Statistical methods for updating meta-analyses

Mark Simmonds

Centre for Reviews and Dissemination
University of York, UK

With:
Julian Elliott, Joanne McKenzie, Georgia Salanti, Adriani Nikolakopoulou, Julian Higgins

THE UNIVERSITY of York
Centre for Reviews and Dissemination

Problems with standard updating

- Type I error
  - Inappropriate significance
    - Rises with more updates
    - Small sample sizes
    - Poor estimation of parameters (heterogeneity)
    - Later update might reverse findings
- Type II error
  - No effect, or just not significant?
    - When is it safe to conclude no meaningful effect?

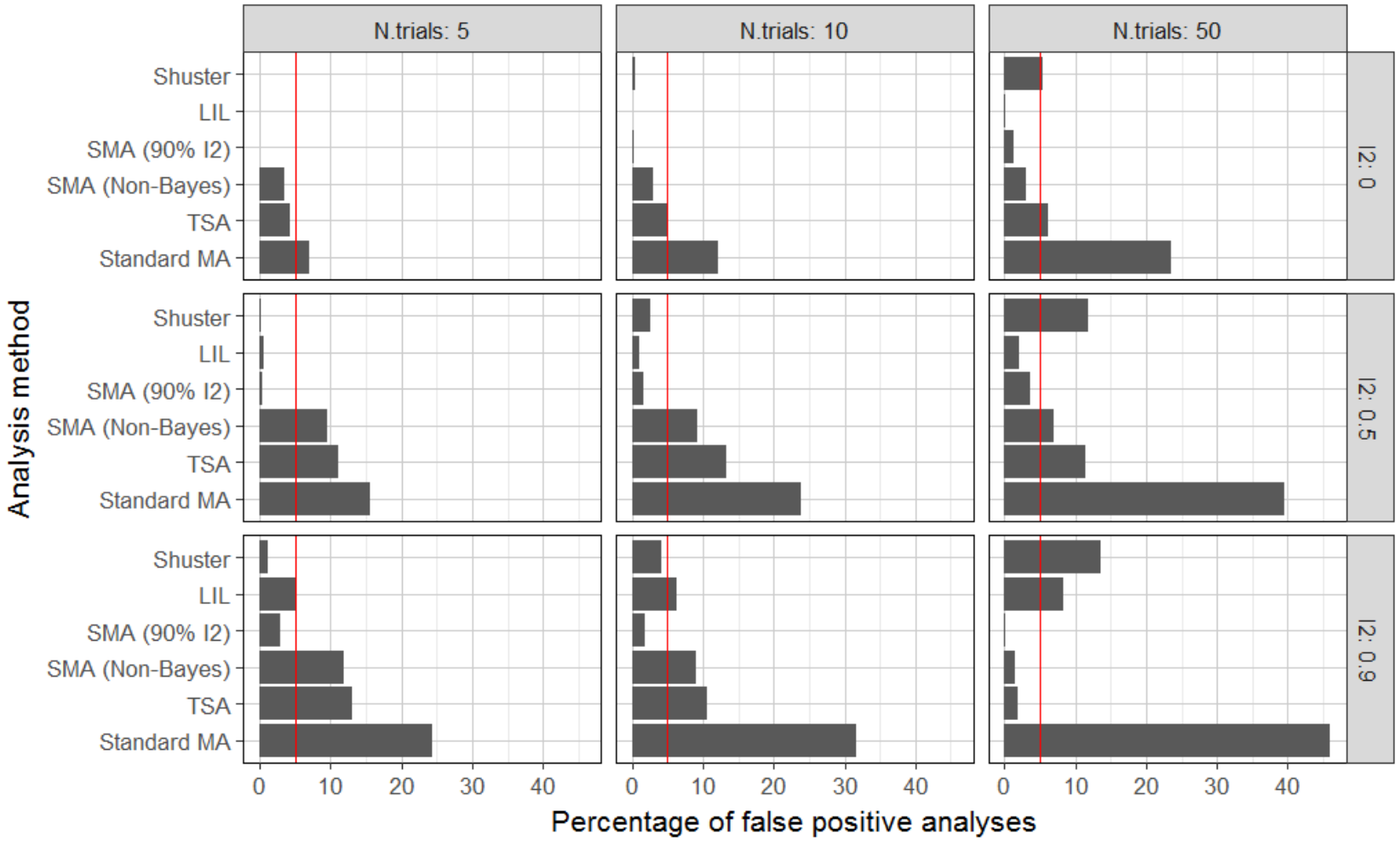
Controlling error

- Control Type I and Type II error
  - Sequential Meta-Analysis (SMA, Higgins et al)
    - With or without “Approximate Bayesian” heterogeneity
  - Trial Sequential Analysis (TSA, Copenhagen)
- Control Type I error
  - Law of Iterated Logarithm (LIL, Hu et al)
  - “Shuster-Pocock” method (Shuster)

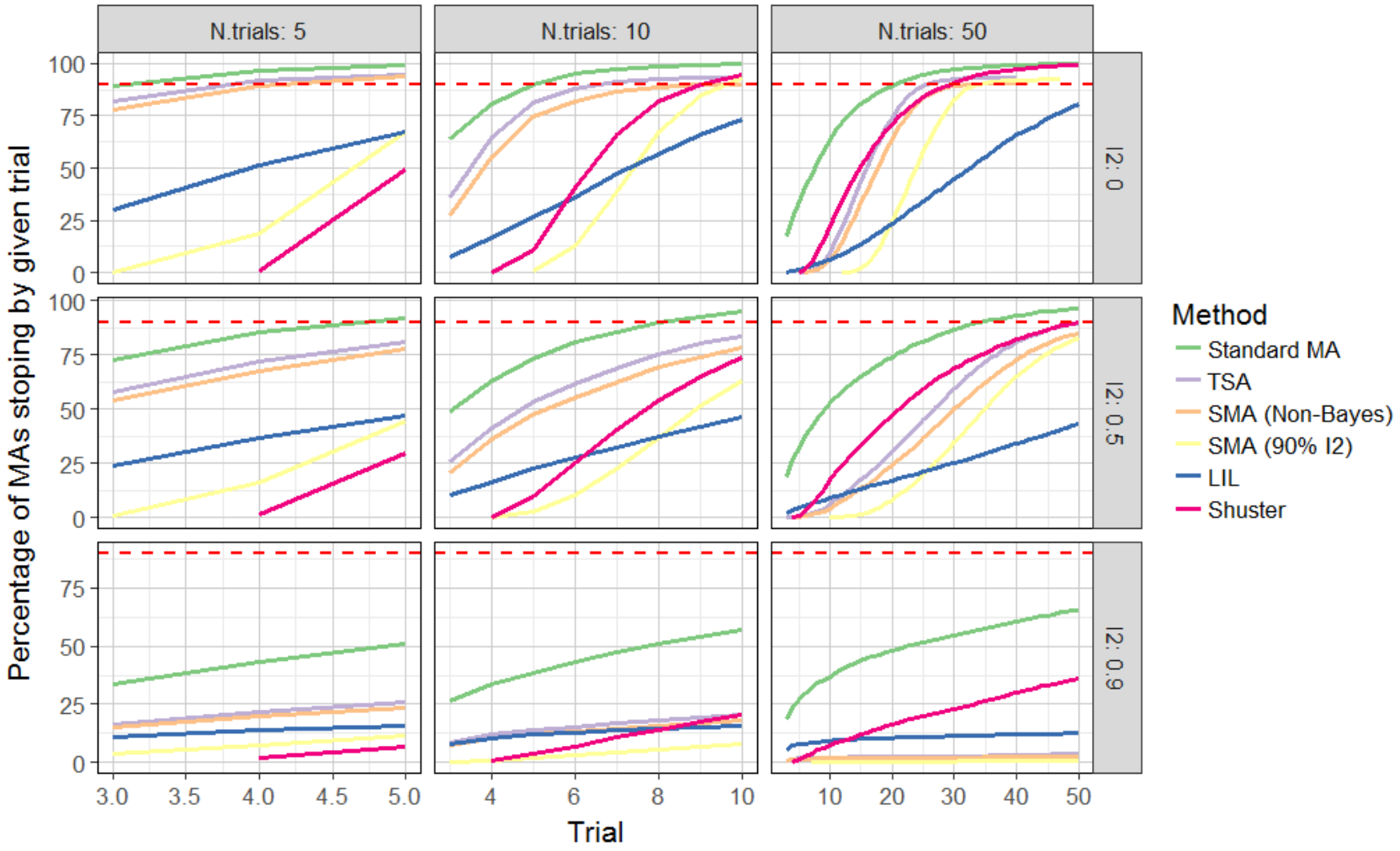
Simulation study

- Simulated meta-analyses varying:
  - True treatment effect: 0 or 0.1
  - Number of studies: 5 to 50
  - Heterogeneity: I^2 0 to 90%
- Fixed total sample size of 9000
  - 90% power to detect effect of 0.1 if $I^2 = 50\%$

False positive rates – Type I error



Cumulative power – Type II error



Conventional updated meta-analysis

- Too many inappropriate positive conclusions
  - Elevated Type I error rate
  - But not vastly elevated for most real updated reviews?
- Many analyses showing significant results are based on too little evidence

TSA and SMA

- Both control Type I error well
  - Except with few trials / high heterogeneity
- “Approximate Bayes” heterogeneity not required in most circumstances?
- Control for Type II error
  - But most Cochrane reviews are underpowered
- No obvious choice of one over the other

LIL and Shuster methods

- Control for Type I error
- But too conservative
  - LIL lacks statistical power
  - Shuster lacks power with few trials
- No control of Type II error
- Can’t be recommended at present?