Network meta-analysis: making best use of the evidence for better decisions

Sofia Dias, PhD

Professor of Health Technology Assessment
Centre for Reviews and Dissemination
University of York, York, UK

Meta-Analysis Workshop 2021:
"Exploring the limits of advanced meta-analysis"

With thanks to: Sumayya Anwer, Nicky Welton

Conflicts of interest

I have no actual or potential conflicts of interest in relation to this presentation.

I acknowledge funding from the Medical Research Council, UK (MR/R025223/1)

iRECORD

*i*nferring *R*elative treatment *E*ffects from *C*ombined *O*bservational and *R*andomised *D*ata

Medical Research Council

Evidence synthesis for decision making

- Meta-analyses are regularly used to decide which medical technologies to fund or recommend for a particular target population
  - Health Technology Assessments (HTA) are conducted following a strict process of evidence gathering (usually a systematic literature review), critique and synthesis
  - Most HTAs will involve a meta-analysis of studies comparing the interventions of interest for the decision problem
- It is important to include all relevant evidence in order to make the best possible decisions given the available data at the time
- An accurate and fair quantification of uncertainty is also required

Evidence synthesis in HTA

- Often **multiple treatments** are of interest
- In many countries, HTA includes a decision model that accounts for the costs and benefits of the technologies being evaluated.
- This requires coherent treatment effect estimates that will lead to a coherent decision and an accurate quantification of the uncertainty in these estimates.
- Network meta-analysis (NMA) is an extension of standard meta-analysis to incorporate direct and indirect evidence on multiple treatment comparisons in a coherent way
  - NMA methods now well established
  - Respects randomisation
  - Increases precision and robustness of results as multiple sources of evidence are used to estimate the same relative treatment effect.

Brief History of NMA

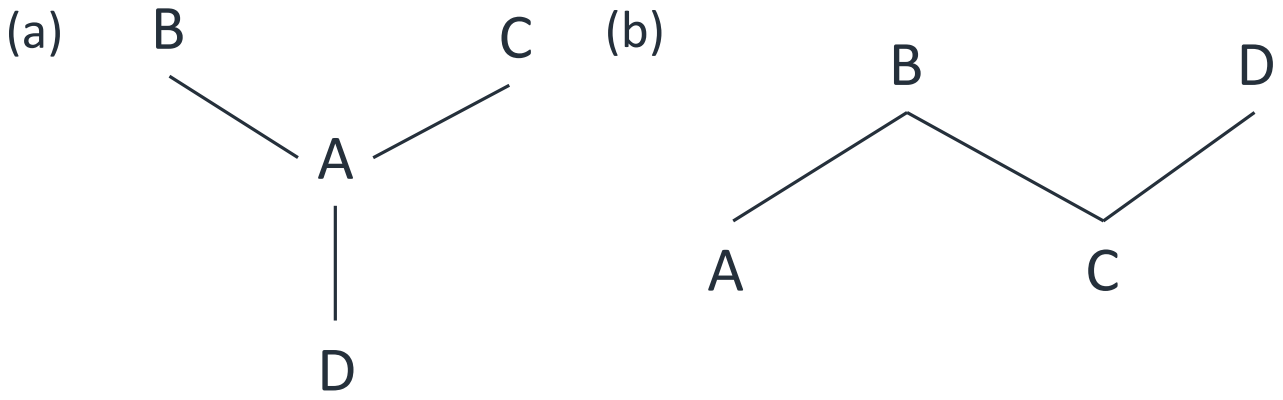
- The original idea was published in 1996
  - Higgins & Whitehead, Stat Med
- It was later picked up by the Health Technology Assessment community due to need to compare multiple treatments
  - Lu & Ades, Stat Med. 2004
  - Initially used in the UK (NICE)
  - Now also used in other countries (Canada, Germany, USA ...)
- Regular feature of NICE technology appraisals and clinical guidelines
- Bayesian simulation (MCMC) particularly appealing for decision analysis and this is regularly used in HTA
  - Also have additional flexibility to adapt models

NICE: National Institute for Health and Care Excellence; MCMC: Markov chain Monte Carlo

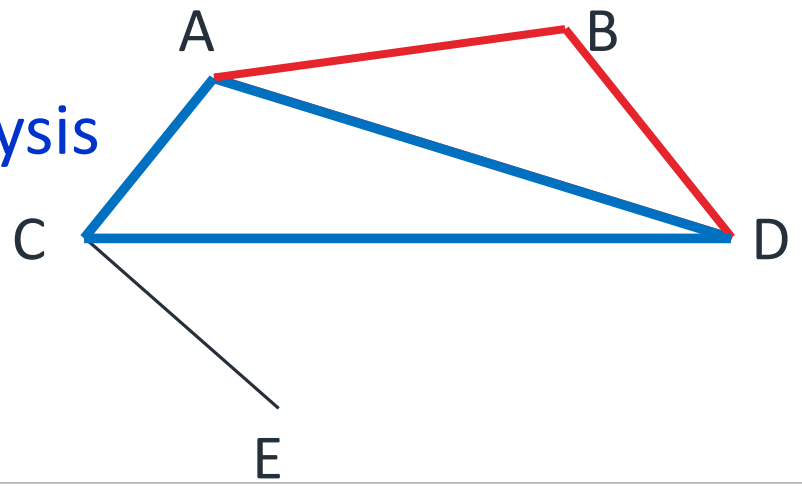
What is Network Meta-analysis?

Pair-wise MA

Indirect Comparisons



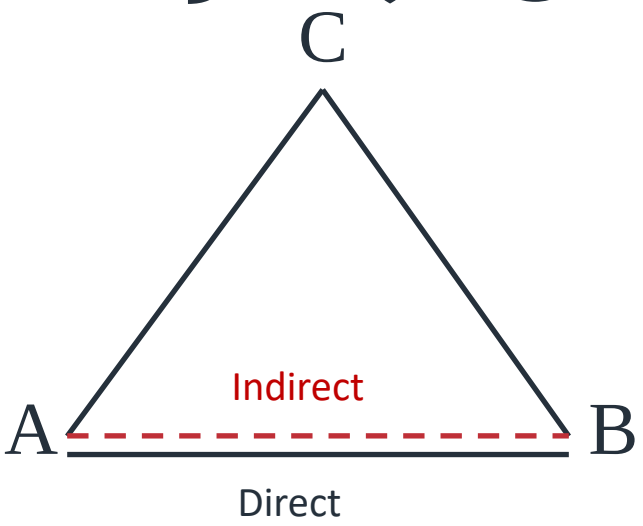
Network meta-analysis



- The existence of "evidence loops" means that there is both DIRECT evidence and INDIRECT evidence on the same contrast
- More data → estimates more precise, more robust (less sensitive to any one source of data)
- Possible to estimate additional parameters.

Network Meta-Analysis (NMA)

- Suppose interested in AB comparison
  - We have both direct and indirect evidence
  - Two sources of evidence to estimate the treatment effect → more evidence, more precision
- Additional source of evidence can be used to estimate additional parameters, eg a bias parameter
  - Ability to estimate parameters will depend on network structure → more connections and more loops mean is better



MA: Generic common effect model

The underlying model is

$$\theta_{ik} = \mu_i + d_{12} I_{k \neq 1}$$

continuous measure of the treatment effect for arm *k* of trial *i* (eg logOR)

effect of baseline treatment in trial *i* (nuisance parameter)

treatment effect of treatment *2* relative to treatment 1

Prior distributions

$$d_{12} \sim N(0, 100^2)$$

$$\mu_i \sim N(0, 100^2) \quad i = 1, \ldots, N$$

MA: Generic random effects model

The underlying model is

$$\theta_{ik} = \mu_i + \delta_{ik}$$

continuous measure of the treatment effect for arm *k* of trial *i* (eg logOR)

effect of baseline treatment in trial *i* (nuisance parameter)

trial-specific treatment effect of treatment in arm *k* relative to treatment in arm 1

Prior distributions

$$\tau \sim U(0,5)$$

$$d_{12} \sim N(0,100^2)$$

$$\mu_i \sim N(0,100^2) \quad i = 1,\ldots,N$$

$$\text{Set } \delta_{i1} = 0$$

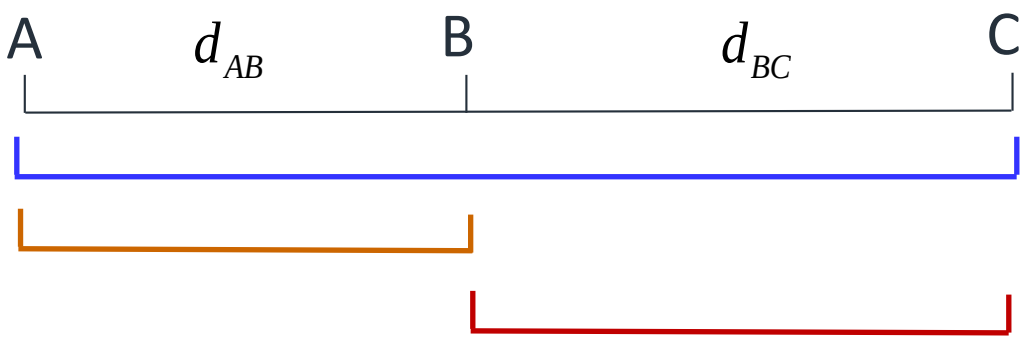
$$\delta_{i2} \sim N\left(d_{12}, \tau^2\right)$$

NMA: consistency

Suppose a decision maker is considering which of 3 treatments A,B,C is best for a specific (perhaps non homogeneous) group of patients.

The following statement about the **true** treatment effects must be correct:

$$d_{BC} = d_{AC} - d_{AB}$$



Extending the exchangeability assumption made in MA of studies comparing two treatments to any number of treatments, implies this is also true of the estimated relative effects.

Basic & Functional parameters

- Eg four treatments
- Take treatment 1 as the reference treatment
- Then the treatment effects (eg log odds ratios) of 2, 3, 4 relative to 1 are the **basic parameters**
- Given them priors: $d_{12}, d_{13}, d_{14} \sim N(0,100^2)$
- Remaining contrasts are **functional parameters**

$$d_{23} = d_{13} - d_{12};\quad d_{24} = d_{14} - d_{12};\quad d_{34} = d_{14} - d_{13}$$

CONSISTENCY EQUATIONS

Basic & Functional parameters

- Functional parameters fully determined by basic, i.e. can infer functional from basic
- Any information on functional parameters tells us indirectly about basic parameters
- For example: RCTs of 3 vs 2

$$\theta_{i1} = \mu_{i2} \quad \text{for "baseline/control" arm (treat 2)}$$

$$\theta_{i2} = \mu_{i2} + d_{23} = \mu_{i2} + d_{13} - d_{12} \quad \text{for "comparator" arm (treat 3)}$$

Treatment effect always written in terms of the basic parameters

- Fewer parameters to estimate → more precision

NMA: Generic random effects model

The underlying model is

$$\theta_{ik} = \mu_i + \delta_{ik}$$

continuous measure of the treatment effect for arm *k* of trial *i* (eg logOR)

effect of baseline treatment in trial *i* (nuisance parameter)

trial-specific treatment effect of *k*-th treatment relative to baseline treatment

Prior distributions

$$\tau \sim U(0,5)$$

$$d_{1k} \sim N(0,100^2) \quad k = 2,\ldots,nt$$

$$\mu_i \sim N(0,100^2) \quad i = 1,\ldots,N$$

Set $\delta_{i1} = 0$

$$\delta_{ik} \sim N\left(\underbrace{d_{1t_{ik}} - d_{1t_{i1}}}_{\text{consistency equations}}, \tau^2\right)$$

Validity of Pairwise MA and NMA

- Pooled treatment effects and resulting decisions rely on integrity of evidence on which they are based
- RCTs considered the gold-standard evidence to inform relative treatment efficacy
- NMAs informed mainly by RCTs but data on some comparisons may be sparse and results imprecise
  - NRS can be used to strengthen evidence but concerns about additional bias
- Need to balance the desired increase in precision with the potential for introducing bias (and heterogeneity)
- Exchangeability assumption is violated if deviations in treatment effect (eg due to bias from different designs) can be differentiated in advance
- Different models have been proposed to deal with this

NRS: non-randomised studies

Bias Model

- Assumes NRS data are at risk of bias and RCT data are not at risk of bias
- True relative treatment effect for treatment Y compared to X
  - Unbiased studies (RCT) → δ_{XY}
  - Biased studies (NRS) → $\delta_{XY} + \beta$
- Estimate bias and adjust for it
- NRS evidence is automatically adjusted and down-weighted by the bias term
  - in a Bayesian framework the additional uncertainty is propagated to all the estimates automatically
- Where does information on β come from?

Bias Model: info on β from…

- External sources
  - Guess
  - Formal elicitation (Turner *et al* JRSS A, 2009)
  - Evidence-based: compare RCT and NRS evidence in similar meta-analyses to inform β (Welton *et al* JRSS A, 2009)
- **Internal sources** (Dias *et al* JRSS A, 2010)
  - Evidence-based
  - Estimate bias due to type of study *within* same set of trials and adjust for it
- External evidence on bias can also be included in the form of prior distributions for β (e.g. from meta-epidemiological data or expert opinion) which can be updated by the data

Adjustment in NMA

- Consistency equations mean there is more power to estimate bias parameters in NMA than in pairwise meta-analysis
- Model is as before but with an added 'bias' term

$$\theta_{ik} = \mu_i + (\delta_{ik} + \beta_{ik} X_i) \qquad \text{where } \delta_{i1} = \beta_{i1} = 0$$

- NMA model on δ's as usual
- Bias assumed **exchangeable** across NRS: $\beta_{ik} \sim N(b, \kappa^2)$

X_i = study design indicator (1=NRS; 0=RCT)

b = mean bias for NRS

κ^2 = between study variance in bias

Modelling Bias: assumptions

$$\beta_{ik} \sim N(b, \kappa^2)$$

- Assumptions required for b
- Eg:
  - Mean bias the same for all NRS
  - Mean bias depends on further study characteristics, eg NRS adjusted for confounders or not, RCTs at risk of bias or not
- Number of 'bias' parameters that can be estimated will depend on network structure
  - Need evidence redundancy

Possible Models

- Separate synthesis model
  - RCT and NRS evidence are synthesised separately (no information sharing).
- Lumped Model (shared parameter model)
  - Share heterogeneity and treatment effect parameters: no difference assumed in effects estimated by RCT and NRS evidence
    - Useful as a comparison model, but not suitable for inference
- Multi-level/ Hierarchical Models
  - Assumes both types of studies are estimating a parameter from a common distribution (exchangeability)
  - does not account for potential bias in NRS evidence and does not specifically downweight it compared to RCT evidence.
- Bias Models
  - Proposed to incorporate RCT evidence at risk of bias with evidence not at risk → extend to NRS
- Prior-based Models
  - Allow for explicit down-weighting of NRS evidence

Hierarchical Model

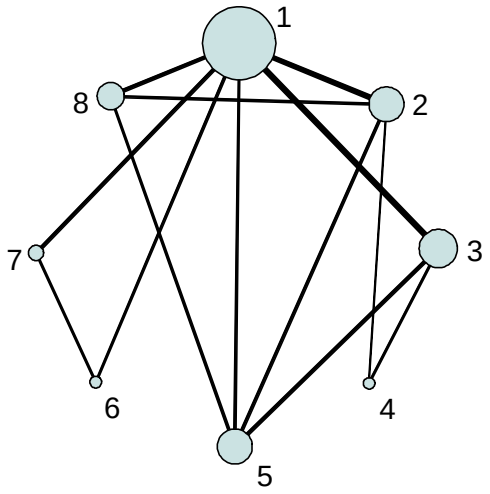
- Assumes exchangeability between parameters informed by RCT and NRS
- Obtain estimate of overall relative treatment effects across both types of evidence
- Also obtain model-based estimate of relative treatment effect for RCT evidence, accounting for the NRS evidence under the model assumptions (shrunken estimate)
- Which of these estimates to choose?
  - Can argue that RCT-specific estimate more meaningful and will usually be more precise

Example Dataset

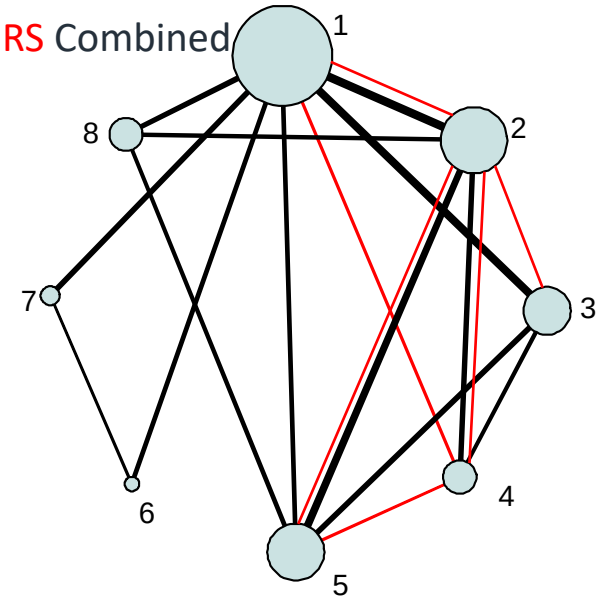
In-stent Restenosis (Efthimiou et al. 2017)

- Outcome: Need for target-lesion revascularisation (binary, odds ratio)
- Aggregate data from
  - 28 RCTs (N=5917, each study had between 17-259 patients)
  - 6 NRSs (N=1019 patients, each study had between 26-165 patients)
- Comparison of 8 different treatments for the treatment of coronary in-stent restenosis

RCTs



RCT+ NRS Combined



Codes for interventions:
1= balloon angioplasty
2= sirolimus-eluting stents
3= drug-coated balloons
4= everolimus-eluting stents
5= paclitaxel-eluting stents
6= rotablation
7= bare metal stents
8=vascular brachytherapy.

NRS = non-randomised study

Efthimiou et al. Combining randomized and non-randomized evidence in network meta-analysis. Stat Med. 2017; 36:1210-1226

General Approach

- NMA with RCT data only
- Check robustness/precision of treatment effect estimates
- NRS evidence available
  - Assumed of sufficient quality and relevance to incorporate in synthesis
  - Provides additional evidence on key comparisons

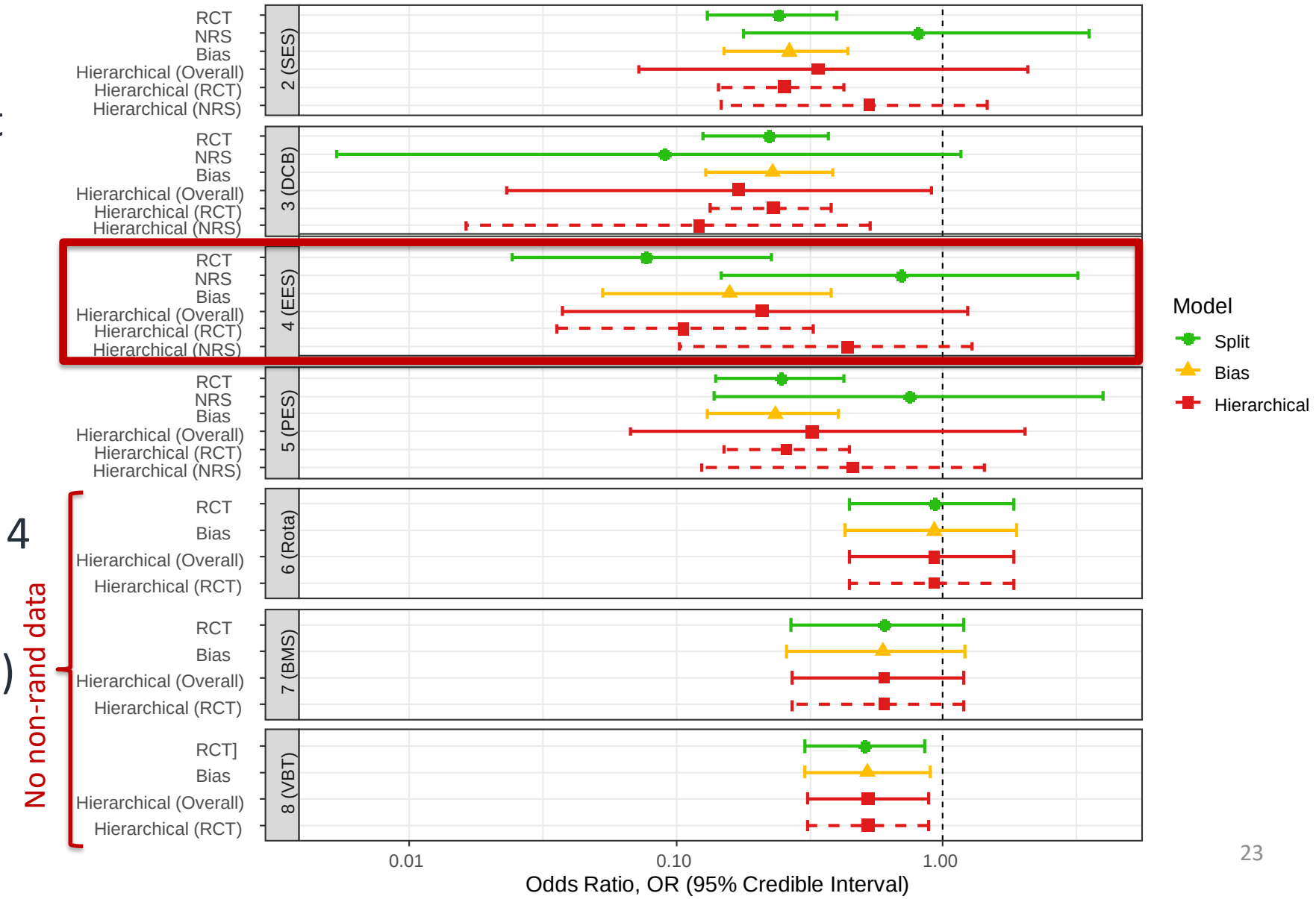
- Will compare separate models, hierarchical model and bias model

Comparing Treatment Effects

- Bias model tends to be similar to RCT estimates but "corrects" for assumed bias in NRS
- Most results similar to hierarchical model RCT estimates but increased "correction" for conflict in comparison with treatment 4
- Bias estimated as

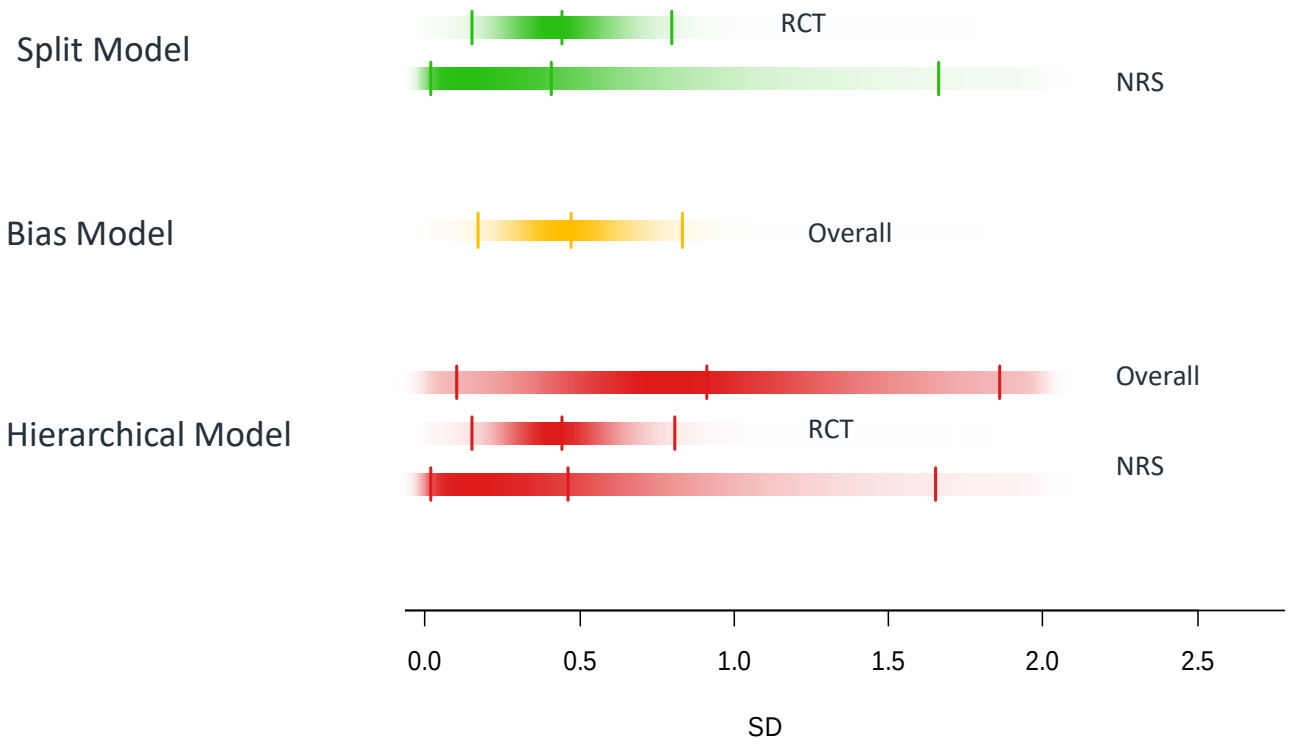
  0.53 95%CrI(-0.48, 1.73)

on a log-scale



Between-study heterogeneity

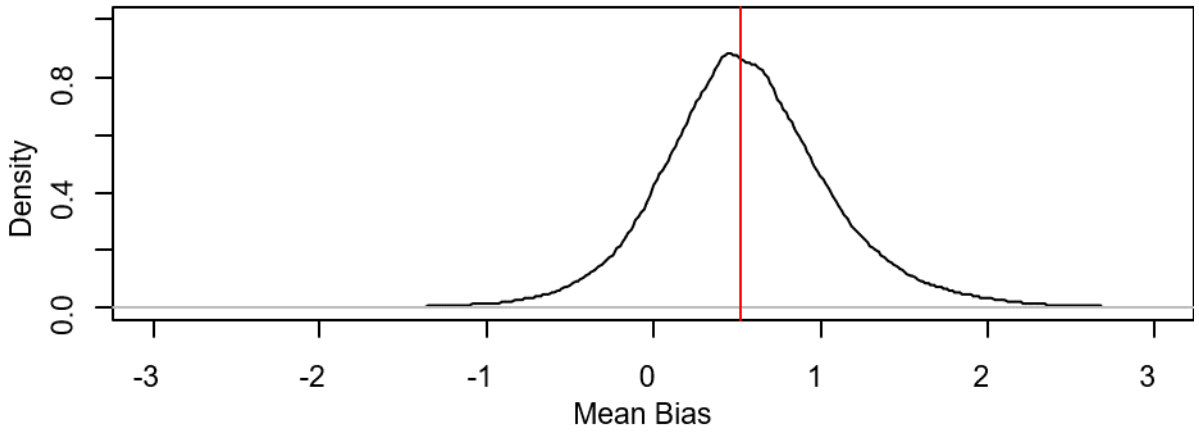
Data source	Split Model	Bias Model	Hierarchical Model
RCT	0.44 (0.15, 0.80)		0.44 (0.15, 0.81)
NRS	0.41 (0.02, 1.66)		0.46 (0.02, 1.65)
Overall			0.91 (0.10, 1.86)
RCT + NRS		0.47 (0.17, 0.83)	



Bias Parameters (Bias model)

- The bias in the NRS can be estimated (but imprecise in this example)
- variability around the mean bias is not well estimated
  - Consider informative prior distributions?
  - Used Unif(0,3)
- Positive (moderate) bias estimated
  - Suggests NRS underestimating treatment effect of everolimus-eluting stents (4) compared to balloon angioplasty (1)

Mean bias	0.53 (-0.48, 1.73)
Variability in Bias (SD)	0.80 (0.04, 2.47)



Bias modelling: Reflections

- It is possible to estimate and adjust for the additional bias present in NRS
  - In networks of moderate size where NRS add new evidence to several comparisons
  - Model performance will vary by network structure
- Assumed RCTs **not at risk** of bias and NRS all at the **same risk** of bias
  - Model can be extended to have different types of bias within study type, according to study quality assessment (requires multiple studies of each type)
- Bias estimation and adjustment should be considered as a **sensitivity analysis**
  - Bias adjustment can have a big impact on relative effects
  - Between trial heterogeneity in treatment effects usually reduced in bias-adjusted models
- Expert opinion and empirical evidence can be used to supplement estimation of bias parameters

Related interventions: class effects

- Decision making framework requires discrete classification of treatments
  - By name of intervention, dose, mode of administration, duration etc
- Must be able to assign a cost and benefit to each treatment in the economic model
- But this can results in very large, yet sparse networks…
- Sometimes these interventions fall into classes
- Treatments in the same class can be assumed to have similar (**exchangeable**) effects
  - Based on mode of action, pharmacological composition, etc
- A class model borrows strength across treatments in the same class
  - The original treatment definitions are retained

Class models

1. No class effects
   - Standard NMA of separate interventions
2. Random class effects
   - Effects of interventions in a class (compared to the reference class) are distributed around a common class mean with a within-class variance
   - Treatment effects are shrunk to class mean (like a random effects meta-analysis)
3. Fixed class effects
   - All treatments in a class have the same relative effect compared to the reference
   - Equivalent to doing a NMA of classes
4. Or, any combination of these…

NMA: exchangeable class model

Standard NMA model:

$$\delta_{ik} \sim N\left(d_{1t_{ik}} - d_{1t_{i1}}, \sigma^2\right) \qquad k > 1$$

$$d_{1k} \sim N(0, 100^2)$$

Treatments effects within classes exchangeable

$$\delta_{ik} \sim N\left(d_{1t_{ik}} - d_{1t_{i1}}, \sigma^2\right) \qquad k > 1$$

$$d_{1k} \sim N\left(m_{C_k}, \tau^2_{C_k}\right) \qquad m_c \sim N(0, 100^2),\ c = 1, \ldots, \text{Nclasses}$$

Within-class mean effect

Within-class variance

C_k – indicates class of treatment k

Informing the within-class variances

1. From data
   - Vague prior distribution
   - Require enough elements in a class and within-class comparisons made in different studies
2. Assume common across some/all classes
3. Assign informative prior distribution
   - Based on previous meta-analyses?
4. Combination of above…

NMA: fixed class model

Standard NMA model:

$$\delta_{ik} \sim N\left(d_{1,t_{ik}} - d_{1,t_{i1}}, \sigma^2\right) \qquad k > 1$$

$$d_{1k} \sim N(0, 100^2)$$

Treatments effects within classes are the same

$$\delta_{ik} \sim N\left(d_{1,t_{ik}} - d_{1,t_{i1}}, \sigma^2\right) \qquad k > 1$$

$$d_{1k} = m_{C_k} \qquad m_c \sim N(0, 100^2)$$

Within-class mean effect

Equivalent to NMA using the class as "treatment"

C_k – indicates class of treatment k

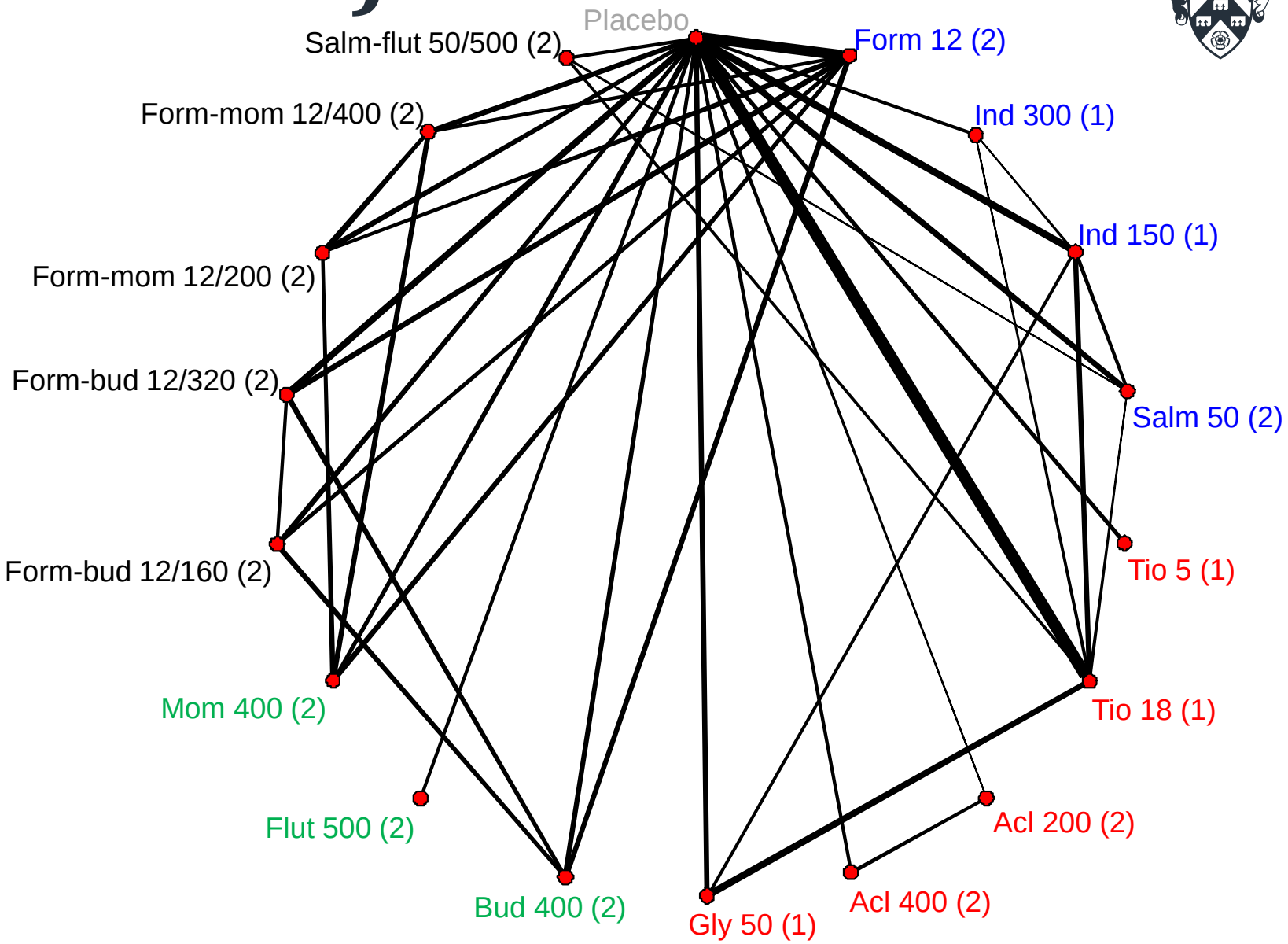
Example: Chronic Obstructive Pulmonary Disease (COPD)

- Treatments for COPD
  - Cochrane Review: Kew at al 2014
- Outcome: mean difference on St George's Respiratory Questionnaire (SGRQ) total score at 6 months
- 25 trials, 18 treatments compared
  - 39 pairwise comparisons made
- Treatments belong to 5 classes: **Placebo**, Long-active β_2-agonists (**LABA**), Long-acting muscarinic antagonists (**LAMA**), Inhaled corticosteroids (**ICS**) and **LABA+ICS**.
- Use exchangeable class model

COPD Treatments and Classes

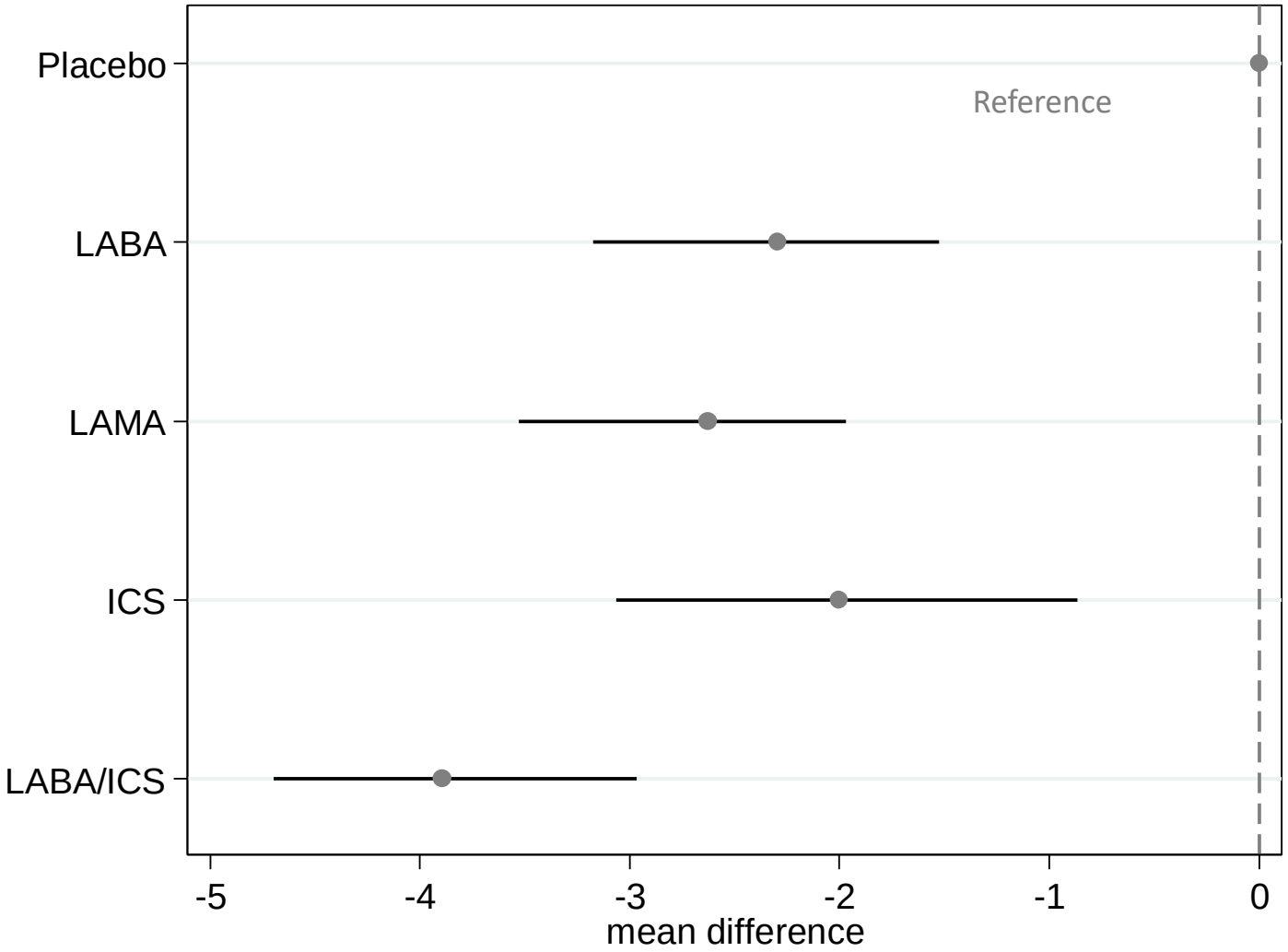
Class	Drug	Dose/individual node	Code for figure
PLACEBO	Placebo		Placebo
LABA	Formoterol	9-12 x2	Form 12 (2)
		18-24 x2	Form 24 (2)
	Indacaterol	300 x1	Ind 300 (1)
		600 x1	Ind 600 (1)
		150 x1	Ind 150 (1)
	Salmeterol	50 x2	Salm 50 (2)
LAMA	Tiotropium	5 x1	Tio 5 (1)
		10 x1	Tio 10 (1)
		18 x1	Tio 18 (1)
	Aclidinium bromide	200 x1	Acl 200 (1)
		200 x2	Acl 200 (2)
		400 x2	Acl 400 (2)
	Glycopyrronium bromide	50 x1	Gly 50 (1)
ICS	Budesonide	200 x2	Bud 200 (2)
		320-400 x2	Bud 400 (2)
		750-1000 x2	Bud 750+ (2)
	Fluticasone	250 x2	Flut 250 (2)
		500 x2	Flut 500 (2)
	Mometasone	400 x2 or 800 x1	Mom 400 (2)
LABA/ICS	Formoterol/budesonide	9-12/160 x2	Form-bud 12/160 (2)
		9-12/320 x2	Form-bud 12/320 (2)
	Formoterol/mometasone	9-12/200 x2	Form-mom 12/200 (2)
		9-12/400 x2	Form-mom 12/400 (2)
	Salmeterol/fluticasone	50/250 x2	Salm-flut 50/250 (2)
		50/500 x2	Salm-flut 50/500 (2)

SGRQ (6 months)



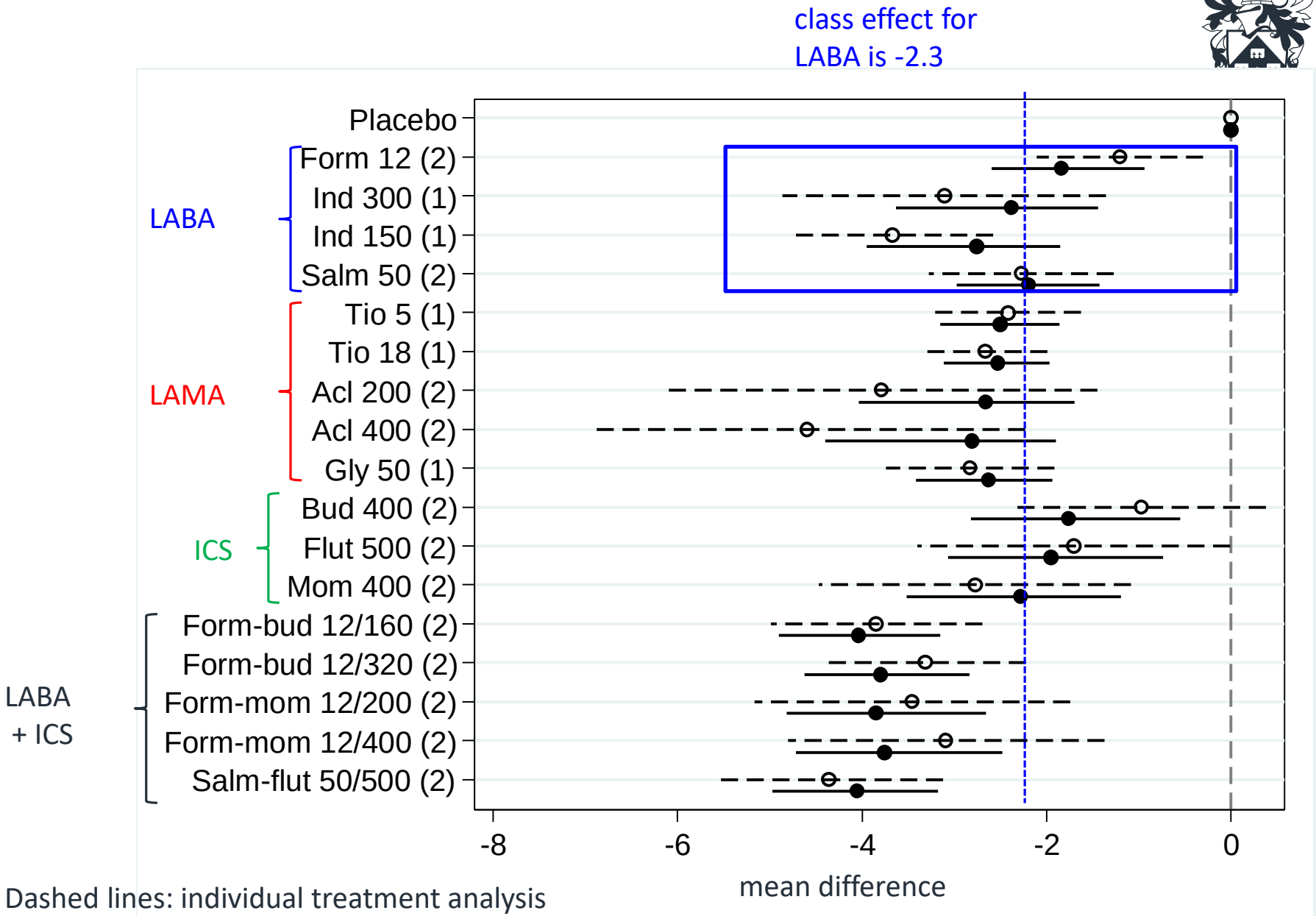
Placebo: Grey
LABA: Blue
LAMA: Red
ICS: Green
LABA/ICS: Black

SGRQ 6 months: Class effects



Within-class standard deviation: 0.44 95%CrI(0.03,1.09)

SGRQ 6 months: FE NMA with and without class effects



Dashed lines: individual treatment analysis
Solid lines: class analysis

Class models: Reflections

- Often clinical decisions are made on a 'class-basis' with the specific treatment being prescribed depending on patient preferences, availability or cost
- Class models can make sense but need to have *a priori* clinical plausibility as usually very few data to check → Clinical advice is key
- Useful when data are sparse
  - Informing within-class variances tricky for classes with few elements
- Borrowing of strength within class and individual treatment effects are retained for the decision model.
- Models also proposed for synthesising treatments at different doses
  - Caution!
  - "Class" exchangeability assumption is violated since we expect that higher doses will have different effects

Conclusions (1)

- NMAs allow more evidence to be combined so can estimate relative effects more precisely
  - Exchangeability assumption across all studies (which implies consistency of relative effects) needs to be satisfied
  - This is impacted by quality and design of included studies
- Advanced modelling can estimate additional parameters, adjust results and/or borrow strength across studies and comparisons
  - The number of parameters that can be estimated depends on the assumptions being made and the network structure: how many studies per comparison and how many loops

Conclusions (2)

- Complex synthesis models (including NMA) have been developed to allow for better decisions under current knowledge
  - Need to have a target population and target interventions → focussed decision problem
  - Key to estimate parameters and their uncertainty using the **best available** evidence
- Complexity of the models can increase but assumptions need to be transparent and validated
- Data availability can be a problem if many parameters to estimate, although some modelling can also mitigate for lack of data.
  - Time and resource implications for data collection, eliciting expert opinion, model validation, communicating findings
  - is it worth it?

Thank you

@sdias_stats
sofia.dias@york.ac.uk