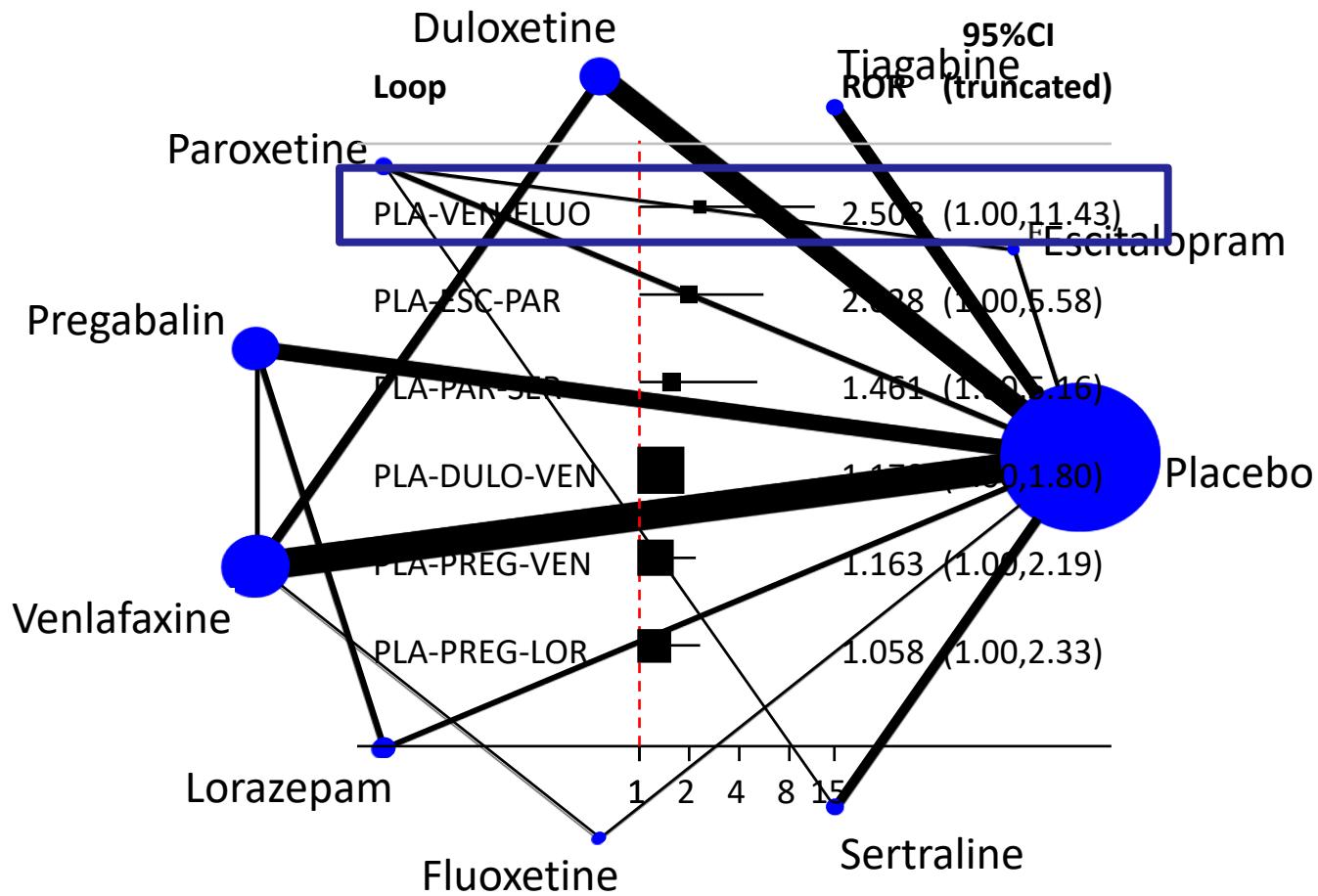


Appropriate methods to evaluate inconsistency

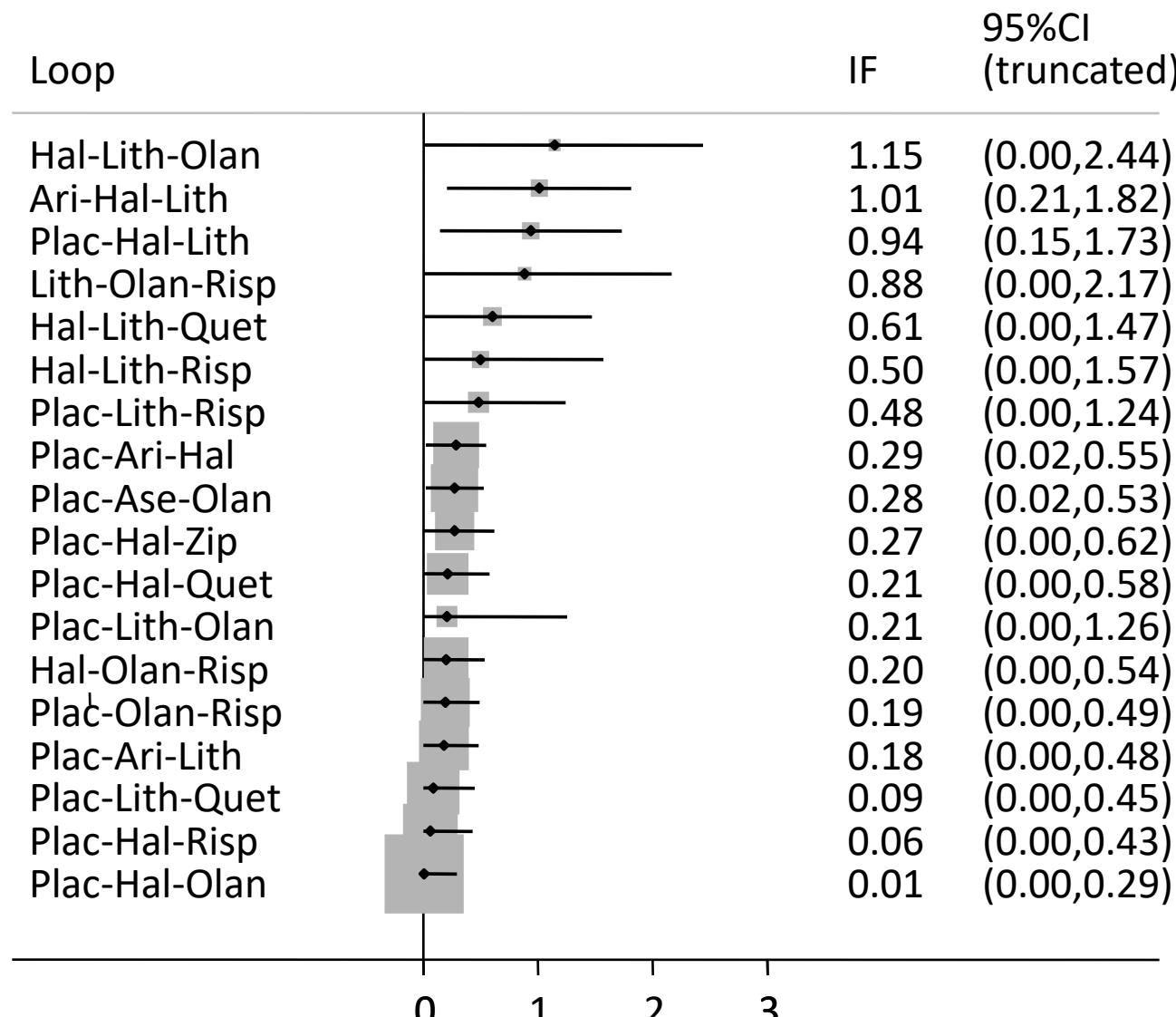
- Local tests
 - focus on a loop



Extend this to all loops



All loops in the acute mania data



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



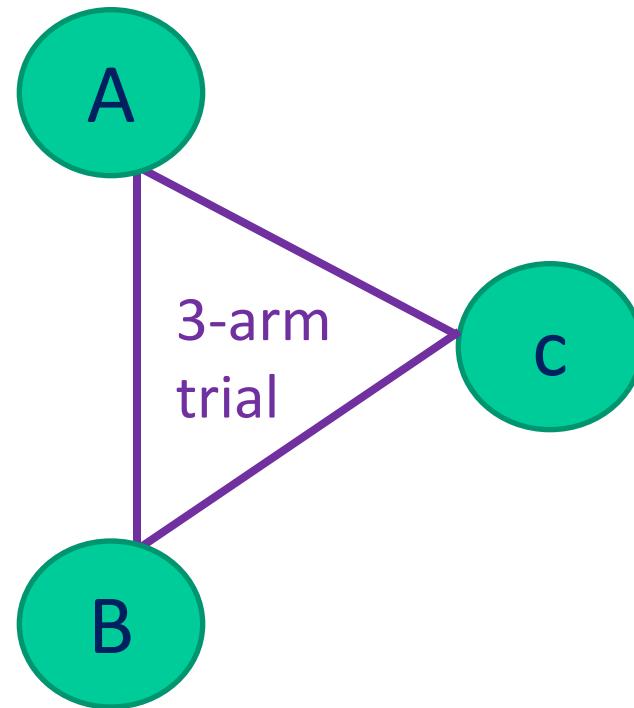
- ✓ Studies comparing more than two treatments
- ✓ Treatment effects in multi-arm studies are correlated
- ✓ E.g. in an ABC study, the relative treatment effects AB and AC are not independent, because they share the same arm, A

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



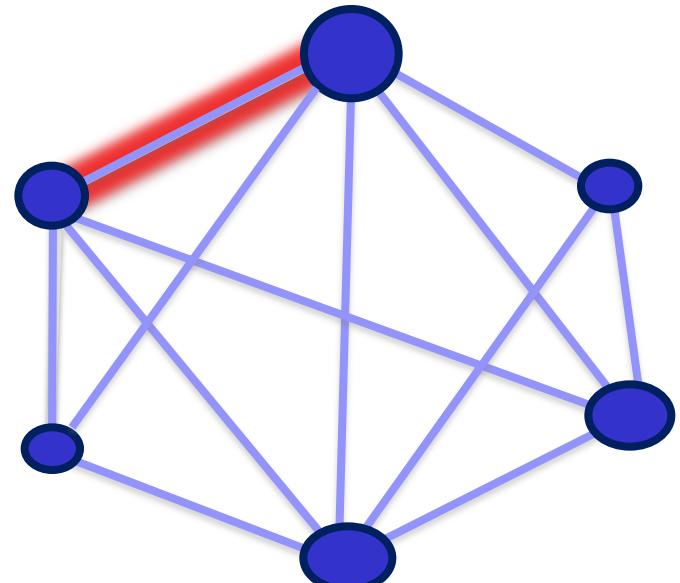
a detour: multi-arm studies

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



Appropriate methods to evaluate inconsistency

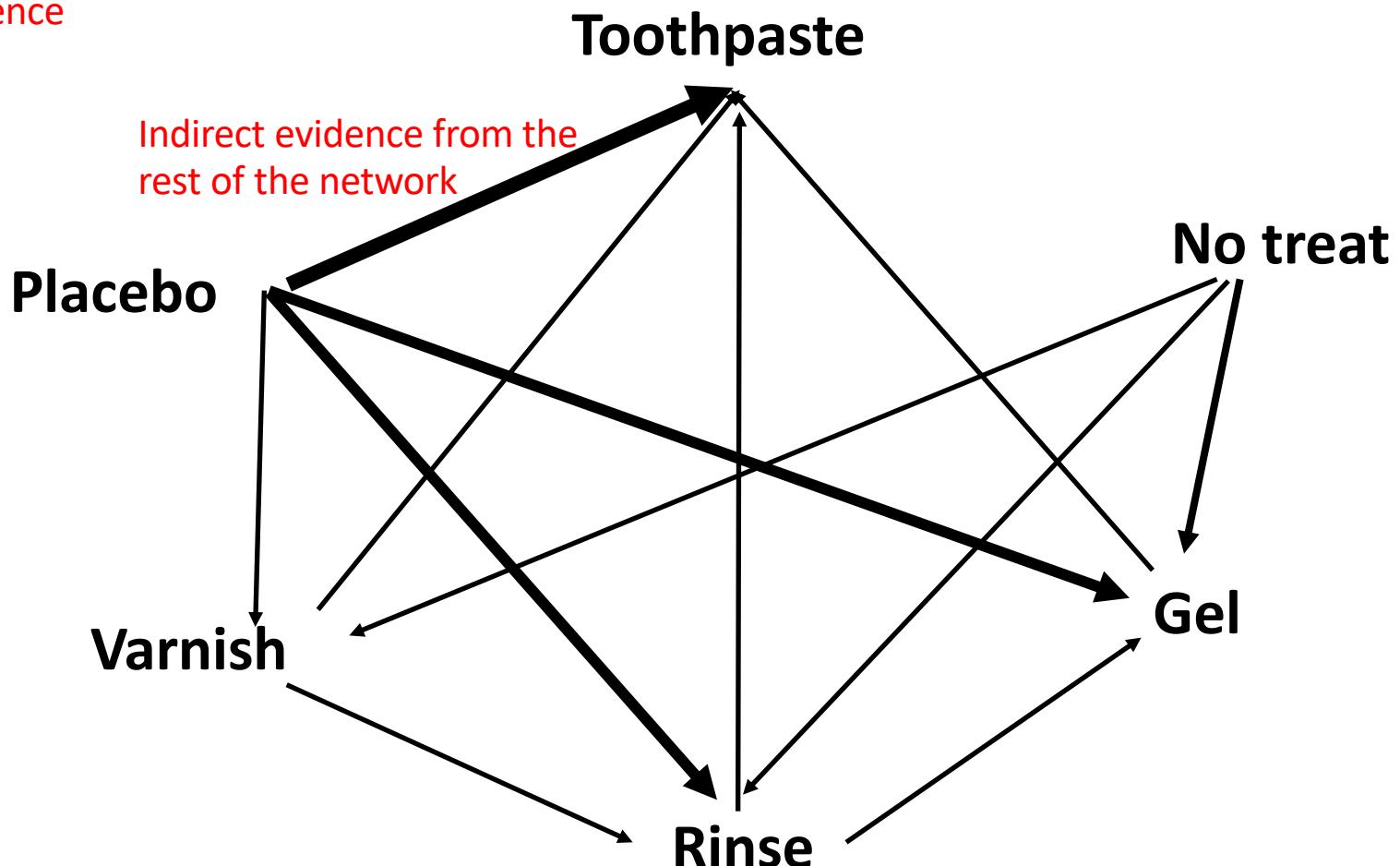
- Local tests
 - focus on a comparison



SIDE (*Separating Indirect from Direct Evidence*)

or “node-splitting” or “side-splitting”

Direct
evidence



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

SIDE2: Back calculation method

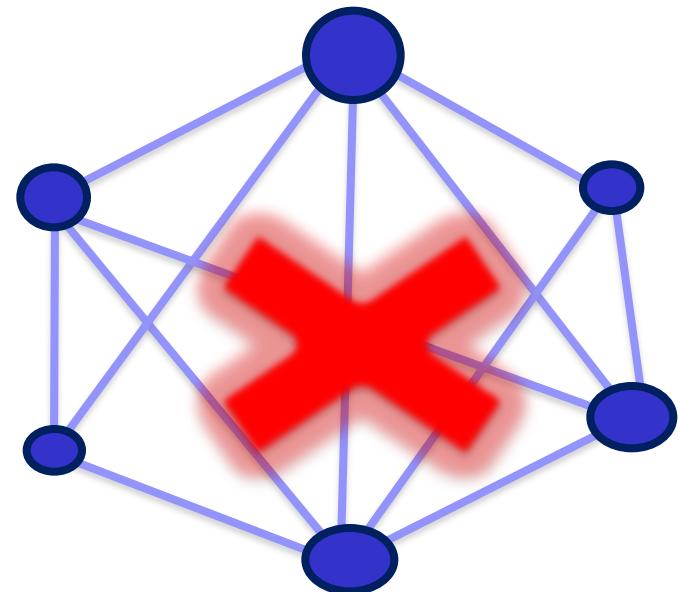
- This is an approach **very similar to SIDE**
- Using this method, for each treatment comparison we partition the evidence in the network in two independent parts: **direct** and **indirect**



Easy to use, very quick in R (`netsplit` command)

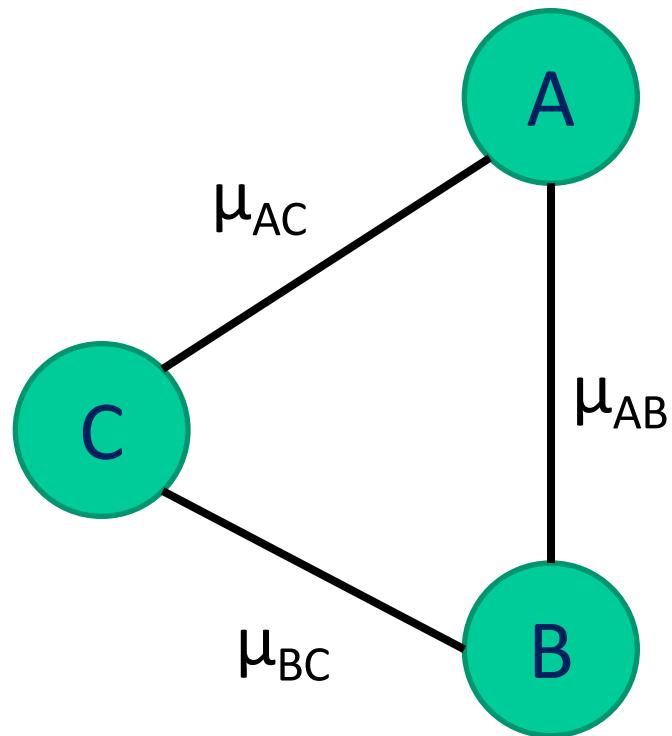
Appropriate methods to evaluate inconsistency

- Global tests



Modelling inconsistency

The consistency model

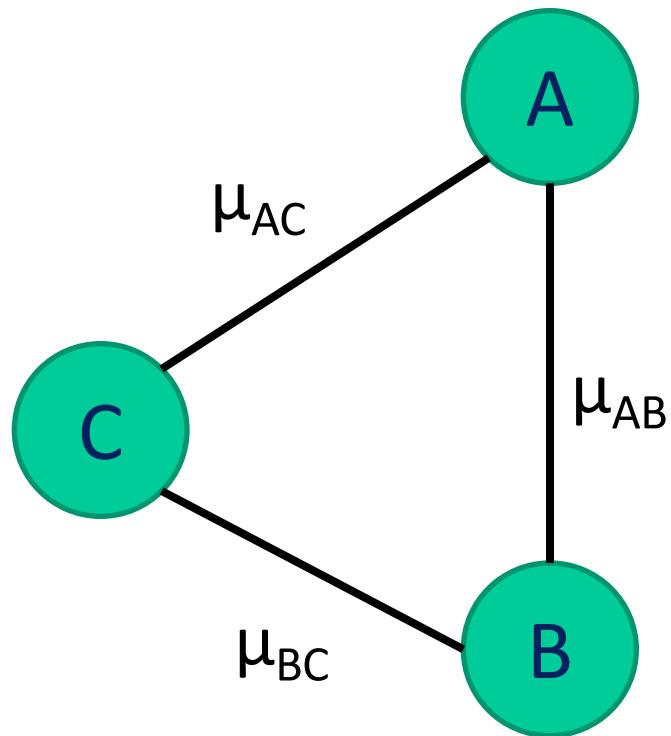


Consistency equation

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

Modelling inconsistency

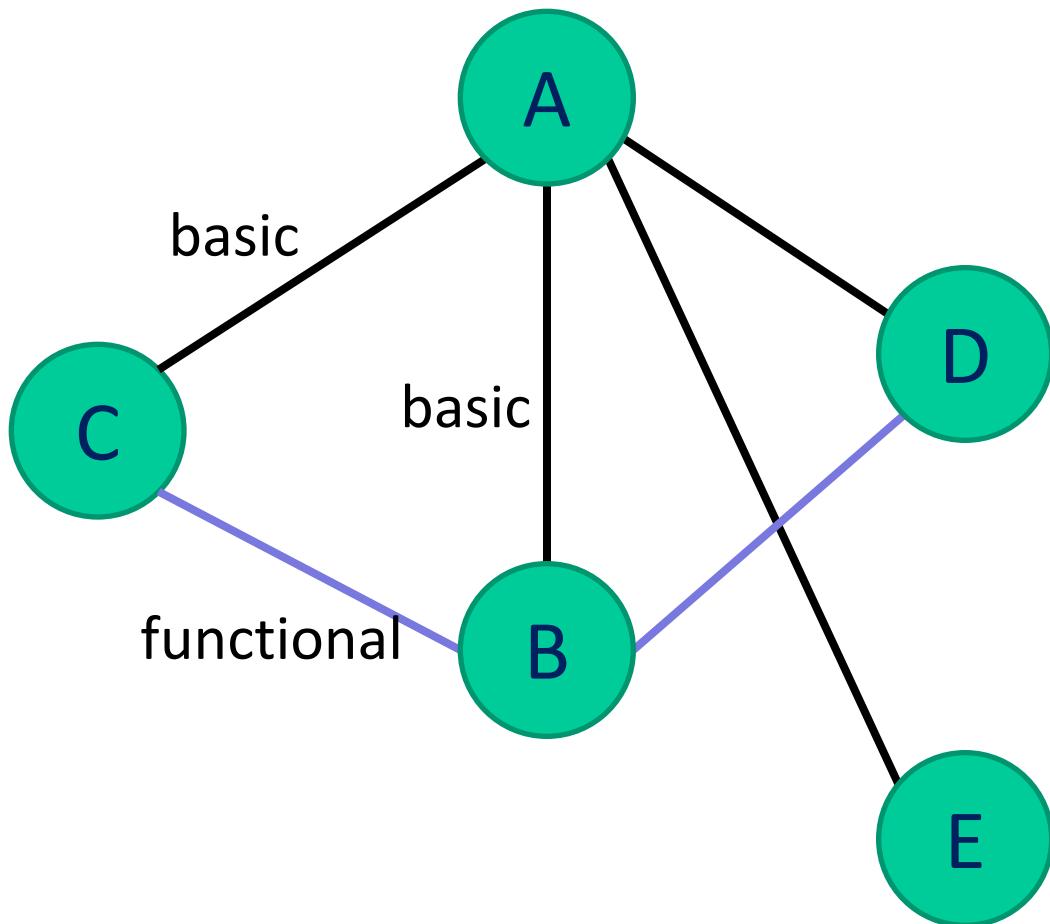
The inconsistency model



Inconsistency equation

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

Inconsistency parameters



Consistency equations:

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

$$\mu_{BD} = \mu_{AD} - \mu_{AB} + w_{BD}$$

$$\mu_{DC} = \mu_{AC} - \mu_{AD}$$

$$\mu_{BE} = \mu_{AE} - \mu_{AB}$$

$$\mu_{DE} = \mu_{AE} - \mu_{AD}$$

$$\mu_{CE} = \mu_{AE} - \mu_{AC}$$

How many inconsistency parameters?

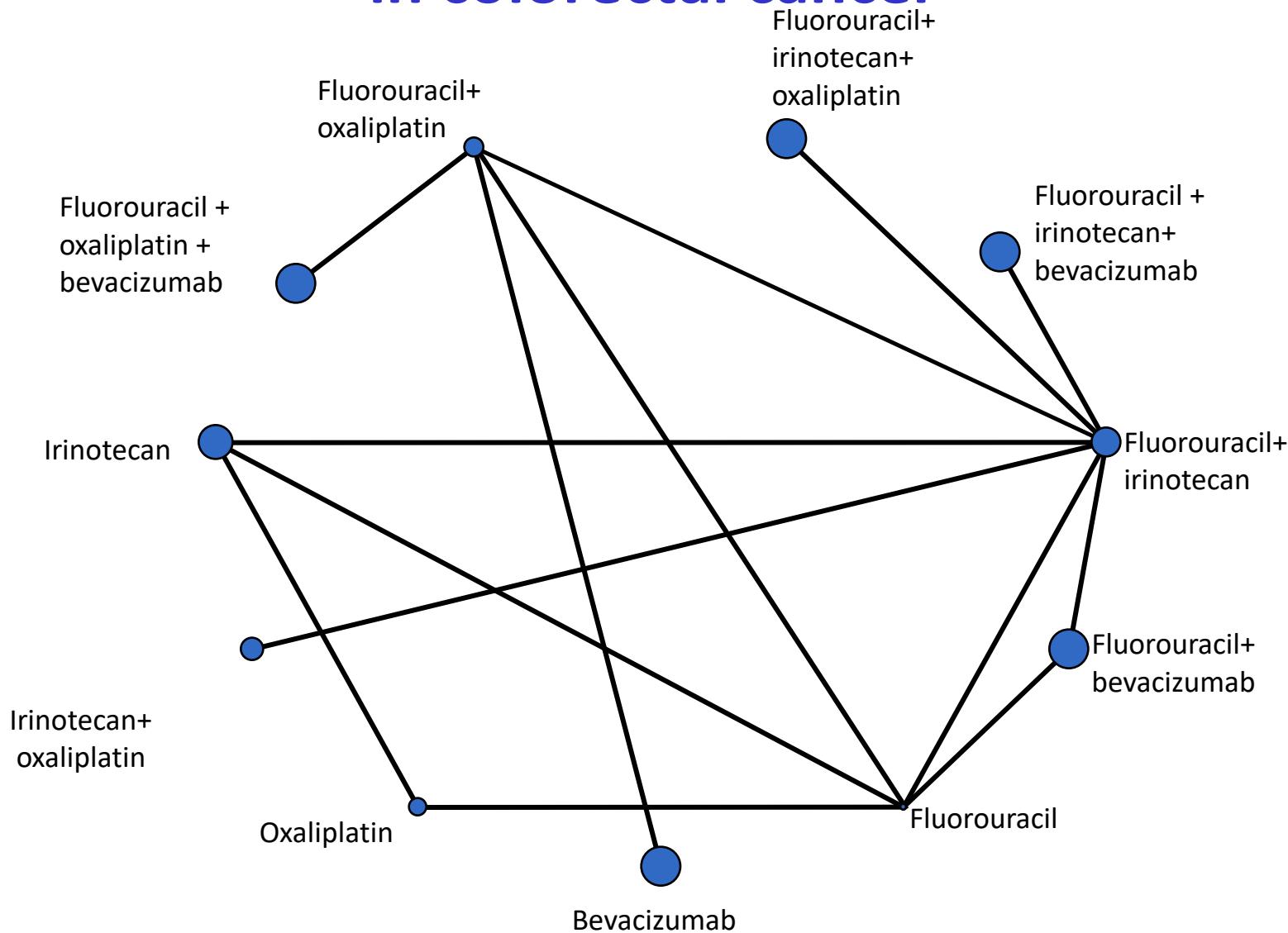
- Number of inconsistency parameters required to allow inconsistency in every closed loop =

#comparisons with data – (#treatments – 1)

$$N_{\text{comp}} - (T - 1)$$

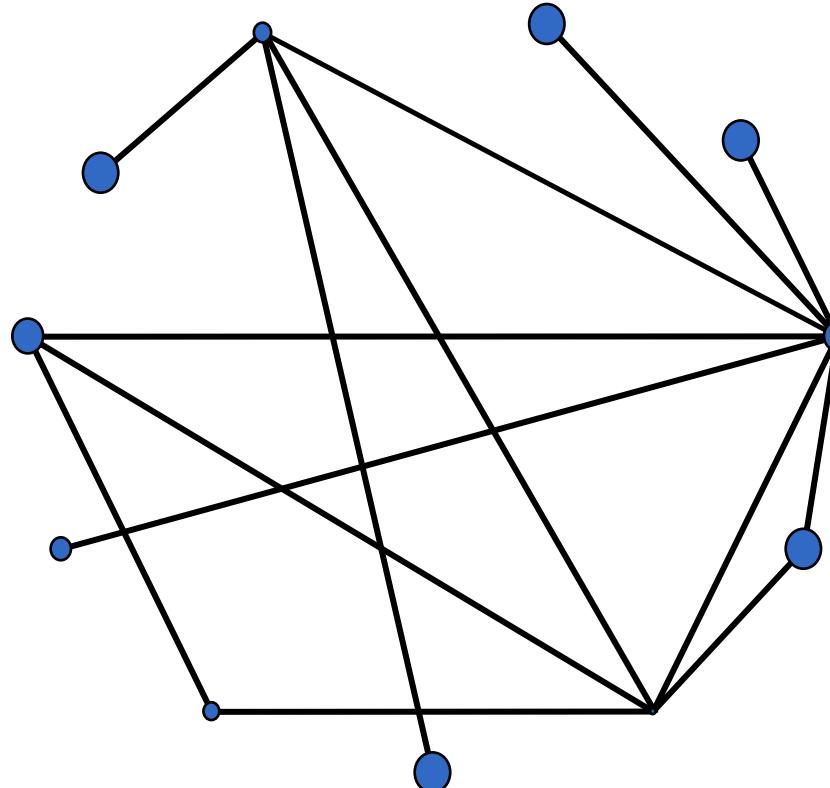
- If basic parameters are chosen such that we have direct evidence on each of them, then this is equivalent to the number of functional parameters that have data (Lu & Ades' approach)
- (Only applies if there are no multi-arm studies)

Example: Survival with 11 chemotherapy regimens in colorectal cancer

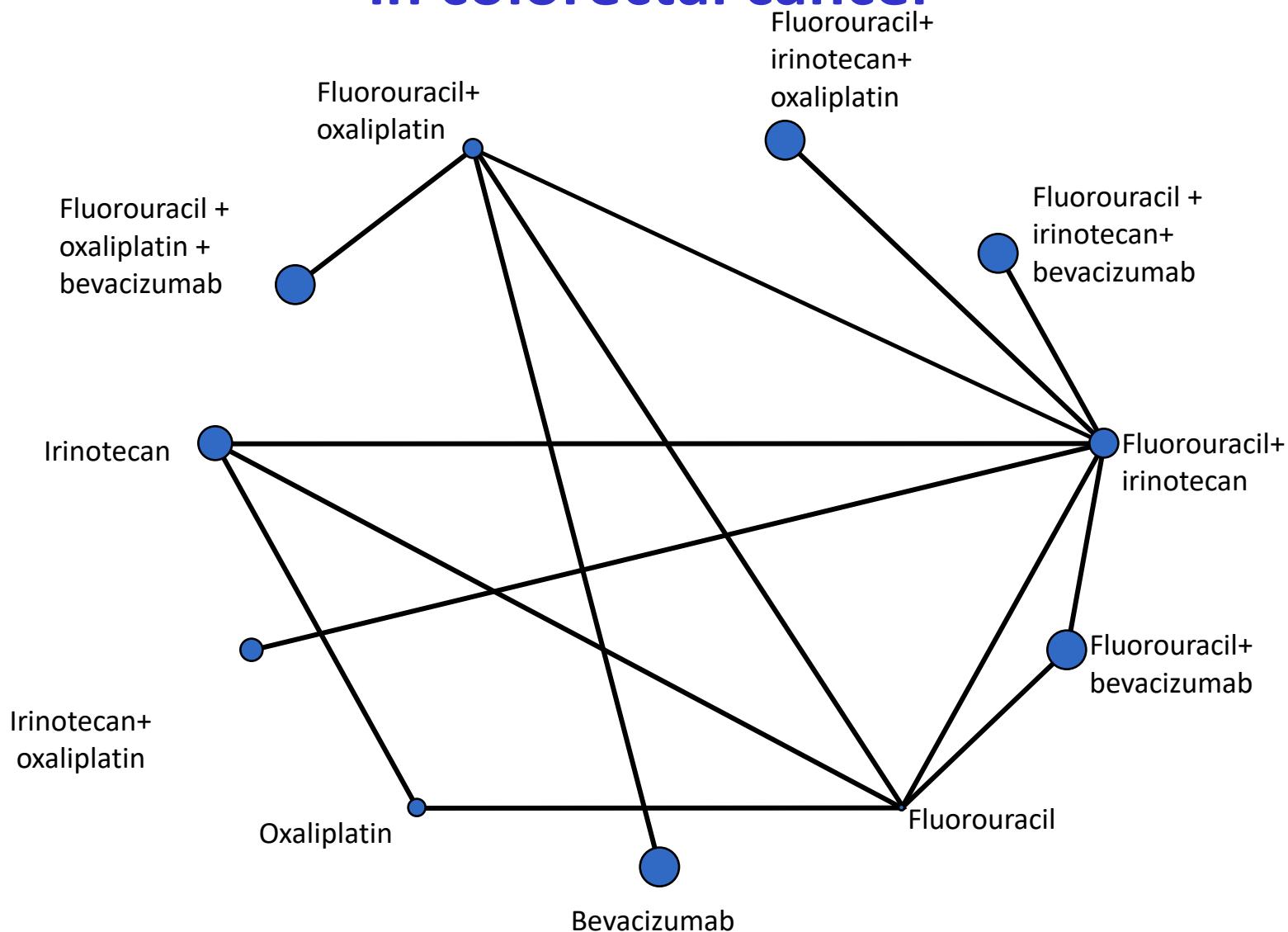


How many inconsistency factors (i.e. closed loops)?

- A. 1
- B. 2
- C. 3
- D. 4
- E. 5
- F. 6
- G. more

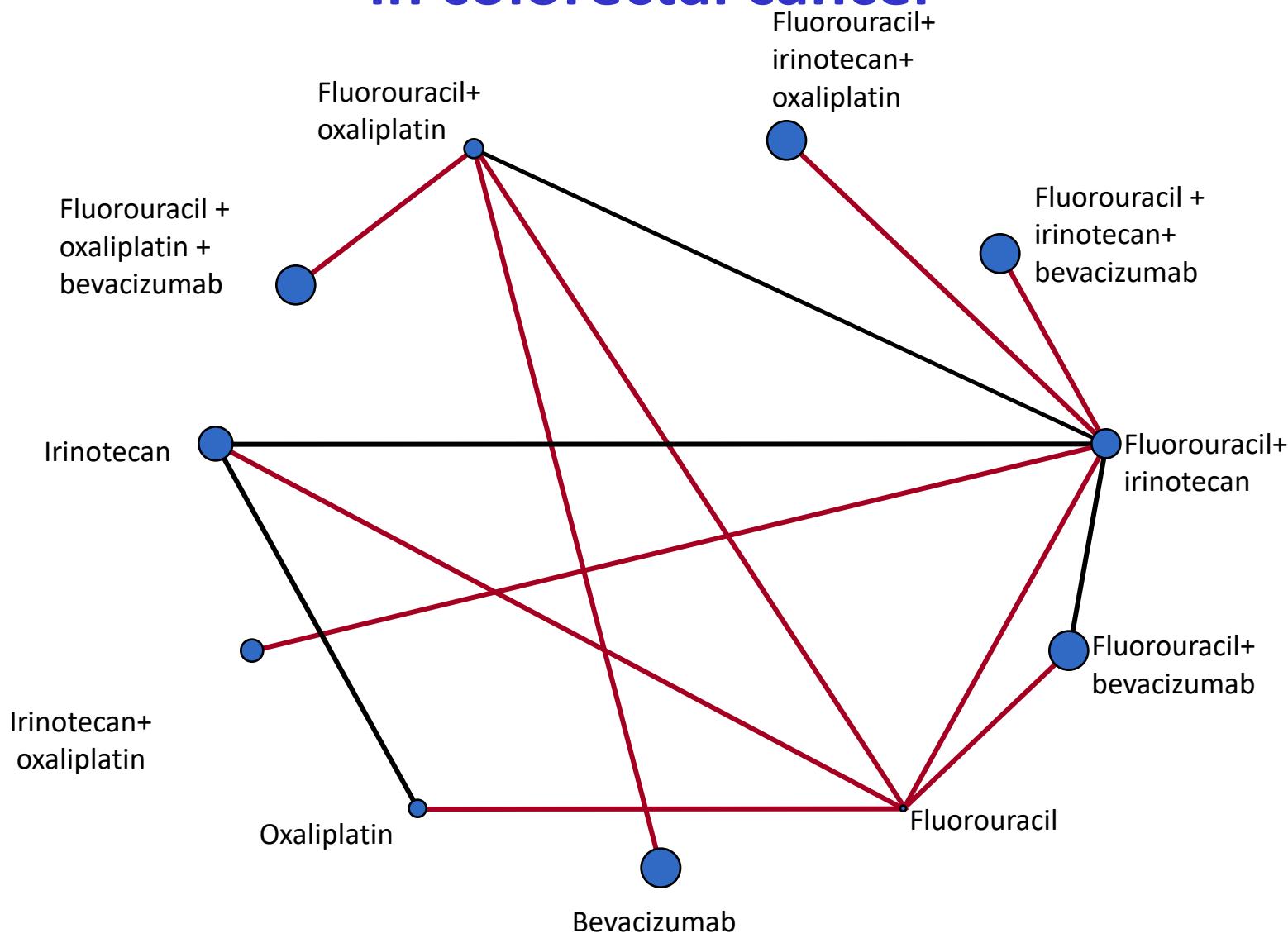


Example: Survival with 11 chemotherapy regimens in colorectal cancer



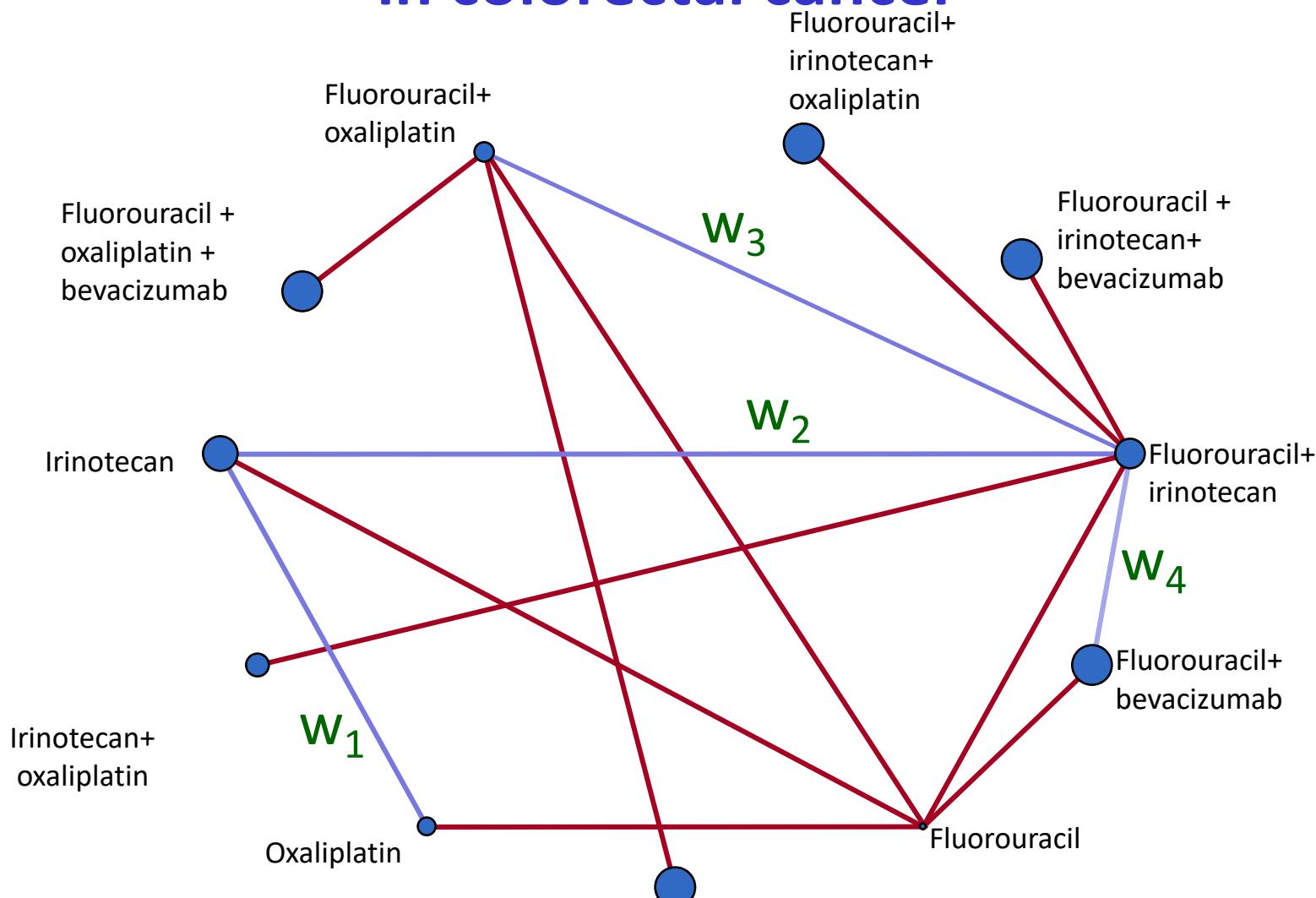
Ncomp = 14, T = 11 Number of inconsistency factors = 4

Example: Survival with 11 chemotherapy regimens in colorectal cancer



10 basic parameters with data

Example: Survival with 11 chemotherapy regimens in colorectal cancer



10 h

On a logHR scale the inconsistency terms are estimated
 $w_1=-0.08$, $w_2=-0.07$, $w_3=-0.06$, $w_4=-0.03$

rs

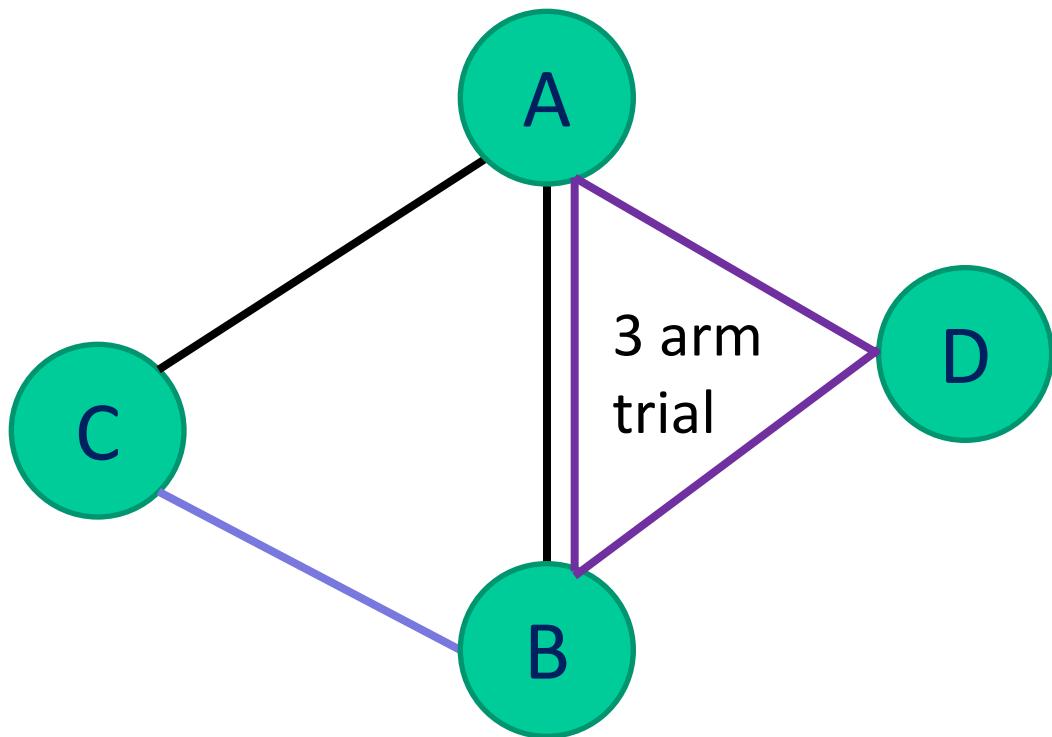
A complication

- But what if a loop is due to a multi-arm study?
- Or if a loop is a mixture of single-arm and multi-arm studies?
- 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

Design by treatment interaction model

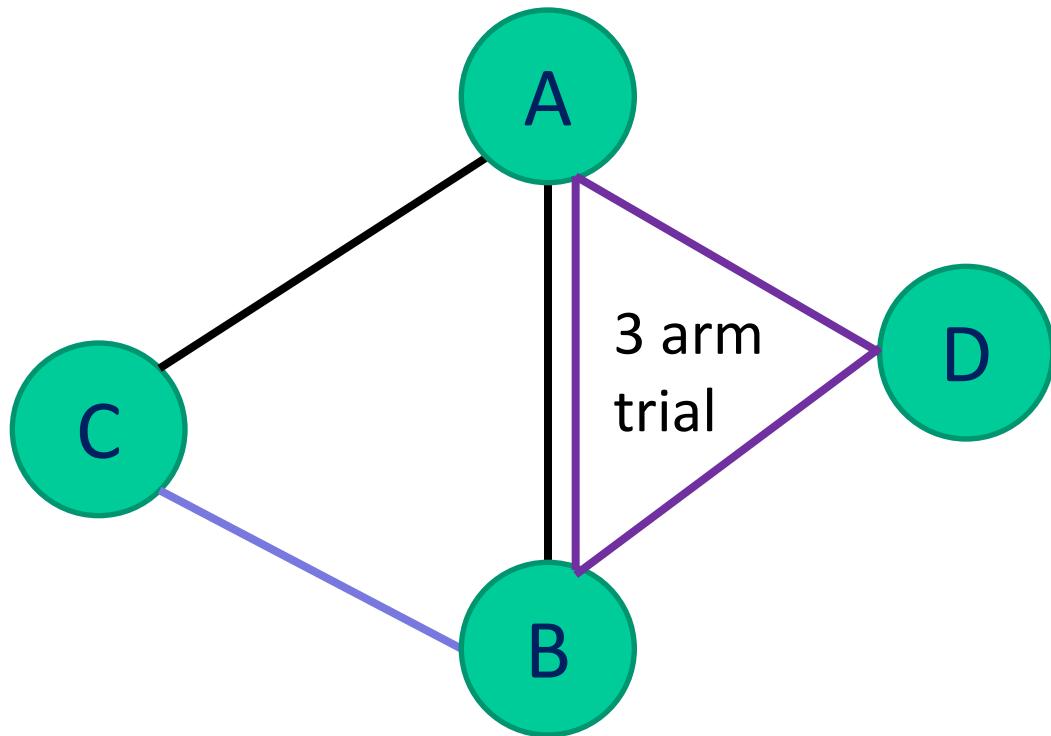
- Inconsistency can be between ‘designs’ as well as between direct and indirect evidence
- Design:= the treatments compared in a trial
 - ABC is a different design from AB or BC
- Design inconsistency: when the relative effectiveness of A versus B is different across designs
 - μ_{AB} is different when estimated in AB or ABC studies
- We then have more inconsistency factors

'Design by treatment interaction'



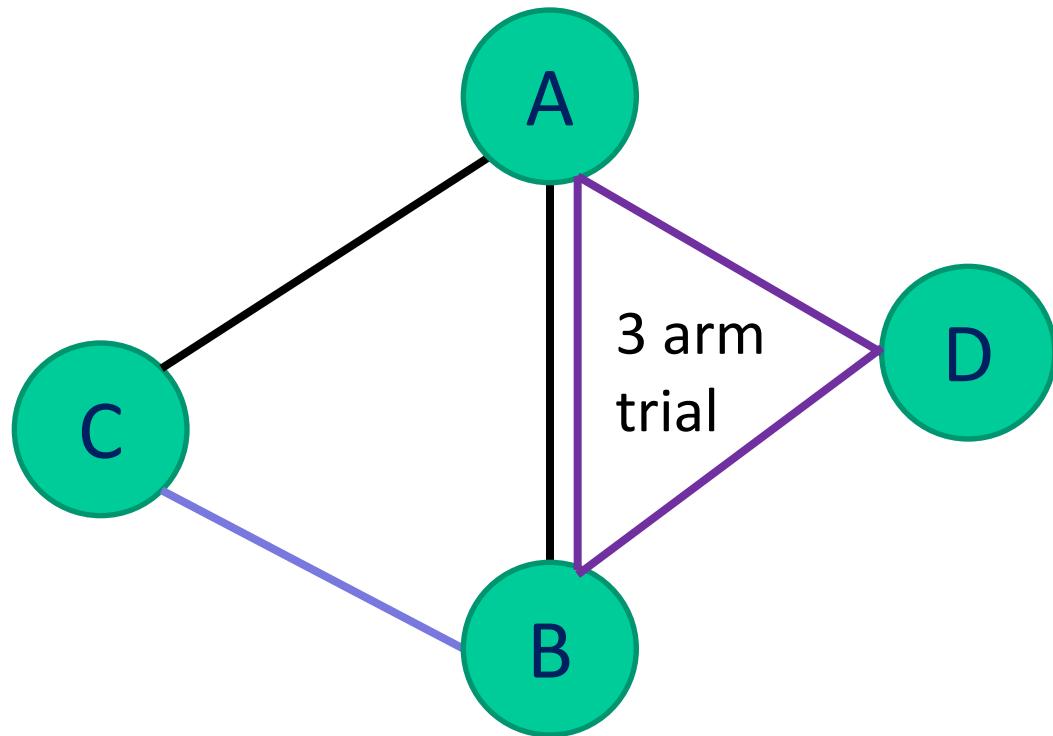
How many inconsistency factors due to closed loops?

- A. 1
- B. 2
- C. 3
- D. 4

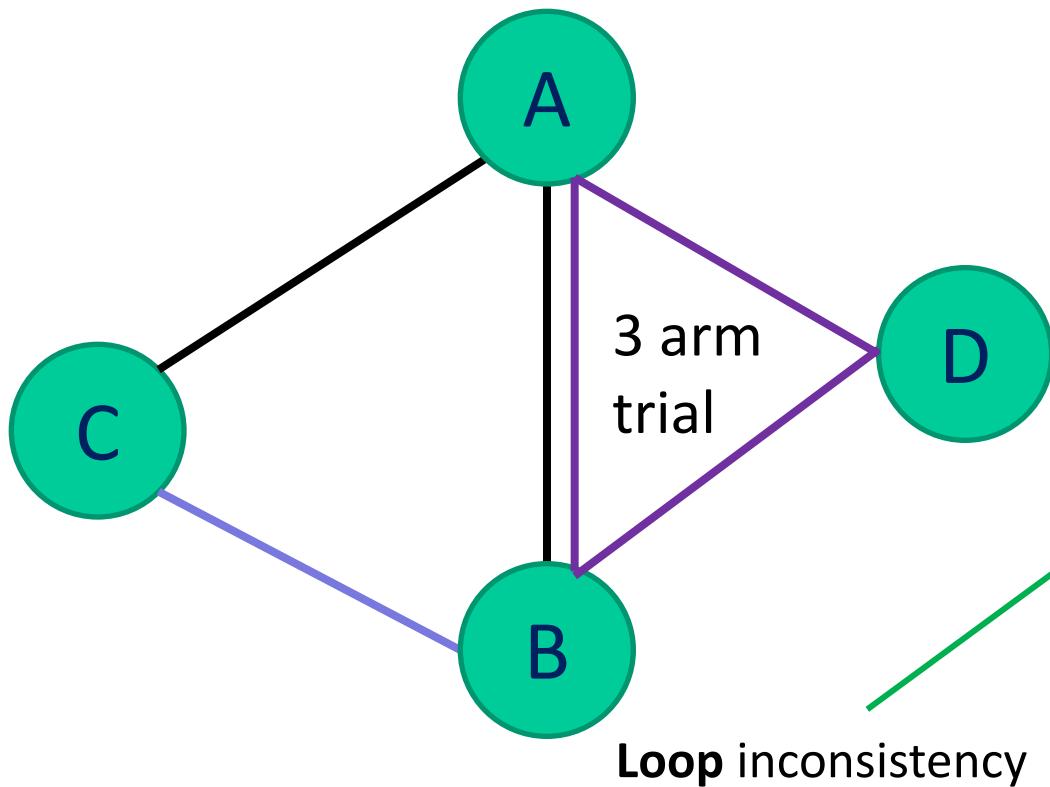


How many inconsistency factors due to design?

- A. 1
- B. 2
- C. 3
- D. 4



'Design by treatment interaction'



(in)consistency model:

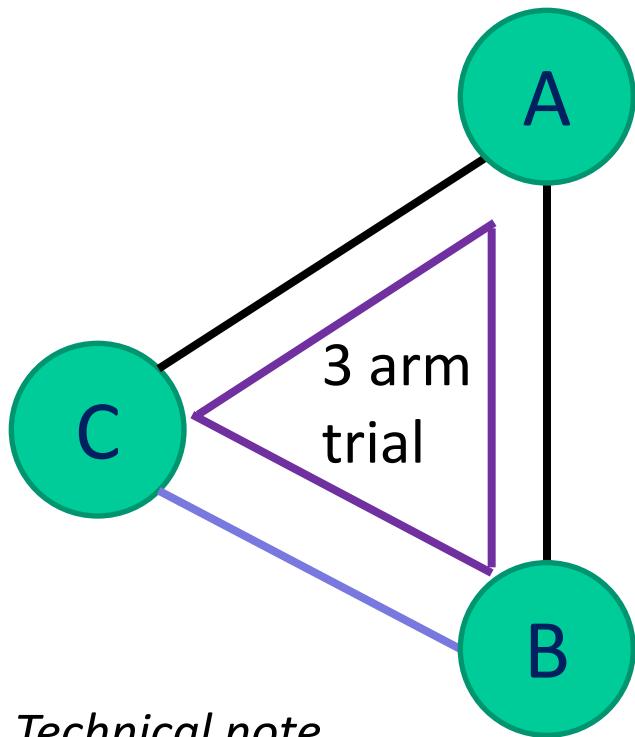
$$\mu_{BC} = \mu_{AC} - \mu_{AB} \quad + w_{BC}$$

$$\mu_{BD} = \mu_{AD} - \mu_{AB} \quad + w_{AB}$$

$$\mu_{DC} = \mu_{AC} - \mu_{AD}$$

Design inconsistency for disagreement between the three arm ABC trial(s) and the AB trial(s)

‘Design by treatment interaction’



(in)consistency model (*conceptually*)

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

$$\mu_{AB} \text{ from AB} = \mu_{AB} \text{ from ABC} + w_{AB}$$

$$\mu_{AC} \text{ from AC} = \mu_{AC} \text{ from ABC} + w_{AC}$$

Technical note

- **network inconsistency** models add three covariates and doing network meta-regression
- there are lots of different ways to parameterize the same model

Global test (1)

White 2012 RSM

We collect all design (d) -specific w_d based on the DBT model in a column vector \mathbf{G} .

In order to assess the presence of inconsistency in the entire network we can assess the null hypothesis $H_0: \mathbf{G} = \mathbf{0}$ using a χ^2 -test with b degrees of freedom

$$W = \mathbf{G}' \mathbf{Z}^{-1} \mathbf{G}$$

where \mathbf{Z} is the $b \times b$ variance-covariance matrix of \mathbf{G} .

Different parameterizations for the DBT model may result in different w_d values, but the overall W test is the same for all parameterizations (14,15).

Global test (2)

Krahn 2013

Overall Q statistic: Q_{total}

- ▶ Measures the total heterogeneity / inconsistency in the network
- ▶ Design is a plausible source of inconsistency, i.e., effect of A versus B can be different across designs ($A:B$, $A:B:C$, $A:B:D$, ...)

Decomposition of Q_{total} by design (Krahn et al., 2013):

$$Q_{total} = Q_B + Q_W$$

- ▶ Q_B : Inconsistency in treatment effects between different designs
(design inconsistency)
- ▶ Q_W : Heterogeneity of study results within the same design
- ▶ Decomposition of Q_B into parts coming from each design possible
- ▶ Decomposition of Q_W into parts coming from each study possible

Design-by-treatment interaction model

- Allows for inconsistency factors to represent
 - loop inconsistency
 - inconsistency between designs
- **Fixed effects:** Can interpret the inconsistency factors individually
 - this is what **network meta** does
- **Random effects:** Give the inconsistency parameters (w) a random-effects distribution across loops
$$w_j \sim N(0, \sigma^2)$$
 - Compare σ^2 with τ^2 (heterogeneity) to infer about inconsistency
- The inconsistency parameters can usually be interpreted in different ways

Design-by-treatment interaction model

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?

Inconsistency as ‘design-by-treatment interaction’

Advantages

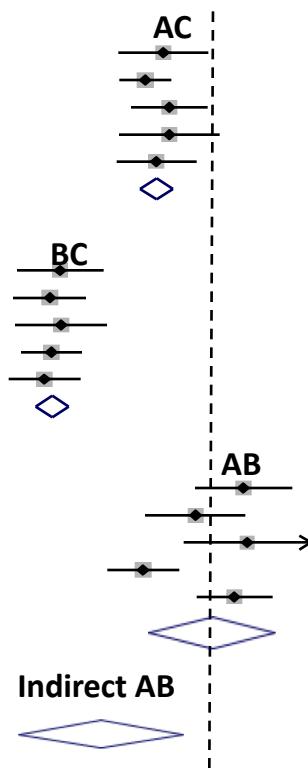
- The hypothesis of consistency in the network can be checked using a **global test**
- Unlike the Lu and Ades model, this global test does not have problems in the presence of **multi-arm studies**

Disadvantages

- 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

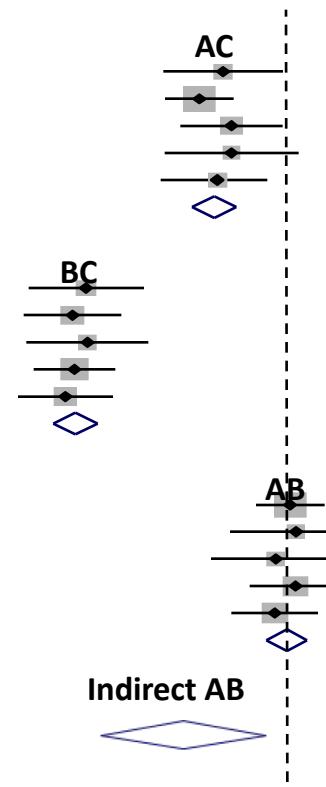
High heterogeneity may mask inconsistency

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

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

Empirical evidence

- Song (2011) examined 112 independent 3-treatment networks and detected 16 cases of statistically significant discrepancies.
- Veroniki et al (2013) examined 315 loops and up to 10% were inconsistent
 - Depends on the estimator of heterogeneity
 - Inconsistency more probable in loops with comparisons informed by a single study
- Veroniki et al (2013) examined 40 networks and one in eight was found statistically inconsistent

What if we don't find any inconsistency?

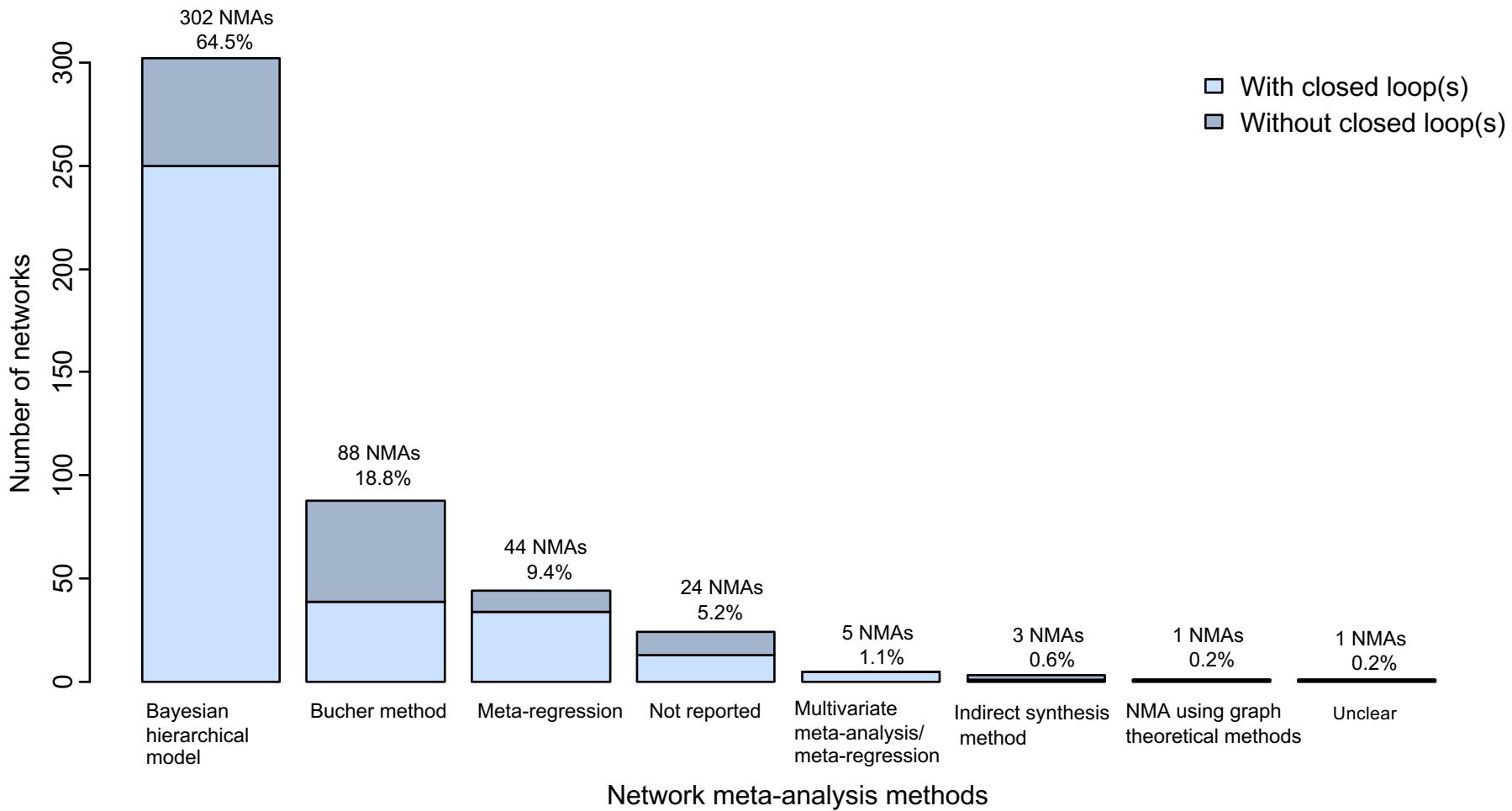
- The absence of significant inconsistency **does not mean** there is consistency
 - ✓ Issues of power and the trade-off with heterogeneity may limit the usefulness of the consistency tests
- Being unable to test for inconsistency **does not mean** there is transitivity
 - ✓ Because inconsistency can only be assessed in loops

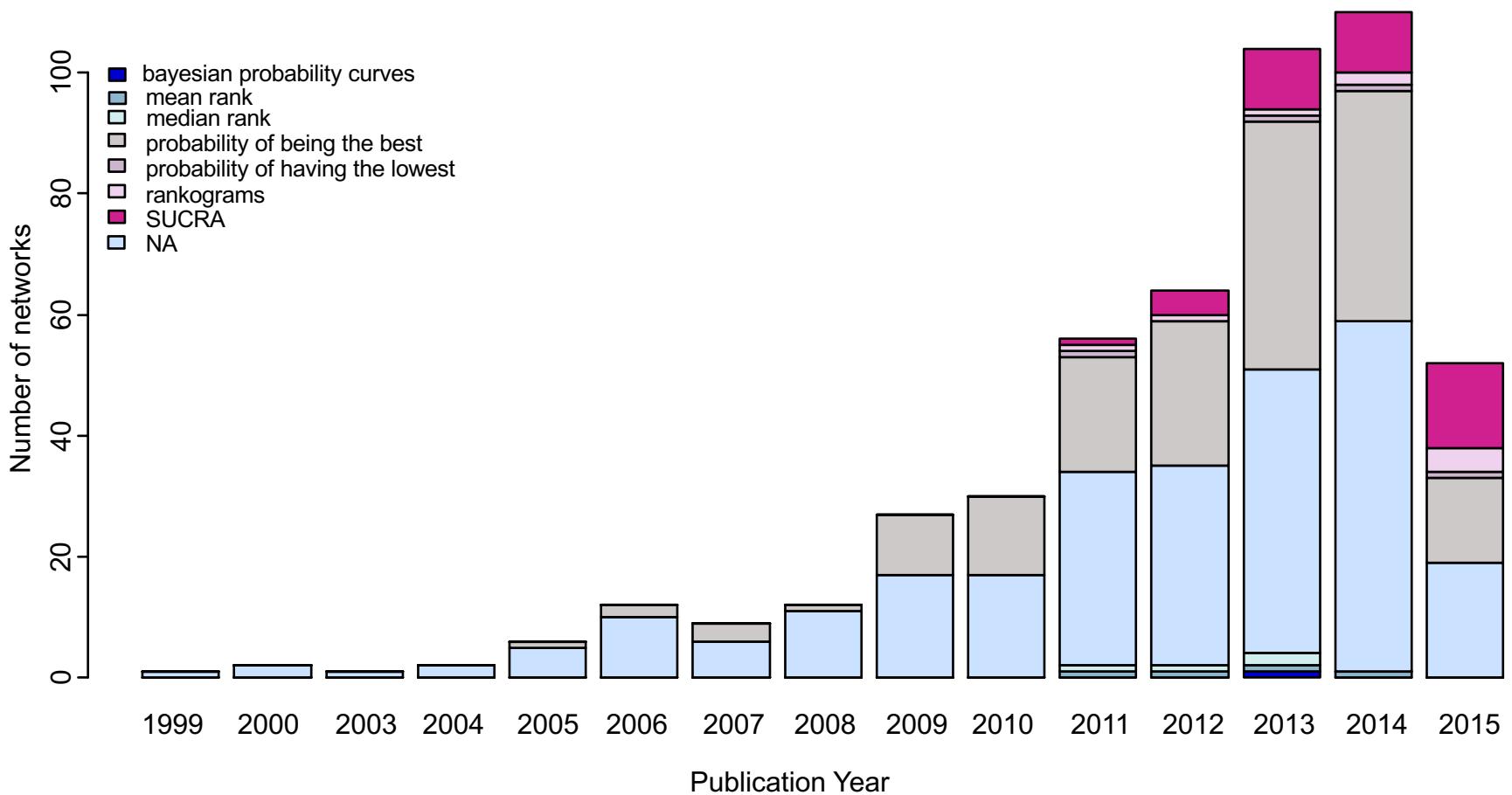
What if we don't find any inconsistency?

- **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 to do when inconsistency is found?

Action	Inconsistency
Check the data	Using simple loop inconsistency you can identify studies with data extraction errors. Inconsistency in loops where a comparison is informed by a single study are particularly suspicious for data errors.
Try to bypass	Empirical evidence suggests that different effect measures of dichotomous outcomes do not impact on statistical inconsistency (Veroniki et al. 2013)
Resign to it	Investigators may decide not to synthesize the network in the presence of excessive inconsistency.
Encompass it	Apply models that relax the consistency assumption by adding an 'extra' loop-specific random effect (Higgins et al. 2012, Lu & Ades 2006).
Explore it	Split the network into subgroups or use network meta-regression to account for differences across studies and comparisons. Specify the variables in the protocol, including bias-related characteristics.





Exploring inconsistency

- Might consider
 - splitting intervention nodes in the network
 - removing parts of the network
 - including study-level characteristics in a network meta-regression

Network meta-regression

- Can account for
 - Patient-level covariates (beware of ecological bias)
 - Risk of bias assessments
 - e.g. Dias et al, 2010
 - Small study effects
 - e.g. Chaimani & Salanti, 2012
 - Sponsorship bias
 - e.g. Cipriani et al, 2009

Which of the following is correct?

Intransitivity will always lead to inconsistency

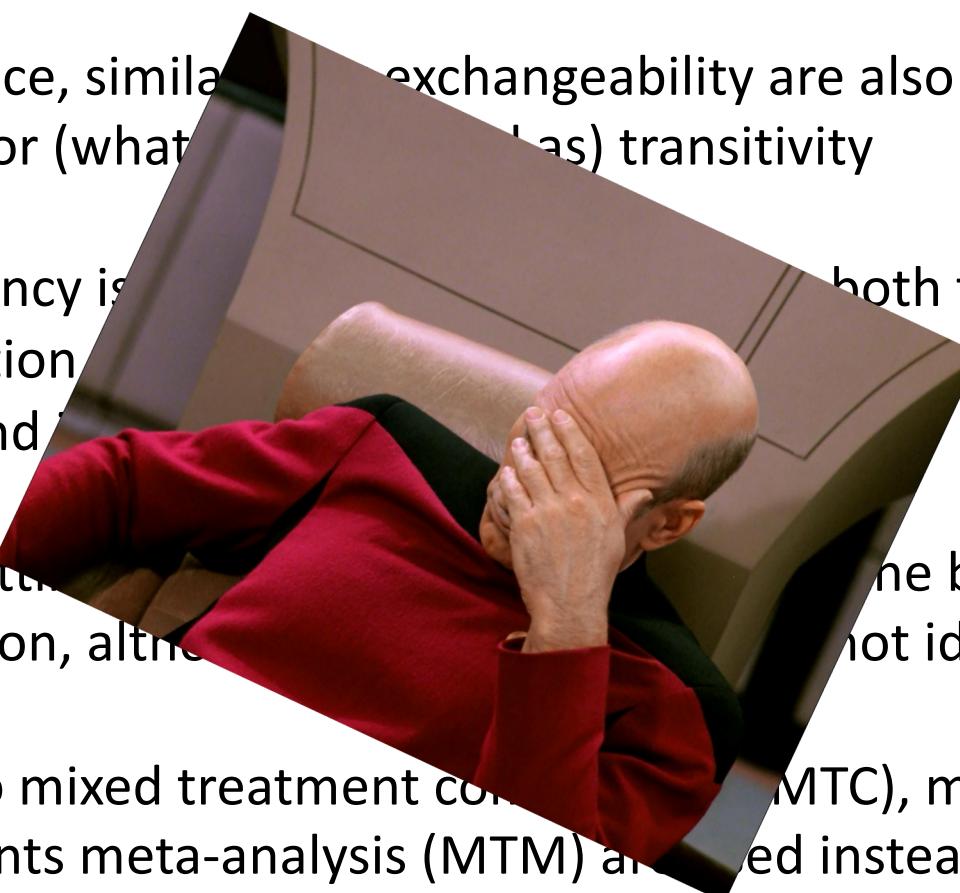
Intransitivity will always lead to inconsistency if there are closed loops

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, exchangeability are also used as a term for (what we call) transitivity
- Consistency is often used as an assumption that both the direct and indirect estimates are equal between studies
- Side splitting is a method to do the back-calculation, although the two methods are not identical
- And also mixed treatment comparison (MTC), multiple treatments meta-analysis (MTM) are used instead of network meta-analysis



References

- Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J Clin Epidemiol* 2010; **63**: 875-82.
- Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Research Synthesis Methods* 2012; **3**: 161–176.
- Cipriani A, Furukawa TA, Salanti G et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; **373**: 746–758.
- Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013; **159**: 130-7.
- 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
- Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomised evidence using mixed treatment comparison meta-analysis. *Journal of the Royal Statistical Society Series A* 2010; **173**: 613–629.
- Donegan S, Williamson P, D'Alessandro U, Tudor Smith C. Assessing key assumptions of network meta-analysis. *Research Synthesis Methods* 2013; **4**: 291-323.
- Golfinopoulos V, Salanti G, Pavlidis N, Ioannidis JPA. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: a meta-analysis. *Lancet Oncology* 2007; **8**: 898-911.

References

- Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies *Research Synthesis Methods* 2012; **3**: 98–110.
- Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Medicine* 2013; **11**: 159.
- Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; **101**: 447–459.
- Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statistics in Medicine* 2007; **26**: 1237-54.
- 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.
- Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of Clinical Epidemiology* 2009; **62**: 857-64
- Veroniki AA, Vasiliadis HS, Higgins JPT, Salanti G. Evaluation of inconsistency in networks of interventions. *International Journal of Epidemiology* 2013;
- White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Research Synthesis Methods* 2012; **3**: 111-125