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

Advertisement Feature

Next steps in antidepressant selection

What to consider when first line pharmacotherapy does not succeed

Treatment of depression frequently involves switching antidepressant – either because of lack of efficacy or adverse events – to find a drug that works well for the individual patient.[1]

As clinicians treating patients with depression are aware, the first pharmacological treatment selected may not achieve remission of symptoms, and a number of treatment steps may be needed.[2] However, successive trials of therapy can result in lower remission rates and higher relapse rates.[2] (Fig 1)

NICE guidelines (CG90) for drug treatment recommend initial use of a generic Selective Serotonin Re-uptake Inhibitor (SSRI) but if response is limited or absent, or if side effects occur, consider switching to an alternative antidepressant.[3] When switching antidepressants, NICE recommends considering, initially, a different SSRI or a better tolerated newer-generation antidepressant.[3] Use of the SSRI Cipralex (escitalopram) in the care pathway in such circumstances is consistent with national guidelines (NICE CG90, BAP).[3,4]

An independent meta-analysis of trials conducted in nearly 26,000 patients with major depression showed that Cipralex was one of two antidepressants judged to have achieved the best possible balance between efficacy and acceptability.[5] These data were subsequently reviewed by the evidence-based medicine journal, Bandolier, which concluded that Cipralex was among the top two antidepressant 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 of 8 weeks' treatment.[1] (Fig 2)

An independent Cochrane review found Cipralex to be superior to citalopram ($p \leq 0.02$) in achieving acute response and remission in major depression (after 6-12 weeks).[6]

In their health economic analysis, NICE found Cipralex to be on of the most cost-effective SSRIs (after sertraline) in both moderate and severe depression.[3] In a UK primary care record database study, usage of Cipralex in patients with severe depression was associated with fewer hospitalisations (all causes) compared with generic SSRIs and venlafaxine.[7] The overall cost of treatment with Cipralex was no higher than with generic SSRIs, and was significantly lower ($p<0.0001$) than with venlafaxine in patients with severe depression.[7]

Using Cipralex in patients who have not responded to initial therapy makes clinical and financial sense.

More information on depression and Cipralex can be found at:
www.challengingdepression.co.uk

Lundbeck

Cipralex® escitalopram

"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)”

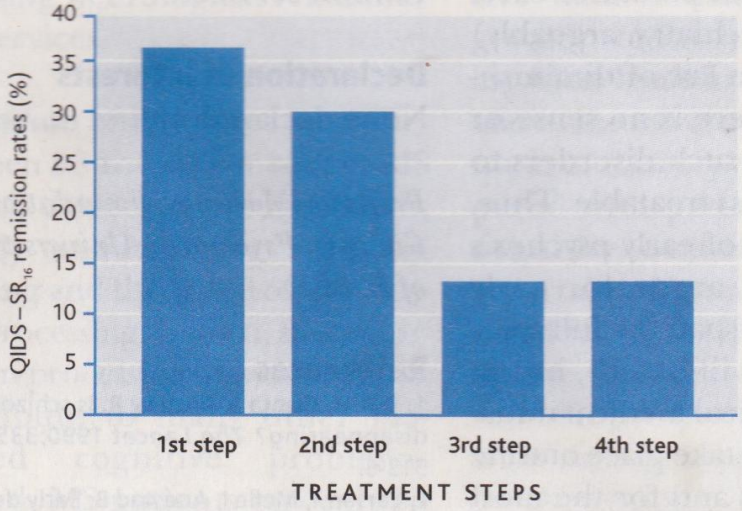


Figure 1. Acute remission rates by treatment step. Adapted from STAR*D, Rush et al.

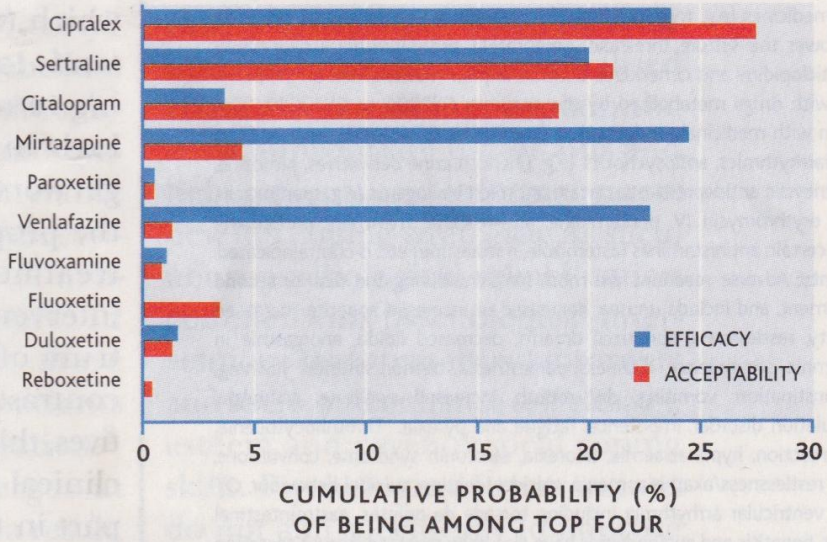


Figure 2. Probability of being among the top 4 drugs for both efficacy (at least 50% reduction in depression score, or considered much or very much improved on clinical global impression) and acceptability (all cause withdrawal) at mean of 8 weeks' treatment. Adapted from Bandolier, 2009[1]

References

1. Bandolier Review. Assessing relative efficacy of antidepressants. 2009. Available at www.medicine.ox.ac.uk/bandolier/booth/mental/cipriani.html. Accessed February 2012
2. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps. STAR*D report. Am J Psychiatry 2006;163:1905-1917
3. NICE. CG90 Depression in adults: full guidance. Available at: http://www.nice.org.uk/nicemedia/live/12329/45896/45896.pdf Accessed February 2012
4. Anderson IM,Ferrier IN, Baldwin RC et al. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association of Psychopharmacology guidelines. J Pscyhopharmacol 2008;22:343-396
5. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009;383:746-58
6. Cipriani A, et al. Escitalopram versus other antidepressive agents for depression (reviéw). Cochrane DB Syst Rev 2009;2: CD006532
7. Wade AG, Saragoussi D, Despiégal N, et al. Healthcare expenditure in severely depressed patients treated with escitalopram, generic SSRIs or venlafaxine in the UK. Curr Med Res Opin 2010;26:1161-1170

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, Atsuo Nakagawa, Ichiro M Omori, Hugh McGuire, Michele Tansella, Corrado Barbui

Lancet 2009; 373: 746–58

Published Online January 29, 2009 DOI:10.1016/S0140-6736(09)60046-5

See Comment page 700

Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Italy (A Cipriani PhD, C Barbui MD, Prof M Tansella M D); Department of Psychiatry and Cognitive-Behavioral Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan (Prof T A Furukawa MD, N Watanabe PhD, I M Omori PhD); Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Greece (G Salanti PhD); Department of Psychiatry, University of Oxford, UK (A Cipriani, Prof J R Geddes M D); MRC Biostatistics Unit Institute of Public Health, University of Cambridge, UK

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

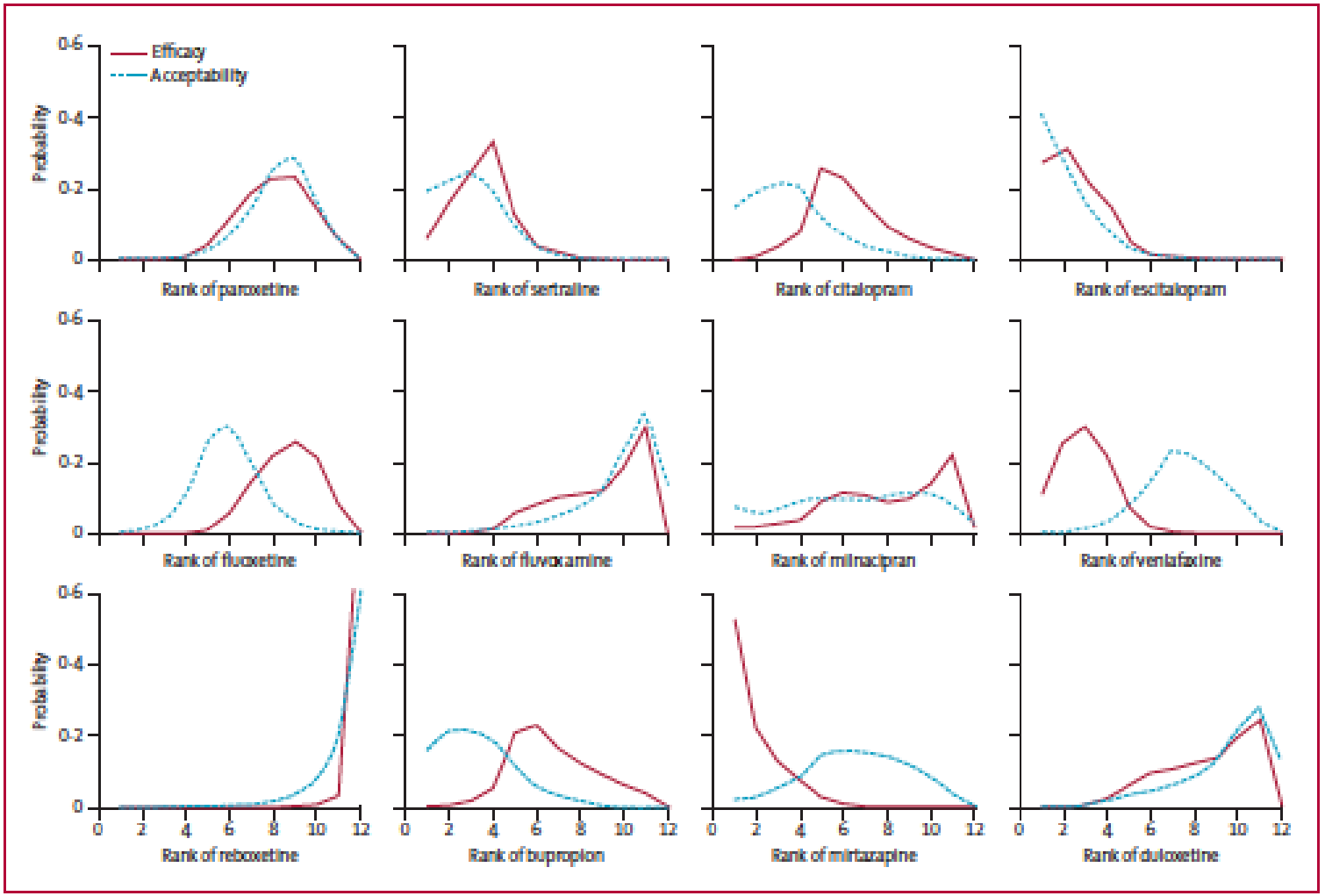
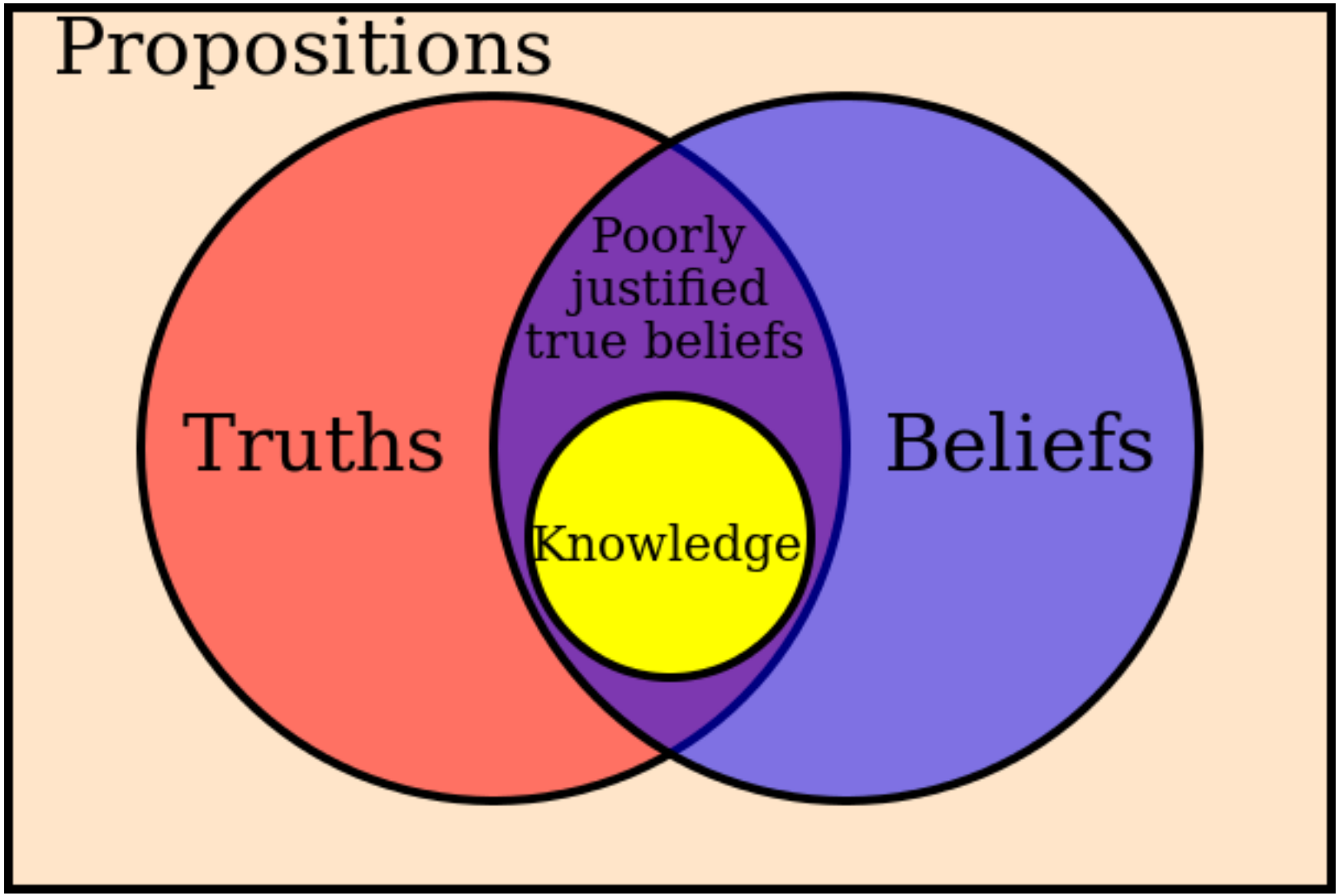


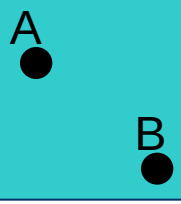
Figure 4: Ranking for efficacy (solid line) and acceptability (dotted line)
Ranking indicates the probability to be the best treatment, the second best, the third best, and so on, among the 12 antidepressants.

But do these claims represent knowledge?



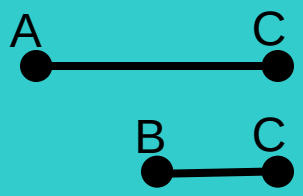
A taxonomy of comparisons

Direct Comparison (head to head RCT)



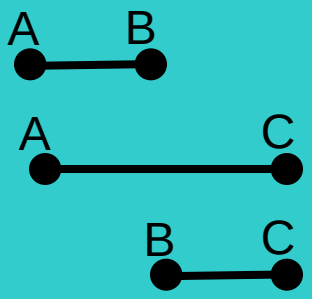
‘Naïve’ or ‘Unadjusted’ Indirect Comparison:

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)

An example of network meta-analysis

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.[1 2] 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,[3–5] even though indirect comparisons produce relatively imprecise estimates.[6] 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

WILL AND DEN MCINTYRE/PL

Angioplasty balloon device used to unblock and widen arteries

Medical Research Council Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Bristol BS8 2PR

Deborah M Caldwell
research associate

A E Ades
professor of public health science

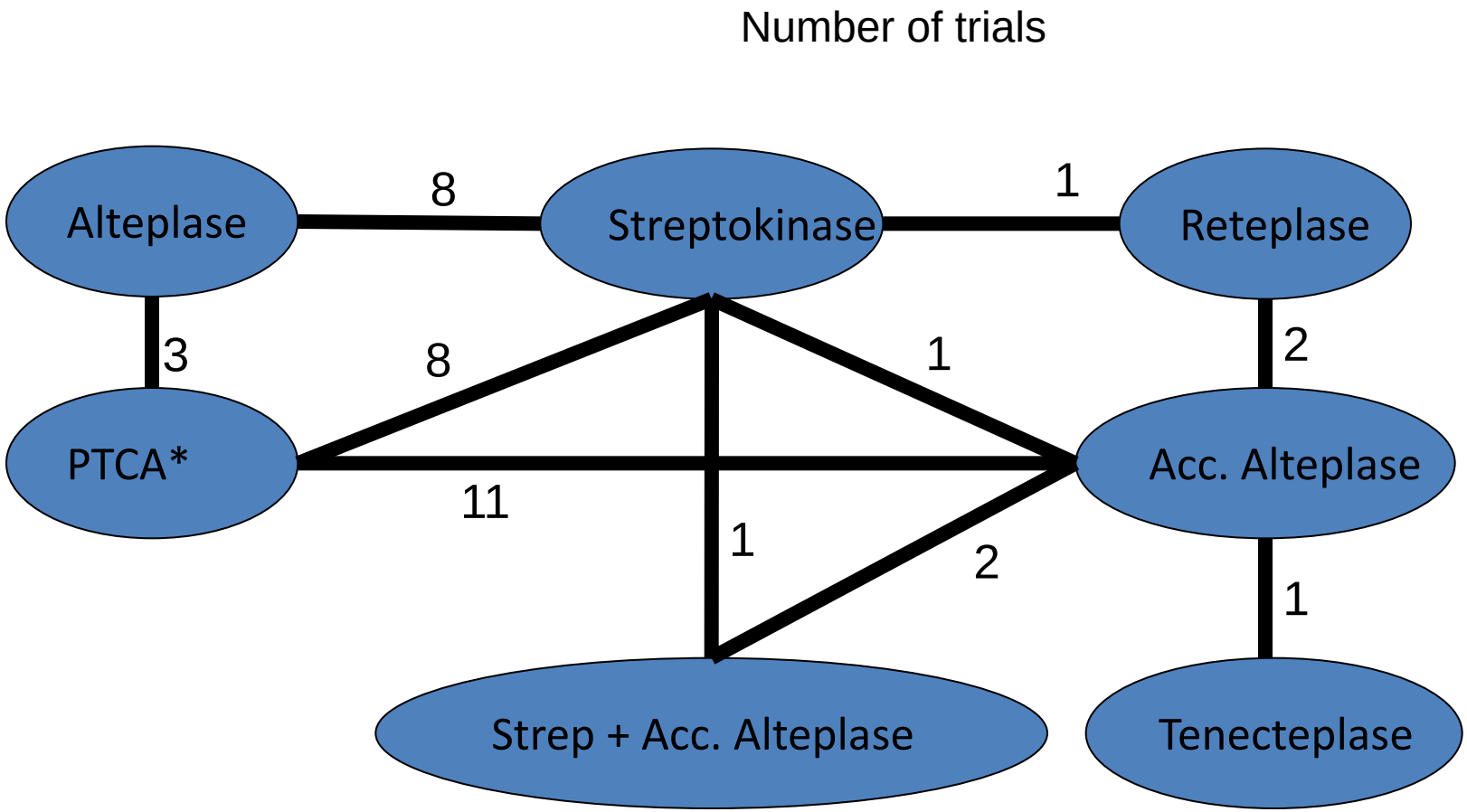
Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR

J P T Higgins
senior research associate

Correspondence to:
D M Caldwell
d.m.caldwell@bristol.ac.uk

BMJ 2005;331:897–900

The network of trial evidence for thrombolysis and angioplasty after myocardial infarction



*Percutaneous transluminal coronary angioplasty

The trial evidence summarised as a set of pairwise comparisons

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

Treatment:	Compared to: Streptokinase	Alteplase	Acc. Alteplase
Streptokinase			
Alteplase	0.89 (0.54 to 1.14)		
Acc. Alteplase	0.86 (0.78 to 0.94)		
Streptokinase+Alteplase	0.96 (0.87 to 1.05)		1.12 (1.00 to 1.25)
Reteplase	0.95 (0.79 to 1.12)		1.02 (0.90 to 1.16)
Tenecteplase	-		1.01 (0.88 to 1.14)
PTCA	0.49 (0.20 to 0.91)	0.63 (0.25 to 1.29)	0.79 (0.55 to 1.05)

Network meta-analyses provide estimates of treatment effects compared to a common reference

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

Treatment:	Mean (95% CrI)
Streptokinase	1.04 (0.91 to 1.35)
Alteplase	1
Acc. Alteplase	0.88 (0.70 to 1.19)
Streptokinase+Alteplase	1.02 (0.78 to 1.51)
Reteplase	0.92 (0.70 to 1.24)
Tenecteplase	0.90 (0.61 to 1.35)
PTCA	0.65 (0.49 to 0.86)

Direct Comparison: 0.63 (0.25 to 1.29)

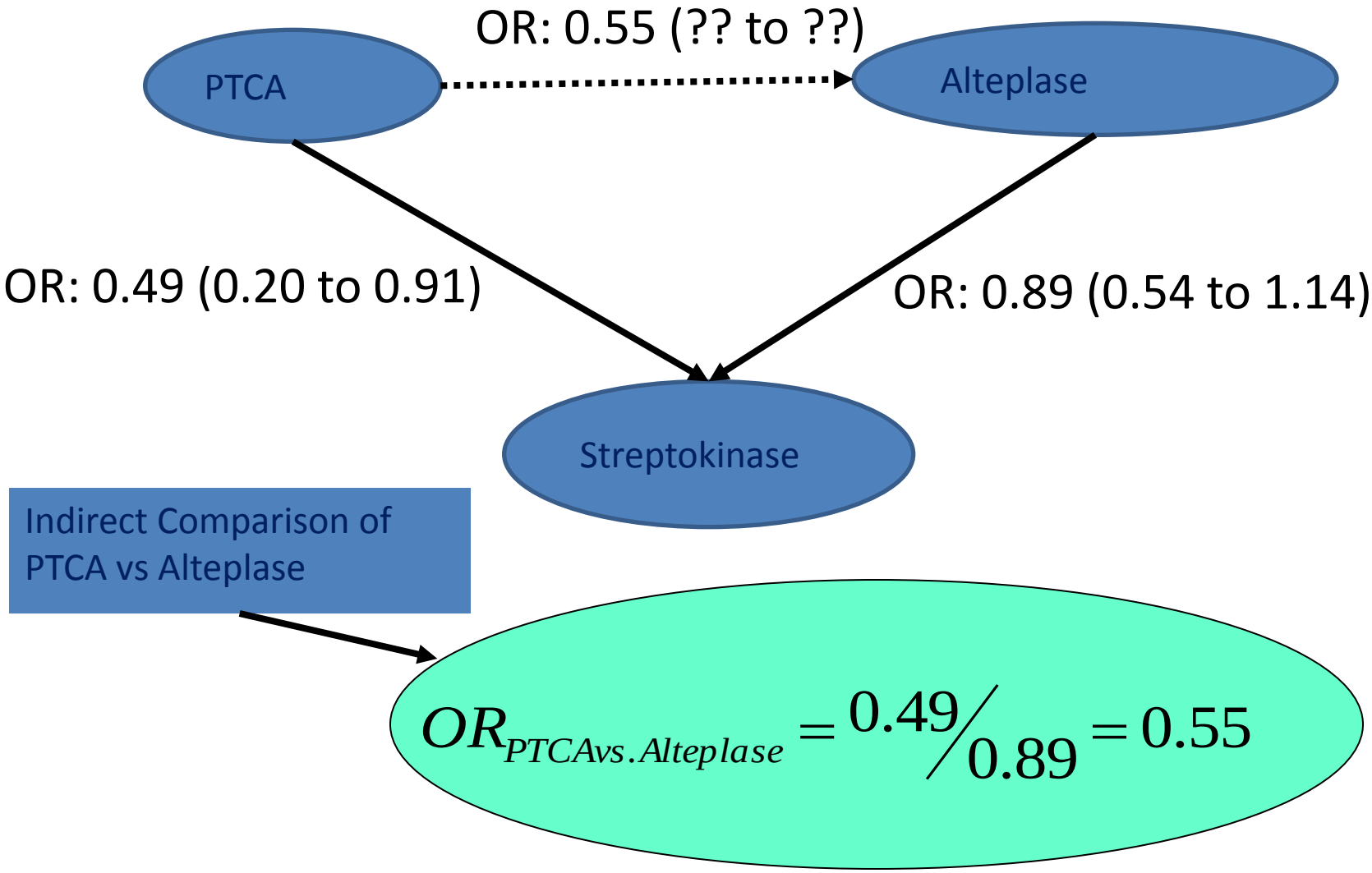
(Bayesian) network meta-analysis provide a readily interpretable summary of joint uncertainty

Table 3 Percentage mortality at 35 days and the probability that each treatment is best (lowest mortality) in multiple treatment comparison analysis*

	Fixed effect model		Random effects model	
	35 day Mortality %	Probability best	35 day Mortality %	Probability best
Streptokinase	6.7	0	6.8	0
Alteplase	6.7	0	6.5	0.003
Accelerated alteplase	5.8	0	5.8	0.001
Streptokinase + alteplase	6.5	0	6.6	0.002
Reteplase	6.1	0	6.0	0.01
Tenecteplase	5.8	0.004	5.8	0.03
Percutaneous transluminal coronary angioplasty	4.4	0.995	4.3	0.95

*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

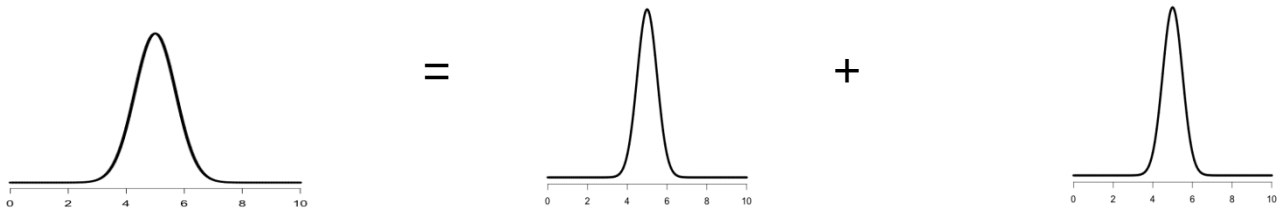


Estimation of uncertainty in an adjusted indirect comparison

Independent Random Variables

$$OR_{AB} = {OR_{AC}}/{OR_{BC}}$$

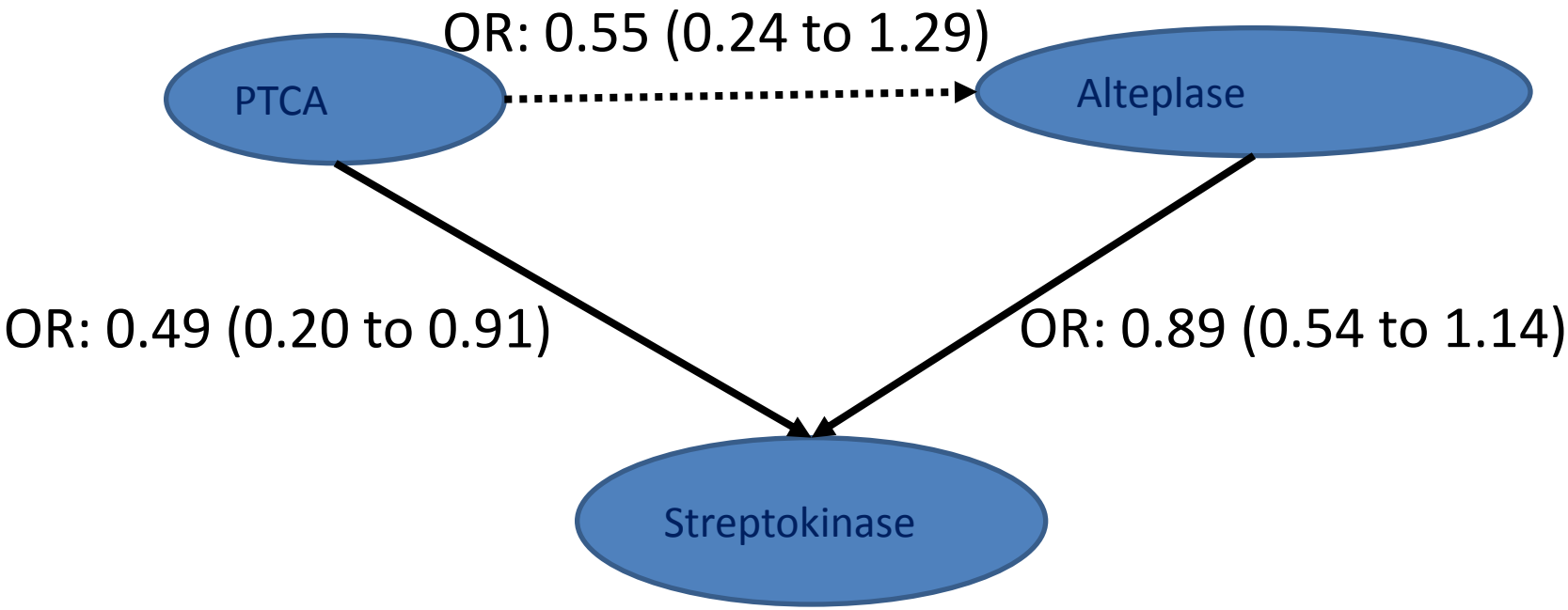
$$\log OR_{AB} = \log OR_{AC} - \log OR_{BC}$$



$$\operatorname{var}\left(\log OR_{AB}\right) = \operatorname{var}\left(\log OR_{AC}\right) + \operatorname{var}\left(\log OR_{BC}\right)$$

$$sd(\log OR_{AB}) = \sqrt{sd(\log OR_{AC})^2 + sd(\log OR_{BC})^2}$$

The final result for the PTCA vs Altelplase AIC



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. $d_{Steprokinase}$, $d_{Reteplease,}$ d_{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. $LOR_{\text{Streptokinase vs. Alteplase}} = d_{Streptokinase}$
  3. $LOR_{\text{Reteplase vs.Alteplase}} = d_{Reteplase}$
  4. $LOR_{\text{Streptokinase vs. PTCA}} = d_{Streptokinase} - d_{PTCA}$
  5. …

The basic assumption underlying network meta-analysis is that:

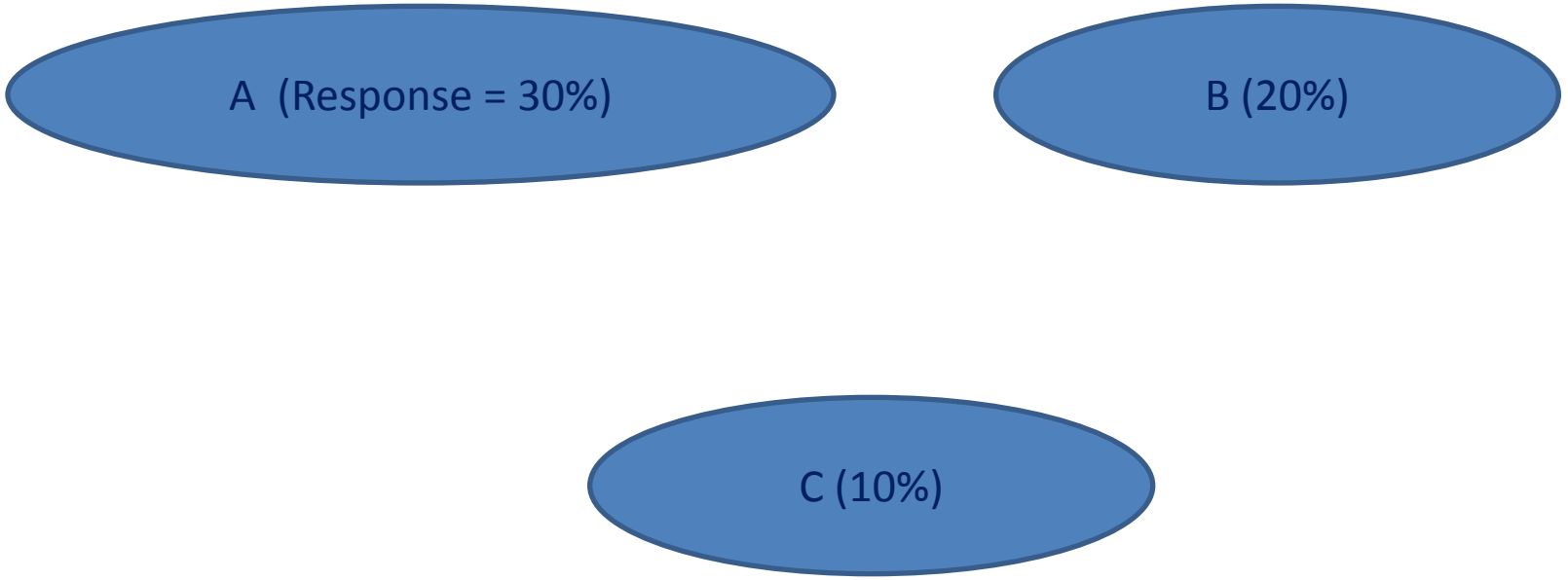
$$\partial_{AB} = \partial_{AC} - \partial_{BC}$$

Referred to as:

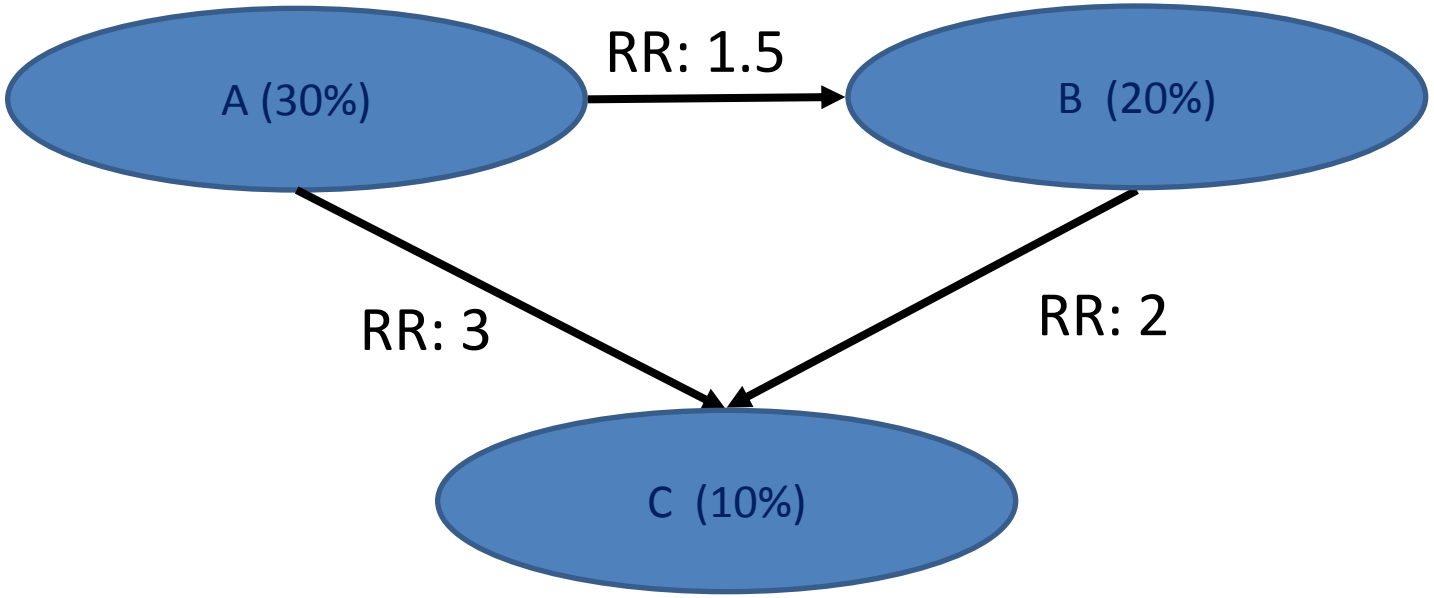
- Consistency
  - Indirect and direct estimates are consistent
- Exchangeability
  - If treatments were exchanged between trials estimates would be the same (allowing for random variation)
- Similarity
  - The trials are similar and comparable
- Transitivity

$$\partial_{AB} = \partial_{AC} - \partial_{BC} \qquad \partial_{AC} = \partial_{AB} - \partial_{CB}$$

Consider a single trial



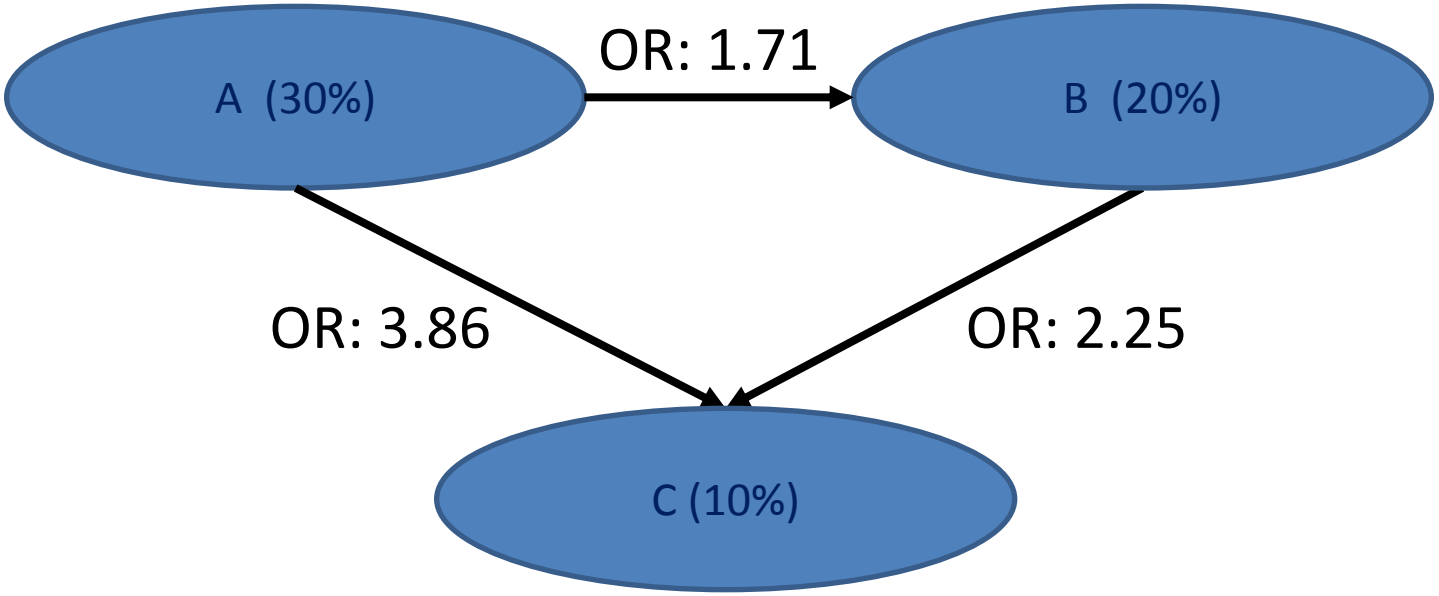
By definition consistent on the relative risk scale



$$RR_{AvsB} = {RR_{AvsC}}/{RR_{BvsC}}$$

$$RR_{AvsB} = {3}/{2} = 1.5$$

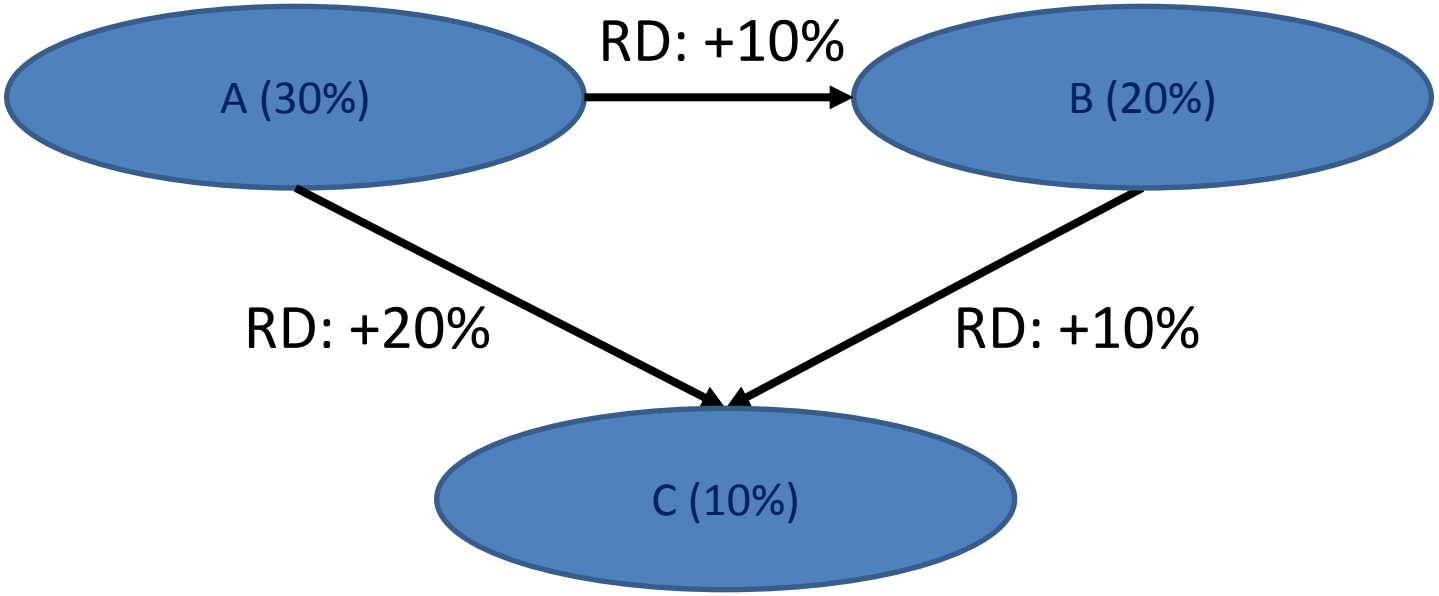
And also on the odds ratio scale...



$$OR_{AvsB} = {OR_{AvsC}}/{OR_{BvsC}}$$

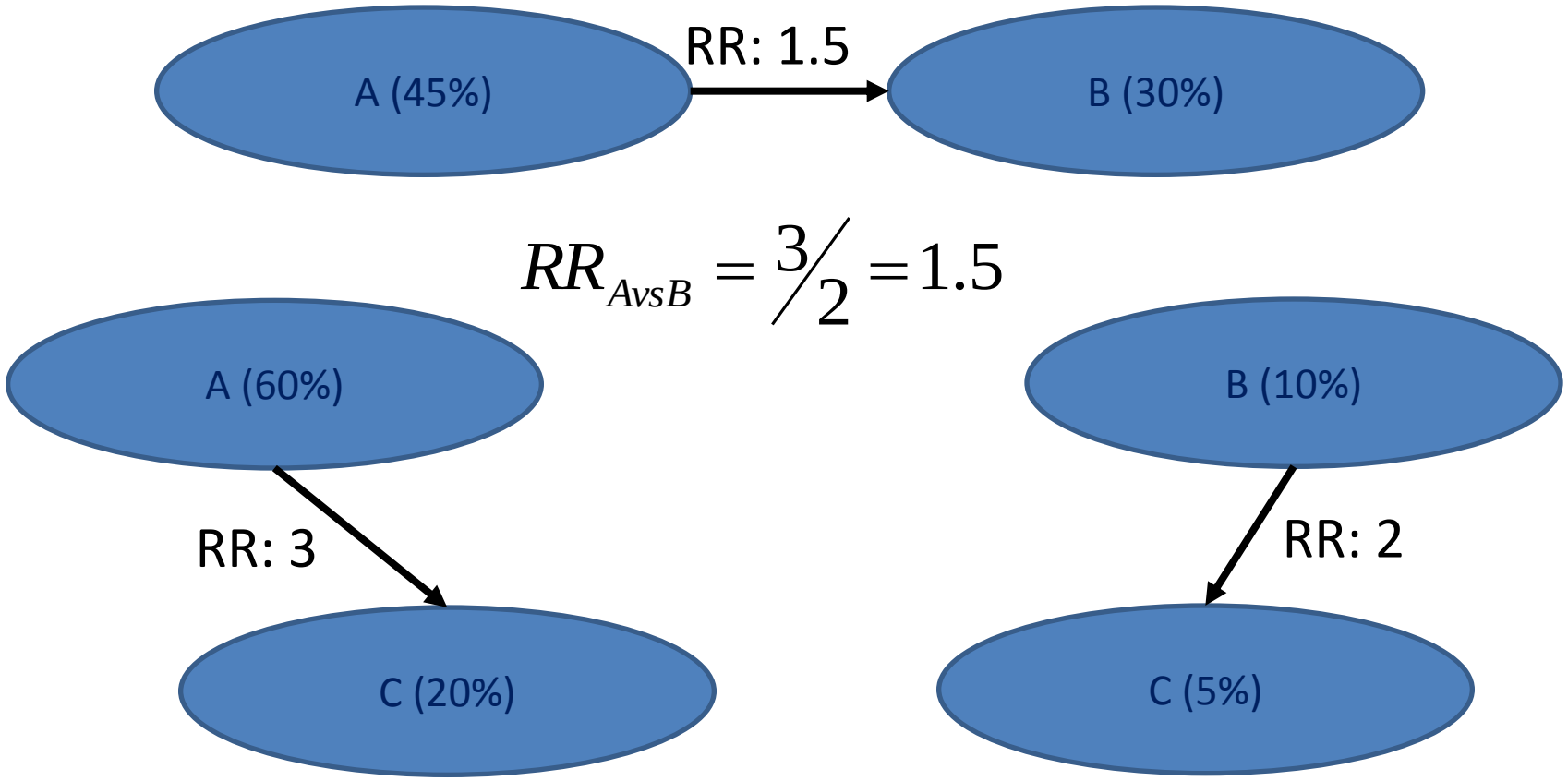
$$OR_{AvsB} = {3.86}/{2.25} = 1.71$$

And on the risk difference (RD) scale...

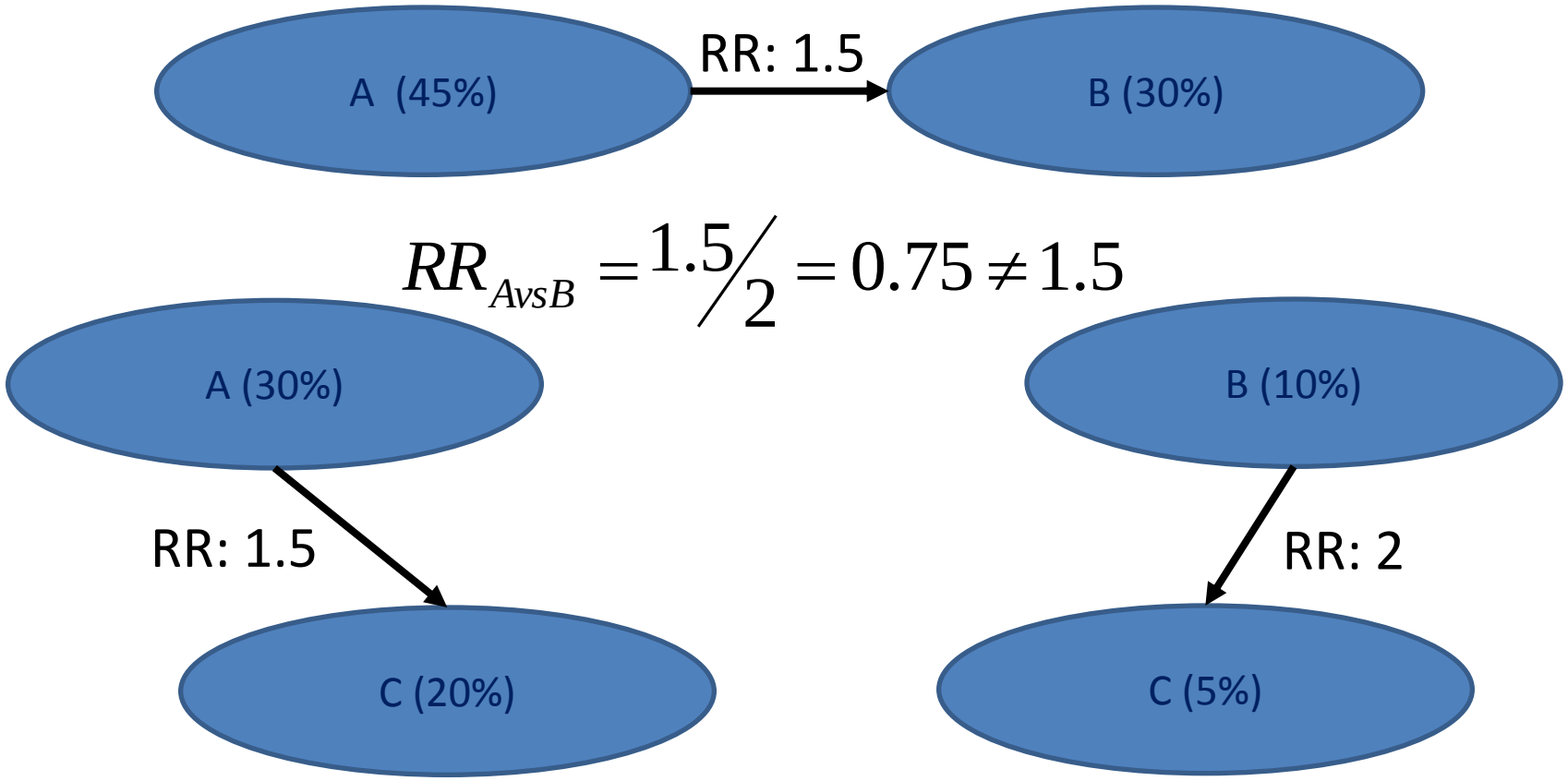


$$RD_{AvsB} = RD_{AvsC} - RD_{BvsC} \qquad RD_{AvsB} = 20\% - 10\% = +10\%$$

Whereas multiple trials may be consistent

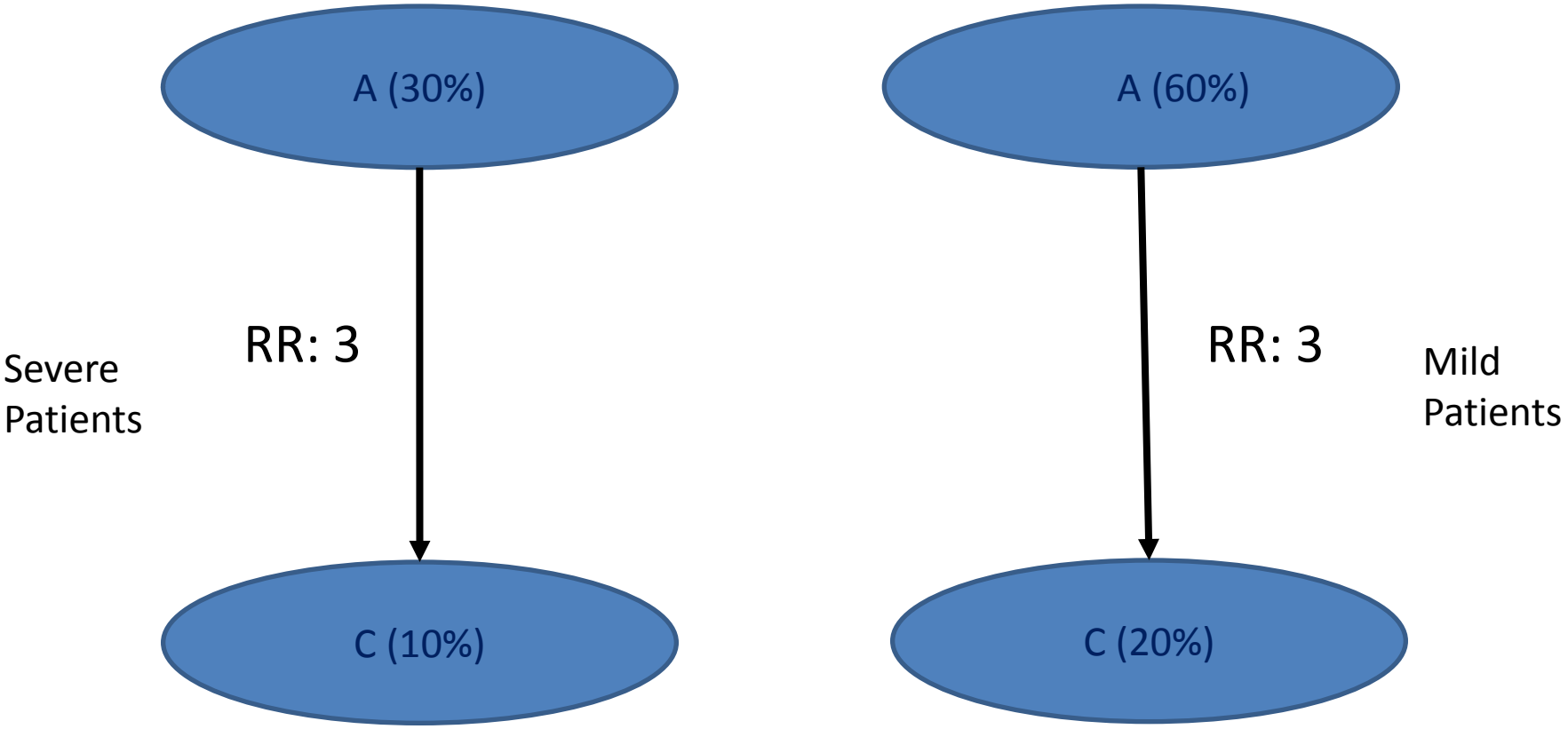


Or may be inconsistent…



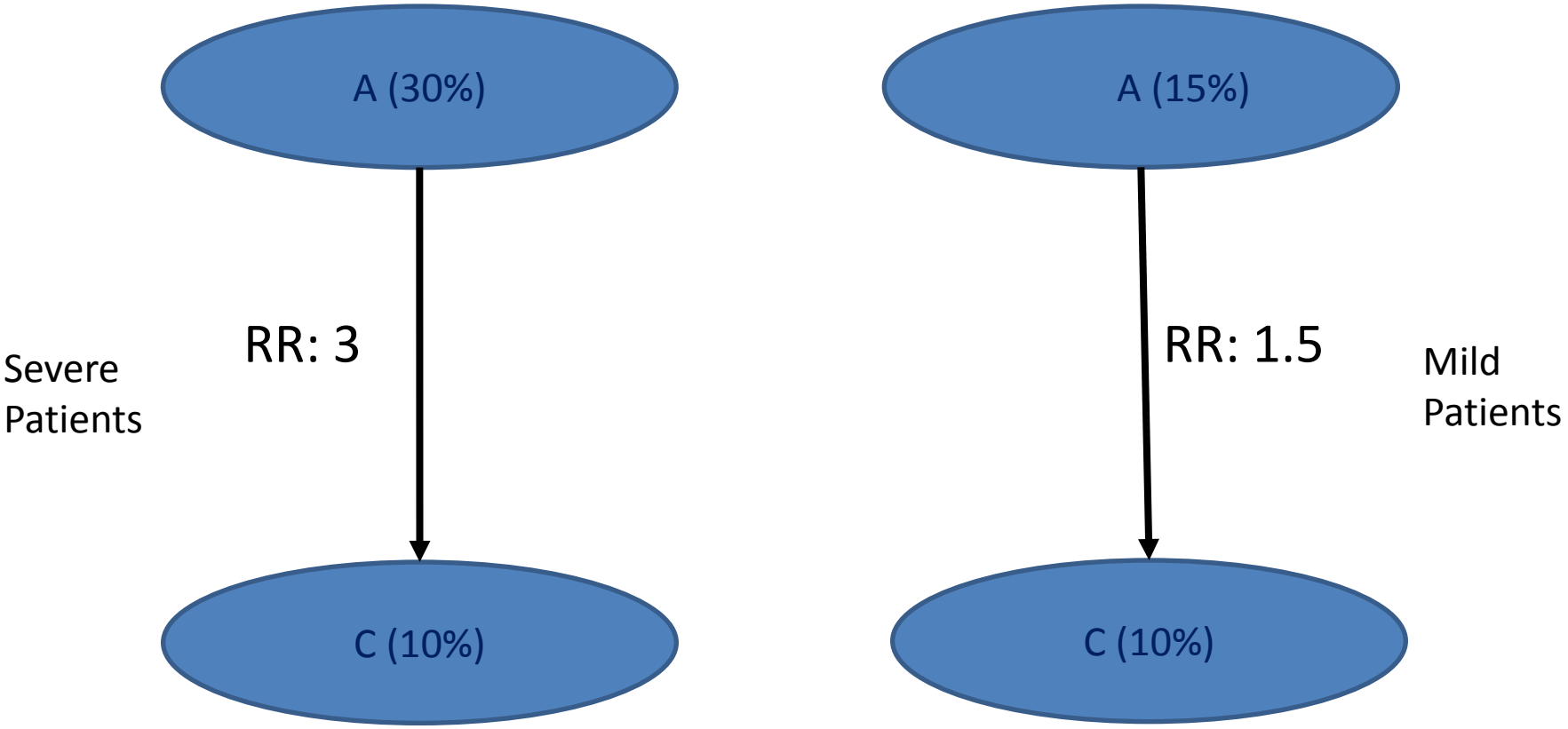
What determines whether networks of multiple trials will be consistent?

Prognostic factors alter response in individual treatment arms



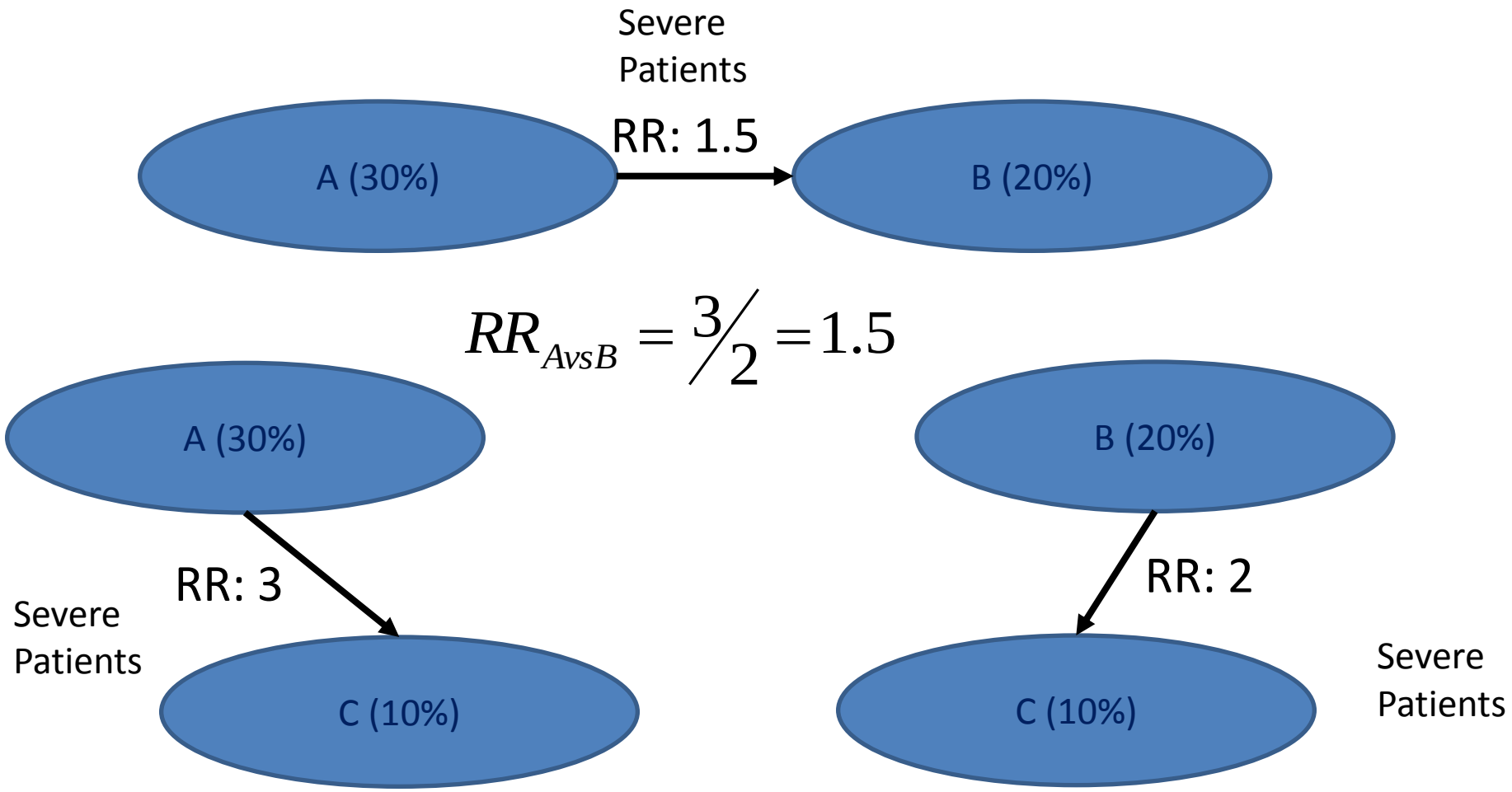
But do not alter the relative treatment effect (on a given scale)

Predictive factors alter response in individual treatment arms



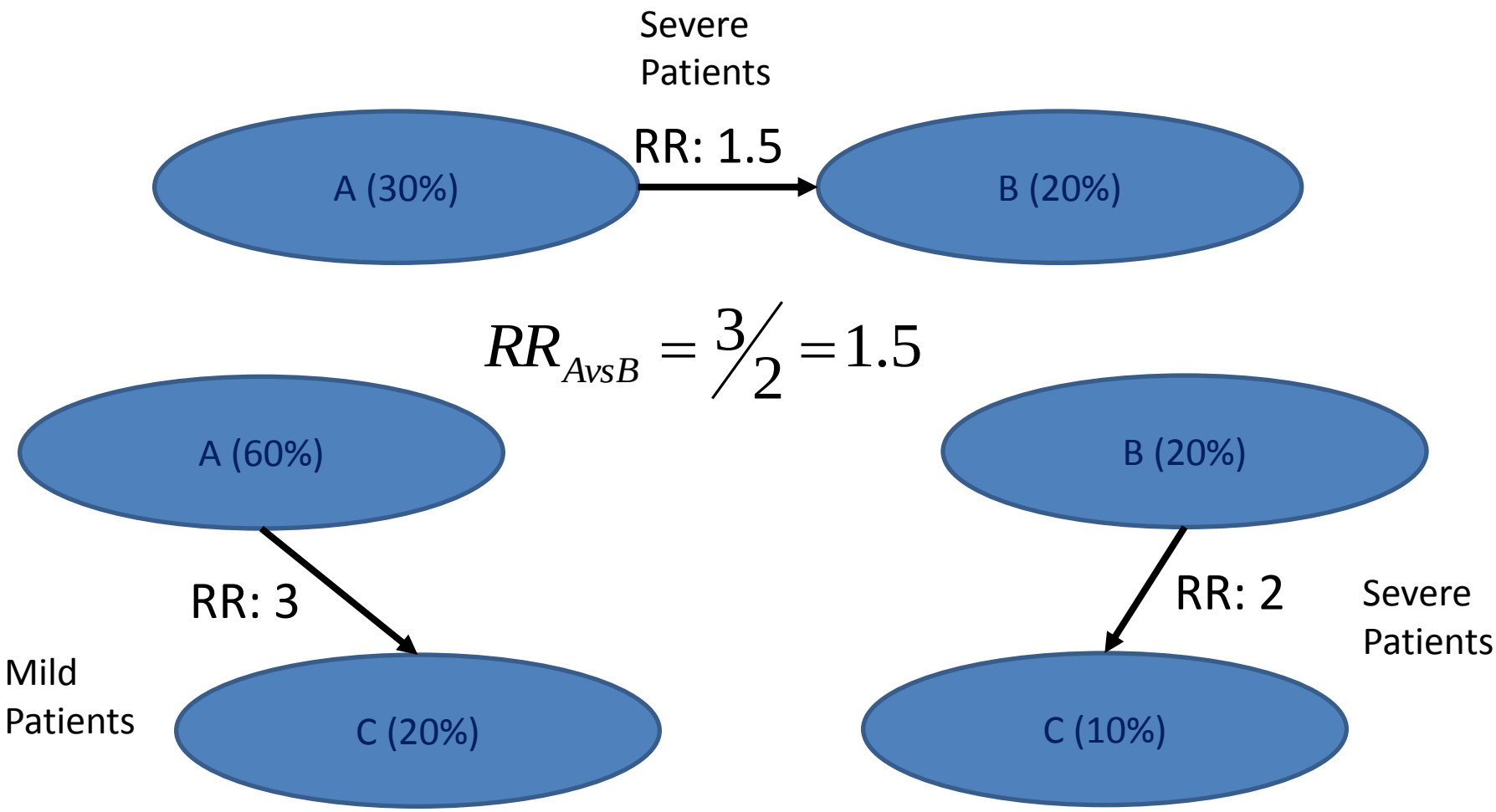
And alter the relative treatment effect (on a given scale)

A completely homogeneous set of trials…



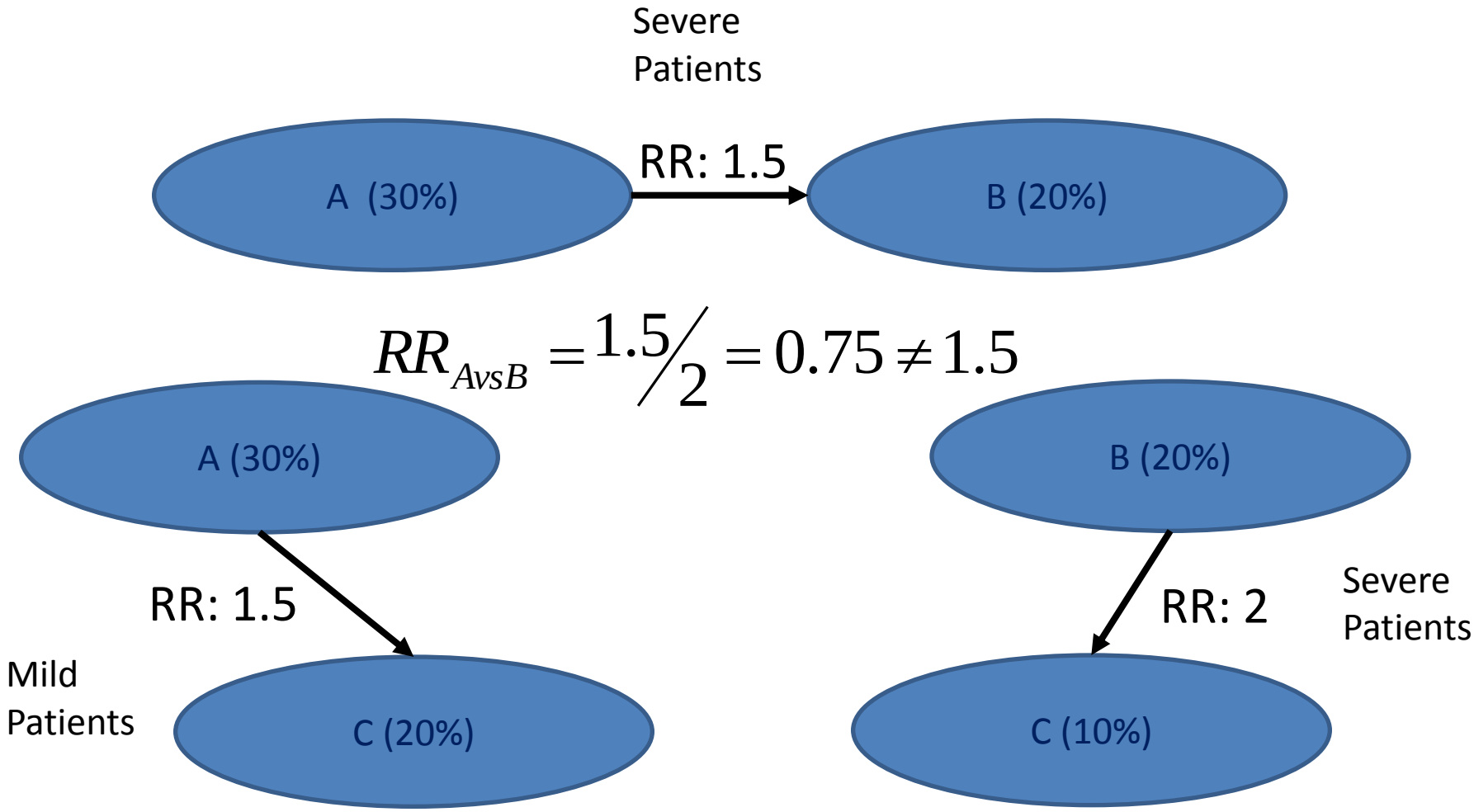
Will behave like a single multi-arm trial and be consistent

A heterogeneous set of trials



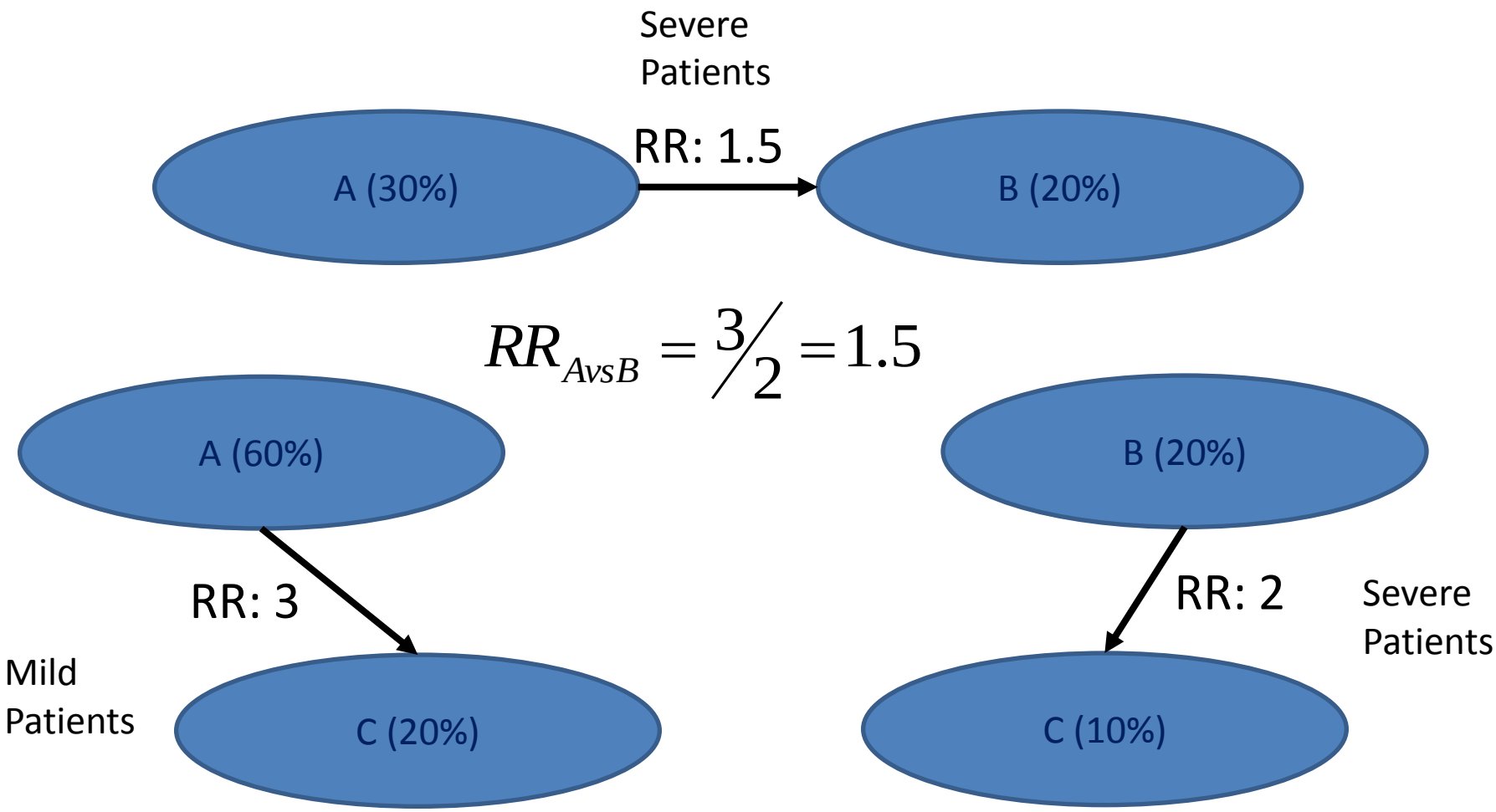
Will still be consistent if they differ in terms of prognostic factors

However, a heterogeneous set of trials



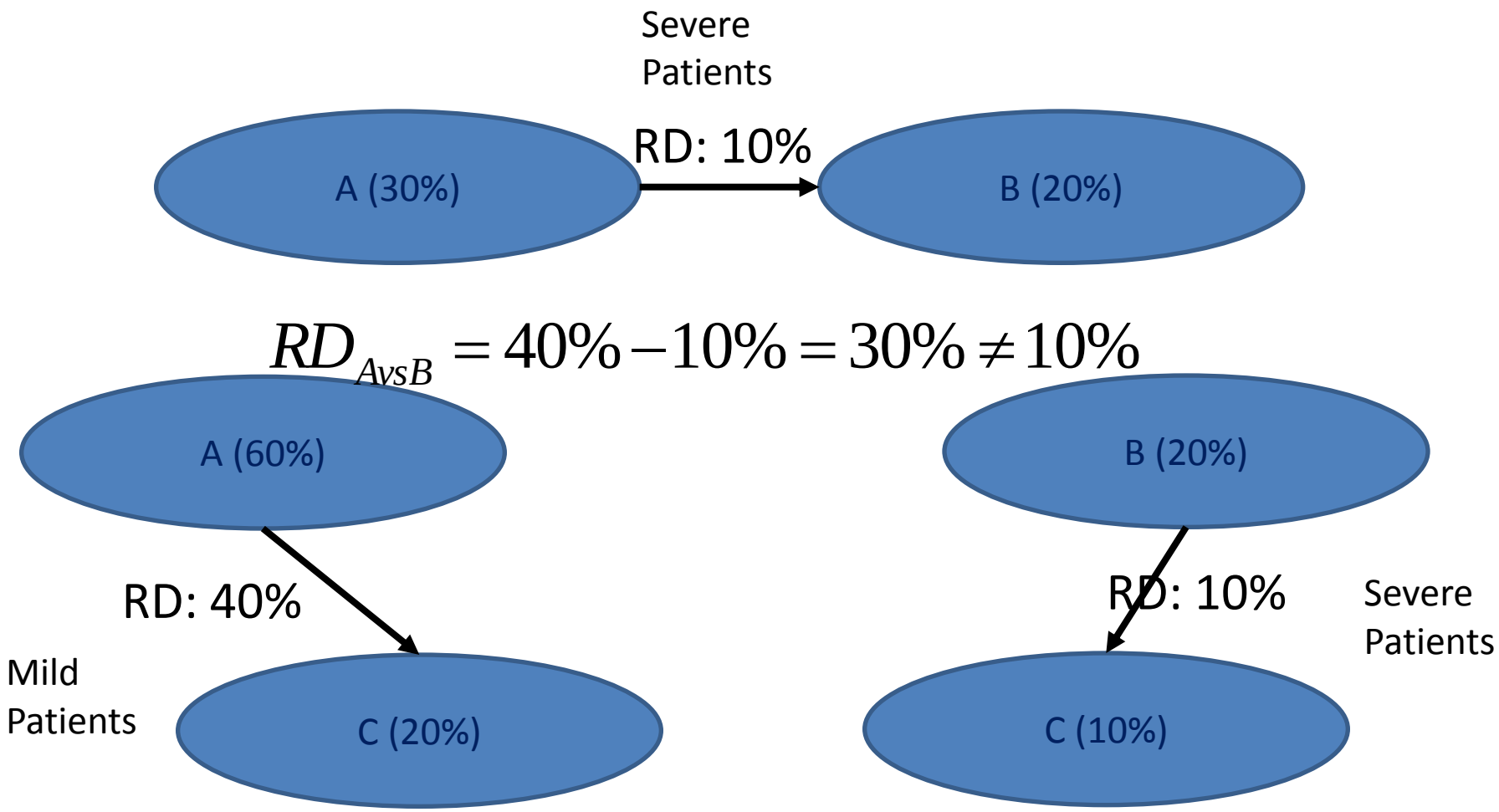
Will be inconsistent if they differ in terms of predictive factors

A heterogeneous set of trials



May be consistent on one treatment effect scale

A heterogeneous set of trials



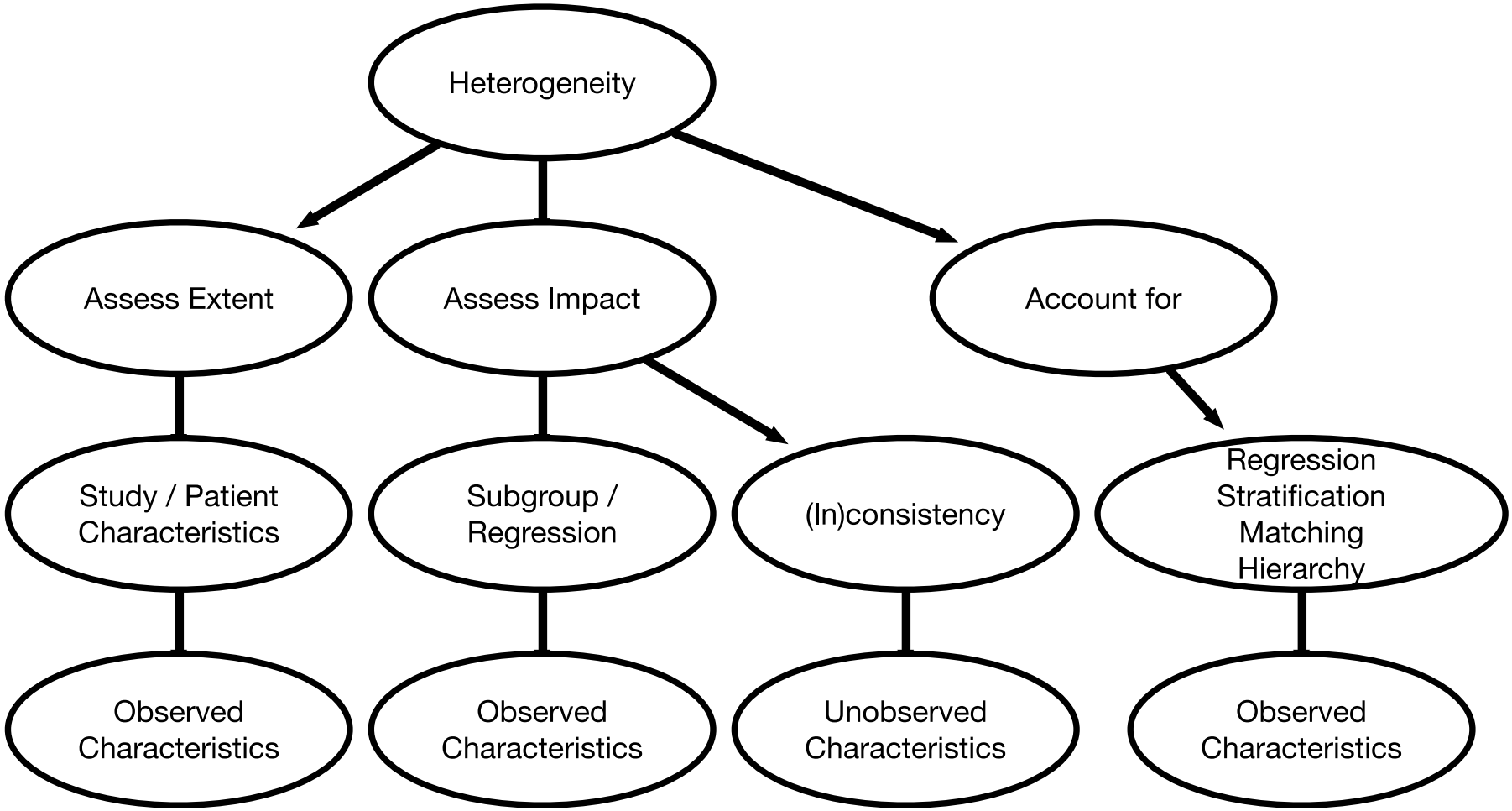
But be inconsistent on a different treatment effect scale

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

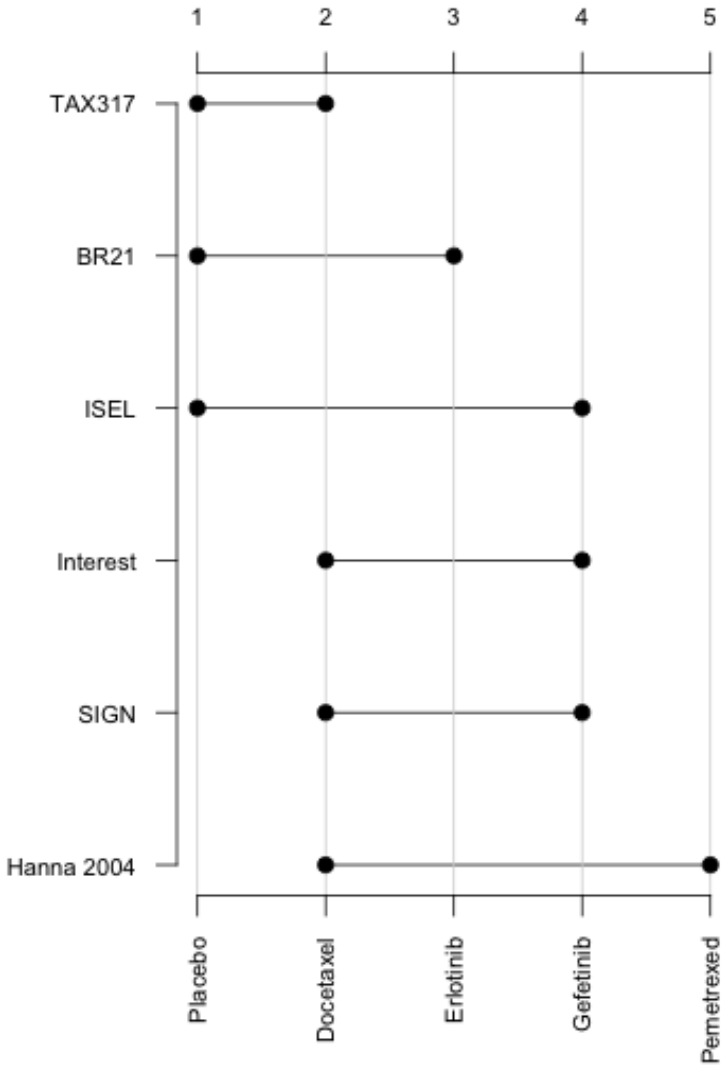


An example network meta-analysis: treatments for advanced NSCLC

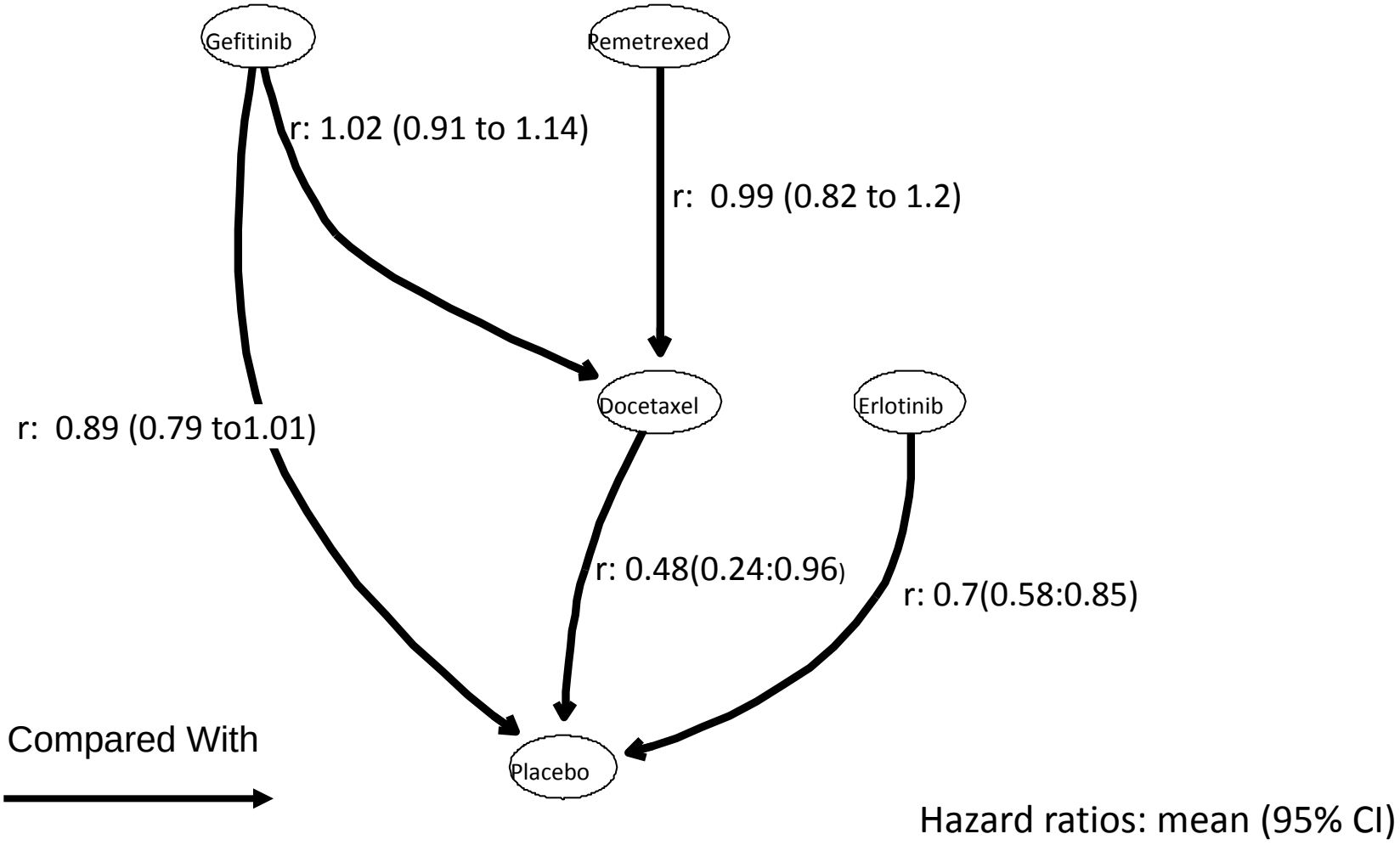
- Comparators
  - Placebo, Docetaxel, Erlotinib, Gefitinib, Pemetrexed
- Continuous Endpoint
  - Hazard Ratio: Overall Survival
- 4,672 patients in 6 studies
- NMA conducted on multiplicative hazard ratio scale: HRAB=HRAC / HRBC

Value Health. 2009 Sep;12(6):996-1003.

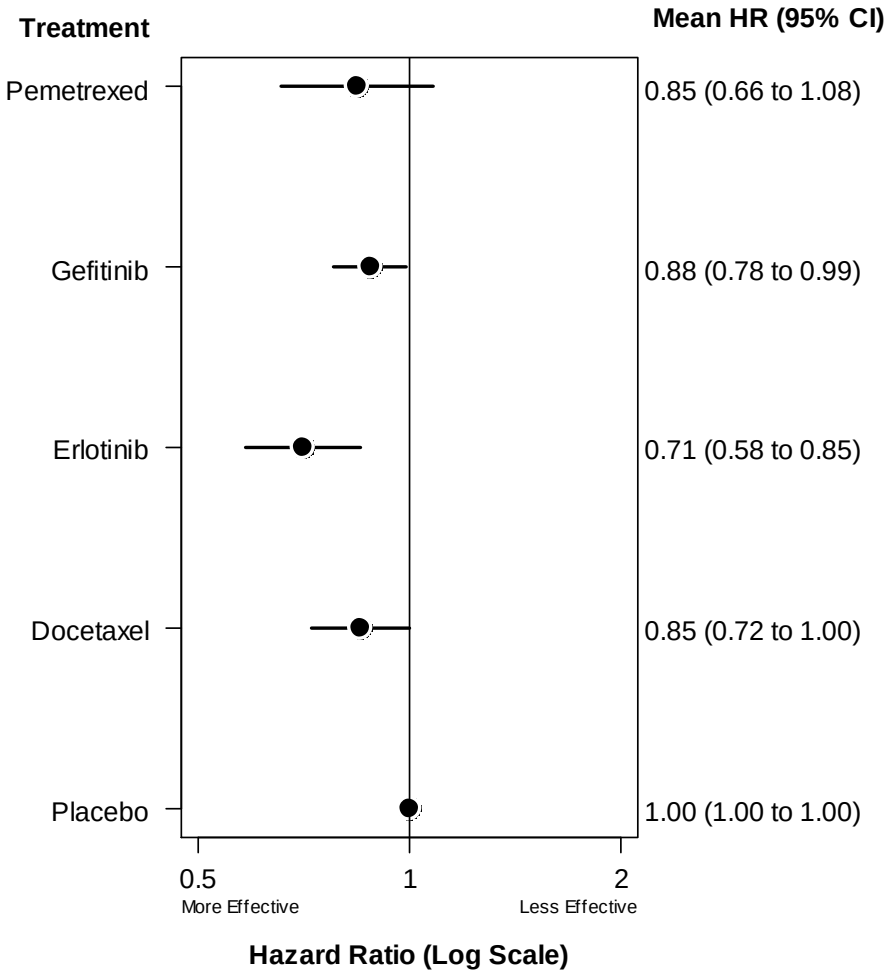
The trials form a connected network



Which includes “loops”



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 I Study characteristics of included trials

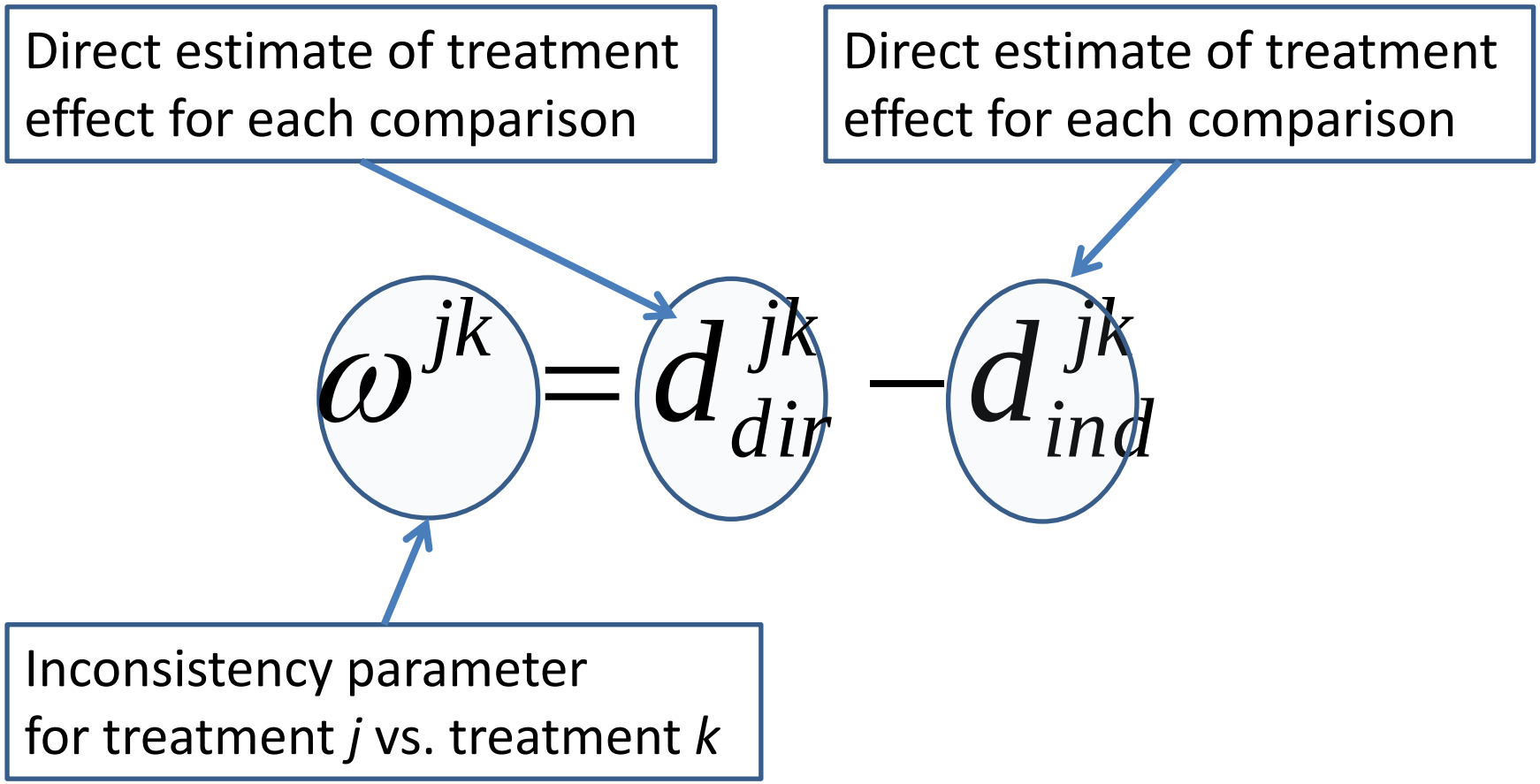
Author and date	Trial name	Trial design	Jadad score	Treatment	Number randomized	Mean treatment duration (months)
Shepherd 2005	BR21	Double-blind	3	Erlotinib	488	Not stated
				Placebo	243	—
Hanna 2004	JMEI	Open-label	2	Pemetrexed	283	4
				Docetaxel	288	4
Shepherd 2000	TAX 317	Not stated	2	Docetaxel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

Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. Journal of Clinical Epidemiology. Elsevier Inc; 2009Aug.1;62(8):857–64

(Some) methods for assessing inconsistency

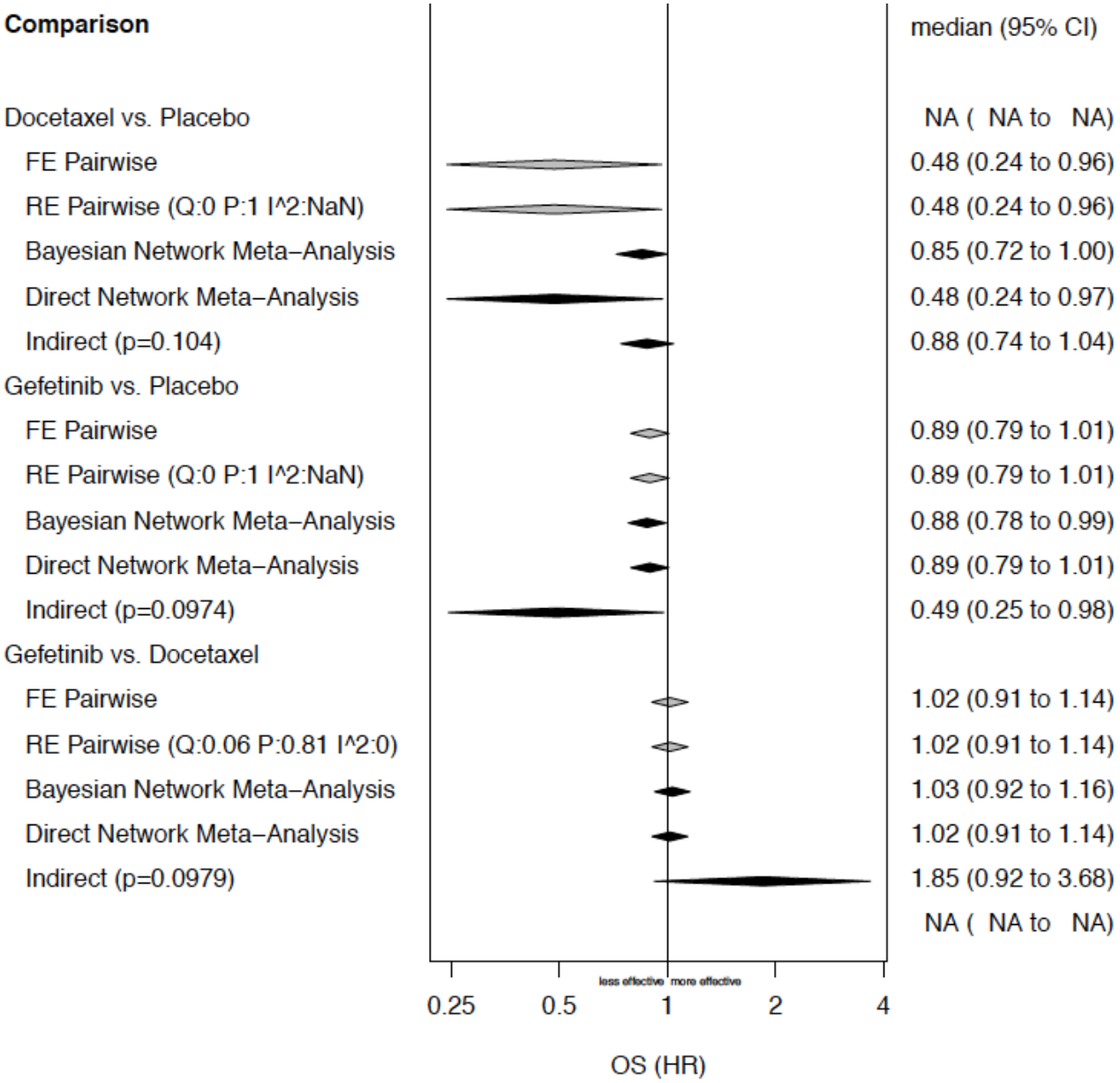
- Node splitting
- Comparison with an inconsistency model
- Treatment by design model

Node Splitting



- The inconsistency parameter represents the discrepancy between direct and indirect estimates.
- Can be tested against null: inconsistency = 0

The results of the node splitting analysis for the NSCLC NMA



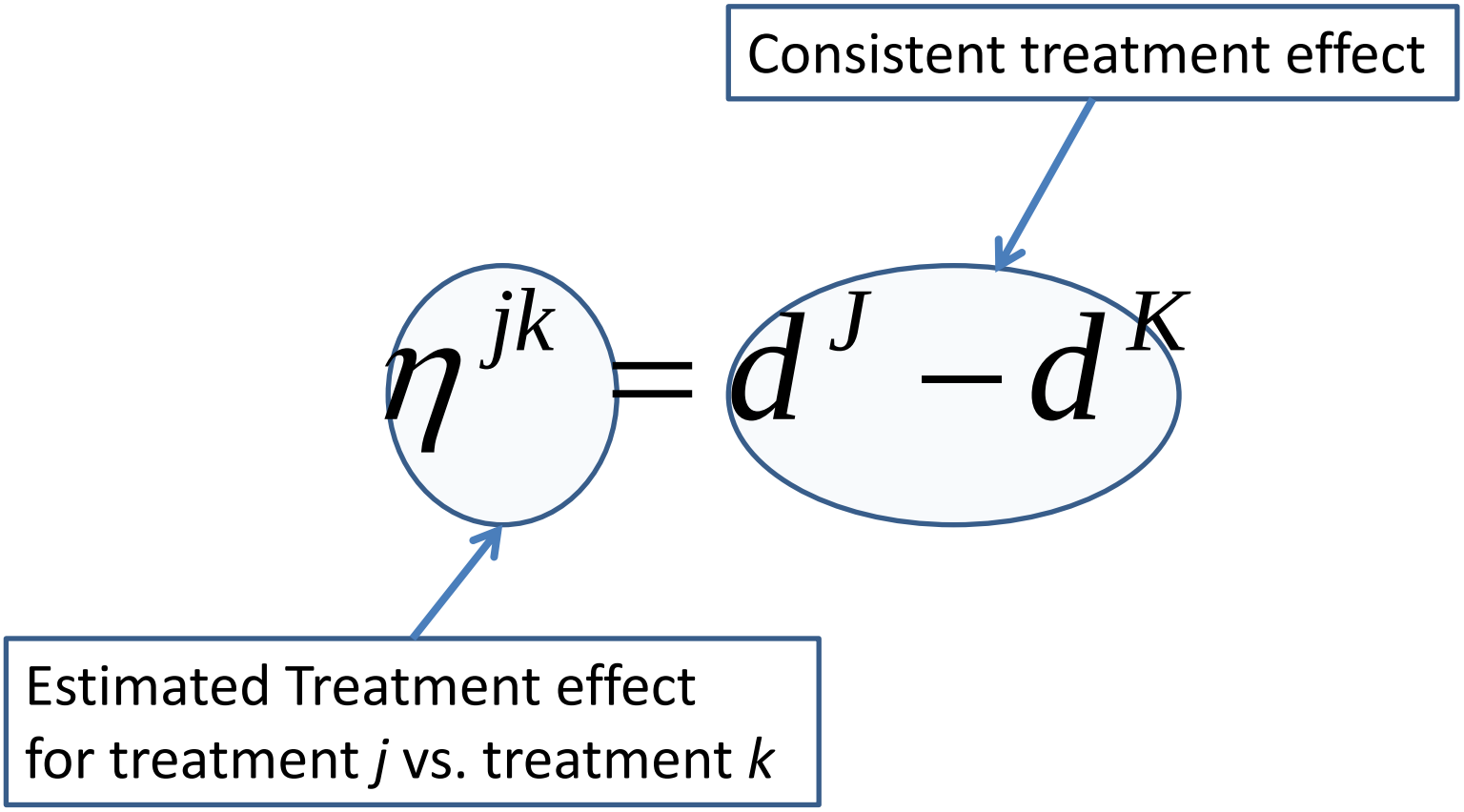
“Inconsistency Model”

Evidence Synthesis for Decision Making 4: Inconsistency in Networks of Evidence Based on Randomized Controlled Trials

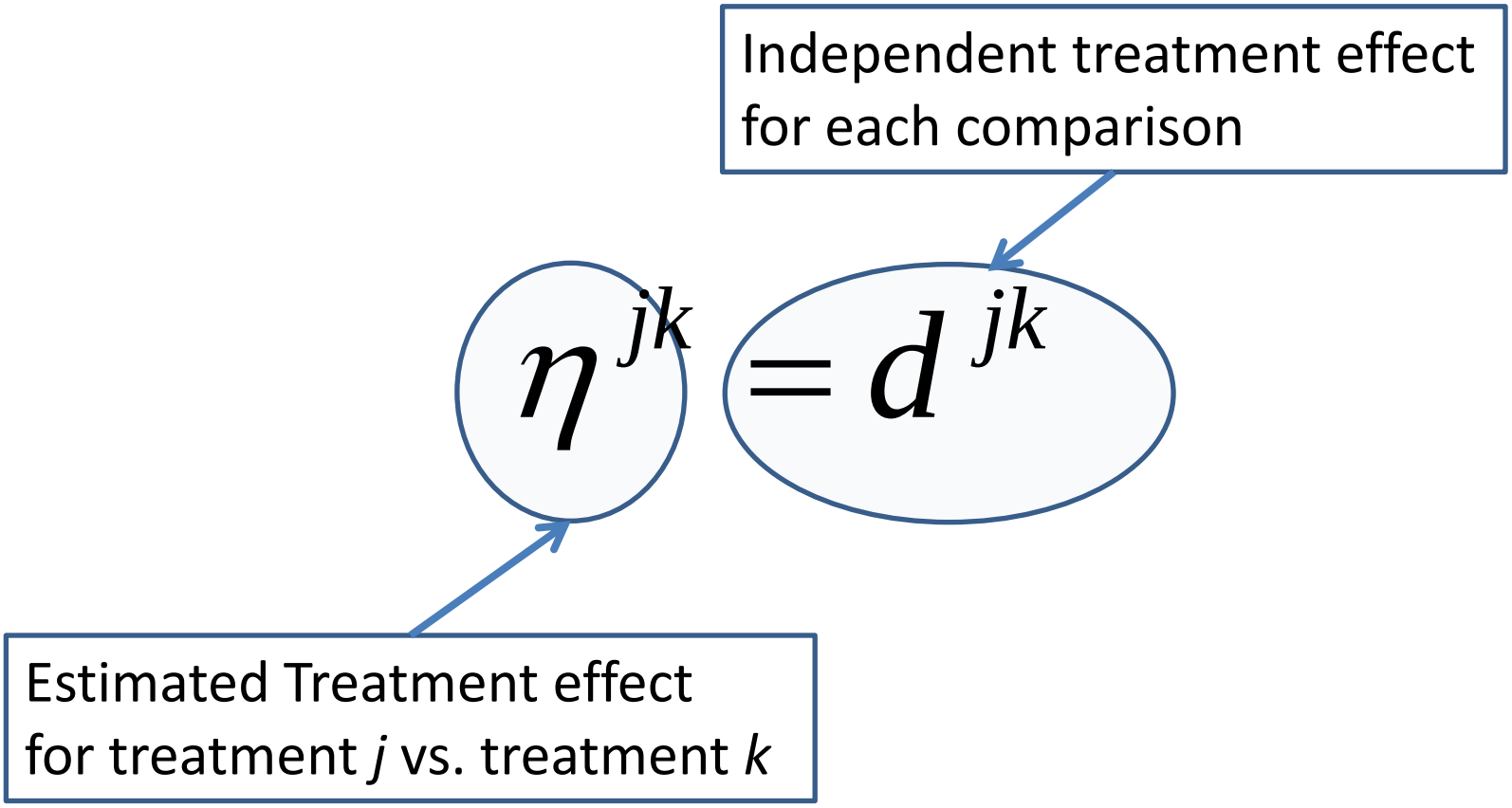
Sofia Dias, PhD, Nicky J. Welton, PhD, Alex J. Sutton, PhD, Deborah M. Caldwell, PhD, Guobing Lu, MSc, A. E. Ades, PhD

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)**

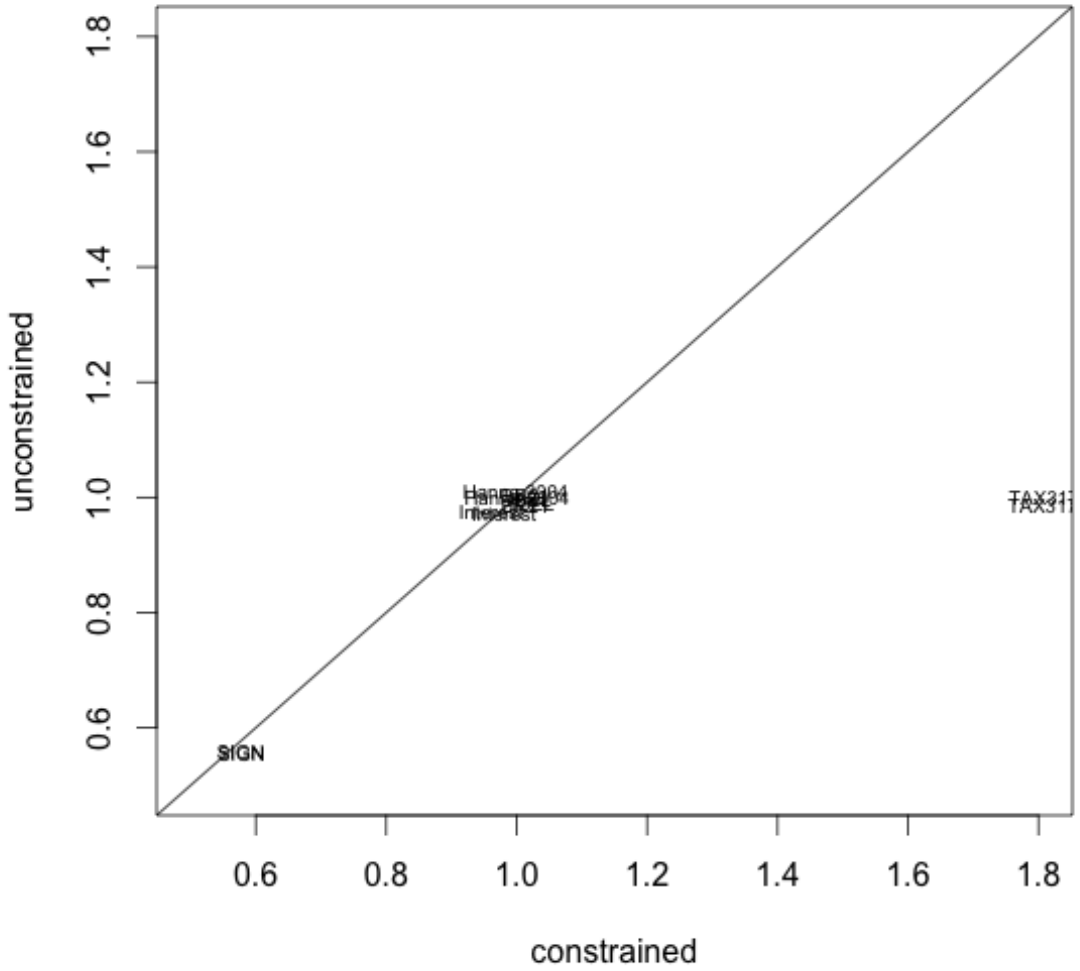
Consistency Model



Inconsistency Model

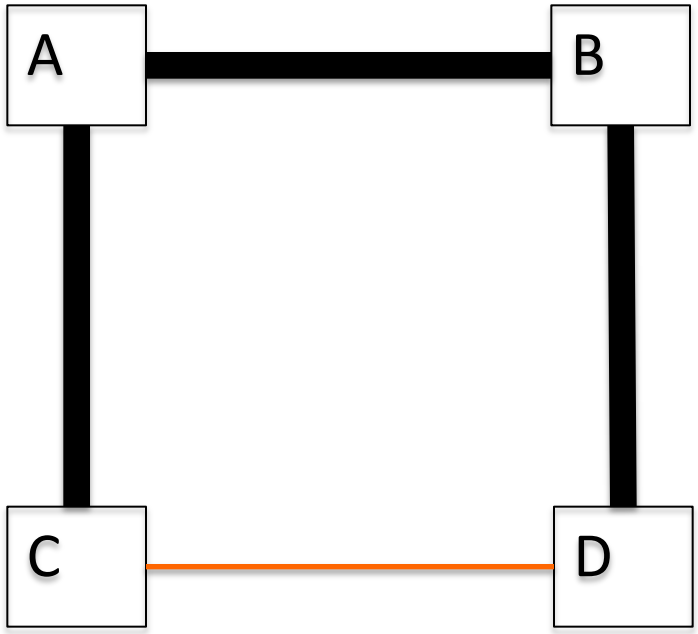


Comparison of posterior residual deviance

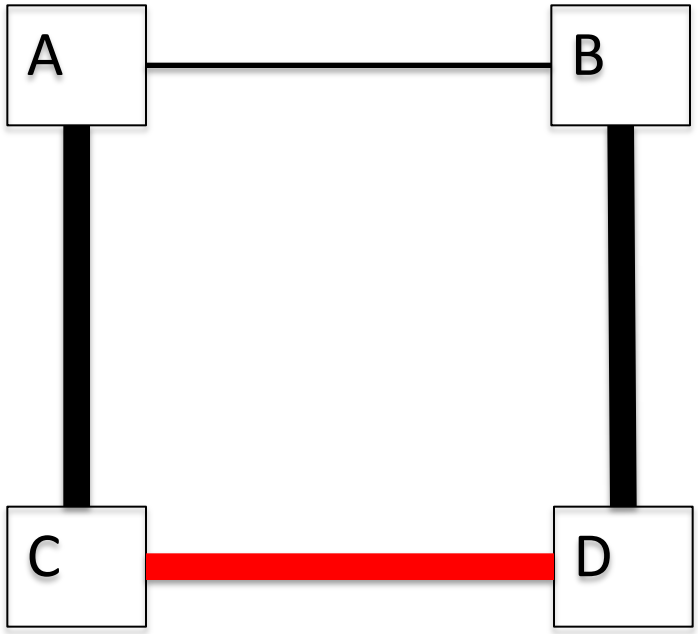


Analysis	DIC (lower = better fit)
Consistency Model	-13.89
Inconsistency Model	-14.75

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

Special Issue Paper

Research Synthesis Methods

Received 22 June 2011, Revised 11 May 2012, Accepted 15 May 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1044

Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies‡

J. P. T. Higgins,[a,b*†] D Jackson,[a] J. K. Barrett,[a] G Lu,[c] A. E. Ades[c] and I. R. White[a]

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

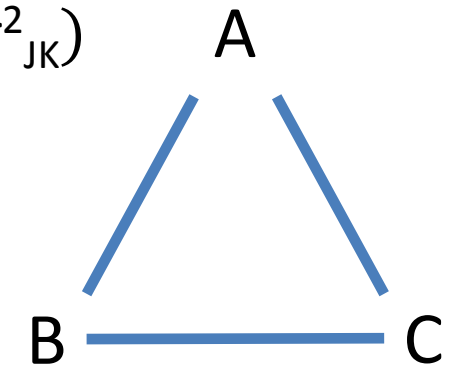
  $\delta^{AB}=\delta^{AC}-\delta^{BC}$

- Heterogeneity (Variation within a comparison)

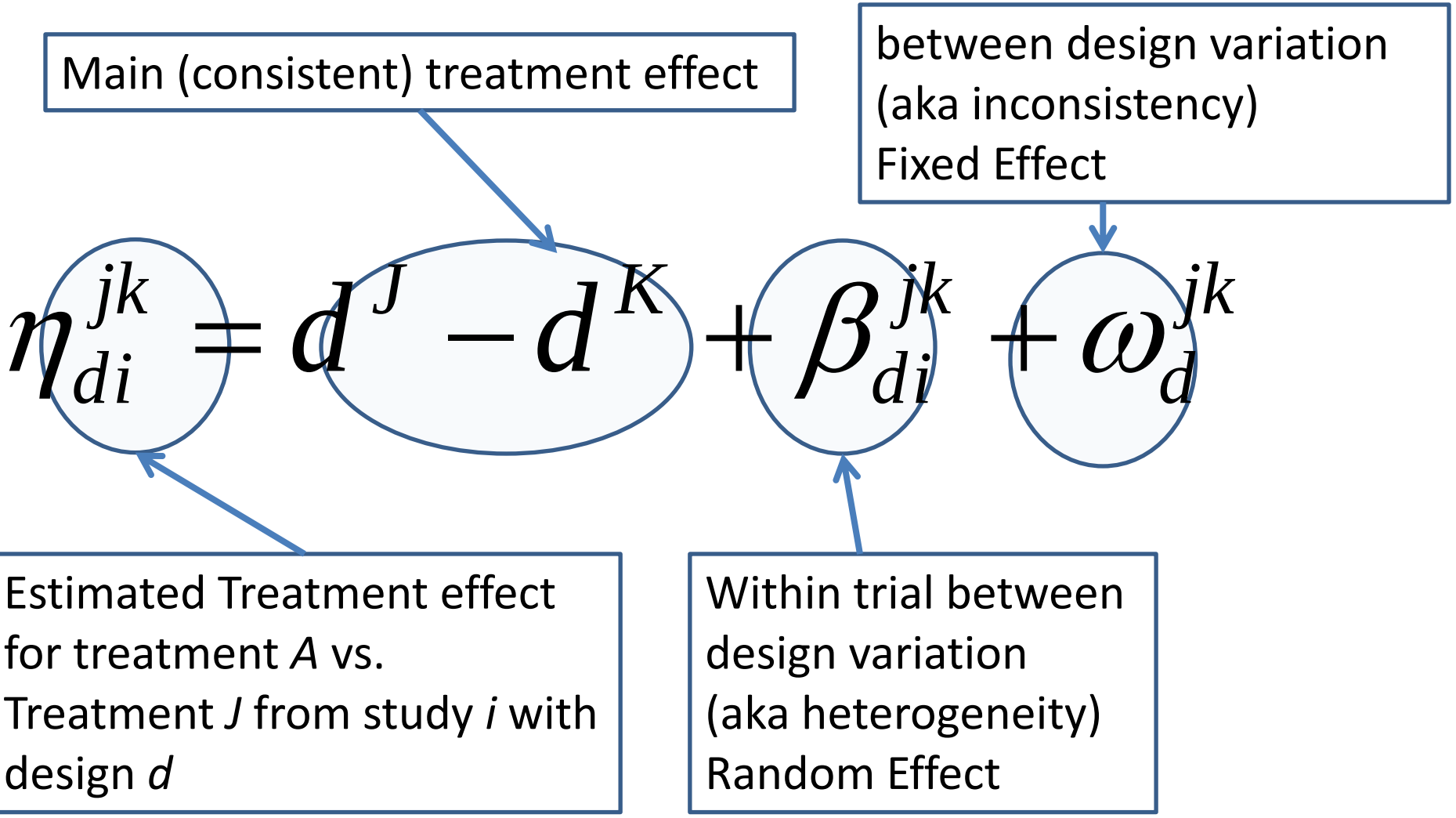
  Treatment specific variance: $\delta_i^{JK}\sim N(\delta^{JK},\tau^2_{JK})$

  Common variance: $\delta_i^{JK}\sim N(\delta^{JK},\tau^2)$

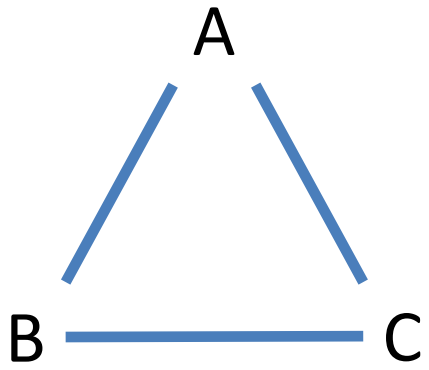
- Design inconsistency
  - Treatment effects vary by study design (design=comparator set)



Design by Treatment Interaction Model

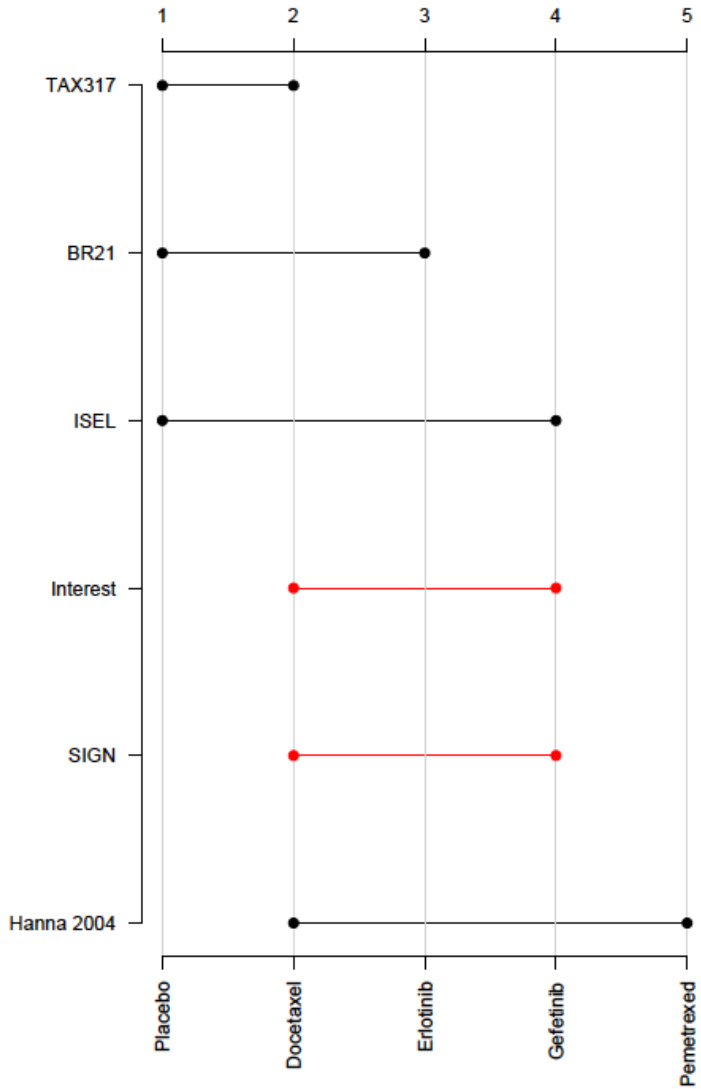


An Example



	Treatment		
Trial Design	A	B	C
ABC	Ref	δ^{AB}	δ^{AC}
AB	Ref	$\delta^{AB}+\omega_2{}^{AB}$	-
AC	Ref	-	$\delta^{AC}+\omega_3{}^{AC}$
BC	Ref	δ^{AB}	$\delta^{AB}+\omega_4{}^{AC}$

Results of the treatment by design analysis for the NSCLC NMA



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

Analysis	DIC (lower = better fit)
Consistency Model	-13.89
Treatment by Design	-14.80

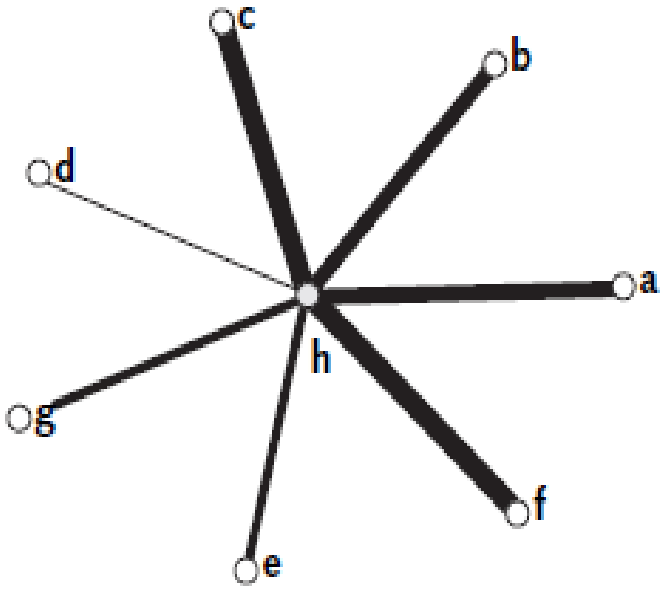
Quick Comparison

Method	Addresses Question	Provides Global Measure
Node splitting	Does each link agree with the rest of network	No
Inconsistency	How does imposing consistency affect fit (globally and per study)	Yes
Treatment by design	What is the difference between treatment effects estimated from different study designs?	Yes

Inconsistency cannot be observed in “star shaped” networks

A

Second-generation antiepileptic drugs in partial epilepsy



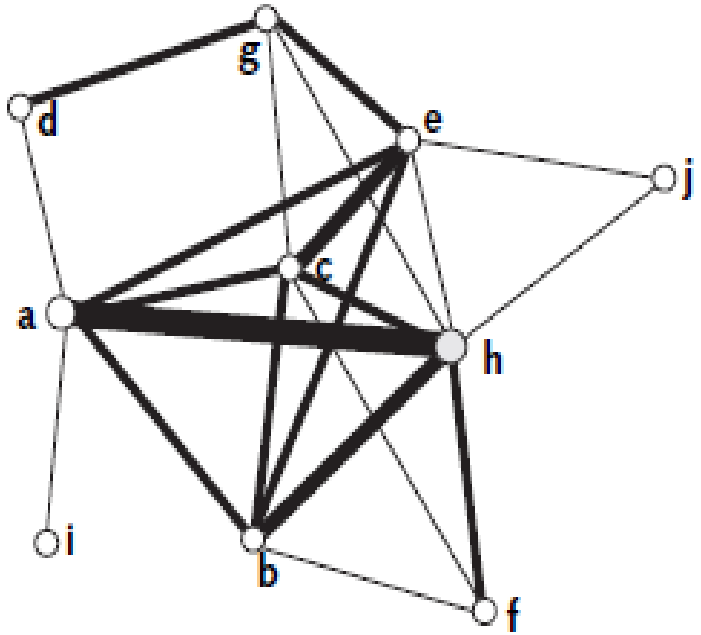
a: levetiracetam, b: gabapentin, c: lamotrigine, d: oxcarbazepine, e: tiagabine, f: topiramate, g: zonisamide, h: placebo

Salanti et al: Ann Intern Med. 2008;148:544-553.

55

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

D First-line antihypertensive therapy



a: diuretics, b: β -blockers, c: dihydropyridine CCB, d: nondihydropiridine CCB, e: ACE-i, f: ARB, g: diuretics or β-blockers, h: placebo/not treated, i: α-blocker, j: ACE-i + diuretics

Which should we trust most?

Salanti et al: Ann Intern Med. 2008;148:544-553.

56

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

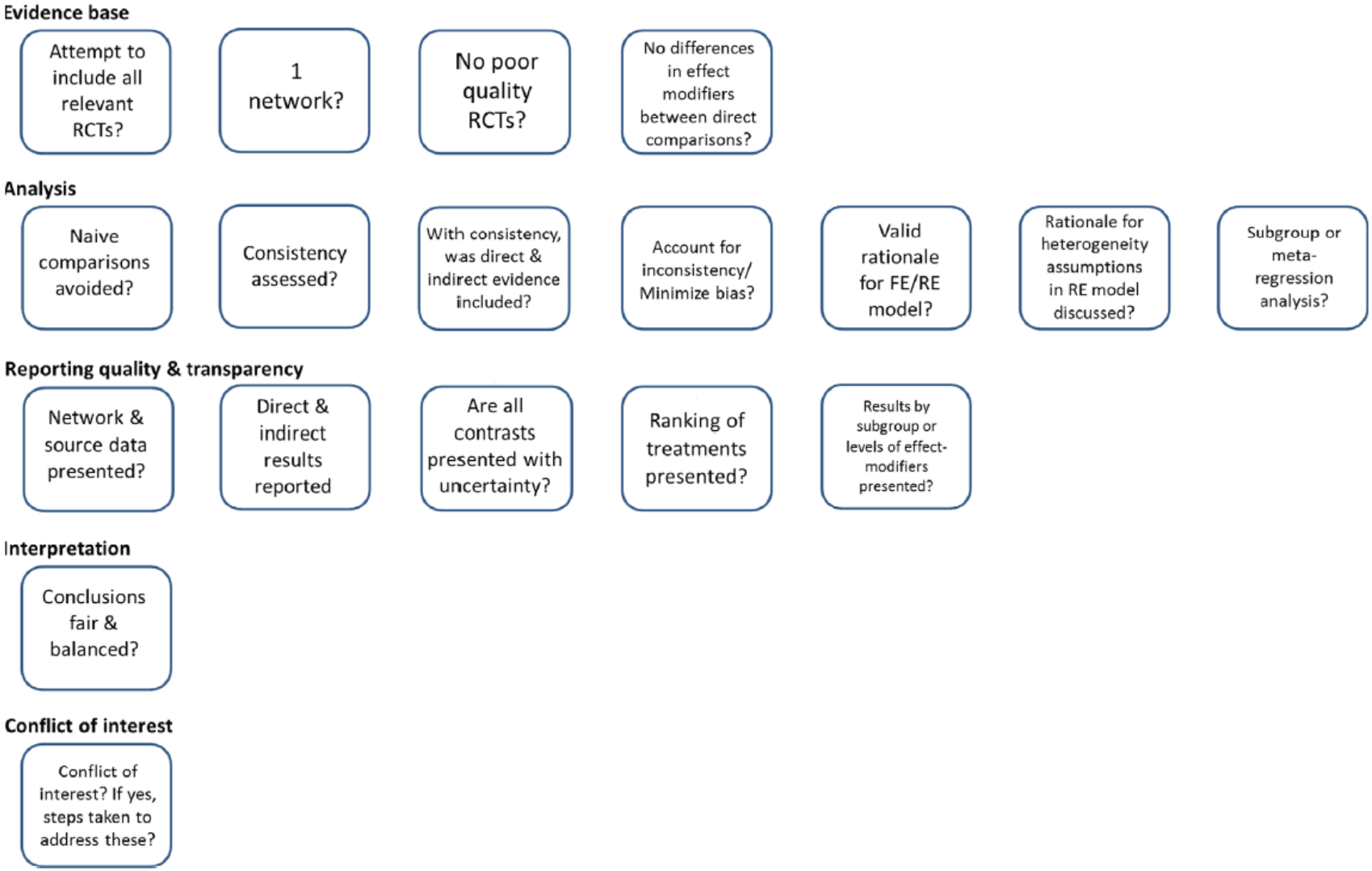


Fig. 1 – Overview of domains related to assessment of the credibility of a network meta-analysis. FE, fixed effects; RCTs, randomized controlled trials; RE, random effects.

But ultimately, the assessment of credibility is a jugdement

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