MRC Clinical Trials Unit

Smarter studies
Global impact
Better health

UCL

A comparison of arm-based and contrast-based approaches to network meta-analysis (NMA)

Ian White <ian.white@ucl.ac.uk>
MRC Clinical Trials Unit at UCL

Cochrane Statistical Methods Group Webinar
14th June 2017

Background

- The choice between arm-based and contrast-based NMA was until recently fairly clear
- Recent work by Hwanhee Hong and others, working with Brad Carlin, has promoted a new concept of arm-based NMA
- There has been heated discussion over pros and cons of this new approach

- I'll set out my understanding of the key issues. Aims:
  - to find some terminology that we can all agree on
  - to recognise similarities and differences, strengths and weaknesses of both approaches
- I'll use well-known data to clarify ideas, and artificial data to illustrate what the methods can do in principle

Plan

1. **What are arm-based and contrast-based NMA?**
2. Models and their key features
3. Breaking randomisation
4. Missing data aspects
5. Estimands
6. Summary

Smoking data (yawn)

study	design	dA	nA	dB	nB	dC	nC	dD	nD
1	ACD	9	140	.	.	23	140	10	138
2	BCD	.	.	11	78	12	85	29	170
3	AB	79	702	77	694	.	.	.	.
4	AB	18	671	21	535	.	.	.	.
5	AB	8	116	19	146	.	.	.	.
6	AC	75	731	.	.	363	714	.	.
7	AC	2	106	.	.	9	205	.	.
..									
20	AD	0	20	.	.	.	.	9	20
21	BC	.	.	20	49	16	43	.	.
22	BD	.	.	7	66	.	.	32	127
23	CD	.	.	.	.	12	76	20	74
24	CD	.	.	.	.	9	55	3	26

successes and participants in arm A …

What are arm-based and contrast-based NMA?

- Term goes back to Salanti et al (2008)
  - Salanti G, Higgins JPT, Ades AE, Ioannidis JPA (2008) Evaluation of networks of randomized trials. Statistical Methods in Medical Research 17: 279–301.
- Arm-based: model the arm-level data
  - #successes + binomial likelihood; *or*
  - log odds of success + approximate Normal likelihood
- Contrast-based: model the contrasts (trial-level summaries; two-stage)
  - log odds ratio + approximate Normal likelihood
- Pros and cons are well known:
  - binomial likelihood for arm-based model is more accurate but usually requires BUGS analysis
  - approximate Normal likelihood for contrast-based model is less accurate but fast e.g. **`mvmeta`** in Stata

I'm going to call these arm-based and contrast-based likelihoods

Why the debate now?

- Hong et al use "arm-based" and "contrast-based" in a new way, referring to different model parameterisations
  - really, different models
  - Hong H, Chu H, Zhang J, Carlin BP (2016) A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods* 7: 6–22.
  - applies only to an arm-based likelihood
- Although much of their work also covers multiple outcomes in NMA, I am going to consider what their work says for a single outcome

Scope of this talk

- Arm-based likelihood
- Binary outcome with treatment effects measured by log odds ratios
- Bayesian analysis with Cochrane-based informative priors from Turner et al (2012)
  - Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT (2012) Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International journal of epidemiology* 41: 818–827.
- Assuming consistency

but all the ideas apply more generally

Plan

1. What are arm-based and contrast-based NMA?
2. **Models and their key features**
3. Breaking randomisation
4. Missing data aspects
5. Estimands
6. Summary

Notation

- Trials: $i = 1, \ldots, n$
- Treatments: $k = 1, \ldots, K$
- R_i = set of treatments included in trial i ("design")
- n_{ik} = number of participants in treatment arm k of trial i
- d_{ik} = number of events in treatment arm k of trial i
  - $d_{ik} \sim Bin(n_{ik}, \pi_{ik})$
- θ_{ik} = parameter of interest in treatment arm k of trial i
  - here the log odds, $\theta_{ik} = \log\left(\frac{\pi_{ik}}{1-\pi_{ik}}\right)$
- e.g. Smoking trial 1:

study	design	dA	nA	dB	nB	dC	nC	dD	nD
1	ACD	9	140	.	.	23	140	10	138

$i = 1,\ R_1 = \{A, C, D\}, d_{1A} = 9,\ n_{1A} = 140$, etc.

General notation for models

I'll use

- superscripts C and A for contrasts and arms
- i for trial; k, k' for treatments
- δ for study-specific parameters
  - hence $\delta^C_{ikk'}$ for contrasts, δ^A_{ik} for arms

I'm going to follow the meta-analysis convention that study-specific effects have mean μ and heterogeneity σ^2:

- contrast parameter $\delta^C_{ikk'}$ has mean $\mu^C_{kk'}$ and heterogeneity SD $\sigma^C_{kk'}$
- arm parameter δ^A_{ik} has mean μ^A_k and heterogeneity SD σ^A_k

I'll take treatment 1 as reference treatment for the NMA

  - but all models are symmetric

Model 1. Lu & Ades (2004) ("LA")

- For each study, denote a baseline treatment b_i
  - usually the first numbered
- Model for study i and treatment arm $k \in R_i$, $k \neq b_i$:

$$\theta_{ik} = \alpha_{iB} + \delta^C_{iBk}$$

  - "B" denotes the use of a study-specific baseline
  - α_{iB} is the log odds in the baseline treatment arm. I'll call it the "study intercept" (also "underlying risk" or "baseline risk")
  - α_{iB} are fixed effects of study
  - δ^C_{iBk} are random treatment effects

$$\delta^C_{iBk} \sim N(\mu^C_{1k} - \mu^C_{1b_i}, \sigma^{C2})$$

  - μ^C_{1k} is the "overall" log odds ratio between treatment k and treatment 1 (of primary interest)
  - σ^{C2} is the heterogeneity variance

Note on "fixed effects"

- "Fixed effects" here refers to a set of parameters that are unrelated to each other
  - as opposed to "random effects" where the parameters are modelled by a common distribution
  - standard statistical meaning of the term
- "Fixed effects" does NOT refer to a meta-analysis model that ignores heterogeneity
  - I'd call that the "common-effect" model
    - Higgins JPT, Thompson SG, Spiegelhalter DJ (2009). A re-evaluation of random-effects meta-analysis. *JRSSA 172*, 137–159.

Heterogeneity in the LA model

- σ^{C2} is the heterogeneity variance
- The above model assumes common heterogeneity variance σ^{C2} across all treatment contrasts
  - LA called this "homogeneous treatment variance"
  - so the heterogeneity is homogeneous!
  - I prefer "common heterogeneity variance"
- Non-common heterogeneity can be allowed:

$$\delta_{iBk}^{C} \sim N(\mu_{1k}^{C} - \mu_{1b_i}^{C}, \sigma_{b_i k}^{C2})$$

  - but tricky to estimate in practice
  - and need to consider "second order consistency"
    - Lu, G., & Ades, A. E. (2009). Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics*, *10*, 792–805.

- I’m now going to extend the LA model in 3 steps to bring us to Hong et al’s arm-based model

Model 2: "LAplus" model

- Avoid study-specific baselines
- $\theta_{ik} = \alpha_{i1} + \delta^{C}_{i1k}$ where $\delta^{C}_{i11} = 0$
  - study intercepts α_{i1} are fixed effects
  - model applies for all k: i.e. this model also describes outcomes in missing arms
  - but model statement in missing arms has no impact
- Now write $\boldsymbol{\delta}^{C}_{i} = (\delta^{C}_{i12}, \ldots, \delta^{C}_{i1K})$
  - model $\boldsymbol{\delta}^{C}_{i} \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$
- Common heterogeneity model: $\boldsymbol{\Sigma}^{C} = \sigma^{C2}\boldsymbol{P}$ where $\boldsymbol{P}$ has ones on the diagonal and halves off the diagonal
- This is only a re-parameterisation of the basic LA model
  - i.e. fit to the data is the same

Bringing in missing data?

- Hong et al claim "Although a standard MTC approach (e.g., Lu and Ades (2006)) models the observed data, we can gain additional information from the incomplete records"
- This is not true: if the missing data are ignorable then modelling the observed data y^{obs} is the same as modelling the complete data (y^{obs}, y^{mis})
- Hong et al's approach is "data augmentation": to draw samples from $(\theta|y^{obs})$, it is sometimes computationally convenient to draw samples from $(y^{mis}, \theta|y^{obs})$
  - Tanner MA, Wong WH (1987) The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* 82: 528–540.
  - NB causes slower mixing in MCMC

More convenient modelling?

Model 2: $\theta_{ik} = \alpha_{i1} + \delta_{i1k}^{C}$
$\boldsymbol{\delta}_i^C = \left(\delta_{i12}^C, \ldots, \delta_{i1K}^C\right)$
$\sim N(\boldsymbol{\mu}^C, \boldsymbol{\Sigma}^C)$

- Hong et al also say "Our own models can more easily and flexibly incorporate correlations between treatments and outcomes"
- I think this is true for non-common heterogeneity:
  - because we describe the heterogeneity parameters via a matrix $\boldsymbol{\Sigma}^C$, we just require $\boldsymbol{\Sigma}^C$ to be positive semi-definite
  - whereas the LA model must enforce "second order consistency" restrictions on the σ_{bk}^{C2}

Model 3 (CB): study intercepts α are random

- Model 2 was
  - $\theta_{ik} = \alpha_{i1} + \delta^{C}_{i1k}$ where $\delta^{C}_{i11} = 0$
  - $\boldsymbol{\delta}^{C}_{i} = (\delta^{C}_{i12}, \ldots, \delta^{C}_{i1K}) \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$
- Model 3 adds a model for the study intercepts: $\alpha_{i1} \sim N(\mu^{\alpha}_{1}, \sigma^{\alpha 2}_{1})$
  - random effects instead of fixed effects
  - again this goes right back to Lu & Ades (2004)
- This means that study intercepts in small studies are shrunk towards an overall mean
  - may gain precision
  - brings concerns about "between-study information" (see later)

Model 4 (AB): Hong's full arm-based model

- Model 3 was
  - $\theta_{ik} = \alpha_{i1} + \delta^{C}_{i1k}$ where $\delta^{C}_{i11} = 0$
  - $\alpha_{i1} \sim N(\mu^{\alpha}_{1}, \sigma^{\alpha 2}_{1})$
  - $\boldsymbol{\delta}^{C}_{i} = (\delta^{C}_{i2}, \ldots, \delta^{C}_{iK}) \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$
- Model 4 is the same plus correlation:
  - $(\alpha_{i1}, \boldsymbol{\delta}^{C}_{i}) \sim N(\boldsymbol{\mu}^{*}, \boldsymbol{\Sigma}^{*})$
- Hong et al parameterised it symmetrically:
  - $\theta_{ik} = \mu^{A}_{k} + \eta^{A}_{ik}$
  - μ^{A}_{k} are fixed effects representing overall mean log odds on treatment k
  - η^{A}_{ik} are mean-zero random effects
  - $\eta^{A}_{i} = (\eta^{A}_{i1}, \ldots, \eta^{A}_{iK}) \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{A})$
- Could have written $\boldsymbol{\theta}_{i} \sim N(\boldsymbol{\mu}^{A}, \boldsymbol{\Sigma}^{A})$

Key feature of model 4: treatment effects are related to study intercepts

Either way, the model has
- one parameter per treatment
- free variation between studies described by a $K \times K$ variance matrix

What's new in model 4?

- Model 4 is
  - $\theta_{ik} = \alpha_{i1} + \delta_{ik}^{C}$ where $\delta_{i1}^{C} = 0$
  - $(\alpha_{i1}^{A}, \boldsymbol{\delta}_{i}^{C}) \sim N(\boldsymbol{\mu}^{*}, \boldsymbol{\Sigma}^{*})$
- Treatment effects δ_{ik} are allowed to correlate with study intercepts α_{i1}
- This sort of model is used to relate treatment effects to underlying risk (baseline risk)
  - Sharp SJ, Thompson SG (2000) Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Stat Med* 19: 3251–3274.
  - Achana FA, Cooper NJ, Dias S, Lu G, Rice SJC, Kendrick D, Sutton AJ (2013) Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Stat Med* 32: 752–771.
- I think the proposal to use a model with treatment effect associated with reference-treatment mean to estimate an overall treatment effect is novel and deserves debate

Summary so far: models for θ_{ik}

Model	Study intercept		Study * treatment	
LA	α_{iB}	~ fixed	+ δ^{C}_{iBk} (0 if $k = b_i$)	$\delta^{C}_{iBk} \sim N(\mu^{C}_{1k} - \mu^{C}_{1b_i}, \sigma^{C2})$
LAplus	α_{i1}	~ fixed	+ δ^{C}_{i1k} (0 if $k = 1$)	$\boldsymbol{\delta}^{C}_{i} \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$
CB	α_{i1}	$\sim N(\mu^{\alpha}_{1}, \sigma^{\alpha 2}_{1})$	+ δ^{C}_{i1k} (0 if $k = 1$)	$\boldsymbol{\delta}^{C}_{i} \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$
AB	α_{i1}	see →	+ δ^{C}_{i1k} (0 if $k = 1$)	$(\alpha_{i1}, \boldsymbol{\delta}^{C}_{i}) \sim N(\boldsymbol{\mu}^{*}, \boldsymbol{\Sigma}^{*})$
or			δ^{A}_{ik}	$\boldsymbol{\delta}^{A}_{i} \sim N(\boldsymbol{\mu}^{A}, \boldsymbol{\Sigma}^{A})$

Treatment effects ($\boldsymbol{\mu}^{C}$ or $\boldsymbol{\mu}^{A}$) are fixed effects in all these models.
LAplus, CB and AB all allow non-common heterogeneity variance.

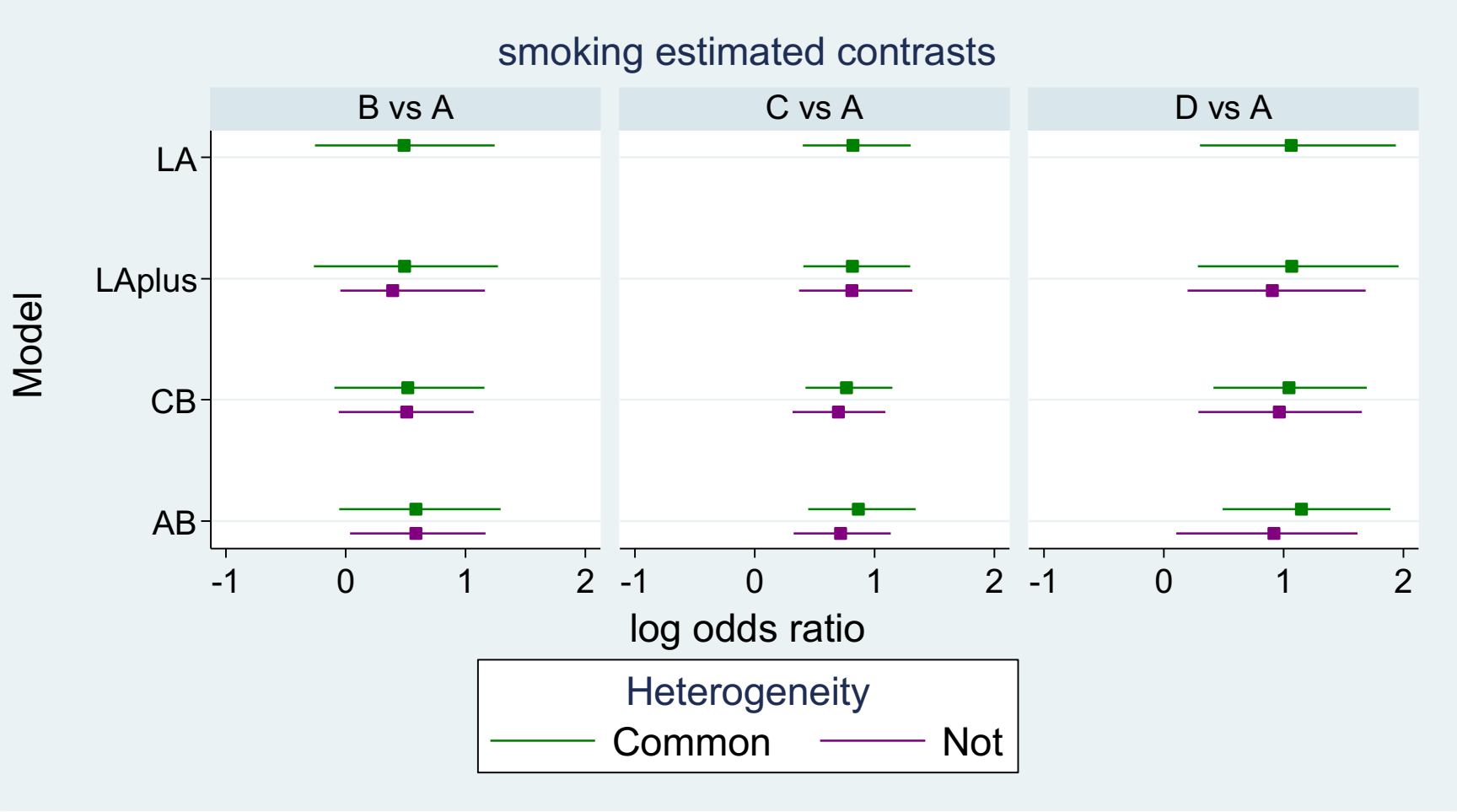
Common-heterogeneity models

Model	Study * treatment		Added assumption for common heterogeneity
LA	δ_{iBk}^{C}	$\delta_{iBk}^{C} \sim N(\mu_{1k}^{C} - \mu_{1b_i}^{C}, \sigma^{C2})$	none
LAplus	δ_{i1k}^{C}	$\boldsymbol{\delta}_{i}^{C} \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$	$\boldsymbol{\Sigma}^{C} = \sigma^{C2}\boldsymbol{P}$
CB	δ_{i1k}^{C}	$\boldsymbol{\delta}_{i}^{C} \sim N(\boldsymbol{\mu}^{C}, \boldsymbol{\Sigma}^{C})$	$\boldsymbol{\Sigma}^{C} = \sigma^{C2}\boldsymbol{P}$
AB	δ_{i1k}^{C}	$(\alpha_{i1}, \boldsymbol{\delta}_{i}^{C}) \sim N(\boldsymbol{\mu}^{*}, \boldsymbol{\Sigma}^{*})$	$\boldsymbol{\Sigma}^{C}$ part of $\boldsymbol{\Sigma}^{*} = \sigma^{C2}\boldsymbol{P}$ *
or	δ_{ik}^{A}	$\boldsymbol{\delta}_{i}^{A} \sim N(\boldsymbol{\mu}^{A}, \boldsymbol{\Sigma}^{A})$	$\boldsymbol{\Sigma}^{A} = \frac{1}{2}\sigma^{C2}\,\boldsymbol{I} + \sigma^{A2}\,\boldsymbol{J}$ (compound symmetry) *

where $\boldsymbol{P} = \begin{pmatrix} 1 & .5 & \cdots & .5 \\ .5 & 1 & \cdots & .5 \\ \vdots & \vdots & \ddots & \vdots \\ .5 & .5 & \cdots & 1 \end{pmatrix}$

* Hong et al used diagonal matrices here, or $\propto$ identity

Results: treatment effects μ^C



Results: heterogeneity SDs σ^C, σ^C_{kl}

Key points from this section

Key differences between Lu-Ades (LA) and arm-based (AB) models are

1. Study intercepts are random
2. Study*treatment effects (i.e. the random heterogeneity) are associated with the study intercepts (underlying risk)

An unimportant difference is

3. Arm-based models describe missing arms as well as observed arms

Should also remember

4. Going beyond common heterogeneity can be tricky in all models

Plan

1. What are arm-based and contrast-based NMA?
2. Models and their key features
3. **Breaking randomisation**
4. Missing data aspects
5. Estimands
6. Summary

Breaking randomisation / Between-study information

- A major concern about random study intercepts is that between-trial information is potentially used in the analysis
  - sometimes called "breaking randomisation"
  - Senn S (2010) Hans van Houwelingen and the Art of Summing up. *Biometrical Journal* 52: 85–94.
    "I consider that in practice little harm is likely to be done"
  - Achana FA, Cooper NJ, Dias S, Lu G, Rice SJC, Kendrick D, Sutton AJ (2013) Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Statistics in Medicine* 32: 752–771.

Artificial data sets

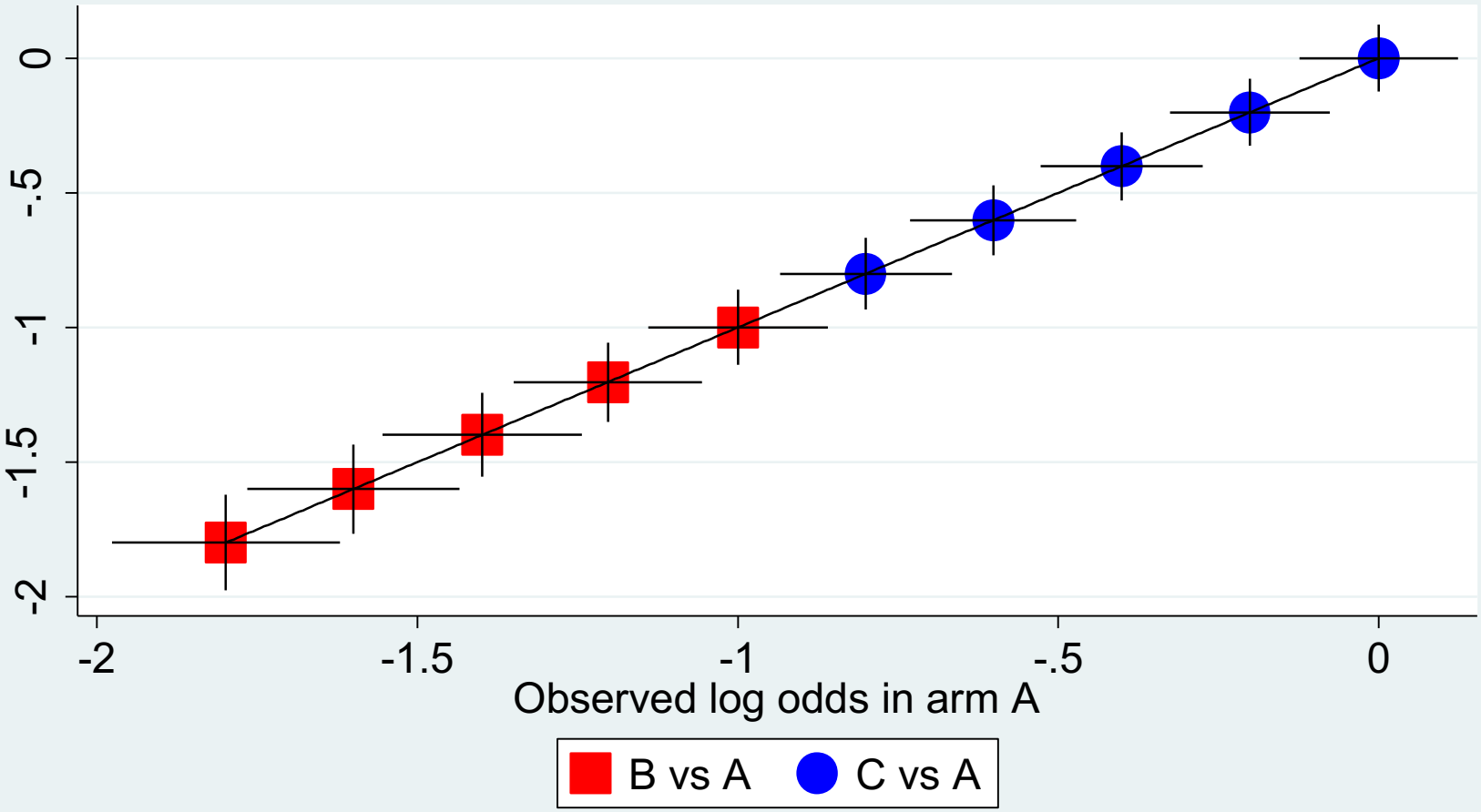
- I'm going to show analyses of artificial data sets chosen to explore what COULD go wrong
- I'll use simple NMAs of 5 A-B studies and 5 A-C studies
- A is reference
- Binary outcome

First example has

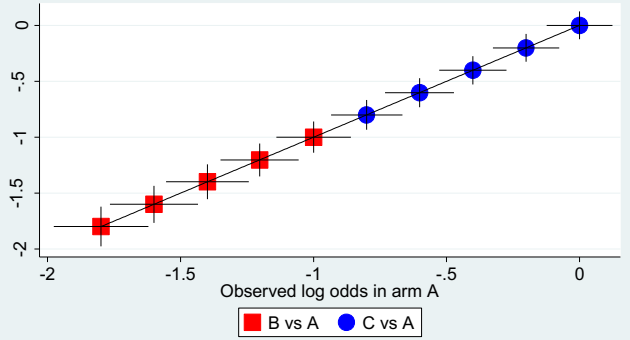
- A-B studies in low risk populations (low odds in arm A)
- A-C studies in high risk populations (high odds in arm A)
- No treatment effects at all
- This is extreme for AB models, because study intercepts in A-C studies will be pulled down and study intercepts in A-B studies will be pulled up
  - hence expect to see C > A> B

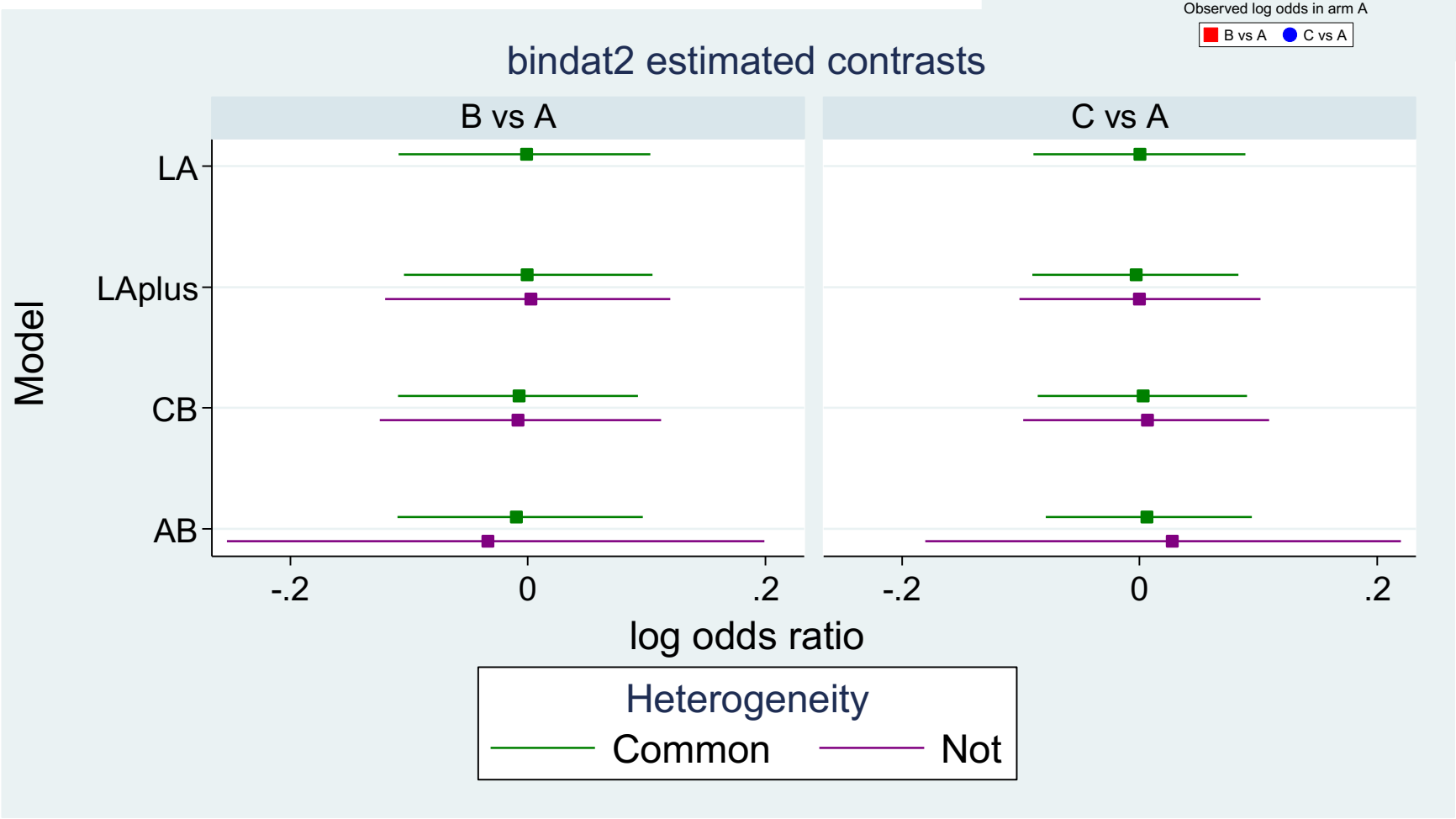
Artificial data 1

L'Abbe plot overlaying B vs A and C vs A
Cross-hairs are 95% CIs for arm-specific log odds
Diagonal is line of equality



Artificial data 1: results





AB with non-common heterogeneity suffers small bias of ± 0.03 (in fact all CB and AB have some tiny bias)

Key points from this section

1. Breaking randomisation is a theoretical problem, but seemingly not a practical problem

Should we be reassured, or is breaking randomisation a "face validity" issue?

Plan

1. What are arm-based and contrast-based NMA?
2. Models and their key features
3. Breaking randomisation
4. **Missing data aspects**
5. Estimands
6. Summary

Missing data aspects

- Again consider a network of treatments A, B and C
- Here we consider all studies as A-B-C studies
  - so C is a "missing arm" in an A-B study
- The problem is conceptually quite clear. If A-B studies differ systematically from A-C studies, say, then bias can occur especially in the B-C comparison.
- Question: does bias occur if A-B studies differ from A-C studies in
  - mean in treatment A?
  - the A-B or A-C treatment effects?
- It's also clear that the problem of missing arms is related to the problem of arm sizes
  - not having a C arm is an extreme case of an A-B-C study whose C arm is smaller than the A and B arms

Is NMA a missing data problem?

- e.g. back to the smoking data: study 1 has a missing B arm, but how many patients were (or weren't?) in it?
- Do we have missing n's as well as missing d's? (treating design features n's as "data"):

study	design	dA	nA	dB	nB	dC	nC	dD	nD
1	ACD	9	140	.	.	23	140	10	138

- Or do we simply have no participants?:

study	design	dA	nA	dB	nB	dC	nC	dD	nD
1	ACD	9	140	0	0	23	140	10	138

- Or do we know the size of the missing arm?:

study	design	dA	nA	dB	nB	dC	nC	dD	nD
1	ACD	9	140	.	140	23	140	10	138

A compromise

- I am going to proceed by assuming that we know the sizes of the missing arms, had they been observed
  - not a bad assumption in many NMAs where most trials randomise equally
  - but clearly not right and open to improvement

- I now ask: what assumptions are (implicitly) made about the missing data by the different models?
- Ignoring the missing data makes an implicit missing at random (MAR) assumption, but there are different sorts of MAR assumption

A contrast-based likelihood

- If our likelihood models contrasts y_{AB}, y_{AC} then our analysis is valid provided that y_{AB}, y_{AC} are MAR
- This means that the probability of particular arms being observed does not depend on the unobserved contrasts, given the observed contrasts
  - "contrast-MAR"
- E.g. for a study i of design AB,
  - $p(R_i = AB|\ y_{iAB}, y_{iAC}) = p(R_i = AB|\ y_{iAB})$

- Note: some authors claim contrast-MAR requires MCAR
  - this is true with all two-arm studies
  - not true in general with multi-arm studies

An arm-based likelihood

- If our likelihood models arm-specific outcomes d_A, d_B, d_C then our analysis is valid provided that d_A, d_B, d_C are MAR
- This means that the probability of particular arms being observed does not depend on the unobserved arm outcomes, given the observed arm outcomes
  - "arm-MAR"
- E.g. for a study i of design AB,
  - $p(R_i = AB \mid d_{iA}, d_{iB}, d_{iC}) = p(R_i = AB \mid d_{iA}, d_{iB})$

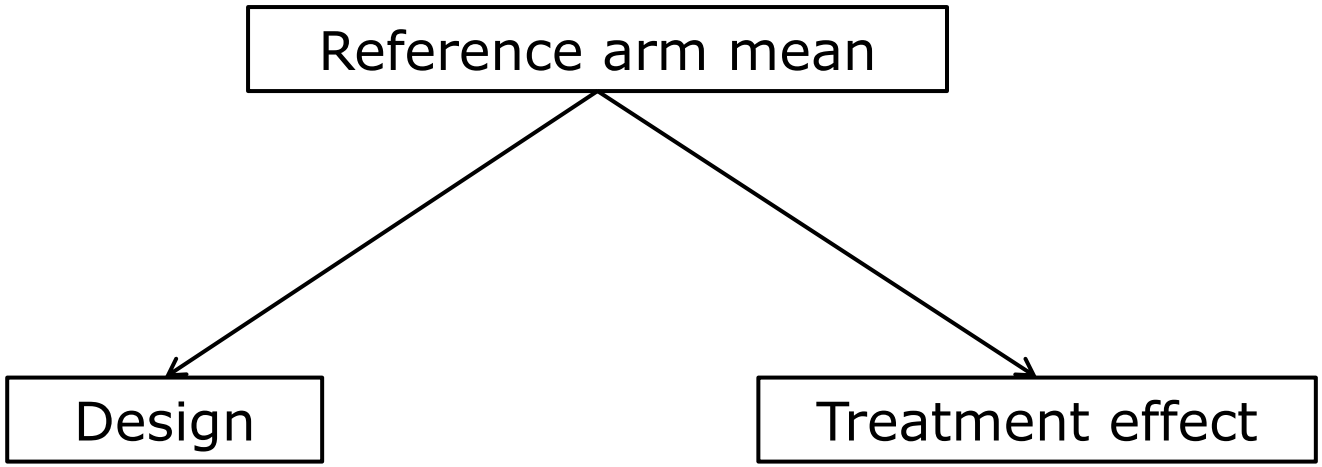
An example of data that are arm-MAR and not contrast-MAR

- Suppose all trials have an arm A
- Suppose (as in Artificial Data 1) that trials with low mean on arm A are more likely to have B as comparator, and trials with high mean on arm A are more likely to have C as comparator
  - and that no other aspect of the likely outcomes affects the design
- Then the data are arm-MAR, because design depends on arm A, which is observed in an arm-based likelihood
- But the data are not contrast-MAR, because arm A is unobserved in a contrast-based likelihood
- Whether bias occurs in a contrast-based likelihood depends on whether A-B or A-C treatment effect is also related to arm A outcome

Model mis-specification

- The above properties of validity under MAR only hold if models are correctly specified
- In particular, what happens if we use an arm-based likelihood to fit models 1-3?
  - i.e. models where the treatment effect is assumed independent of the study intercept?
- It turns out (tentatively) that this is like using a contrast-based likelihood
  - i.e. models 1-3 are only validly fitted under contrast-MAR

Exploration using more artificial data



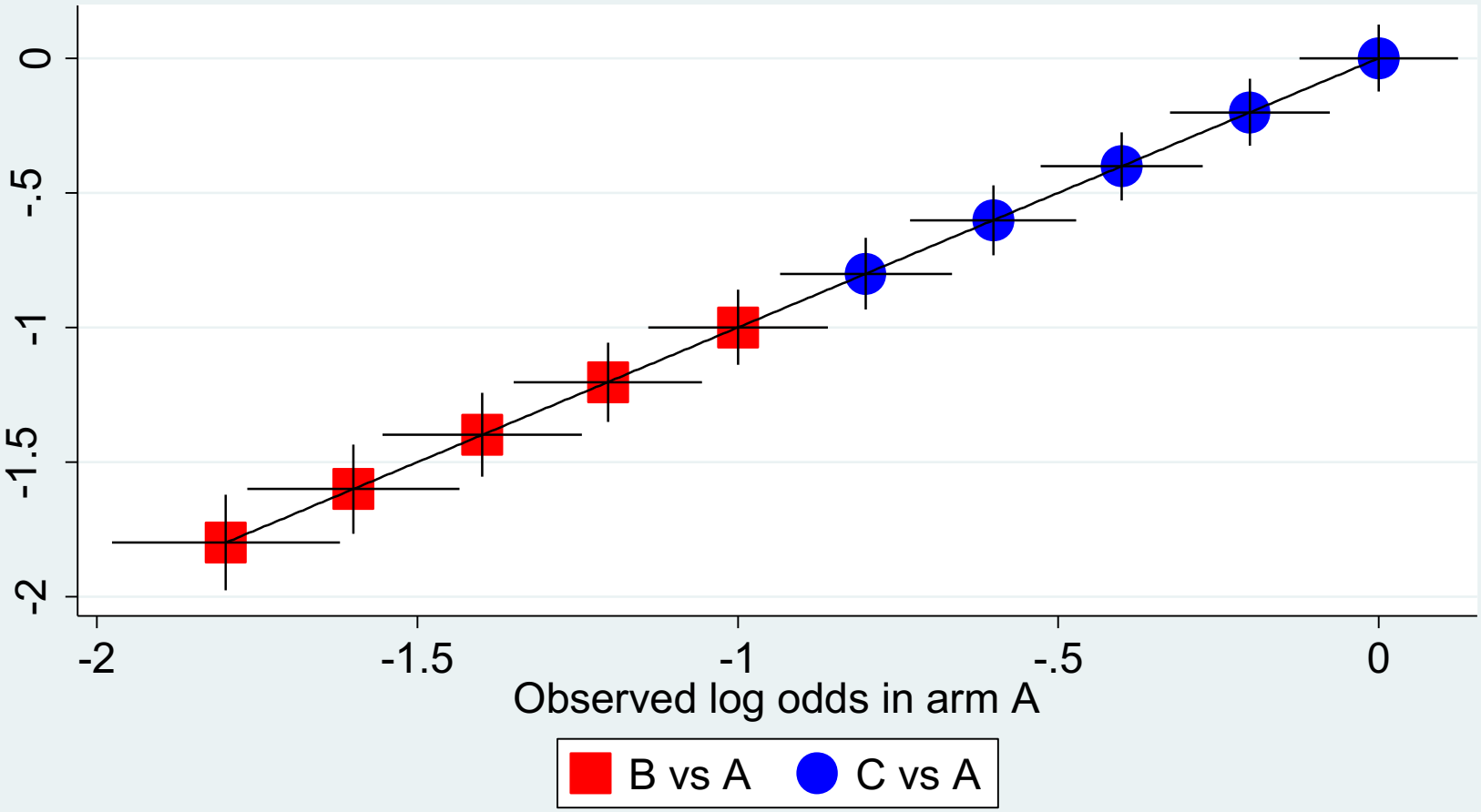
- Bias is likely to occur in models 1-3, if both the above arrows exist

Artificial data 1

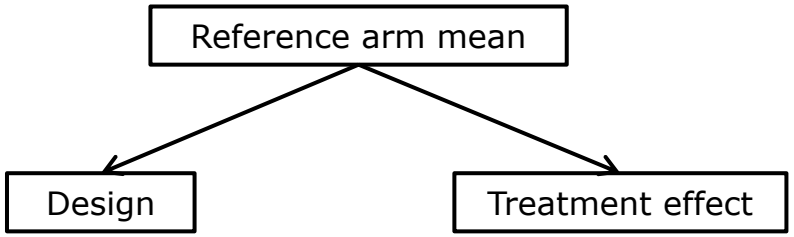
Reference arm mean

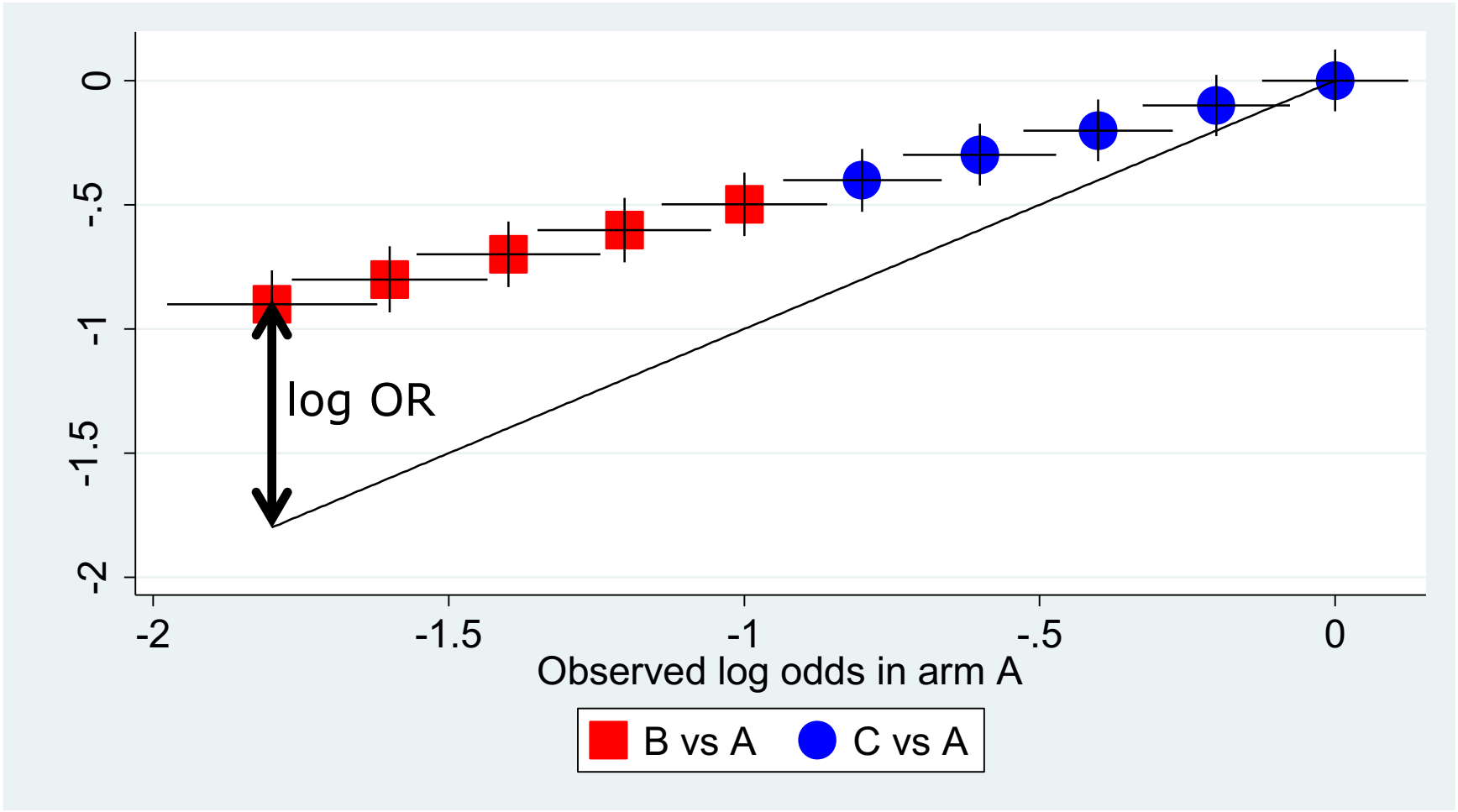
Design

Treatment effect



Artificial data 2



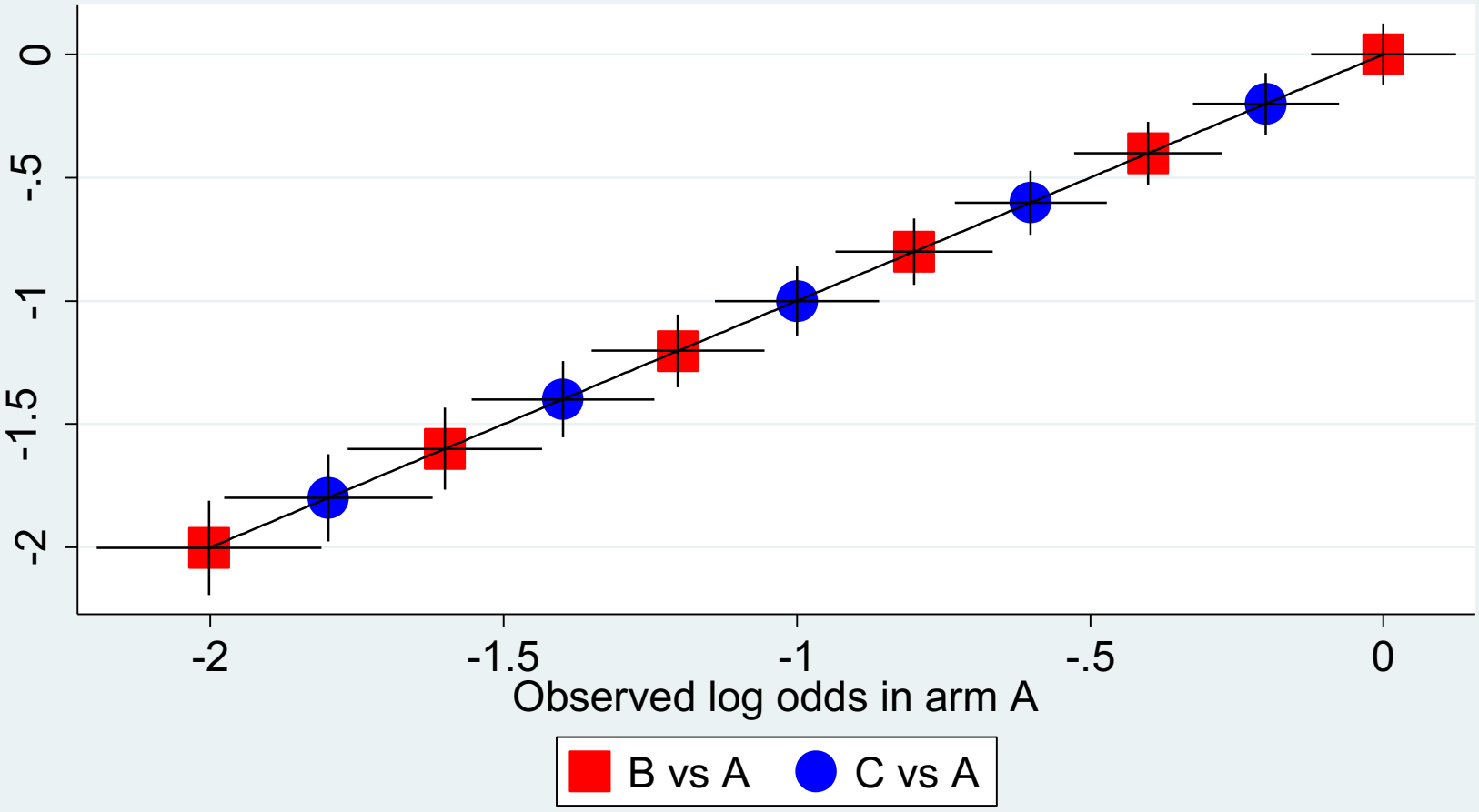


Reference arm mean

Artificial data 3

Design

Treatment effect

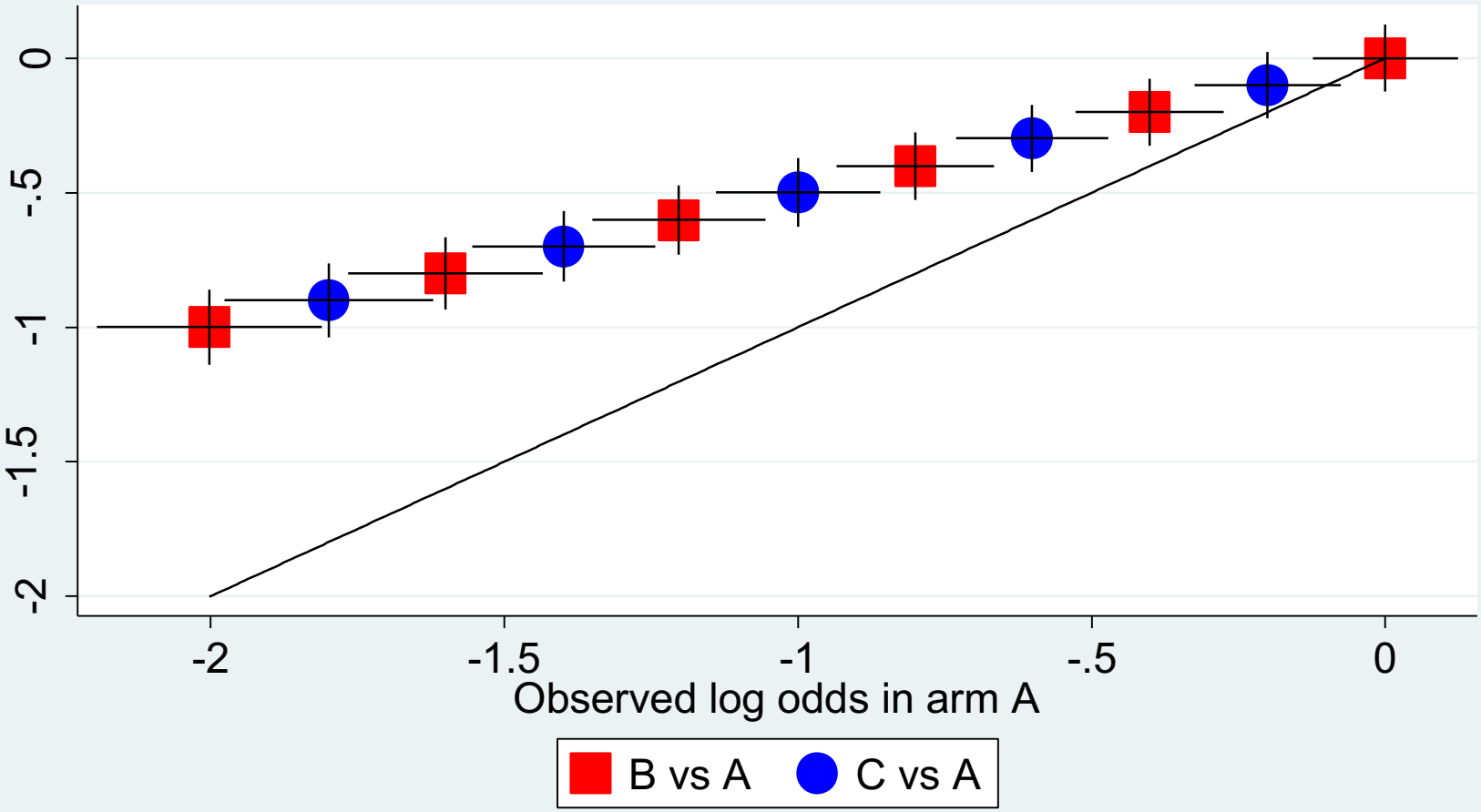


Artificial data 4

Reference arm mean

Design

Treatment effect

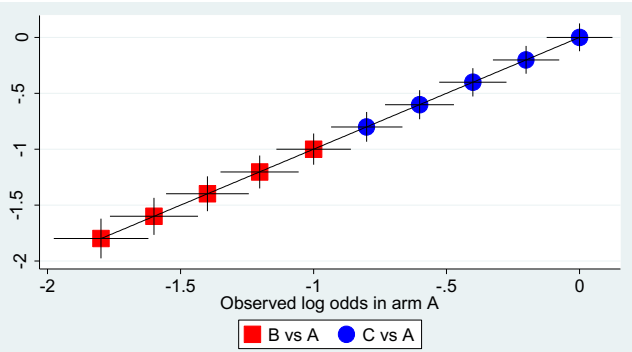


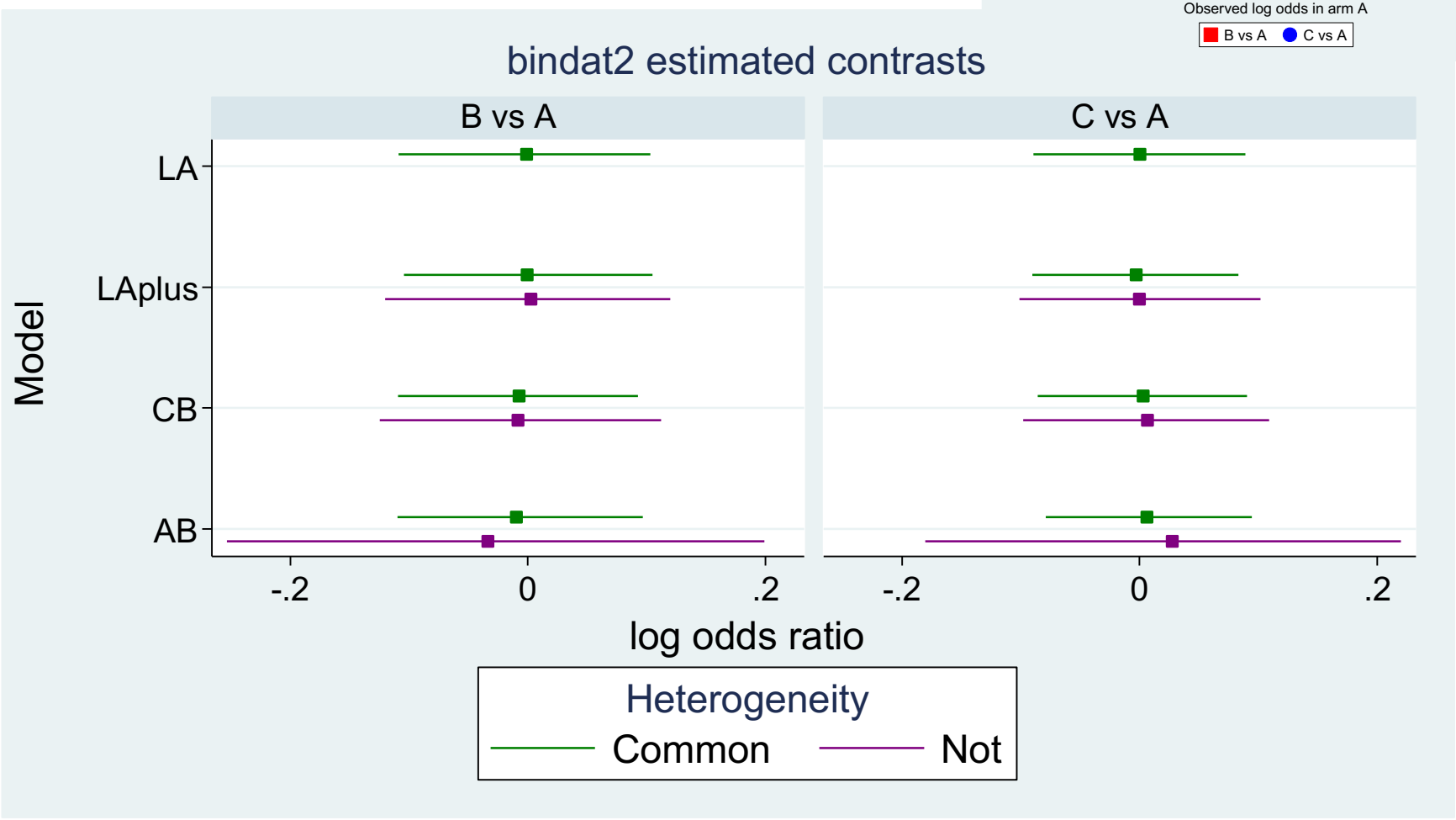
Artificial data 1

Reference arm mean

Design

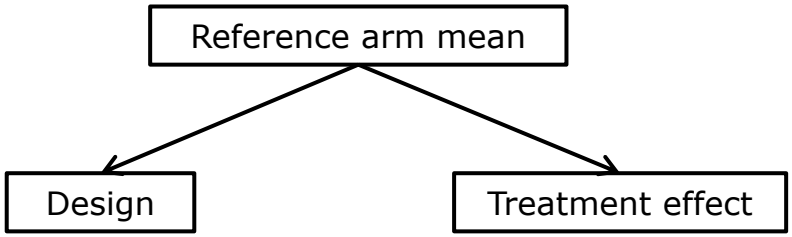
Treatment effect

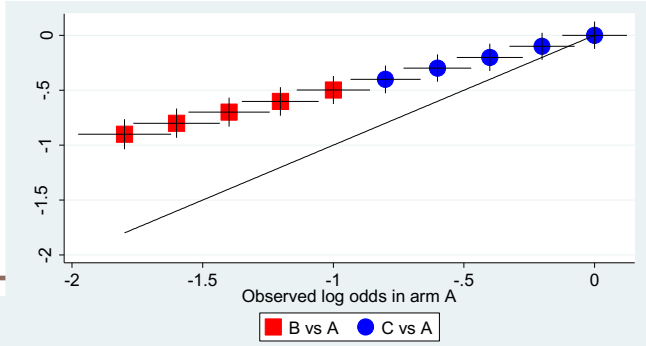


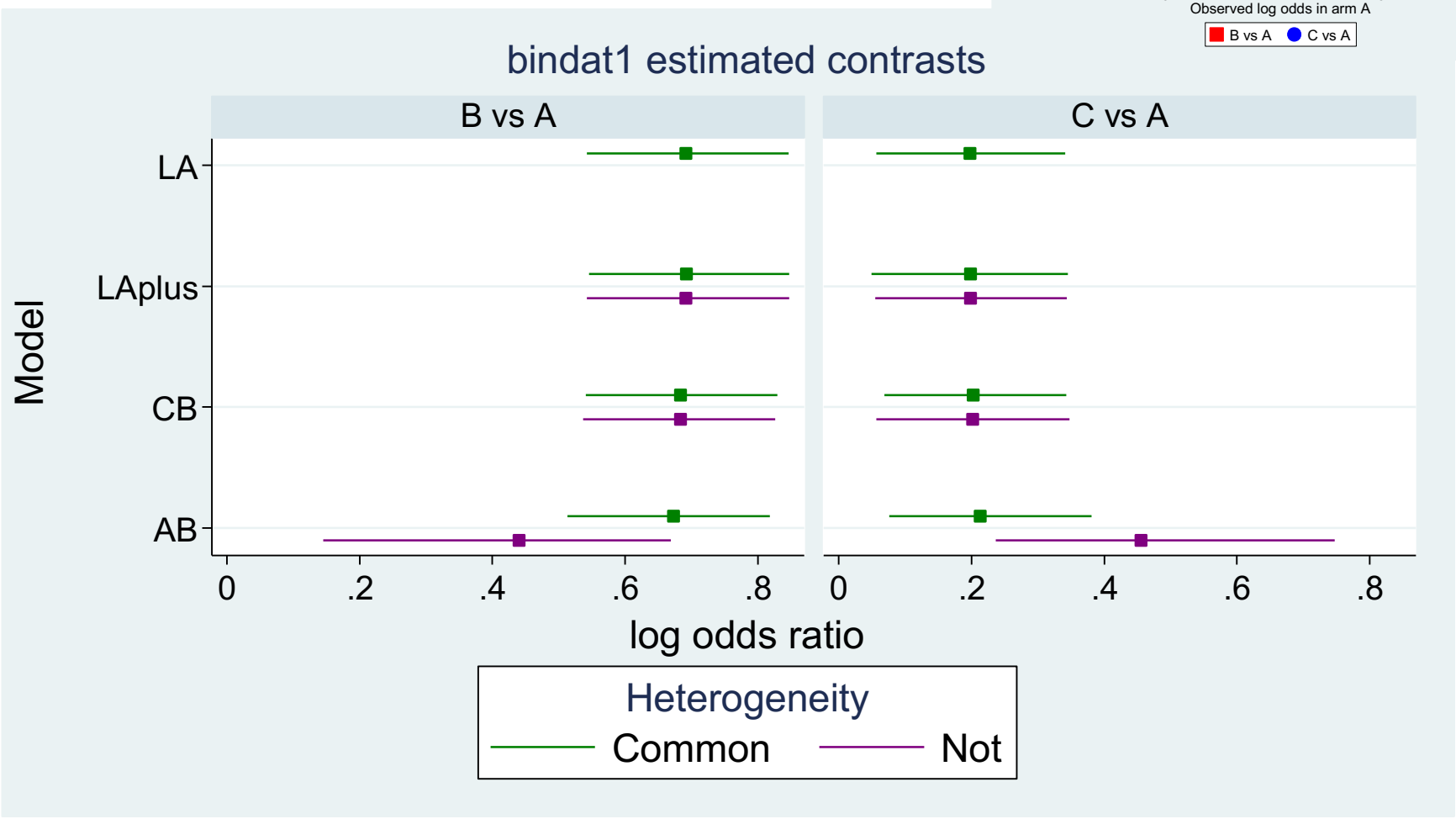


AB with non-common heterogeneity suffers small bias of ± 0.03 (in fact all CB and AB have some tiny bias)

Artificial data 2







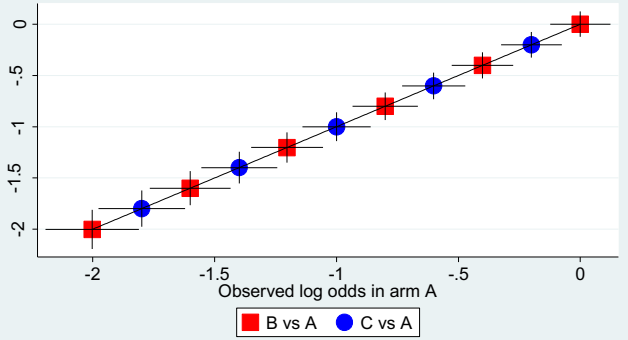
Only AB with non-common heterogeneity can see that B ≈ C

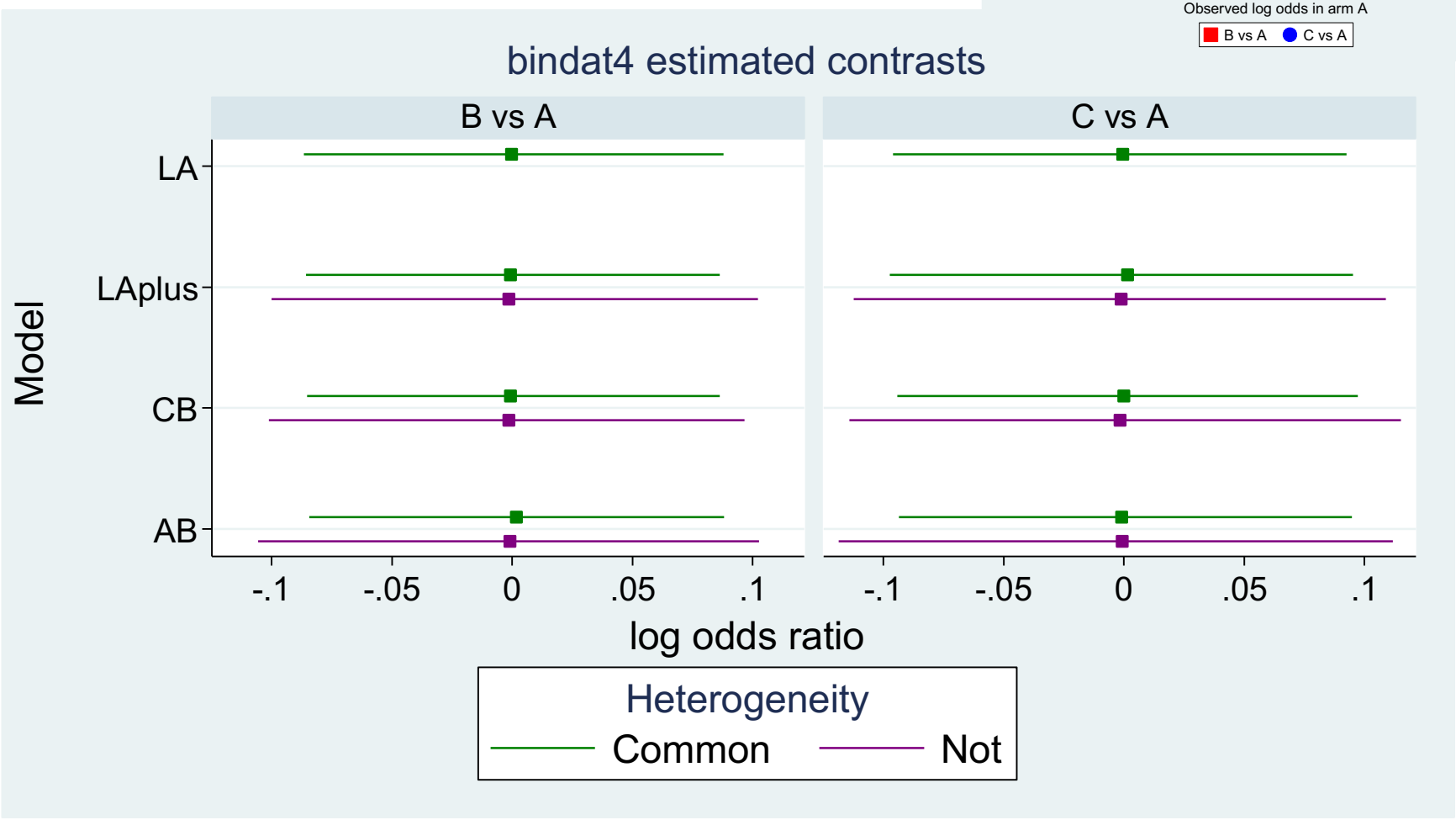
Artificial data 3

Reference arm mean

Design

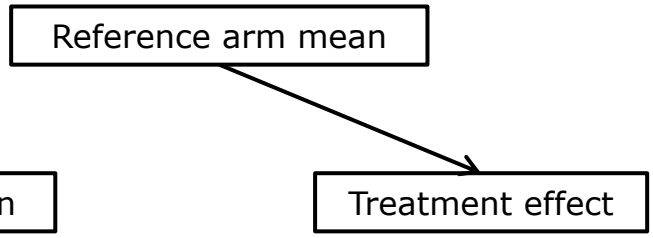
Treatment effect

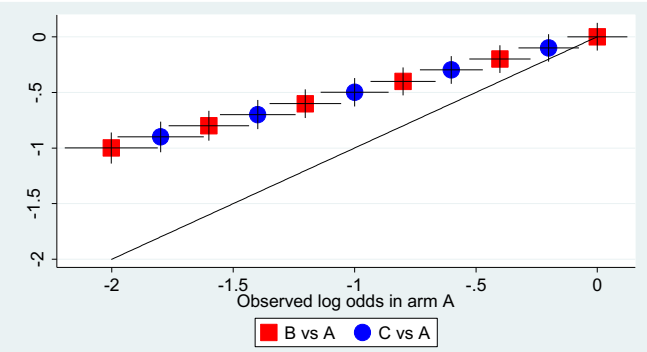


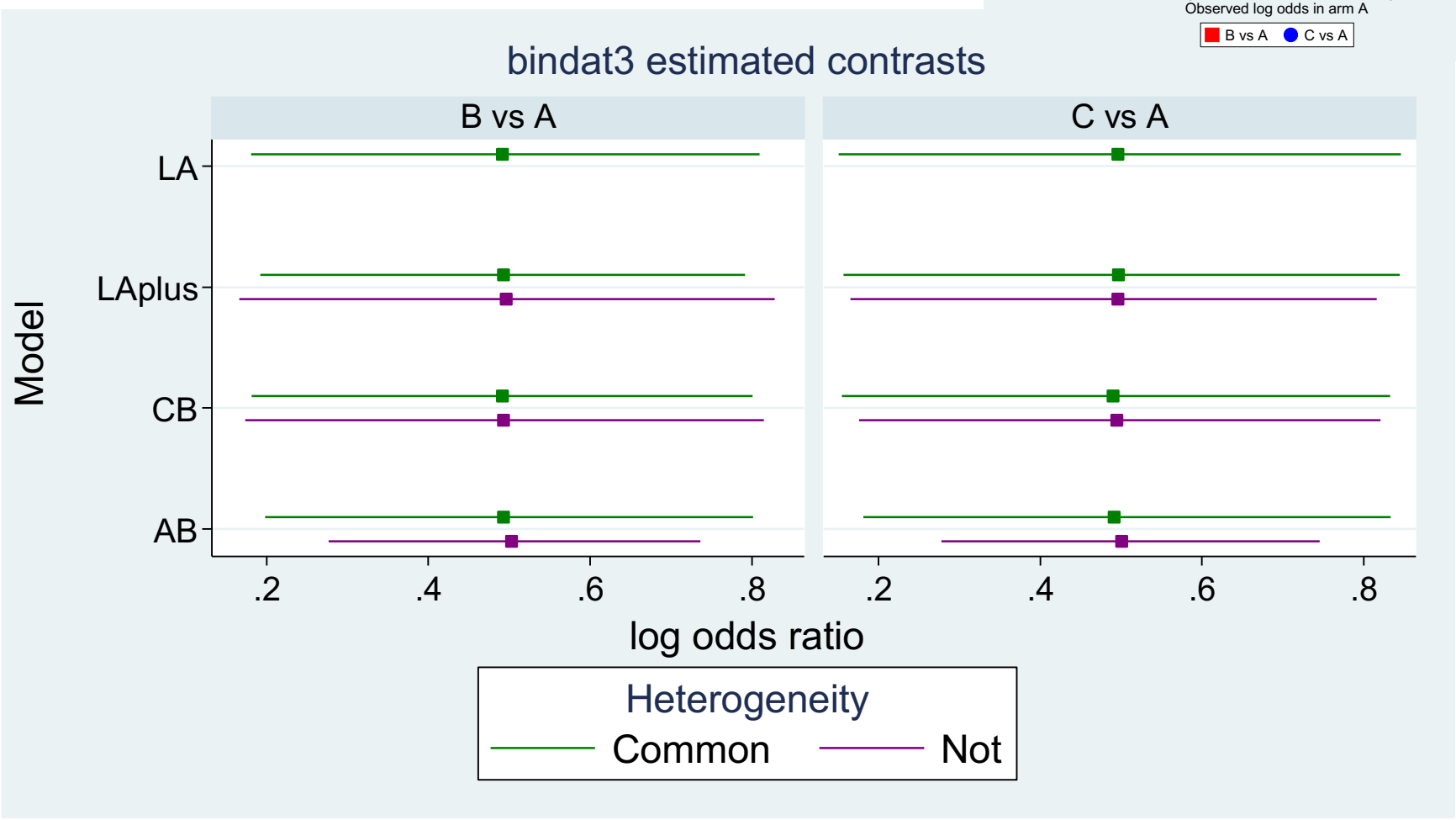


Every method works

Artificial data 4







Every method works

What do we really believe about missing data? [if time]

- Hard to believe the design depends on data actually observed in observed arms
- Easier to believe the design depends on true means in those arms
- So I can imagine making a working assumption that

$$\left[N_i, R_i \middle| \mu_i^A\right] = \left[N_i, R_i \middle| \mu_{R_i}^A\right]$$

  where N_i is the set of sample sizes chosen for the arms in R_i
- Would involve complex modelling as this **isn't** MAR
  - but might be close enough to MAR?

Key points from this section

1. There are datasets where the arm-based model gives very different results from the LA model
   - and arguably better results
2. Such datasets have study intercept (underlying risk) ~ design
   - and study intercept ~ treatment effect
3. However they risk
   - using between-study information
   - sensitivity to choice of effect measure

Plan

1. What are arm-based and contrast-based NMA?
2. Models and their key features
3. Breaking randomisation
4. Missing data aspects
5. **Estimands**
6. Summary

Estimands

- Estimand: the thing we want to estimate (causal inference term)
- Model 1 (LA) estimates the μ_{1k}^{C} $(k = 2, \ldots, K)$ and σ^{C2}
- The μ_{1k}^{C} would commonly be taken as the main estimands
  - "overall" log odds ratios for k vs. 1
  - and of course other contrasts derived from the μ^{C} under consistency: $\mu_{kk'}^{C} = \mu_{1k'}^{C} - \mu_{1k}^{C}$ etc.
  - also rankings, prediction intervals, …

Marginal estimands (1)

- In analysis of longitudinal data, there's a difference between "cluster-specific" (conditional on cluster) and "population-averaged" (marginal) estimands
- Similar issues here
- μ_{1k}^{C} can be interpreted as a treatment effect *conditional on study*
- Zhang et al (2014) show that the parameters μ_k^A have a marginal interpretation that may be of relevance in a public health setting
  - Zhang J, Carlin BP, Neaton JD, Soon GG, Nie L, Kane R, Virnig BA, Chu H (2014) Network meta-analysis of randomized clinical trials: Reporting the proper summaries. *Clinical Trials* 11: 246–262.
- Thus we might compute $\pi_k^A = logit^{-1}(\mu_k^A)$ and report marginal RR or RD

Marginal estimands (2)

- Dias and Ades: "While randomised controlled trials are unquestionably the best data sources to inform relative effects, the data sources that best inform the absolute effects might be cohort studies, a carefully selected subset of the trials included in the meta-analysis, or expert opinion."
  - they wish to apply the model for (relative) treatment effect, derived from NMA, to absolute means/risks in order to estimate absolute changes in mean/risk due to treatment
  - seems right to me
- Dias S, Ades AE (2016) Absolute or relative effects? Arm-based synthesis of trial data. *Research Synthesis Methods* 7: 23–28.

Marginal estimands: 2 questions

1. What estimand do we want, if treatment effect is related to study intercept?
   - insist on reporting treatment effects conditional on study intercept?
     (probably best with qualitative effect modification)
   - or report a summary?
     (appropriate with quantitative effect modification?)
2. To what extent should our models allow for treatment effect related to study intercept, even when there is no evidence for this?
   - just as we expect allowance for heterogeneity, even when there is no evidence for heterogeneity?

Absolute estimands? [if time]

- Hong et al claim "absolute measures of effect will often be of genuine interest, for example, the absolute amount of reduction in blood glucose produced by a given diabetes treatment"
  - they refer to the μ_k^A as "absolute treatment effect estimates"
  - I think this is a misconception, equating an observed change to a causal effect

Key points from this section

- Estimands need careful definition
- Estimands can be computed from either model
- Most estimands require doing some extra work

Plan

1. What are arm-based and contrast-based NMA?
2. Models and their key features
3. Breaking randomisation
4. Missing data aspects
5. Estimands
6. **Summary**

Summary

Model	Non-common hetero-geneity?	Uses between-study information?	Treatment effects relate to reference risk?	Missing data assump-tion	Main estimands	Other possible estimands
LA	Tricky	No	No	Contrast -MAR	Study-conditional contrast	Any
LAplus	Fine	No	No	Contrast -MAR	Study-conditional contrast	Any
CB	Fine	Yes (very little)	No	Contrast -MAR	Study-conditional contrast	Any
AB	Fine	Yes (little)	Yes	Arm-MAR	Marginal means and contrasts	Any

Some points that worry me [if time]

1. Non-common heterogeneity models are implemented in practice with inverse Wishart priors - but often these are more informative than we might wish
2. Symmetry: CB model is asymmetrical across treatments, but LA and AB are symmetrical
3. Is between-study information a matter of bias?
   - i.e. do we only care if it affects results on average over NMAs?
   - or do we care about between-study information changing the results of a specific NMA?

Future research

1. How much does between-studies information matter in practice? When does it matter?
2. Likely missingness mechanisms are that studies are designed based on true study intercepts, not observed ones. What effect does this have?
3. How often does study intercept relate to design?
4. What estimand do we want, if treatment effect is related to study intercept?
5. Can we express our assumptions about arm sizes as we express our assumptions about missing arms?
6. Can we get benefits of LAplus and AB models by having fixed study effects α_i and treatment effects $\delta_i^c \sim \alpha_i$?
7. Why is between-studies information so weak?

Coming soon: **`network bayes`**

Key points

1. Key differences between arm-based and LA models are
   - random study effects
   - random study*treatment effects (i.e. random heterogeneity) that are associated with the study intercepts (underlying risks)
2. Breaking randomisation is a theoretical problem, but seemingly not a practical problem
3. There are datasets where the arm-based model gives very different results from the LA model and arguably better results. Such datasets have study intercept ~ design and ~ treatment effect
4. Estimands can be computed from either model