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

Updating meta-analyses

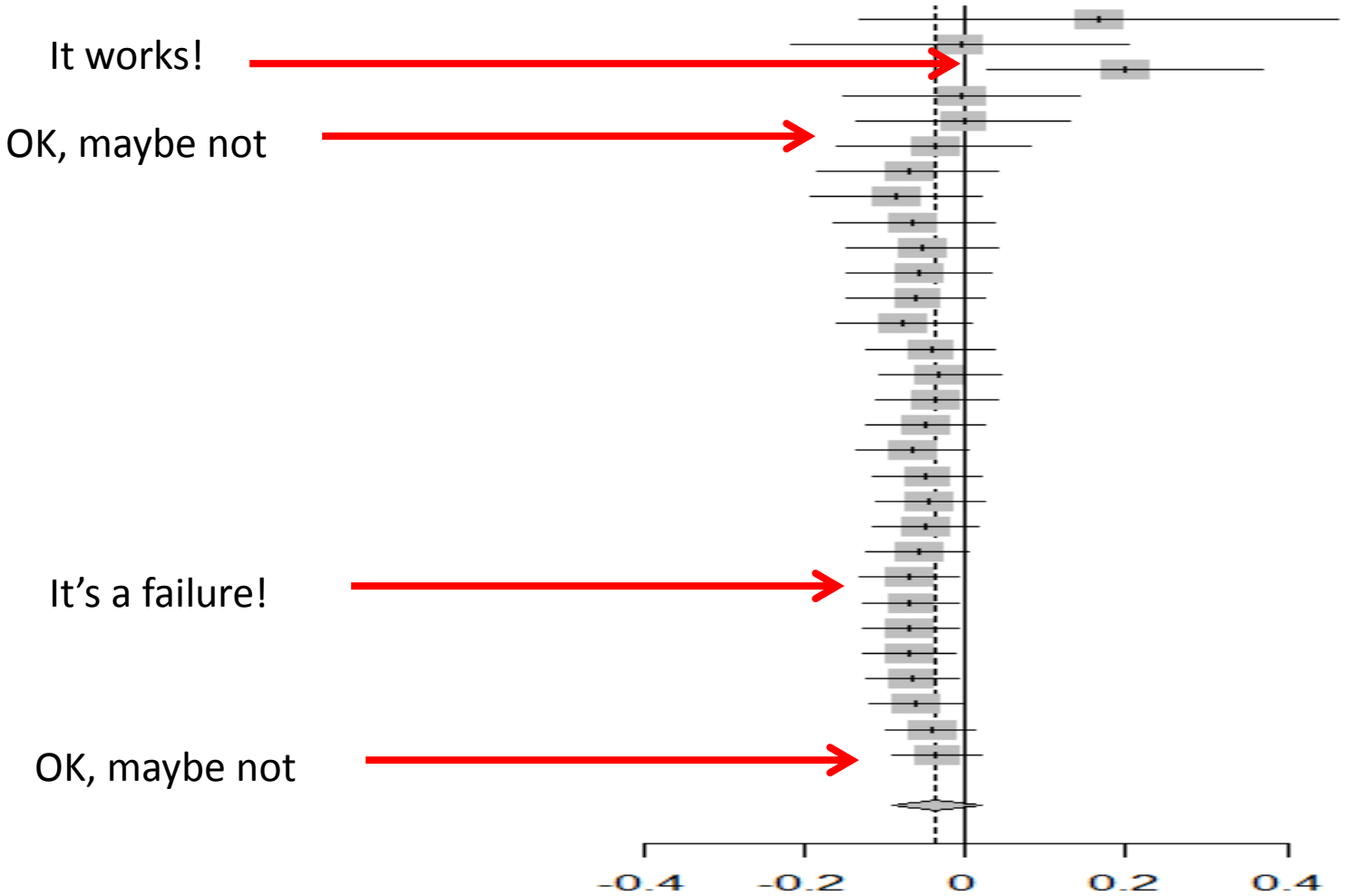
- When should we update a meta-analysis in an LSR?
  - As soon as new studies emerge?
  - When new data might alter our conclusions?

- Updating is time-consuming

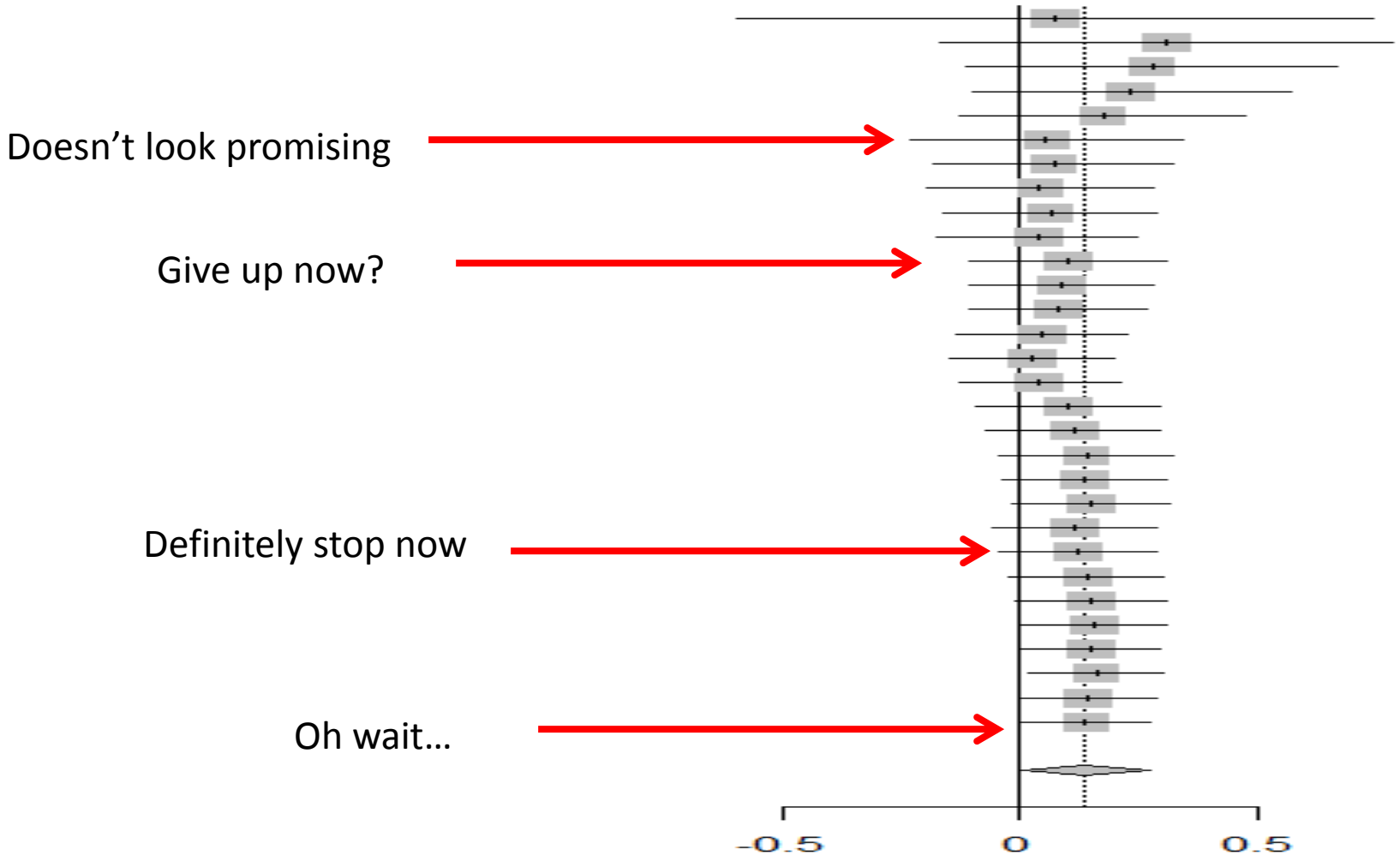
Some issues

- Conclusions can change over time
  - Risk of error if we stop too soon
- Are the results robust?
- When should the next update happen?
- When can we stop updating?

Cumulative MA : Type I error



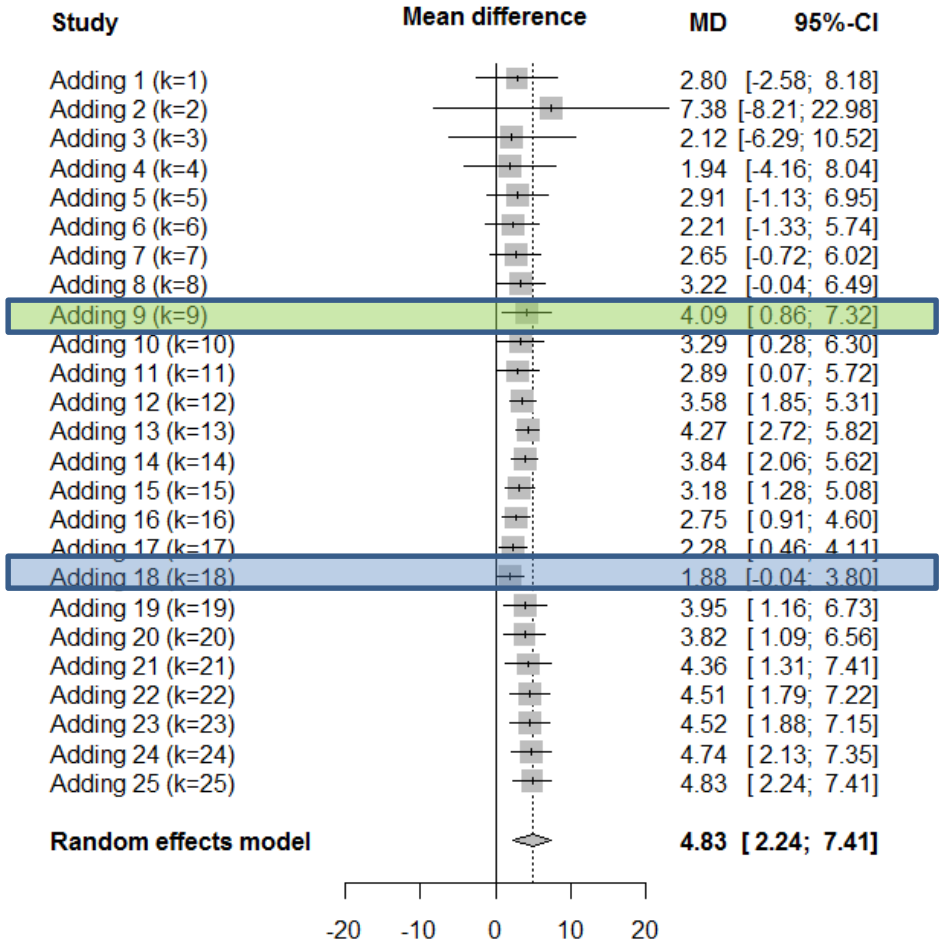
Cumulative MA: Type II error



Controlling error

- Control Type I and Type II error
  - Sequential Meta-Analysis (SMA, Higgins et al)
  - Trial Sequential Analysis (TSA, Copenhagen group)
- Control Type I error
  - Law of Iterated Logarithm (LIL, Hu et al)
  - “Shuster-Pocock” method (Shuster)
- Other methods
  - Fully Bayesian analysis
  - Robustness or stability of analysis
  - Consequences of adding new studies
  - Power gains from adding new studies

Example cumulative meta-analysis



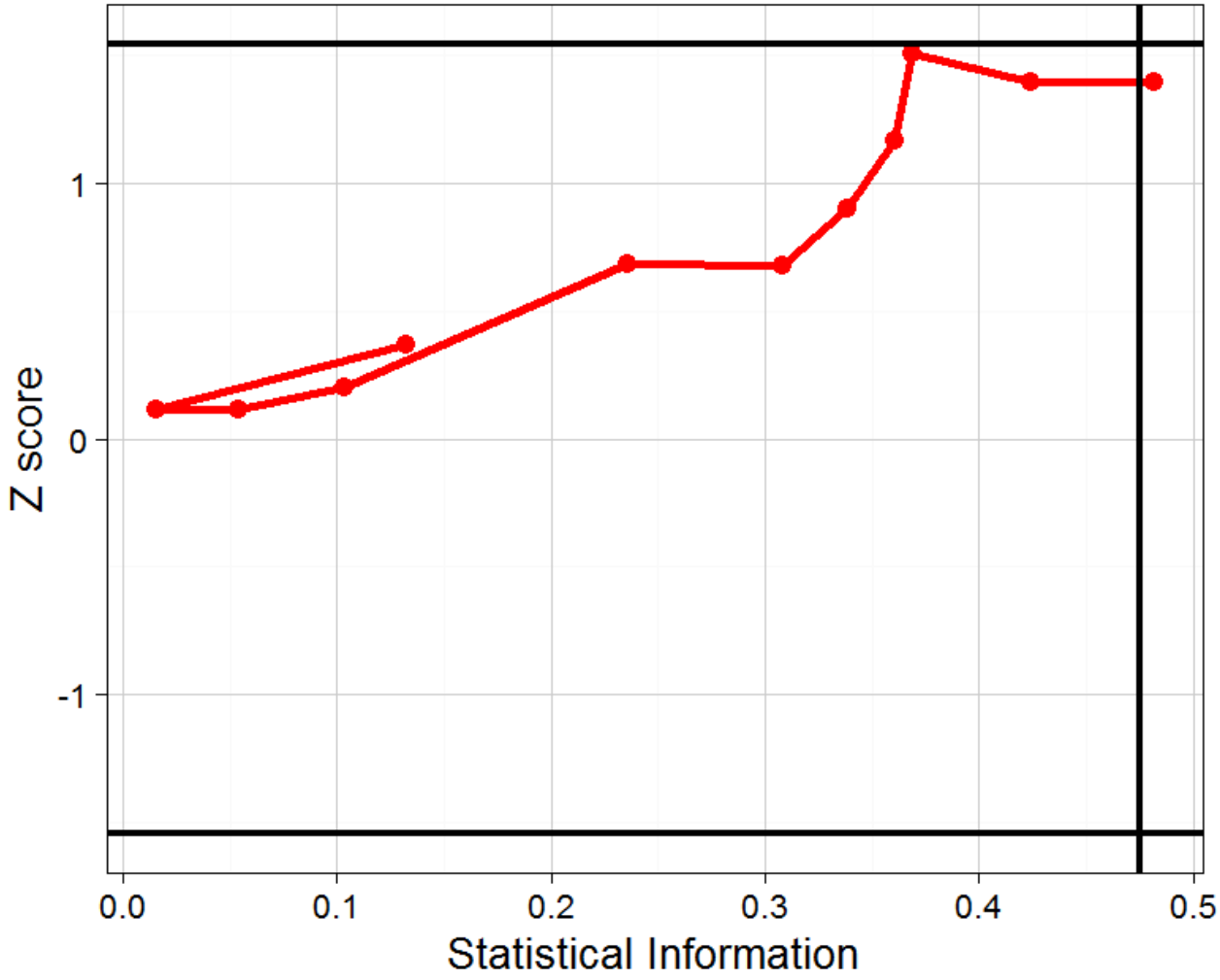
$I^2 = 95\%$

Sequential meta-analysis (SMA)

Higgins, Simmonds, Whitehead 2010

- Calculate cumulative Z score and cumulative Information for each updated meta-analysis
- Stop when a pre-specified boundary is crossed
  - Boundary designed to control type I and II error
- Optional Bayesian estimation of heterogeneity
  - Avoid mis-estimation of heterogeneity with few trials

Sequential meta-analysis



Trial sequential analysis (TSA)

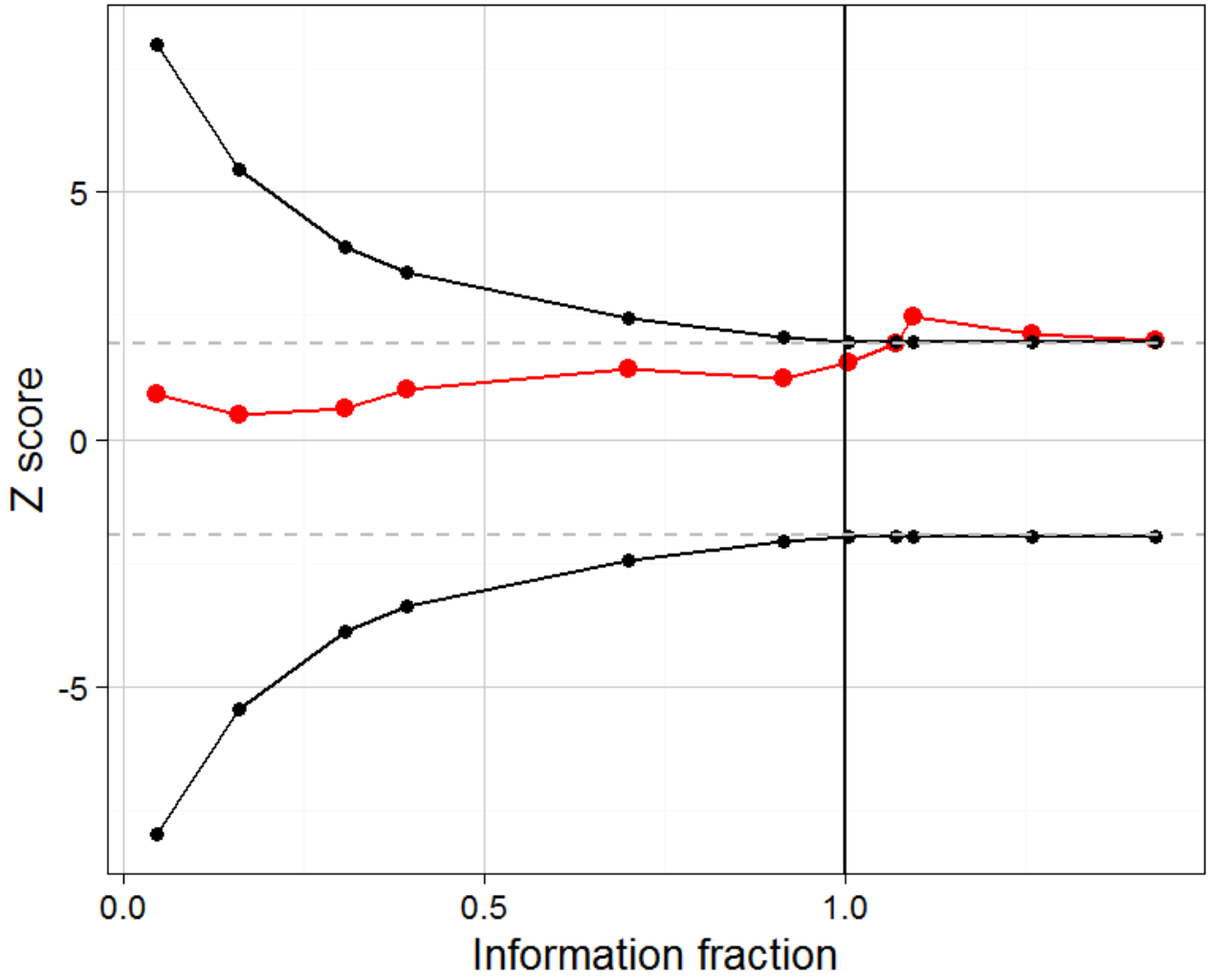
Wetterslev, Thorlund, Brok, Gluud 2008

- Calculate required sample size for the meta-analysis
- Calculate alpha-spending boundaries

- Stop if Z score exceeds the boundary
- Or if sample size is reached

- Sample size must be adjusted for heterogeneity

Example



Law of Iterated Logarithm (LIL)

Lan, Hu, Cappelleri 2007

- Uses an adjusted Z statistic
- $Z^* = \frac{Z}{\sqrt{2\log(\log(N))}}$

- This is bounded as $N \to \infty$
- So controls Type I error

Shuster-Pocock method

Shuster, Neu 2013

- Compares the Z statistic to a t distribution
- Parameters of t distribution are based on Pocock's group sequential boundaries
- Must specify number of meta-analyses performed

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%

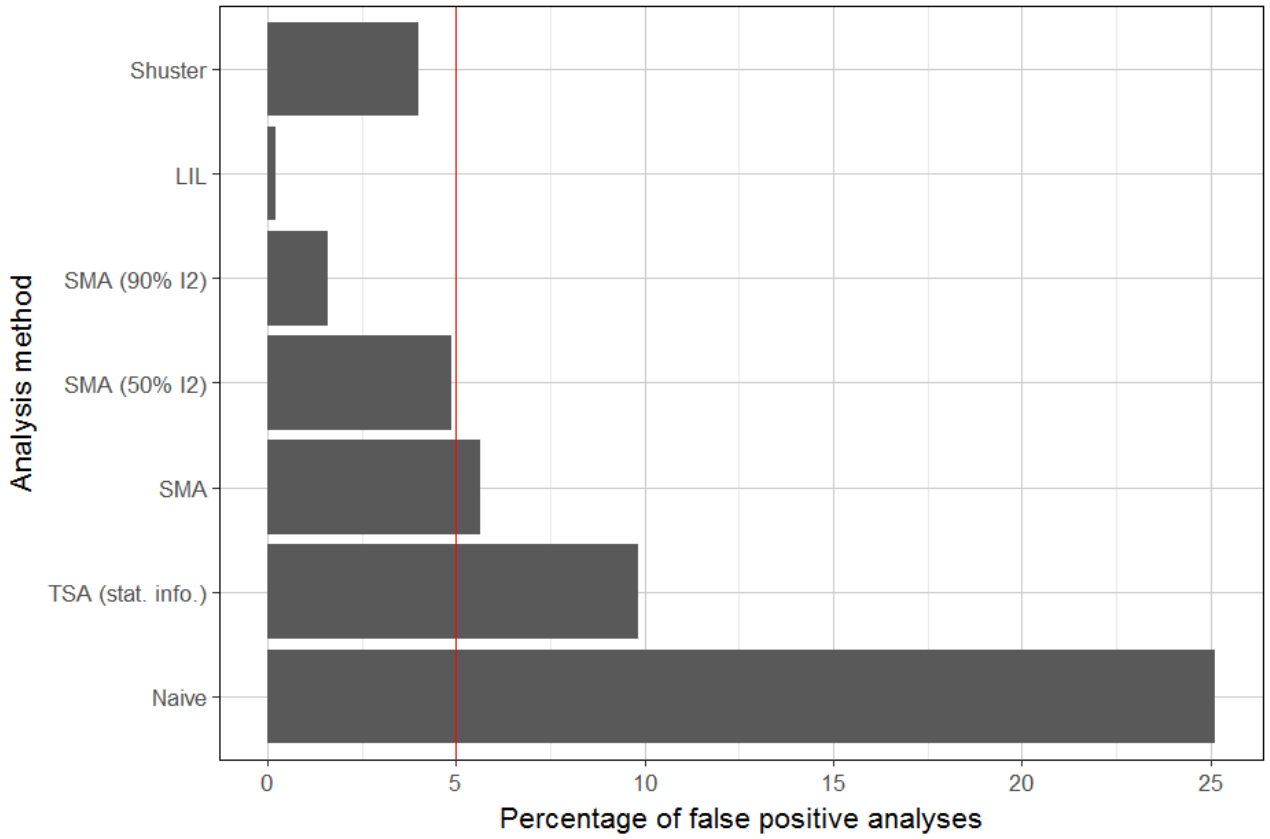
- Fixed total sample size of 9000
  - 90% power to detect effect of 0.1 if $I^2 = 50\%$

Methods applied

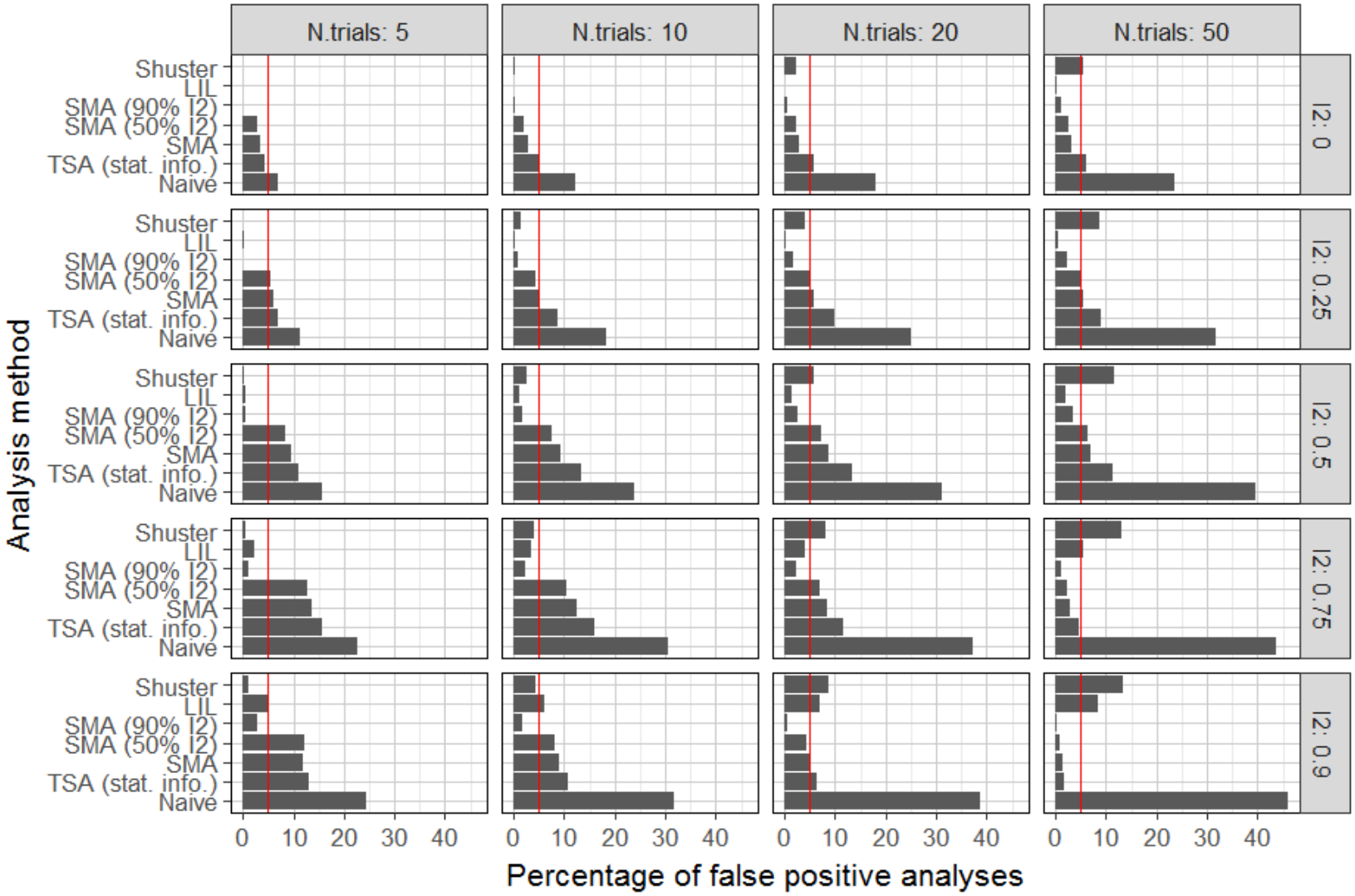
- Naïve analysis (standard cumulative MA)
- Trial Sequential Analysis (TSA)
- Sequential Meta-Analysis (SMA)
  - No prior heterogeneity
  - Prior I^2 of 50% or 90%
- Law of Iterated Logarithm (LIL)
- Shuster method

False positive rates – Type I error

- 20 trials / updates, I^2 = 25%

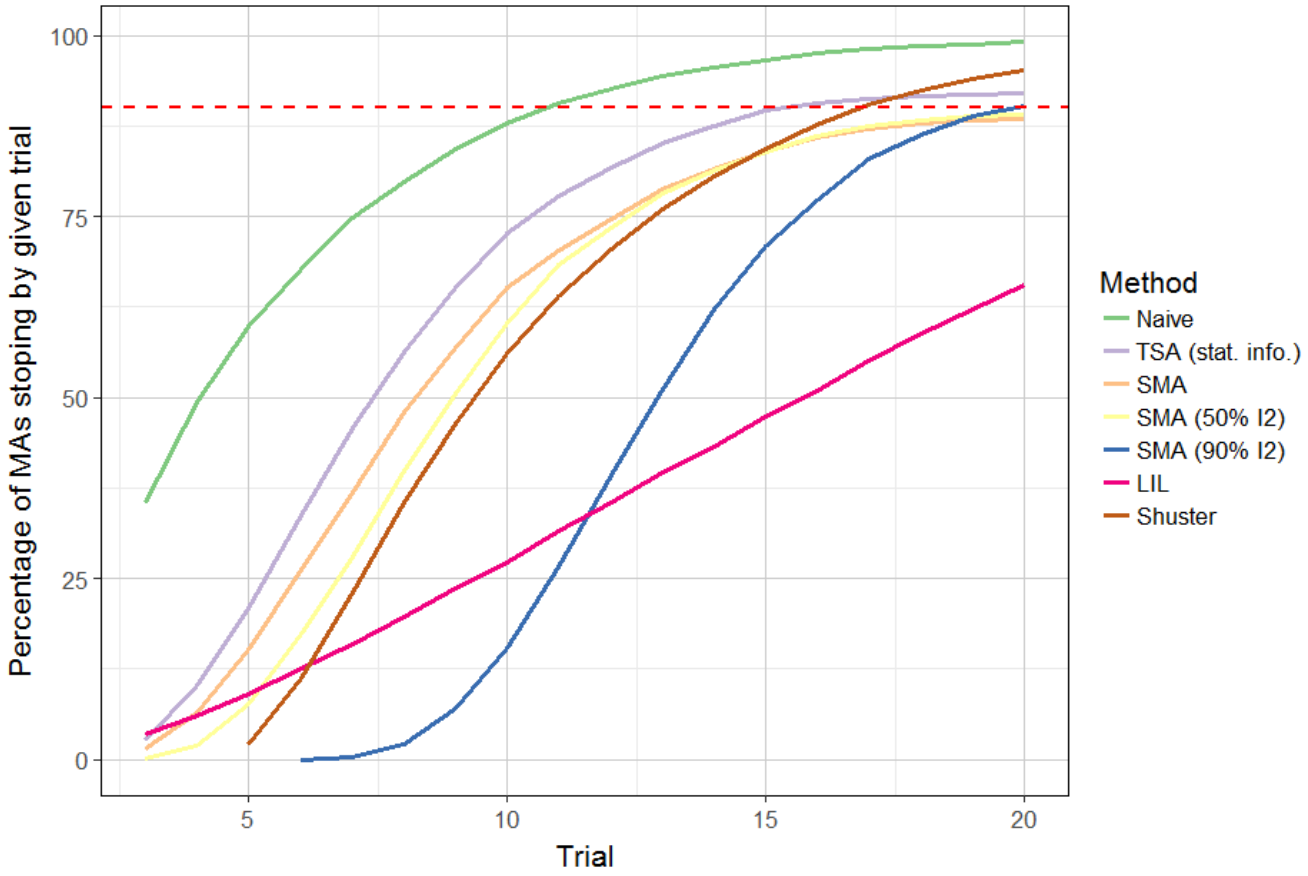


False positive rates – Type I error

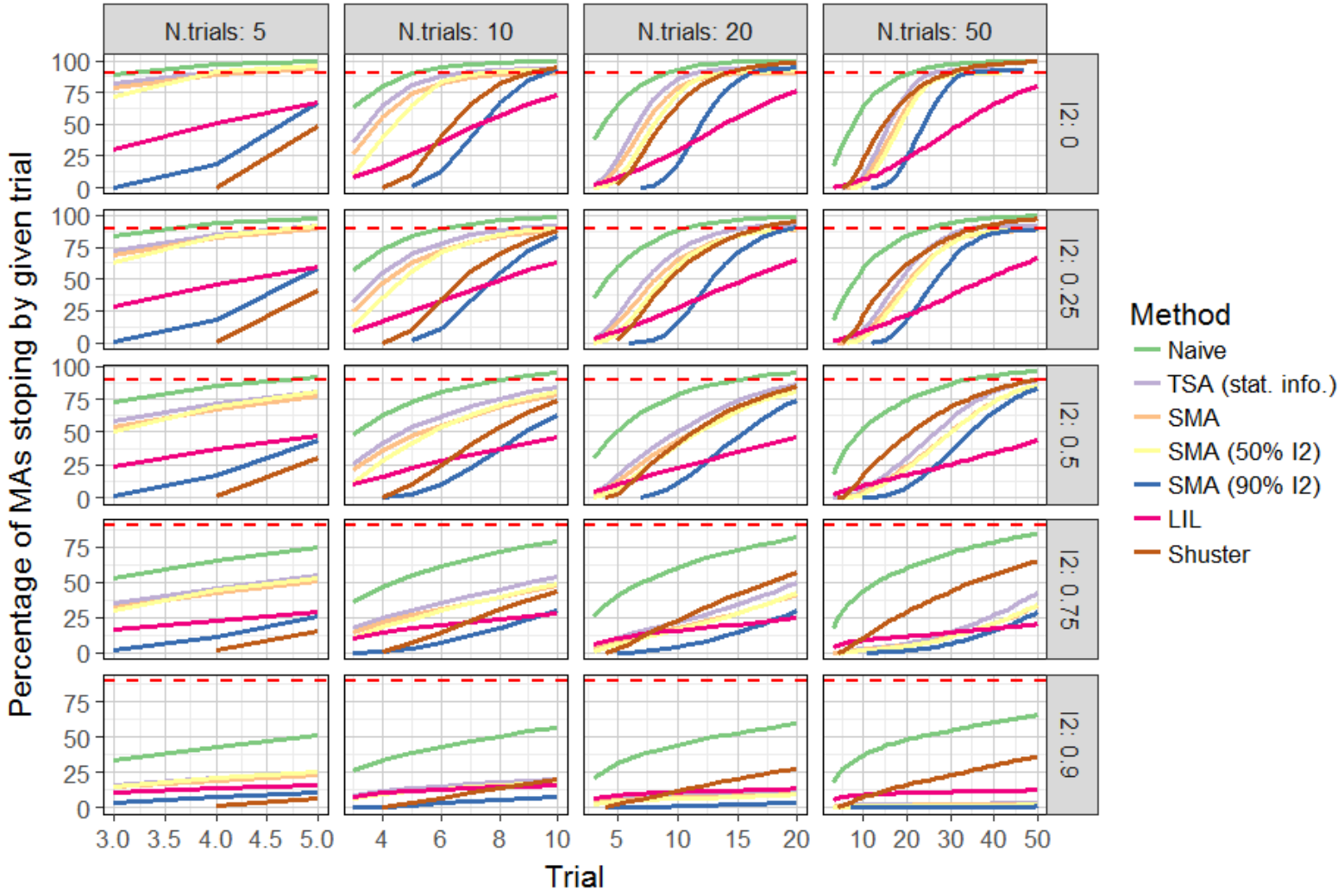


Cumulative power

- 20 trials / updates, I^2 = 25%



Cumulative power



76 Cochrane Reviews

- 76 Reviews: 286 meta-analyses
  - 68% binary data
  - Median 9 trials (IQR 6 to 14)

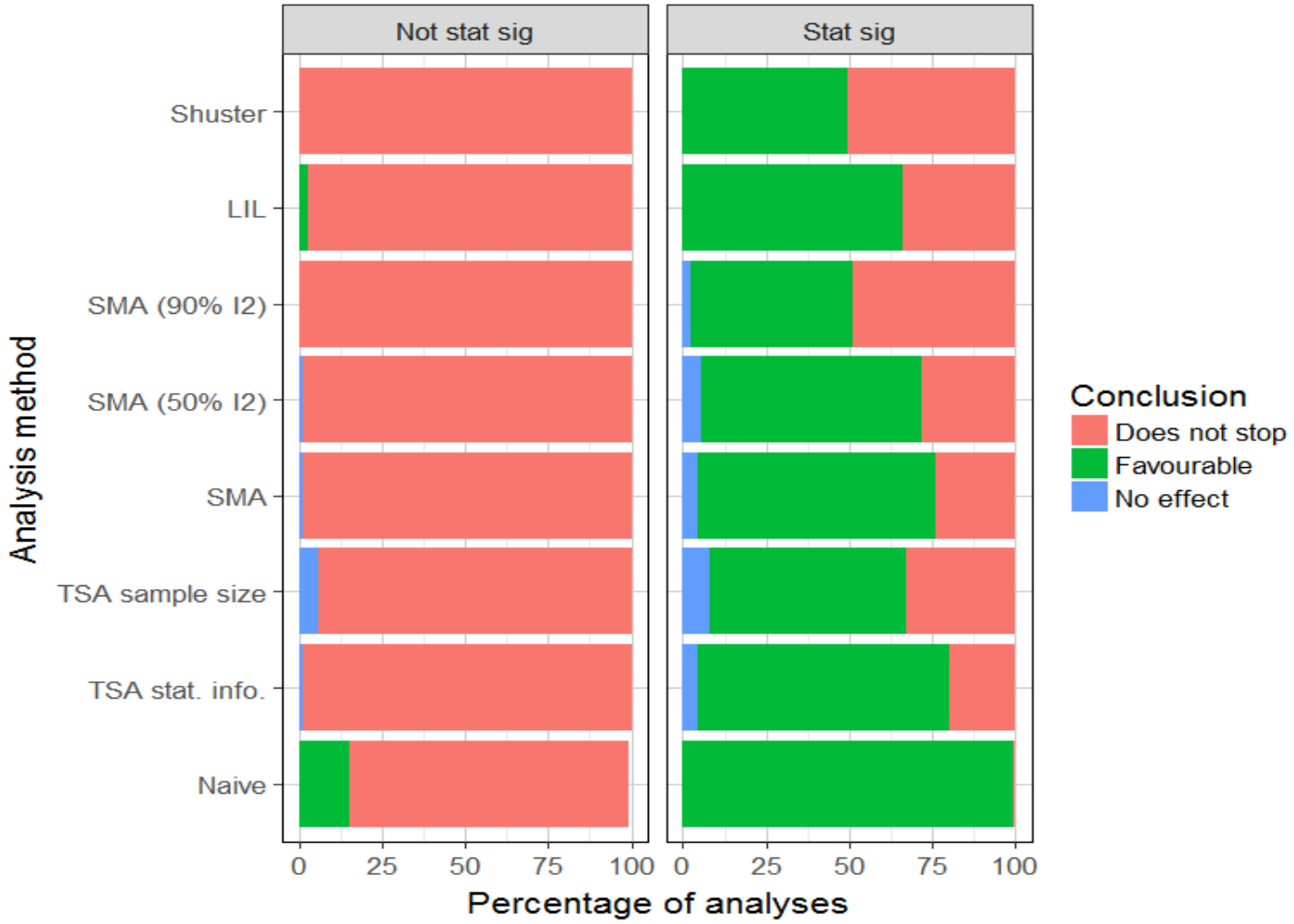
- 62% had a statistically significant result using conventional analysis

Sample size

Effect size	% of meta-analyses with power of at least:	
(5% Type I error)	**80%**	**90%**
As observed	43.5%	38.6%
1	71.6%	67.0%
0.5	43.9%	36.2%
0.25	23.5%	20.4%

- Most reviews are underpowered
- Waiting for required sample size is not realistic

Conclusions of analyses



Conventional “Naïve” 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

Do we need sequential methods?

- Is the problem with “naïve” analysis serious enough in real Cochrane reviews?
- Or in Living Systematic Reviews?
- Do the methods needlessly delay a statistically significant result?

When should they be implemented?

- At protocol stage in all reviews?
- At first update?
- Only once a statistically significant result is found?
- Only when evidence is limited?
  - E.g. small total sample size