University of BRISTOL

Network meta-analysis on disconnected evidence networks. What can be done?

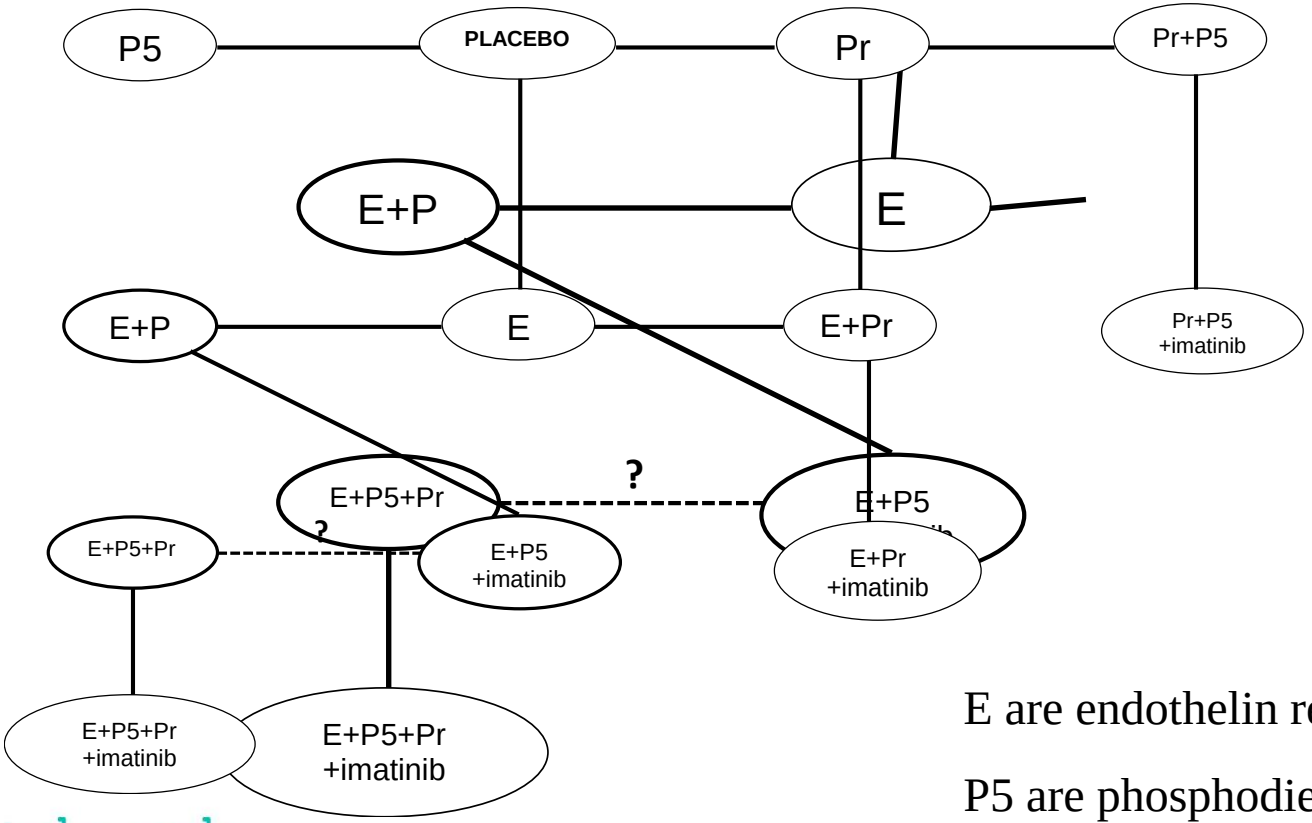
Howard Thom
with Joy Leahy and Jeroen Jansen
University of Bristol
27th August 2020
Howard.thom@Bristol.ac.uk

NIHR | Bristol Biomedical Research Centre

bristol.ac.uk

Disconnected network in imatinib for PAH

- Thom 2015 network meta-analysis (NMA) in pulmonary arterial hypertension (PAH)
- Wanted to compare “E+P5+Pr” vs “E+P5+imatinib”
- Network based solely on RCTs was disconnected.



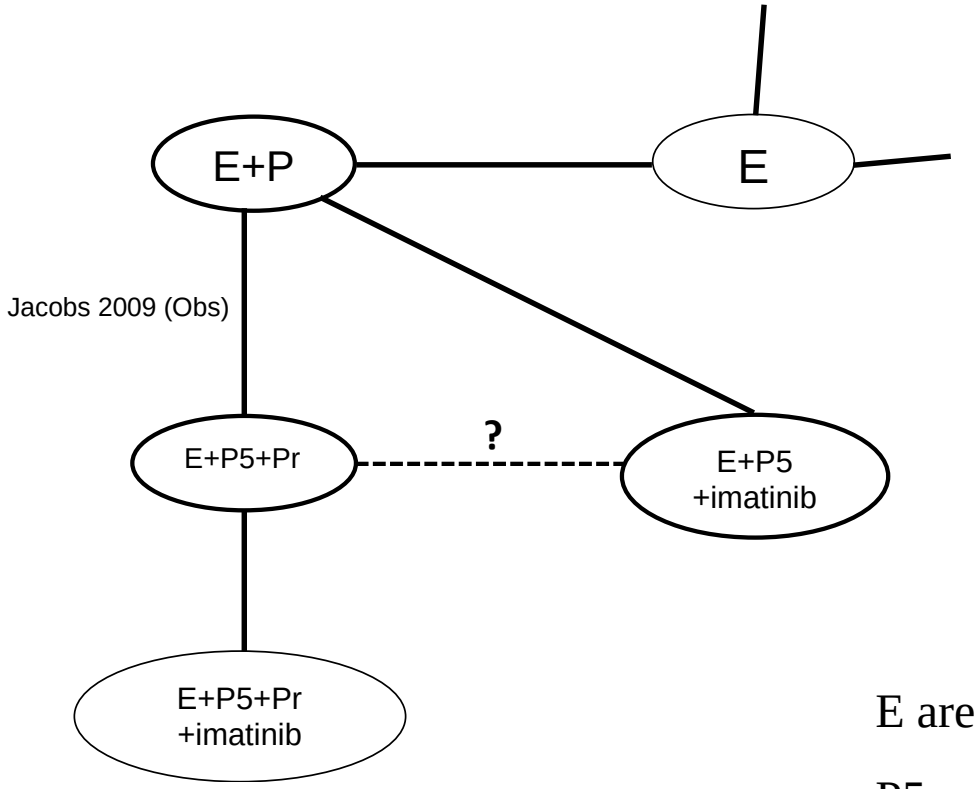
E are endothelin receptor antagonists

P5 are phosphodiesterase-5 inhibitors

Pr are prostacyclin analogues.

Disconnected network in imatinib for PAH

- Completed the network using single-arm observational studies
- This required an assumption of exchangeable or **random** baseline effects

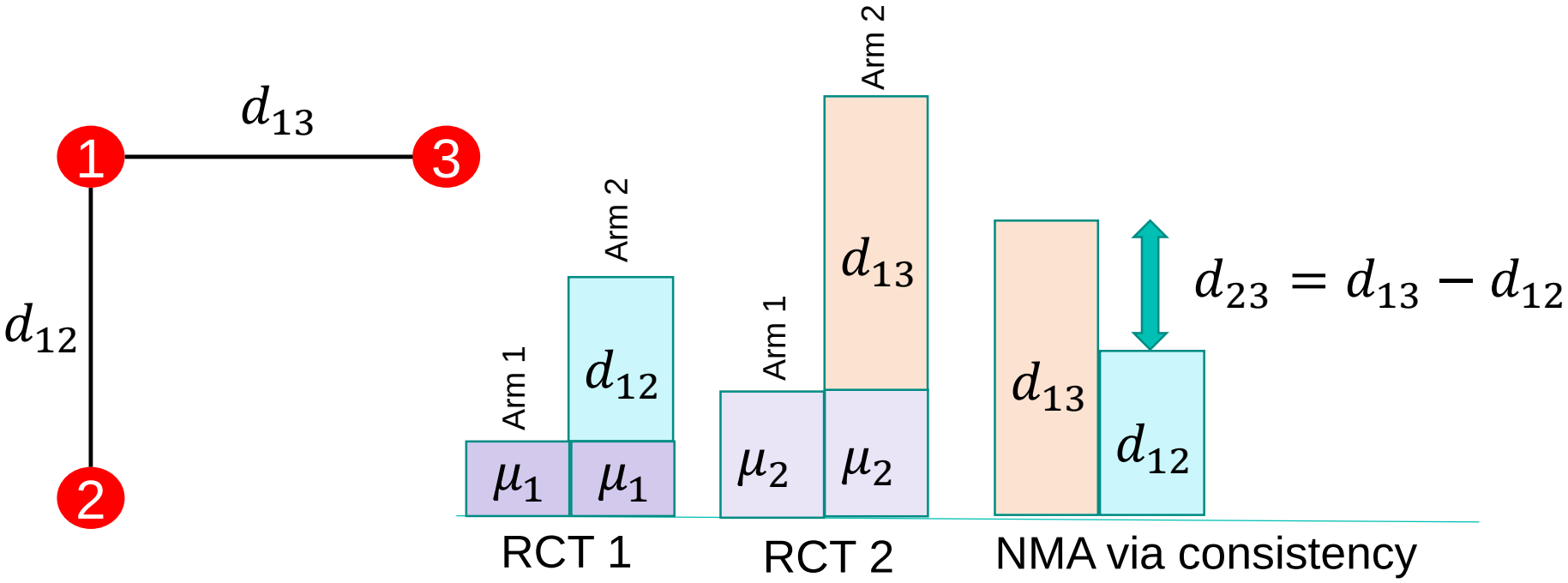


E are endothelin receptor antagonists

P5 are phosphodiesterase-5 inhibitors

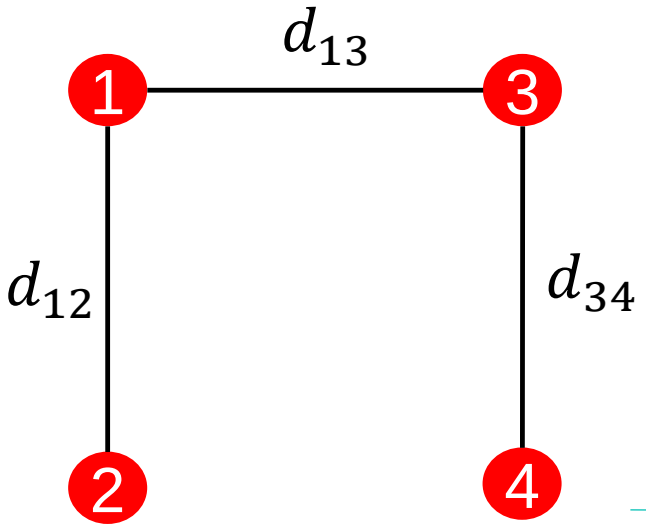
Pr are prostacyclin analogues.

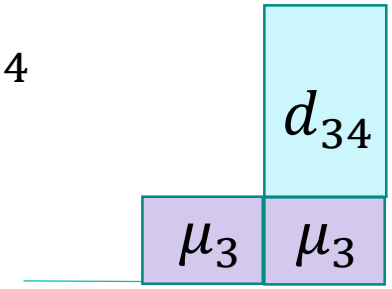
Network meta-analysis with independent baselines



- μ_i is the "baseline effect" representing the (for example) log odds of event on the baseline treatment of RCT i.
- In RCTs 1 and 2 above, it corresponded to "reference treatment" 1.
- All models fit through Bayesian MCMC in OpenBUGS software with vague priors

Independent baselines if networks are connected





RCT 3

$$d_{34} = d_{14} - d_{13}$$

$$\Rightarrow d_{14} = d_{34} + d_{13}$$

NMA via consistency assumption

And also $d_{24} = d_{14} - d_{12}$

- In RCT 3 the μ_3 corresponds to log odds of response on treatment 3.
- Use consistency to link to any treatment in the connected network.

Network meta-analysis with independent baselines

- If modelling binary outcomes for arm k of trial i

$$r_{ik} \sim Binomial(p_{ik}, n_{ik})$$

$$\text{logit}(p_{ik}) = \mu_i + \delta_{i,bk} \quad \text{if } t_{ik} \neq b$$

$$\text{logit}(p_{ik}) = \mu_i \quad \text{if } t_{ik} = b$$

- With

$$\delta_{i,bk} = d_{1t_{ik}} - d_{1t_{ib}} \quad \text{(Fixed effects)}$$

$$\delta_{i,bk} \sim Normal(d_{1t_{ik}} - d_{1t_{ib}}, \sigma) \quad \text{(Random effects)}$$

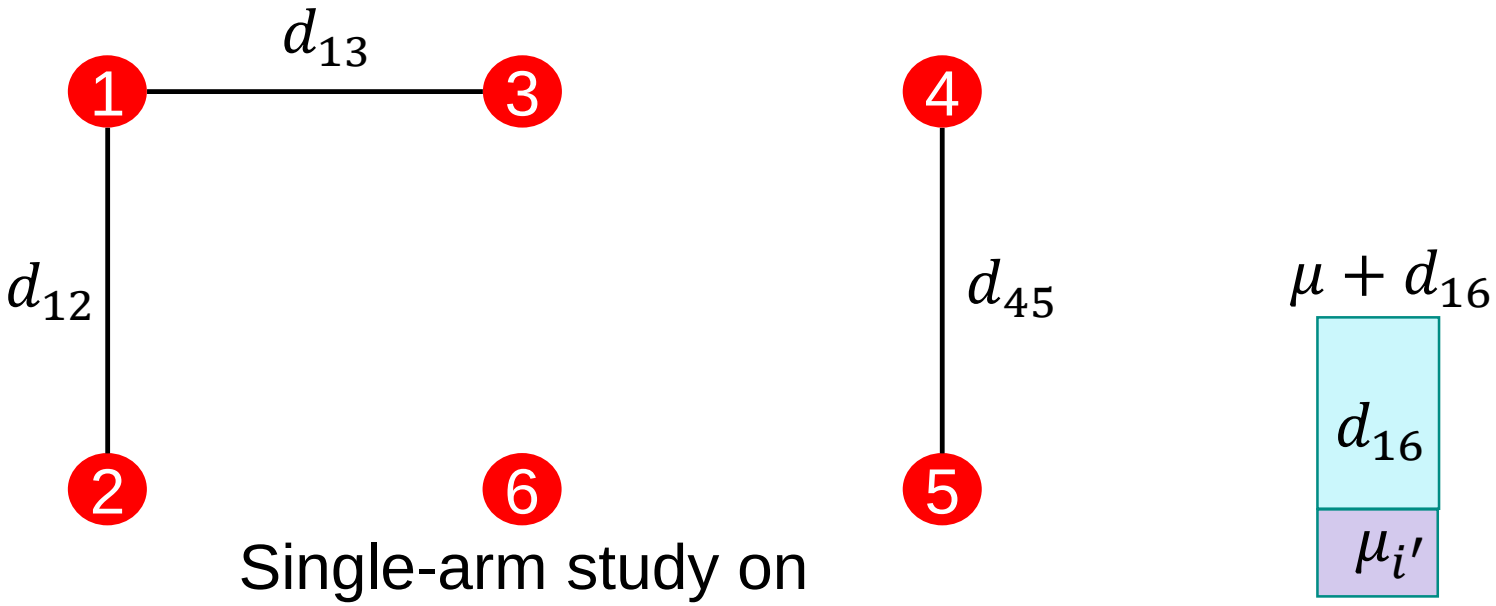
- So d_{1t} are log odds ratios for treatment t vs 1

Network meta-analysis with independent baselines

$$\text{logit}(p_{ik}) = \mu_i + \delta_{i,bk} \quad \text{if } t_{ik} \neq b$$
$$\text{logit}(p_{ik}) = \mu_i \quad \text{if } t_{ik} = b$$

- μ_i are assumed independent across RCTs (nuisance parameters)
- This controls for differences in *prognostic variables* across RCTs
  - These are variables that affect the baseline response
- If not assumed independent (e.g. $\mu_i \sim \text{N}(\text{m}, \sigma_\mu)$) can interfere with randomization (i.e. estimates of d_{12} and d_{13} may be biased)
  - 'Placebo' response may improve over time. Random baseline effects pull older placebo response up and push new placebo response down.
  - Biases against older treatments
  - However, work has demonstrated bias can be limited in practice (Beliveau 2017)

Disconnected networks



Single-arm study on treatment 6

Although consistency tells us

$$d_{45} = d_{15} - d_{14}$$

We don't know d_{15} or d_{14} as neither 5 nor 4 are connected to 1

Why not call it a day?

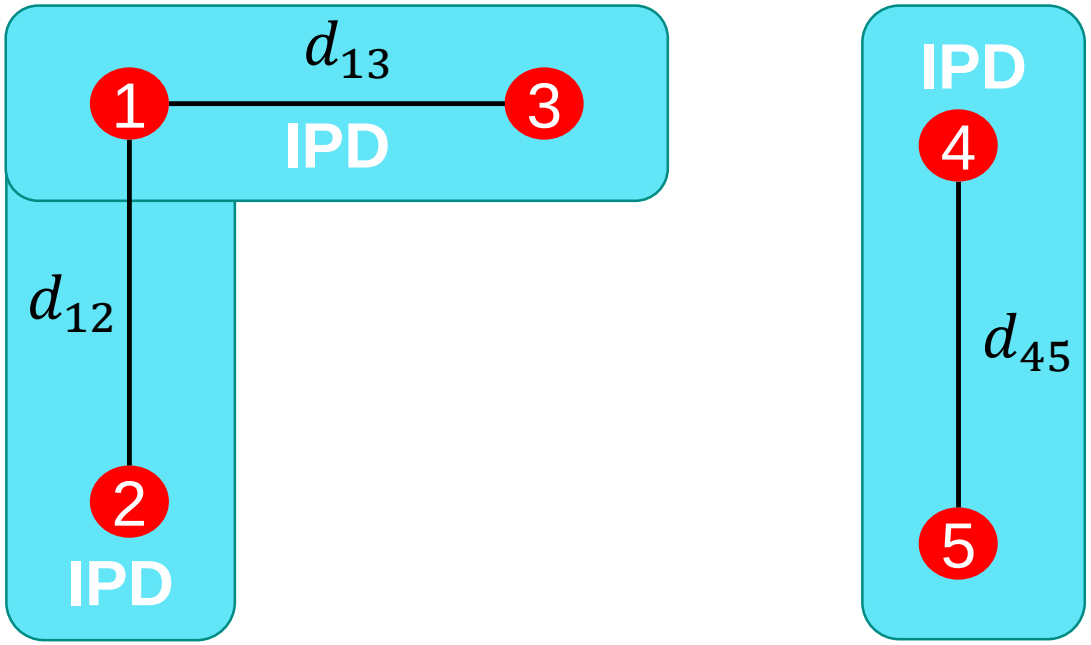
- Wait for RCTs to compare 4 or 5 with 1, 2, or 3?
  - Rare disease, so running RCTs practically impossible.
  - Treatments 1, 2, and 3 may be very old. Perhaps not ethical to include in an RCT.
- Healthcare decision makers (e.g. NICE in the UK) don't have that luxury.
- Not comparing 4 or 5 due to lack of evidence is an implicit decision that 1, 2, or 3 are better.

NICE
National Institute for Health and Care Excellence

bristol.ac.uk

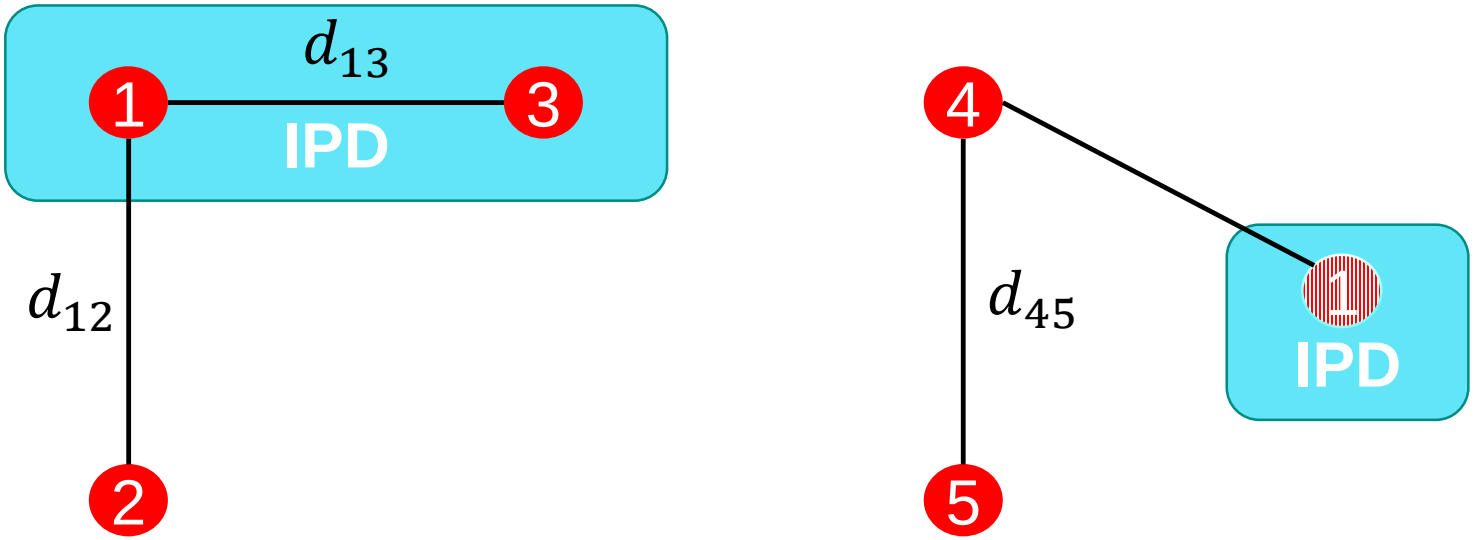
NIHR | Bristol Biomedical Research Centre

If individual patient data are available for all trials



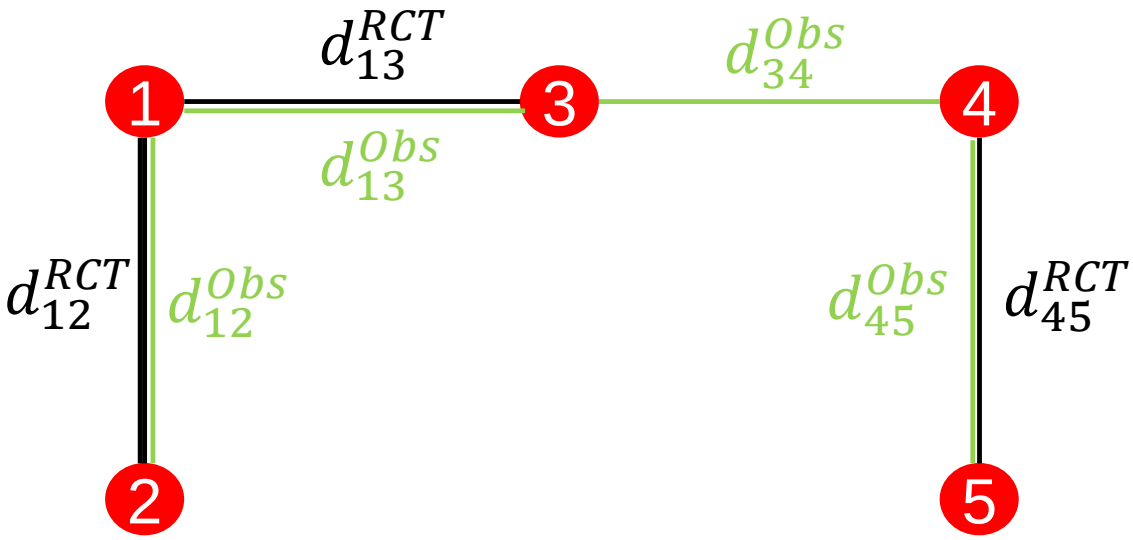
- If we have individual participant data (IPD) on trials can balance populations in three ways
  - Propensity score matching or weighting
  - Regression adjustment, which predicts outcomes in common populations
  - Doubly robust estimation (weighting + regression adjustment)
- These methods can only be implemented if we have IPD from all studies
- Rarely have IPD from all trials…

If individual patient data available for subset of trials



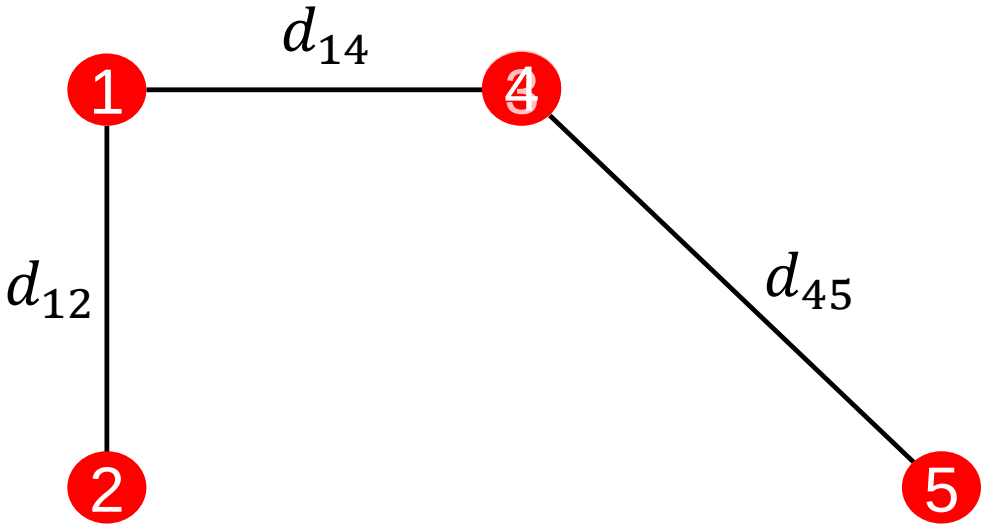
- Companies usually have IPD from at least the trial on their drug
- Then use IPD on treatment 1 to predict response in population of 4v5 RCT.
  - Propensity score reweighting: Matching Adjusted Indirect Comparison (MAIC)
  - Outcome Regression: Simulated Treatment Comparison (STC)
- However, governments and academics rarely have any IPD.

Non-randomized comparative evidence



- Could combine RCT with observational evidence
  - Retrospective registry studies (e.g. hip replacement surgery - Fawsitt 2019)
- Simple approach is to assume $d_{34}^{Obs} = d_{34}^{RCT} = d_{34}$
- Or if sufficient data use hierarchical models $d_{ab}^{RCT} \sim N(D_{ab}, \omega^2)$ and $d_{ab}^{Obs} \sim N(D_{ab}, \omega^2)$ (Schmitz 2013)
- However, cohort or registry studies not always available, especially for novel therapies.

Node merging or class effects models



- Or just assume treatments 3 and 4 are the same, perhaps as they are members of the same pharmacological class
  - Or related in a class effects model (e.g. Owen 2015)
- This doesn't solve the problem of comparing treatments 3 and 4 as there is still no evidence (e.g. comparing DOACs in atrial fibrillation)
- Maybe not justified to assume 4=3 (or 4=2, 4=1, 5=3, 5=2, or 5=1)

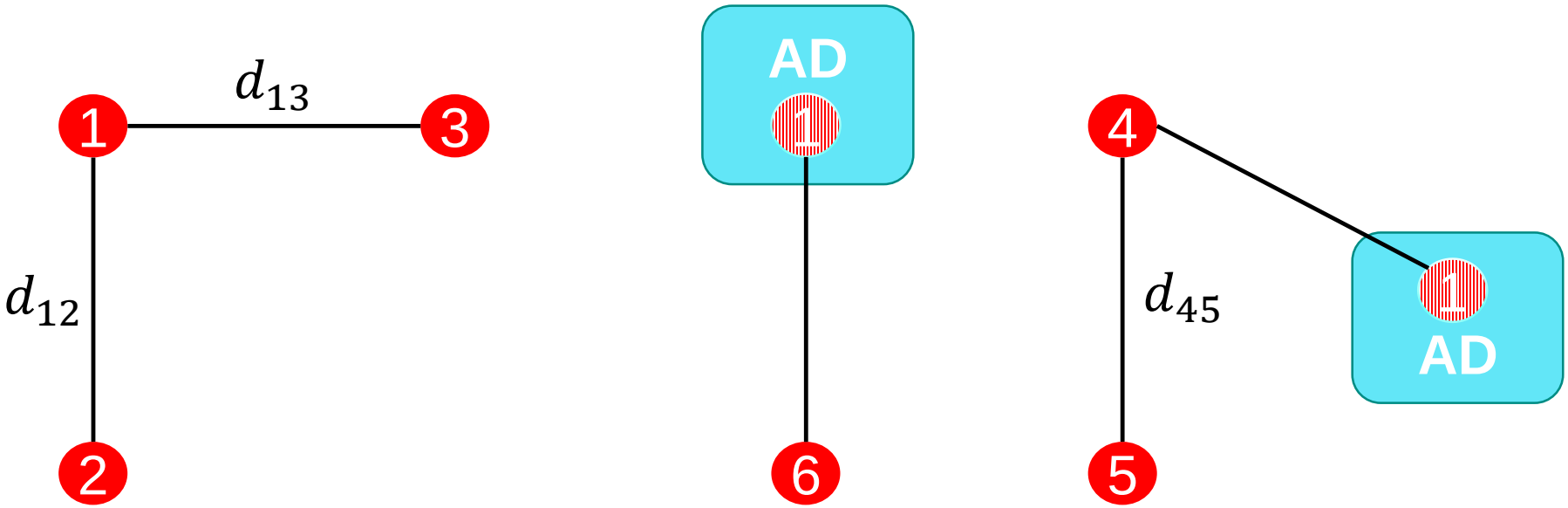
- No individual patient data
- No observational comparative evidence
- Can't merge treatments

What **can** you do?

We have explored two methods using only aggregate RCT data (AD)…

Connecting study i'

- Study i' is either a single arm or a disconnected RCT
- In both cases continue to assume μ_i independent in connected network, preserving randomization
- Generate $\mu_{i'}$ for single-arm study or disconnected RCT using
  - Reference prediction (RP) - *a refined random effects on baseline*
  - Aggregate level matching (ALM)



Could also predict treatment 2 or 3

Reference prediction (RP)

- On the connected component of the network, proceed with independent baselines NMA, avoiding interference with randomization
- Perform a meta-analysis of μ_i from RCTs with the reference treatment.
  - So these μ_i represent response on the same treatment
  - Keep the data separate from independent baselines NMA to avoid interference with randomization
- Fit a model with or without covariates

$$\mu_i \sim \mathrm{N}(\mathrm{m}, \sigma_\mu)$$
$$\mu_i \sim \mathrm{N}(\mathrm{m} + \beta x_i, \sigma_\mu)$$

- Predict response in disconnected RCT or single-arm study

$$\mu_{i'}^{pred} \sim \mathrm{N}(\mathrm{m}, \sigma_\mu)$$
$$\mu_{i'}^{pred} \sim \mathrm{N}(\mathrm{m} + \beta x_{i'}, \sigma_\mu)$$

Aggregate level matching (ALM)

- Choose the best matching RCT i with any baseline treatment t_{ib}
  - e.g. Euclidean distance on age, gender, baseline severity
- Fit a standard independent baselines model first and use the mean μ_i from that best matching RCT
- Make it less precise by modelling

$$\mu_{i'}^{plug-in} \sim \mathrm{N}\left(\mu_i, \sigma_{\mu_i}\right)$$

- Where σ_{μ_i} is SD of estimated μ_i
- Randomization again preserved in connected portion as independent baseline NMA used.

Random effects on treatment

- Simulation studies and artificial data examples found fixed effects to give precise estimates but with poor coverage, we therefore recommend random treatment effects
- Recall random effects models

$$\delta_{i,bk} \sim Normal(d_{t_{ib}t_{ik}}, \sigma)$$

- For a study comparing treatments 1 and 2, for example

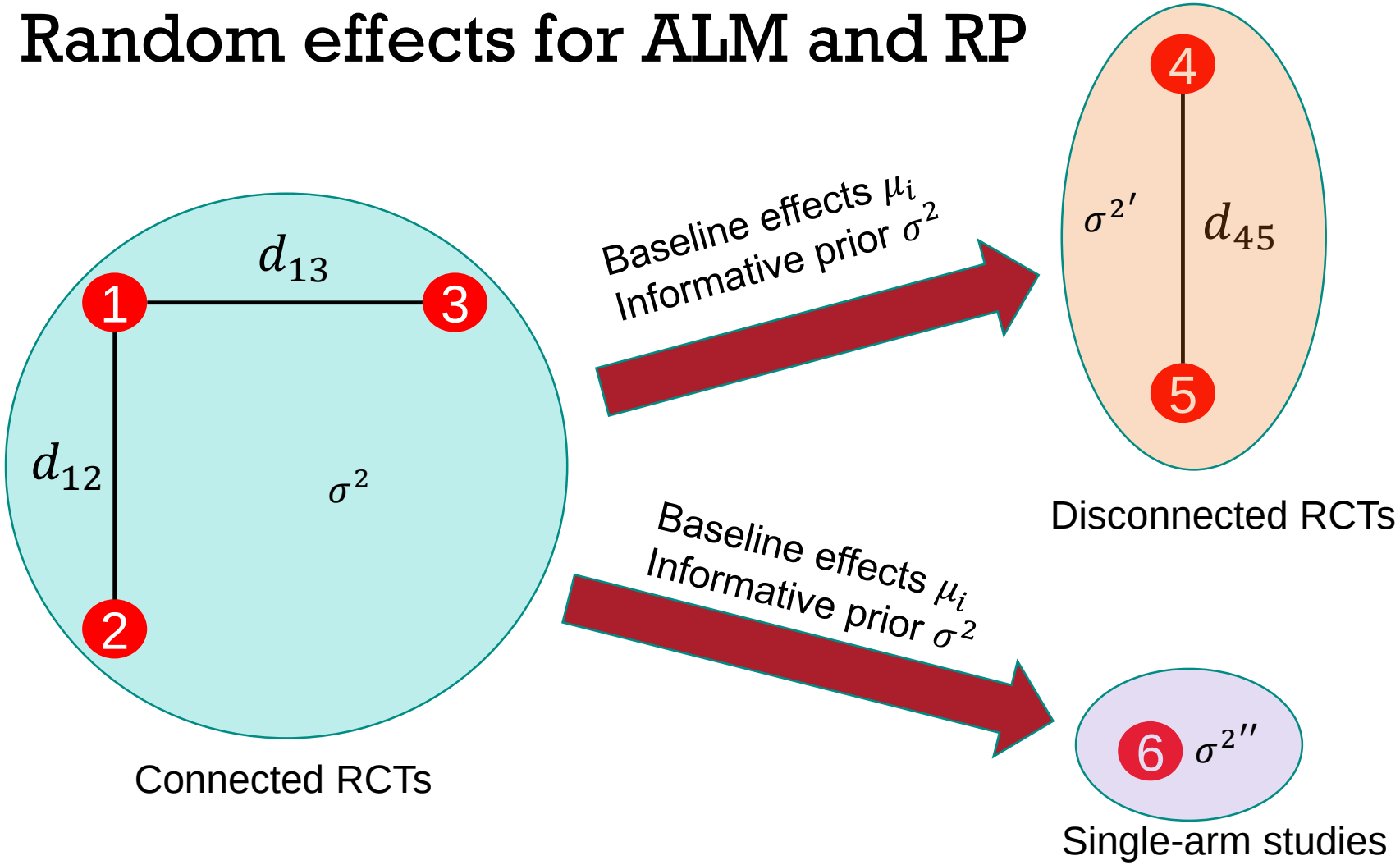
$$\delta_{i,12} \sim Normal(d_{12}, \sigma)$$

- The heterogeneity variance σ^2 represents extent of variation between study-level relative effects on each contrast (e.g. $\delta_{i,12}$)
- NMA commonly assumes same σ^2 for all contrasts as evidence sparse
  - Need at least two studies on one contrast for σ^2 to be identifiable

Random effects

- Two problems with assuming the same σ^2 for the connected RCTs, disconnected RCTs, and single-arm studies
  - Variation potentially different. Likely larger in single-arm studies
  - Allowing disconnected RCTs or single-arm studies to influence estimation of σ^2 in connected RCTs will interfere with randomization
- We therefore assume different σ^2 for each
- Can overcome identifiability issues by using σ as an informative prior for σ'
  - Or by using Turner informative priors

Random effects for ALM and RP



Multi-arm correction for standard NMA

- Trials with more than 2 arms have more than one relative effect $\delta_{i,bk}$
- They are correlated as they are all relative to the same baseline arm on treatment b.
- Under reference prediction for disconnected RCTs, this must be adapted as relative effects $\delta_{i,k}$ are always relative to the reference 1 and not a common baseline arm of the RCT. Similar correction applies to ALM
- The multivariate Normal distribution becomes (a_i is number of arms)

$$\overrightarrow{\delta_i} = \begin{pmatrix} \delta_{i,1} \\ \vdots \\ \delta_{i,a_i} \end{pmatrix} \sim N_{a_i-1}\left[\begin{pmatrix} d_{1,t_{i1}} \\ \vdots \\ d_{1,t_{ia_i}} \end{pmatrix}, \begin{pmatrix} \sigma^{2'} & \sigma^{2'}/_2 \dots & \sigma^{2'}/_2 \\ & & \vdots \\ \sigma^{2'}/_2 & \ddots & \sigma^{2'}/_2 \\ \vdots & & \\ \sigma^{2'}/_2 & \dots \sigma^{2'}/_2 & \sigma^{2'} \end{pmatrix} \right]$$

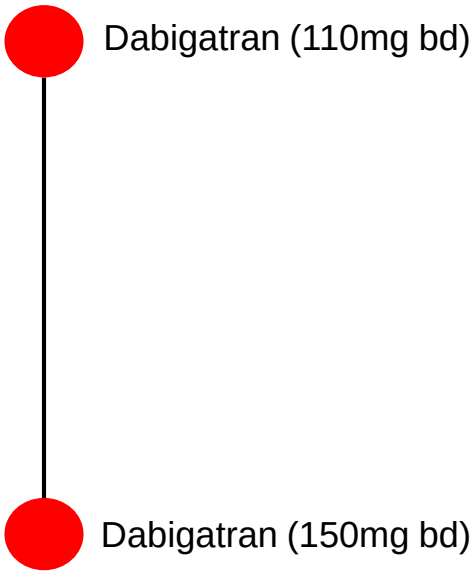
- There are a_i (rather than $a_i - 1$) relative effects
- Note that the disconnected RCT specific heterogeneity variance $\sigma^{2'}$ is being used.

Application to Atrial Fibrillation

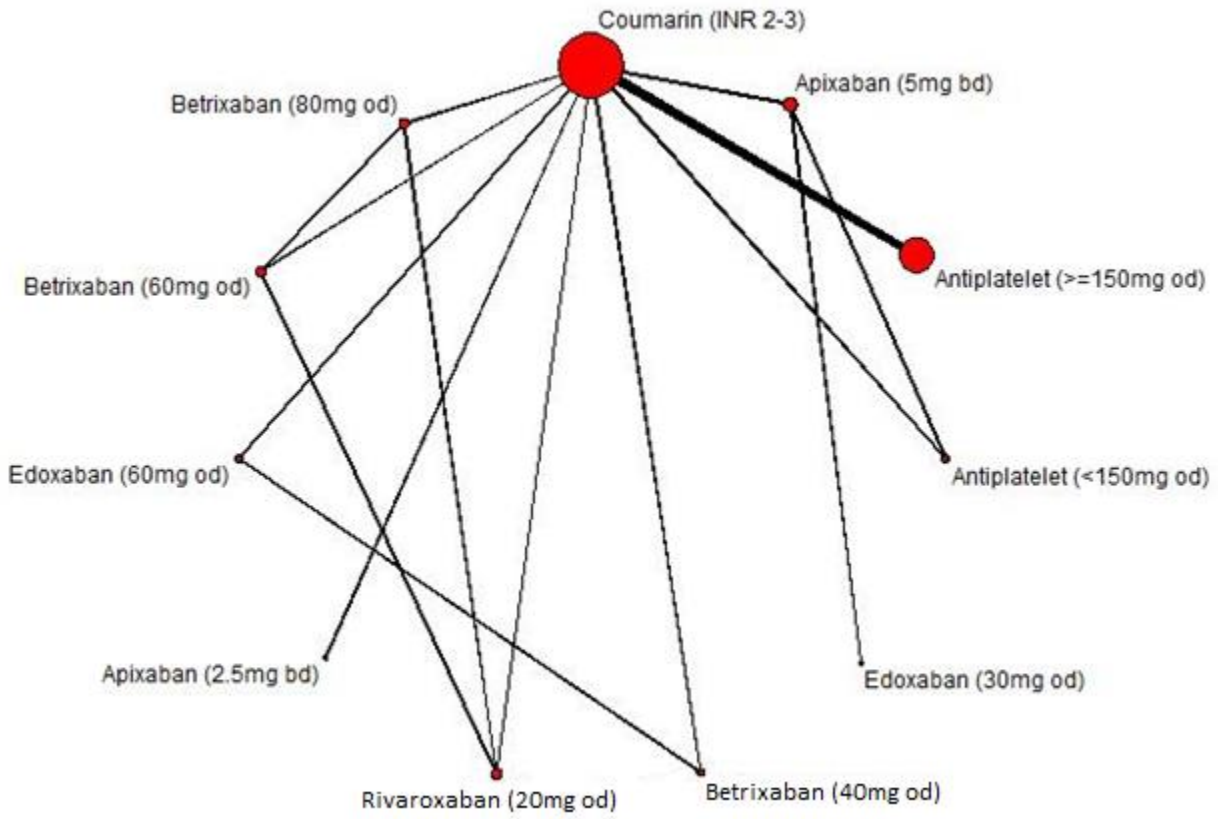
- Start with a single constructed example, then present a simulation study.
  - This is simple test to confirm our methods do what we expect
- Consider key outcome of ischaemic stroke
- Reference treatment was coumarin (INR 2-3), also called Warfarin.
- Network was connected but we will artificially disconnect it

Removing coumarin arms from dabigatran 110mg and 150mg RCT

Dabigatran subnetwork
(Only 1 RCT: RE-LY)



Non-dabigatran subnetwork



bristol.ac.uk

NIHR | Bristol Biomedical Research Centre

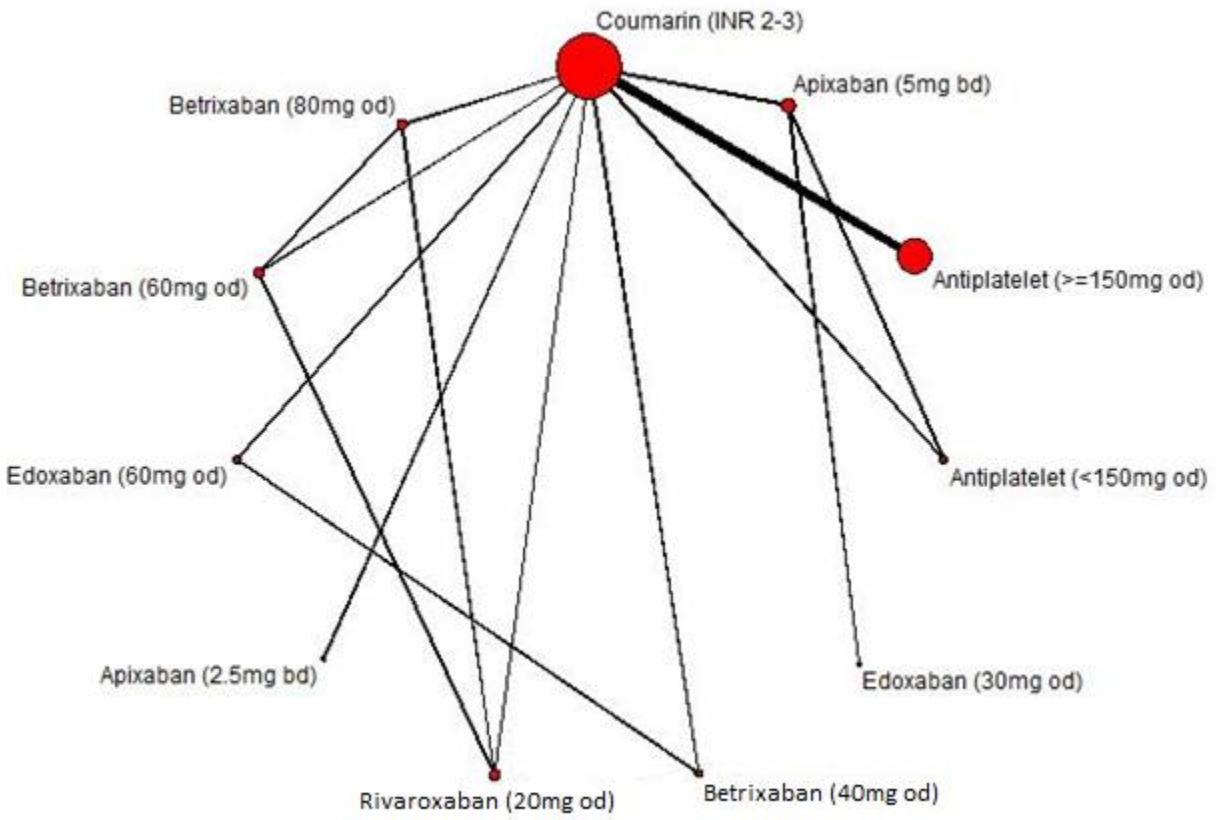
Removing coumarin arms from dabigatran 110mg and 150mg RCT

Dabigatran
(2 single-arm studies)

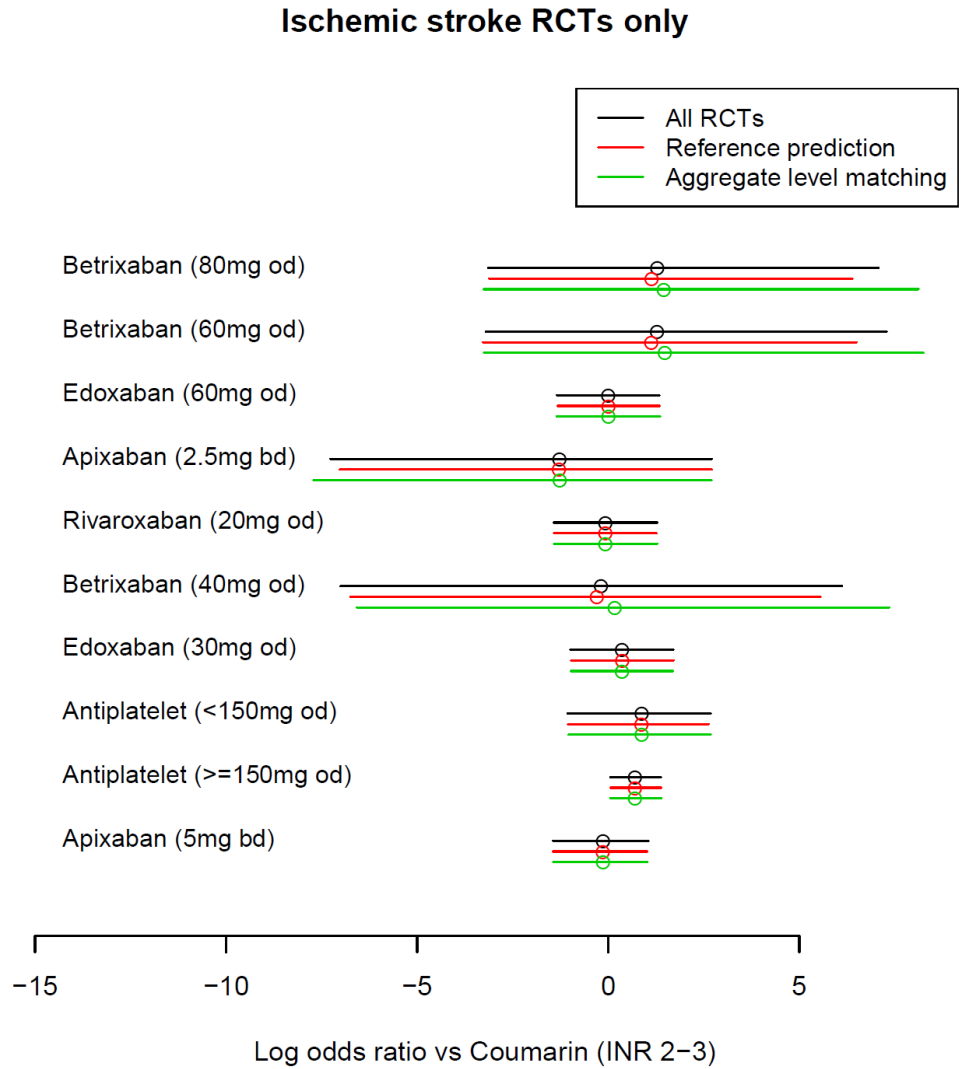
Dabigatran (110mg bd)

Dabigatran (150mg bd)

Non-dabigatran subnetwork

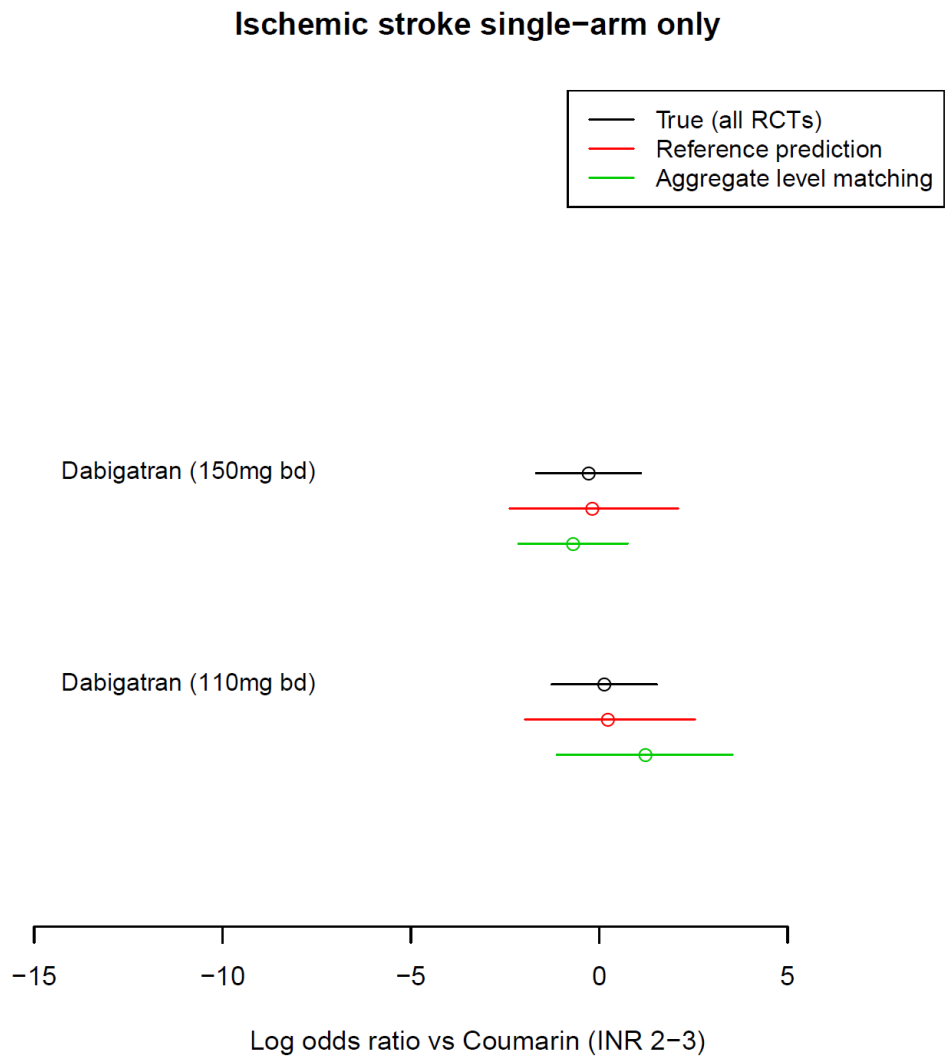


Connecting single-arm studies
Random effects



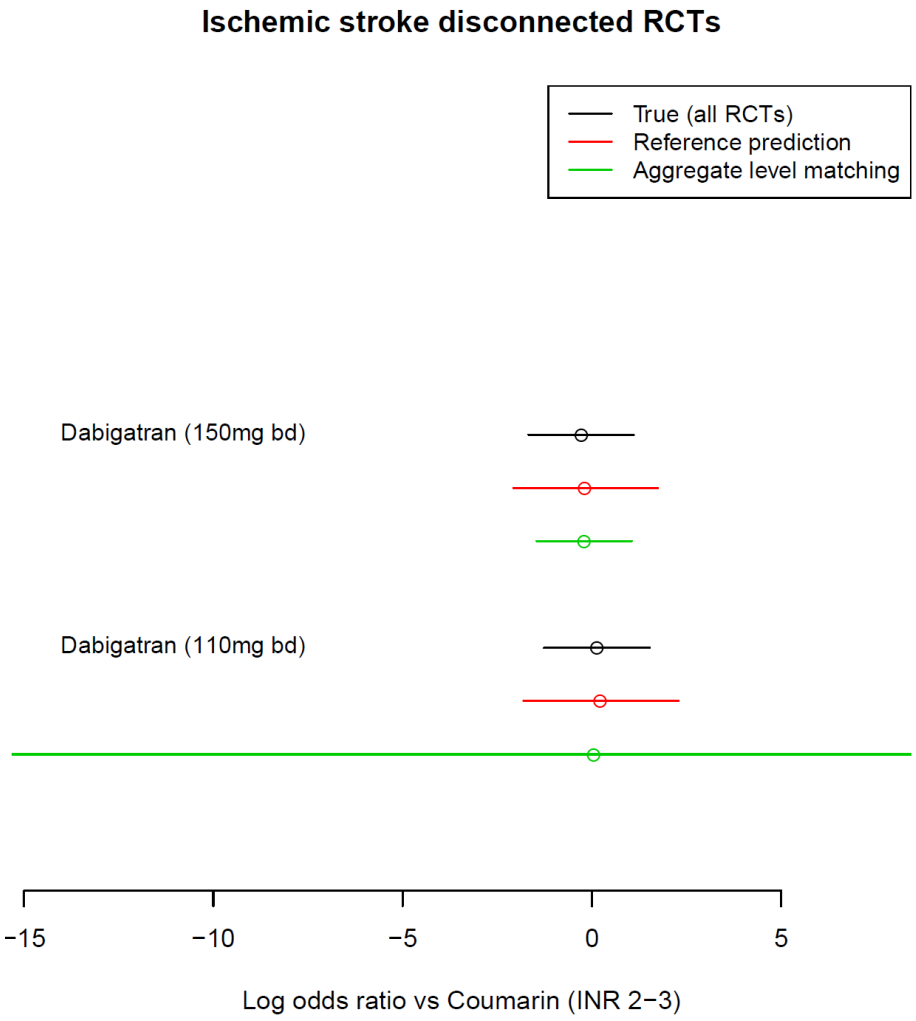
- First consider relative effects in only the connected (non-dabigatran) network
- Point estimates and uncertainty intervals match
- Safe to use

Connecting single-arm studies Random effects



- Reference prediction and ALM appear to have good coverage and are close to truth
- Fixed effects analyses similar but with tighter credible intervals for ALM

Connecting disconnected RCTs Random effects



- Point estimates close
- But credible interval for ALM too wide!
  - Be careful with choice of matching RCT
- Fixed effects similar but poor coverage for ALM

Summary so far

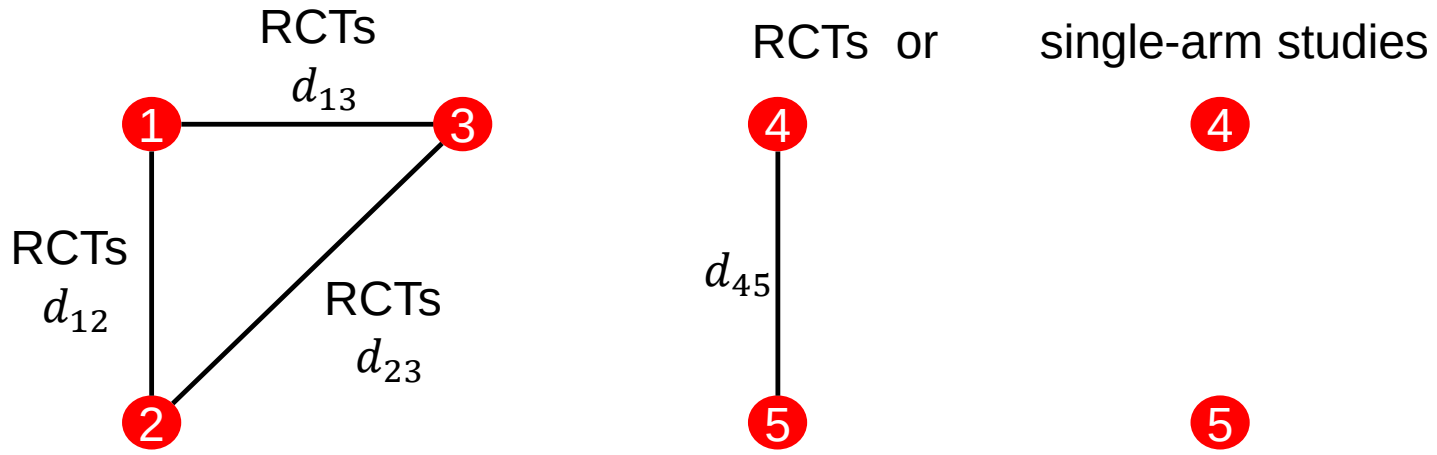
Findings

- Reference prediction and ALM can reproduce treatment effects using single-arm studies and disconnected RCTs and no IPD
- Both avoid interference with randomization in connected RCTs
- Reference prediction may be ‘safer’ as more conservative.
- Random effects requires assumptions on heterogeneity variance

Next step

- Simulation study to assess how ALM and RP would do in other situations.

Basic geometries to explore

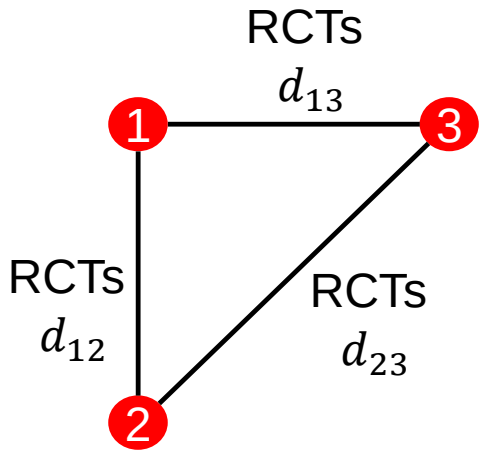


- We vary the number of connected RCTs but **not** the number of disconnected RCTs or single-arm studies.
- We want to vary evidence for reference prediction and ALM, which is only the connected RCTs.
- In all scenarios assume 5 disconnected RCTs (4 vs 5) or 10 single-arm studies (5 on treatment 4 and 5 on treatment 5)

Number of RCTs

Treatments to compare	1 vs 2	1 vs 3	1 vs 2 vs 3
Number of RCTs	2	2	1
	5	5	5
	20	20	10

- Three scenarios for number of RCTs on the connected network
- Expect reference prediction and ALM to improve as more data available on reference treatment.
- Size of trials fixed at 100 patients on each arm
  - Focus is difference across trials



Underlying model for baselines

- Model for simulated baseline response on log odds ratio scale

$$\mu_s \sim Normal(m + X_s\beta + I_{s=disc}\gamma, sd = 1)$$

- m is overall mean, β is covariate effect, X_s is covariate value
- Set $m \sim N(0.5,1)$ determining scale for other parameters
- X_s represents variation in treatment effect across studies

$$X_s \sim N(0.5,1) \text{ for each study}$$

- Different scenarios explored for β (next slide)
- $I_{s=disc}$ is indicator for s being disconnected or a single-arm
- γ is additional variation in baseline response in such studies.
  - Represents differences in prognostic variables between RCTs and disconnected RCTs or single arm studies.
  - Non-zero γ implies that RP and ALM will be biased
  - Consider two scenarios

$$\gamma = 0$$

and

$$\gamma \sim N(0.5,1)$$

Scenarios on prognostic variable

$$\mu_s \sim Normal(m + X_s\beta + I_{s=disc}\gamma, sd = 1)$$

- The β represents strength of prognostic variable
  - Stronger relation suggests better reference prediction or ALM
  - Only consider one prognostic variable
  - This is without loss of generality – unaccounted extras would be captured by m or γ while accounted extras would just be stronger β
- Weak vs strong prognostic covariate $\beta \sim N(0.1,1)$ or $\beta \sim N(1,1)$
  - Covariate assumed not to be an effect modifier. Effect modifiers are a problem for both connected and disconnected networks and our methods to not purport to overcome imbalance in effect modifiers.

Power calculations

- Used expected coverage and expected bias formulae from Morris 2019
- Coverage:
  - If 95% CrI includes 'truth' it is success, otherwise fail.
  - Report proportion of 'success', which is the *coverage probability*.
- Bias: $\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{d}_i - d$
- For $n_{sim} = 100$ MSE of bias 0.015
- For $n_{sim} = 100$ MSE of coverage 0.047
- These are acceptable MSE for coverage and bias but may increase to $n_{sim} = 1000$

Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019

Total number of scenarios

$$\begin{pmatrix} Fixed\ effects \\ Random\ effects \end{pmatrix} \times$$

$$\begin{pmatrix} Disconnected\ RCTs \\ Single-arm\ studies \end{pmatrix} \times \begin{pmatrix} \beta\ weak \\ \beta\ strong \end{pmatrix} \times \begin{pmatrix} \gamma\ zero \\ \gamma\ strong \end{pmatrix} \times \begin{pmatrix} 5\ connected\ RCTs \\ 15\ connected\ RCTs \\ 50\ connected\ RCTs \end{pmatrix}$$

- β is prognostic effect of covariate(s)
- γ is difference in prognostic factors between connected and disconnected components
- This gives 2x2x2x3 = 24 scenarios each for fixed and random effects.
- For each scenario, need to apply connected NMA, reference prediction, and ALM
  - 3 models fit for every scenario
  - Assess fixed and random effects separately.

Results
(Random effects, $n_{sim} = 100$)

Bias on connected ($n_{sim} = 100$)
$\gamma = 0$

Standard NMA	5 RCTs		15 RCTs		50 RCTs	
	β weak	β strong	β weak	β strong	β weak	β strong
RCT only	0.02	0.03	0.00	0.02	0.00	0.00
ALM single	0.03	0.03	0.00	0.02	0.00	0.00
ALM disconnected	0.02	0.04	0.00	0.03	0.00	0.00
RP single	0.03	0.04	0.00	0.03	0.00	0.00
RP disconnected	0.03	0.04	0.00	0.03	0.00	0.00

- $\gamma = 0$ means baseline response in connected RCTs is similar to that in disconnected RCTs and single-arm studies
- Results on log odds ratio scale with true mean d=0.5
- Bias may be zero as MSE is 0.015*
- Bias on connected component largely agrees

*Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019

Bias on disconnected ($n_{sim} = 100$)
$\gamma = 0$

	5 RCTs		15 RCTs		50 RCTs	
	β weak	β strong	β weak	β strong	β weak	β strong
ALM single	0.06	0.03	0.07	0.13	-0.01	-0.06
ALM disconnected	0.15	0.21	-0.04	0.11	-0.10	0.04
RP single	0.02	0.08	0.08	-0.07	0.03	-0.01
RP disconnected	0.00	-0.07	-0.09	-0.15	0.07	0.01

- MSE is 0.015
- Much higher bias than in connected component (as expected)
  - Bias is 10-20% of the mean log odds ratios (d=0.5)
- Similar to fixed effects
- Can't consistently say that ALM or RP are better.
- Also can't say that strong covariates (β) give lower bias

Coverage on connected ($n_{sim} = 100$)
$\gamma = 0$

	5 RCTs		15 RCTs		50 RCTs	
	β weak	β strong	β weak	β strong	β weak	β strong
RCT only	0.99	0.99	0.98	0.96	0.96	0.95
ALM single	1.00	0.99	0.97	0.96	0.96	0.95
ALM disconnected	0.99	0.99	0.98	0.96	0.94	0.95
RP single	1.00	0.99	0.97	0.96	0.96	0.95
RP disconnected	1.00	0.99	0.97	0.96	0.96	0.96

- MSE is 0.047
- Coverage greater than for fixed effects
- Coverage on connected component largely agrees

Coverage on disconnected ($n_{sim} = 100$)
$\gamma = 0$

	5 RCTs		15 RCTs		50 RCTs	
	β weak	β strong	β weak	β strong	β weak	β strong
ALM single	0.73	0.77	0.57	0.55	0.52	0.53
ALM disconnected	0.86	0.86	0.68	0.74	0.73	0.74
RP single	0.89	0.93	0.91	0.88	0.95	0.92
RP disconnected	0.92	0.92	0.87	0.81	0.89	0.84

- MSE is 0.047 – so again maybe this is sufficient?
- RP has very good coverage
  - This is expected as the predictions are so vague
- ALM has worse coverage, in particular for single arm studies
  - However, more than twice as high as for fixed effects models

Bias on disconnected ($n_{sim} = 100$) γ strong

	5 RCTs		15 RCTs		50 RCTs	
	β weak	β strong	β weak	β strong	β weak	β strong
ALM single	0.19	0.25	0.24	0.22	0.22	0.29
ALM disconnected	0.03	0.13	0.15	0.30	0.04	0.20
RP single	0.30	0.17	0.26	0.21	0.19	0.21
RP disconnected	0.19	0.21	0.23	0.14	0.14	0.09

- Much higher bias than in $\gamma = 0$ case
- Now around 50% of mean true log odds ratio (d=0.5)
- This is because this noise in prognostic factors and effect modifiers can't be modelled by ALM or RP.
- Similar to fixed effects

Coverage on disconnected ($n_{sim} = 100$)
γ strong

	5 RCTs		15 RCTs		50 RCTs	
	β weak	β strong	β weak	β strong	β weak	β strong
ALM single	0.69	0.65	0.52	0.52	0.43	0.39
ALM disconnected	0.82	0.84	0.74	0.74	0.66	0.68
RP single	0.82	0.87	0.77	0.75	0.79	0.78
RP disconnected	0.84	0.85	0.70	0.73	0.77	0.76

- MSE is 0.047
- ALM has worse coverage than in $\gamma = 0$ case
  - But twice that in fixed effects models
- RP is slightly worse in single-arm case but otherwise reasonably high.
- Coverage gets worse if there are more RCTs, since predictions are closer to the connected evidence which is (non-zero γ) systematically different.

Simulation study

Findings

- Bias depends on whether the baseline response is similar to connected and disconnected components.
  - Ranged from 10% (good) to 50% (bad), although even 50% gives some indication of treatment effect size.
  - Can't model this unfortunately, but maybe get clinical advice?
- Reference prediction has good coverage in all cases, because predictions are so uncertain
- ALM has poorer coverage in all cases
- Can't say if stronger covariates or more data improve performance.
  - Unsure if extra simulations will resolve this, given the MSE is already low.

Closing remarks

Findings

- Reference prediction and ALM can reproduce treatment effects using single-arm studies and disconnected RCTs
  - ALM poorer coverage but bias can be low
  - RP better coverage and similar bias
- Both avoid interference with randomization in connected RCTs
- Recommend cross-validation to assess reference prediction or ALM
- Reference predictio may be ‘safer’ as more conservative.

Next steps

- In simulation study, maybe vary trial sizes and more simulations?
- Could use external information to inform the reference prediction models?
- Develop IPD models?

Remember

- ALM and reference prediction are methods of last resort for decision makers. High quality RCTs, or at least access to IPD, are still needed.

References

Combining RCT and observational evidence

- Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. Statistics in Medicine. 2013; 32(17):2935-49.
- Fawsitt CG, Thom HHZ, Hunt LP, Nemes S, Blom AW, Welton NJ, et al. Choice of Prosthetic Implant Combinations in Total Hip Replacement: Cost-Effectiveness Analysis Using UK and Swedish Hip Joint Registries Data. Value Health. 2019; 22(3):303-12.

Class effect models

- Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. Value Health. 2015; 18(1):116-26.

Random baseline effects NMA in PAH

- Thom H, Capkun G, Cerulli A, Nixon R, Howard L: Network meta-analysis combining individual patient and aggregate data from a mixture of study designs with an application to pulmonary arterial hypertension. BMC Medical Research Methodology 2015 15:34 DOI: 10.1186/s12874-015-0007-0

Early use of random baseline effects in pairwise meta-analysis

- Li Z, Begg CB: Random Effects Models for Combining Results from Controlled and Uncontrolled Studies in a Meta-Analysis. *Journal of the American Statistical Association* 1994, 89.

Criticism of random baseline effects

- Senn S: Hans van Houwelingen and the Art of Summing up. *Biometrical Journal* 2010, 52:85-94.
- Dias S, Ades T: Absolute or relative effects? Arm-based sythesis of trial data. Research Synthesis Methods 2016, 7 23–28

Guidance on baseline natural history models

- Dias S, Welton N, Sutton A, Ades A: NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2012; http://www.nicedsu.org.uk. *National Institute for Health and Care Excellence* 2012.

Exploration of consequences of random baseline effects in NMA

- Beliveau, A. Goring, S. Platt, R. Gustafson, P. Network Meta-Analysis of Disconnected Networks: How Dangerous are Random Baseline Treatment Effects? Research Synthesis Methods 2017. Accepted

Aggregate Level Matching/Optimal Matching

- Rosenbaum, P. Optimal Matching for Observational Studies. *Journal of the American Statistical Association*. Vol 84, no. 408, 1989.
- Leahy J, Thom H, Jansen J, et al (2019). Incorporating Single Arm Evidence into a Network Meta-Analysis Using Aggregate Level Matching: Assessing the Impact. Statistics in Medicine. https://doi.org/10.1002/sim.8139

Designing simulation studies

- Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019;

Acknowledgments

- *This study was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at the University Hospitals Bristol National Health Service (NHS) Foundation Trust and the University of Bristol, and the NIHR Manchester BRC.*
- *Funding for the DOACs NMA was provided by NIHR HTA grant 11/92/17*

NIHR | Bristol Biomedical Research Centre

Thank you!