Modeling inconsistency as heterogeneity in network meta-analysis

Gert van Valkenhoef

SRSM 2013, Providence RI, USA, June 2013

umcg

university of groningen

Introduction

Network meta-analysis assumes *exchangeability* of trials

- i.e. that trials are *similar* or *homogeneous*
- in terms of **relative-effect modifying covariates**

Variations in absolute effects not as important

- randomization + appropriate effect measure

Violations of exchangeability:

- Heterogeneity – within comparison (pair-wise and network)
- Inconsistency – between comparisons (network only)

Dealing with heterogeneity

In pair-wise meta-analysis

- Up-front screening: filter trials, identify possible confounders
- Visual assessment using forest plots
- Statistical assessment (e.g. I^2)

All of these apply to network meta-analysis

- But now we also have to worry about inconsistency

Methods for detecting inconsistency

Several methods have been proposed:

- Inconsistency factors (Lu & Ades 2006)
  - Single model, but not unique (multi-arm trials)
  - Hard to parameterize
  - Weak local tests, no global test

Methods for detecting inconsistency

Several methods have been proposed:

- Inconsistency factors (Lu & Ades 2006)
  - Single model, but not unique (multi-arm trials)
  - Hard to parameterize
  - Weak local tests, no global test
- Node splitting (Dias et al. 2010)
  - Many models, not unique (multi-arm trials)
  - Easy to parameterize
  - Local tests, no global test

Methods for detecting inconsistency

Several methods have been proposed:

- Inconsistency factors (Lu & Ades 2006)
  - Single model, but not unique (multi-arm trials)
  - Hard to parameterize
  - Weak local tests, no global test
- Node splitting (Dias et al. 2010)
  - Many models, not unique (multi-arm trials)
  - Easy to parameterize
  - Local tests, no global test
- Design inconsistency (White et al. 2012)
  - Single model, global test unique
  - Easy to parameterize
  - Weak local tests, global test
  - Overparameterized?

Inconsistency: conceptual problems

Multi-arm trials pose difficulties

- Heterogeneity / inconsistency distinction unclear
  - Interpretation and/or parameterization may differ
  - Where should the line be drawn?

Inconsistency: conceptual problems

Multi-arm trials pose difficulties

- Heterogeneity / inconsistency distinction unclear
  - Interpretation and/or parameterization may differ
  - Where should the line be drawn?
- Dependencies and overparameterization
  - Inconsistency factors: need to identify dependencies
  - Node-splitting: avoid issue by splitting one-by-one
  - Design inconsistency: more parameters (design inconsistencies)

Inconsistency: conceptual problems

Multi-arm trials pose difficulties

- Heterogeneity / inconsistency distinction unclear
  - Interpretation and/or parameterization may differ
  - Where should the line be drawn?
- Dependencies and overparameterization
  - Inconsistency factors: need to identify dependencies
  - Node-splitting: avoid issue by splitting one-by-one
  - Design inconsistency: more parameters (design inconsistencies)
- Loop inconsistency vs design inconsistency
  - Design inconsistency convenient for parameterization
  - But how should it be interpreted?

Objective

Simplify analysis of inconsistency

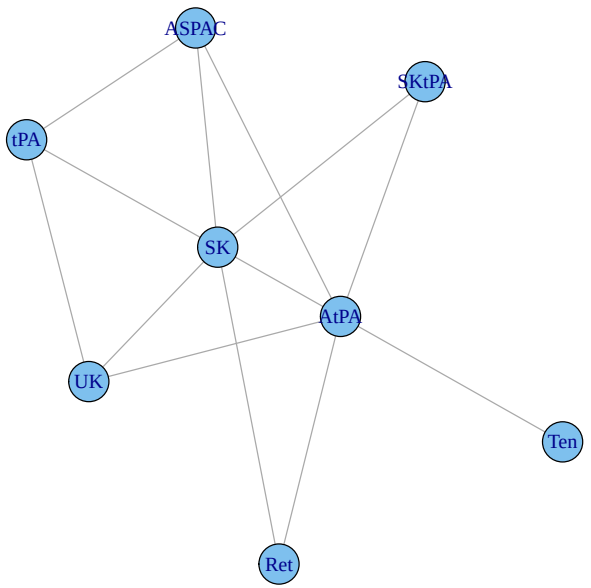
- Modeling inconsistency separately is **very complex**
- Can we do better by *not* modeling it?
  - View both as deviation from posterior mean
  - Allow visual inspection using forest plots
  - Extend the I^2 statistic
- Rely on interpretation rather than models
- No assumptions on nature of inconsistency
  - Viewed as additional source of heterogeneity
- Ensure attention is payed to heterogeneity, not just inconsistency

Outline

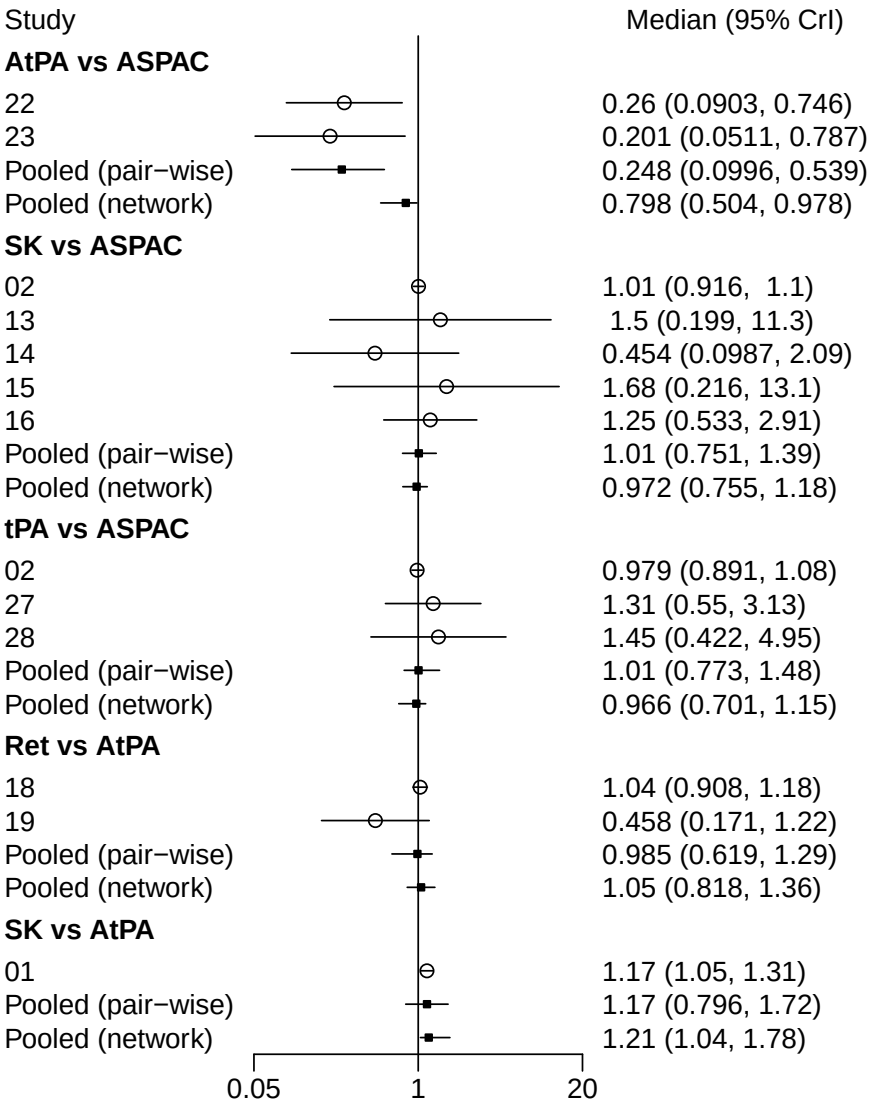
- Example: thrombolytics data
- Technical details
- Discussion

Example: thrombolytics data

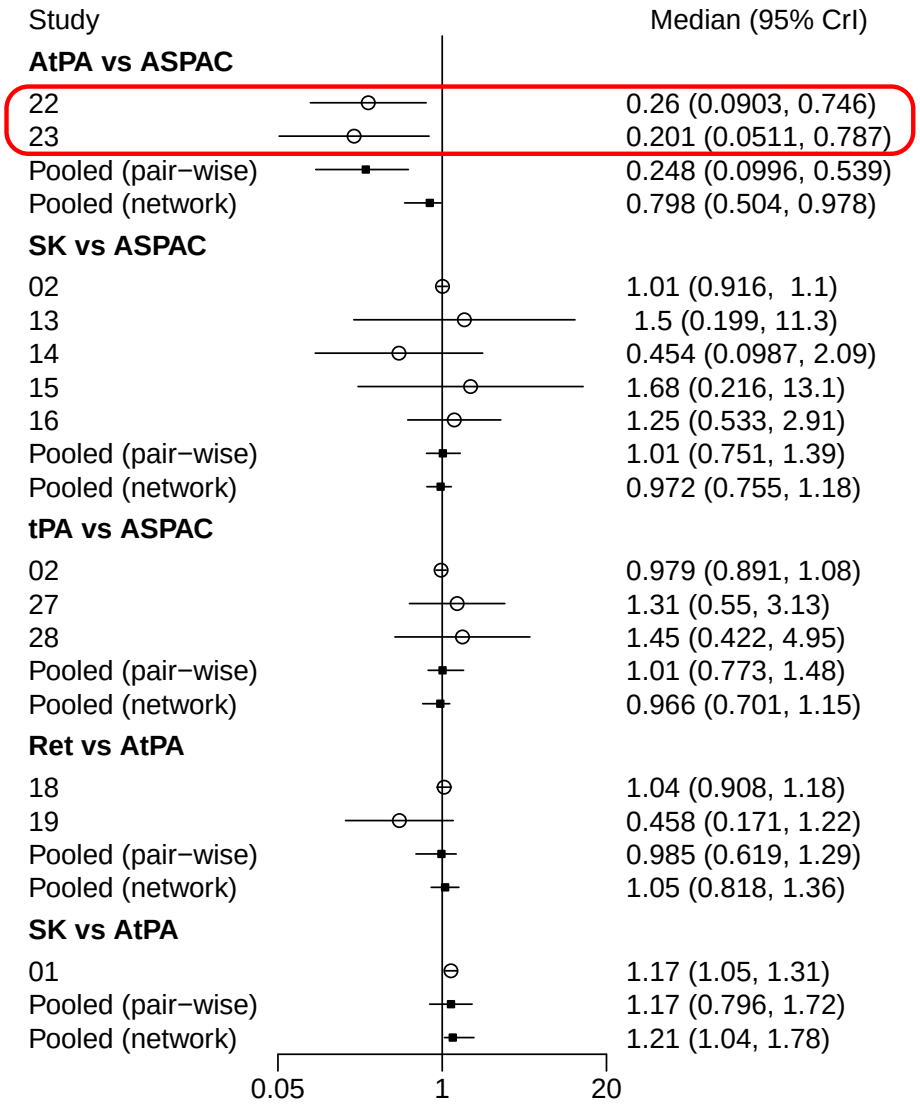
- A 'classic' in methods literature, based on Boland et al., 2003
- 28 trials (26 two-arm and 2 three-arm)
- Comparing 8 anti-thrombolytics
  - 13 out of 28 comparisons have direct evidence
- Outcome: mortality at 30-35 days
- Measure: log odds ratio



New method (random effects): page 1/3

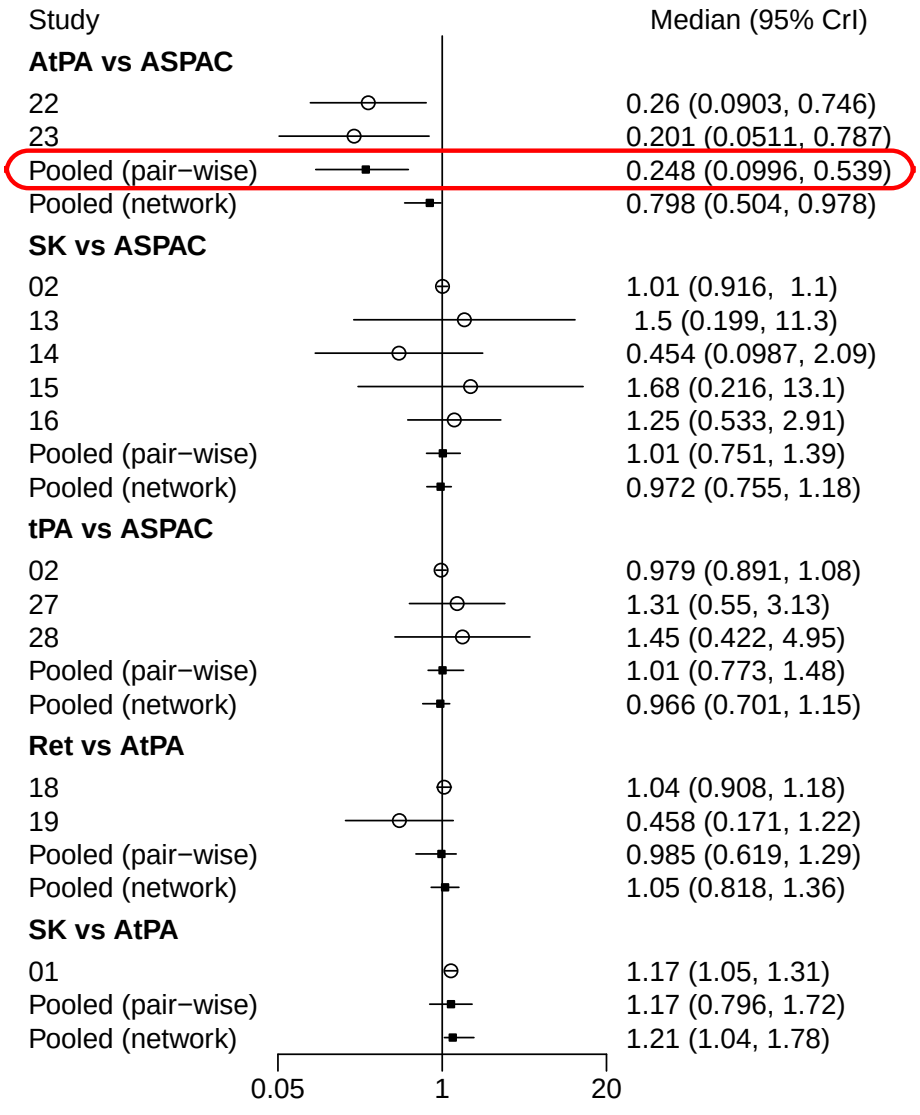


New method (random effects): page 1/3



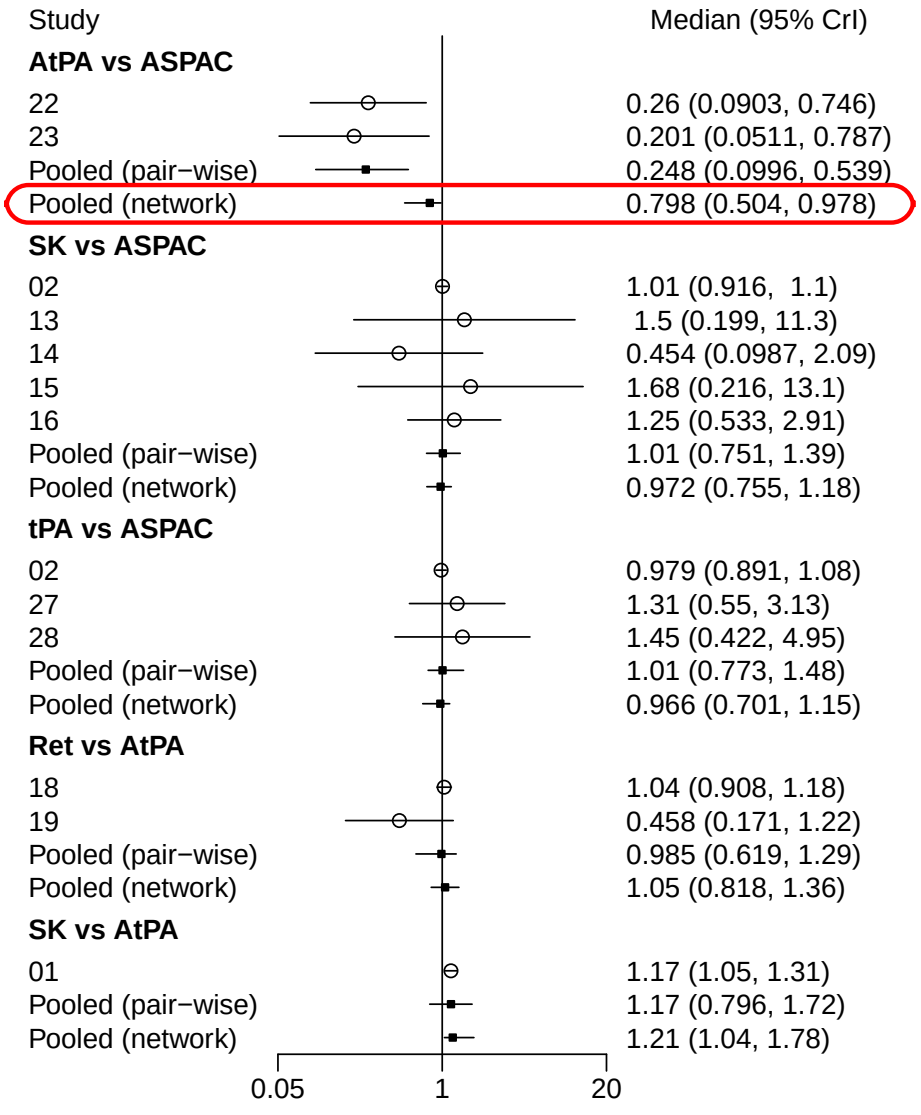
Individual study estimates

New method (random effects): page 1/3



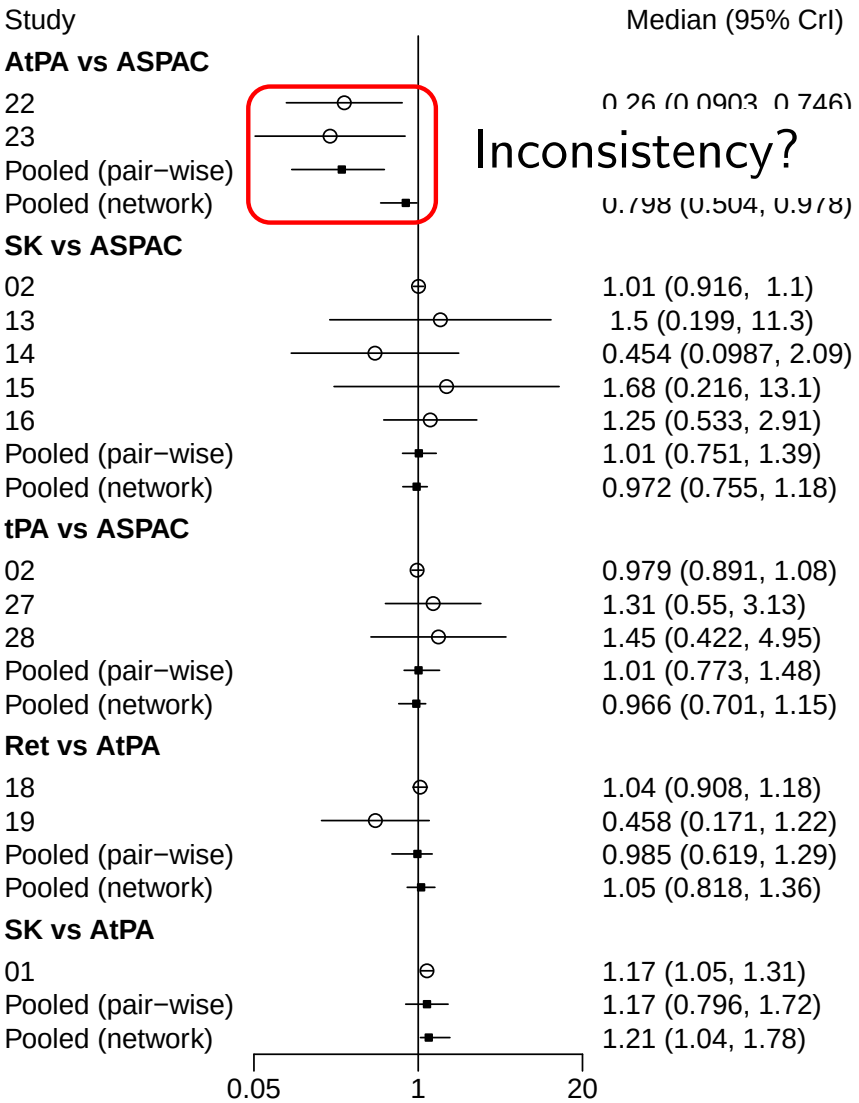
Unrelated mean effects (shared RE variance)

New method (random effects): page 1/3



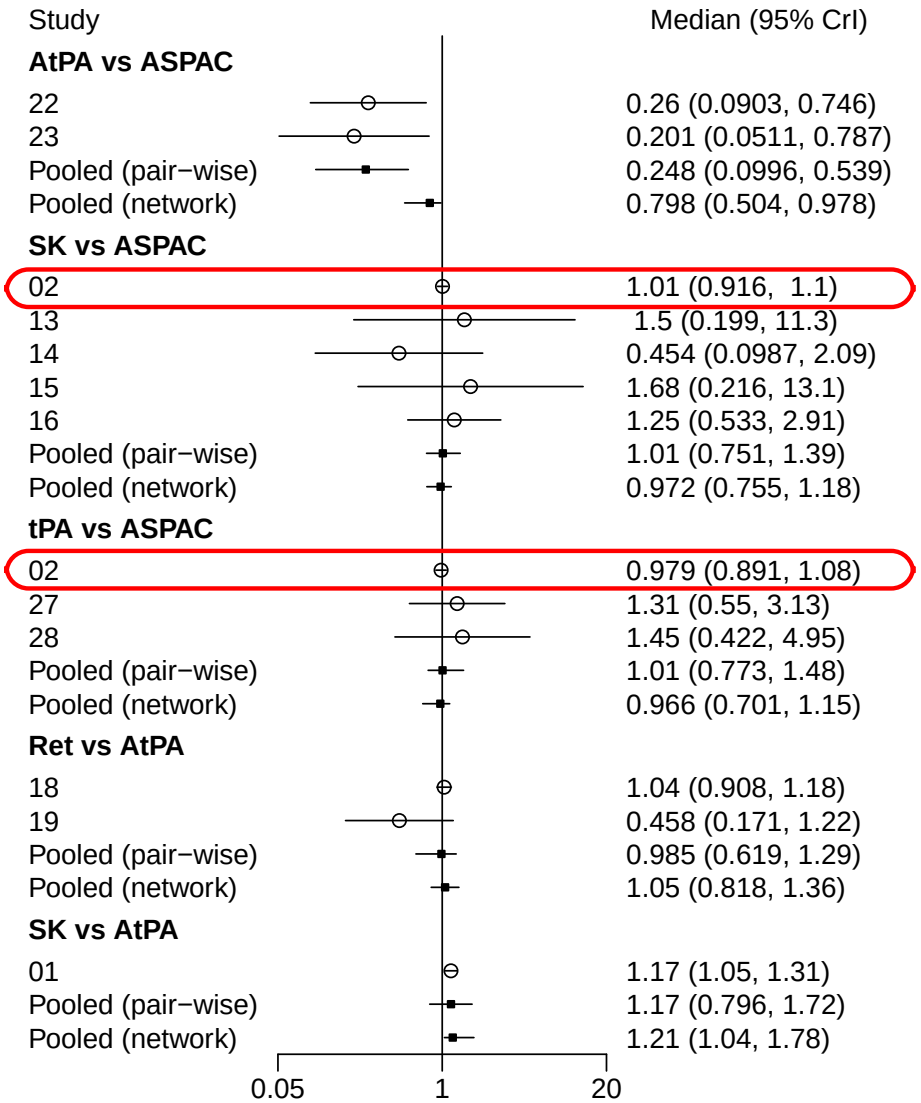
Consistency estimate

New method (random effects): page 1/3



- pair-wise $I^2 = 0\%$
- network $I^2 = 76\%$

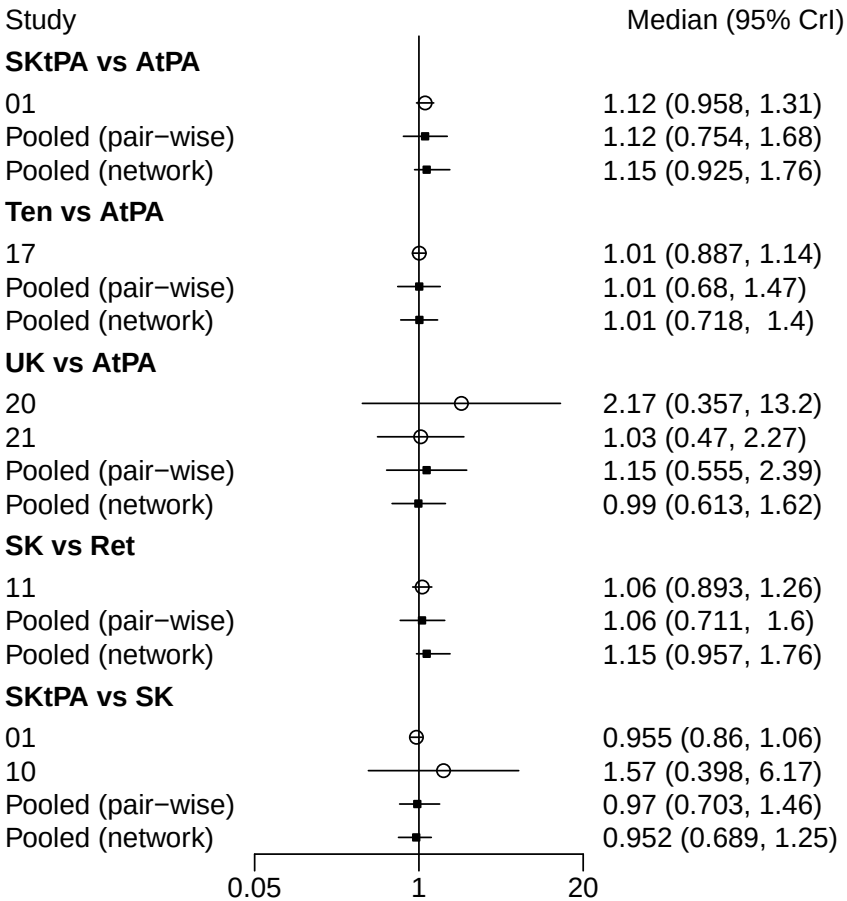
New method (random effects): page 1/3



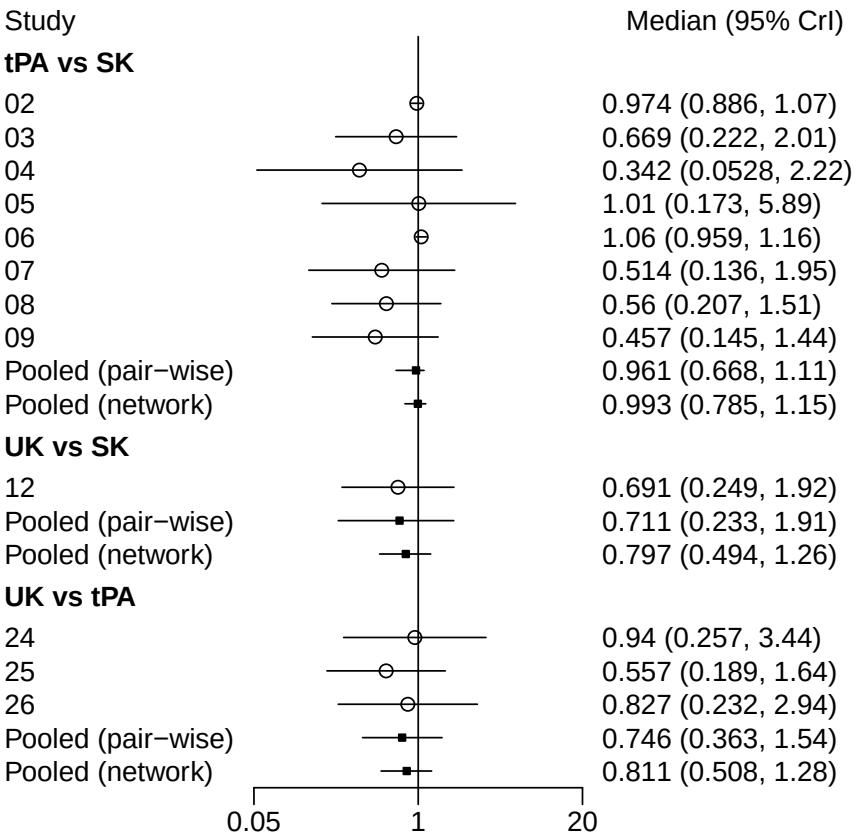
Multi-arm trials

- shown in all relevant comparisons
- automatically downweighted

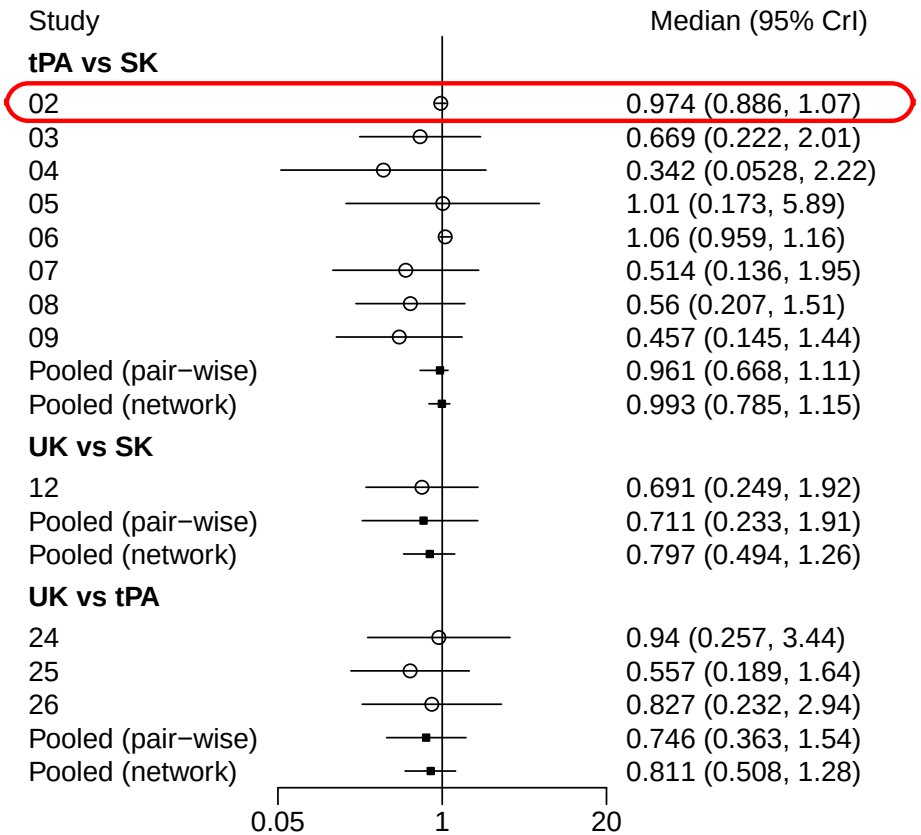
New method (random effects): page 2/3



New method (random effects): page 3/3

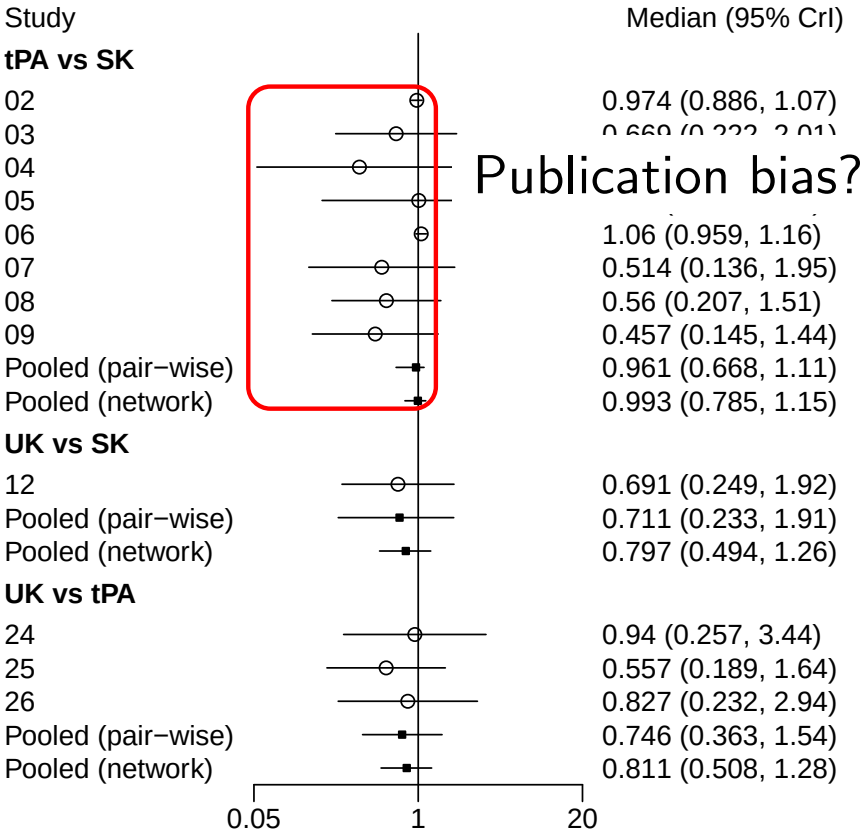


New method (random effects): page 3/3



(final arm)

New method (random effects): page 3/3



Technical details

- All models in Bayesian hierarchical framework
  - But easy to duplicate in frequentist framework
- Three models are estimated:
  1. Unrelated study effects (USE) model
  2. Unrelated mean effects (UME) model (shared variance if RE)
  3. Consistency model
- Multi-arm trials are decomposed
  - UME model uses decomposed data (multi-arm trials only)
  - USE & consistency model use original data
- I^2 is calculated per comparison and globally

Unrelated study effects (USE) model

Why?

- Maximum likelihood estimates not always defined
  - Frequentist 'adjustments' do not reflect what happens in BHM
- Consistency model posteriors shrink towards mean
  - Leads to underestimation of heterogeneity

How?

- Uses same priors as consistency model
- But each study gets its own parameter(s)

Unrelated mean effects (UME) model

Why?

- Estimate within-comparison heterogeneity
- Contrast direct and indirect evidence

How?

- Uses same priors as consistency model
- Uses decomposed ('downweighted') multi-arm trials
  - Multi-arm trials translated to equivalent 2-arm trials
- Uses original data for two-arm trials
- Each comparison gets its own parameter
  - No consistency constraints

Consistency model

Why?

- This is the model of actual interest
- Contrast results to other models

How?

- As you normally would
  - Use original data for two-arm and multi-arm trials

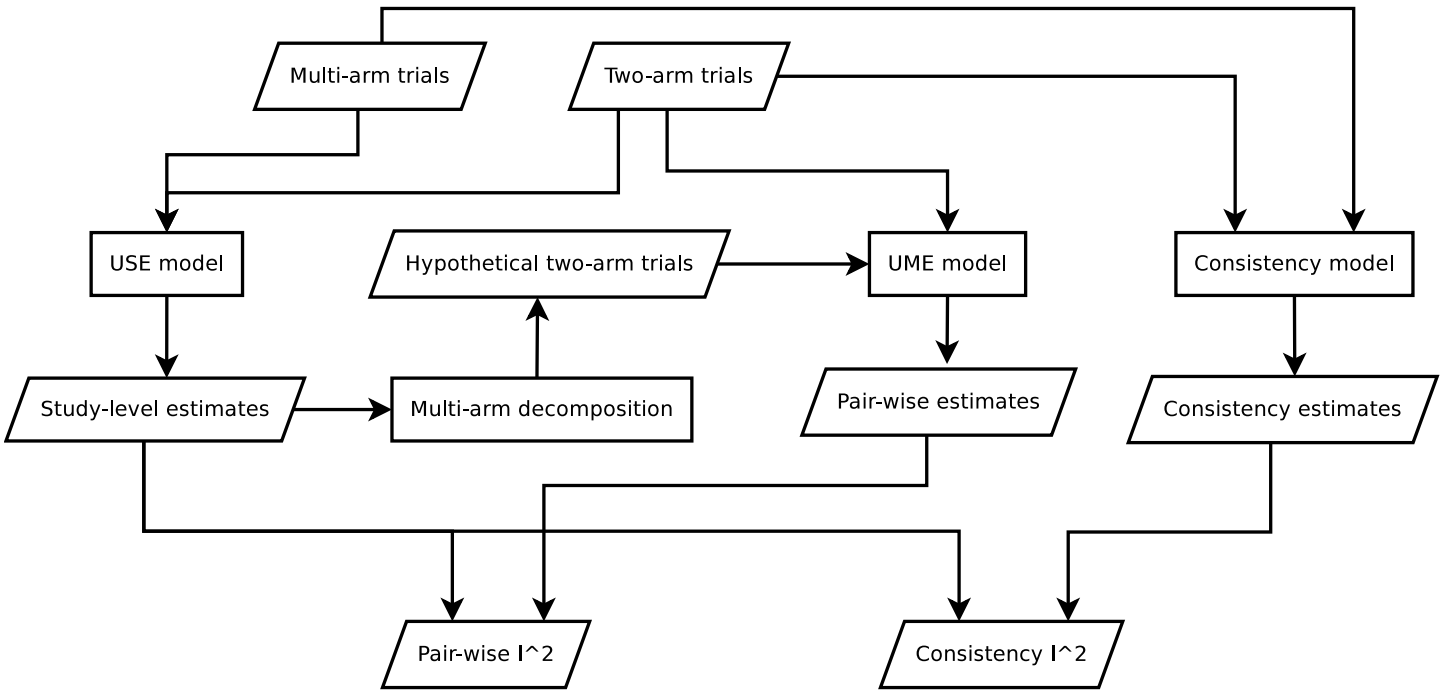
Multi-arm trial decomposition

- Method due to Gerta Rücker (RSM 2012)
  - Based on pseudo-inverse of Laplacian
- *n*-arm trial viewed as result of a network meta-analysis
  - With one two-arm trial per comparison
  - The two-arm trials are inferred from the *n*-arm trial
  - Inverse meta-analysis
- Resulting two-arm trials are not correlated
  - Exactly what we need for UME model
  - But not appropriate for consistency model
- Decomposed estimates incorporate both prior and data
  - 'Double counting' of priors in UME model

I^2 statistic

- Defined as usual (k trials):
  - $Q = \sum_{i=1}^{k} (\mu_i - \mu)^2 / \sigma_i^2$
  - $I^2 = 100 \max(0, (Q - \mathrm{df})/Q)$
- The degrees of freedom (df) is:
  - $k - 1$ for within-comparison I^2
  - $k - 1$ for network I^2 without indirect evidence
  - k for network I^2 with indirect evidence
- The μ is either from the UME or consistency model
- The μ_i and σ_i are from the USE model

Summary


Multi-arm trials
Two-arm trials
USE model
Hypothetical two-arm trials
UME model
Consistency model
Study-level estimates
Multi-arm decomposition
Pair-wise estimates
Consistency estimates
Pair-wise I^2
Consistency I^2

Discussion

Note on computation times

- Number of models does not increase with network size
  - Unlike node-splitting
- Parameterization is unique
  - No need to assess sensitivity to parameterization
- Computationally intensive (MCMC)
  - Takes several minutes for larger networks

Discussion

Summary

- Viewing inconsistency as heterogeneity simplifies issue
- Single visual display provides information on both
- Well-known I^2 statistic is trivial to generalize

Limitation

- 'Double counting' priors on multi-arm trials (UME model)
  - Could the decomposition be part of the UME model itself??
  - Only use USE model when MLE undefined?
  - Any suggestions are welcome!