Statistical methods for reliably 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?

- When new studies emerge?
- When new data might alter our conclusions?

- Updating is time-consuming

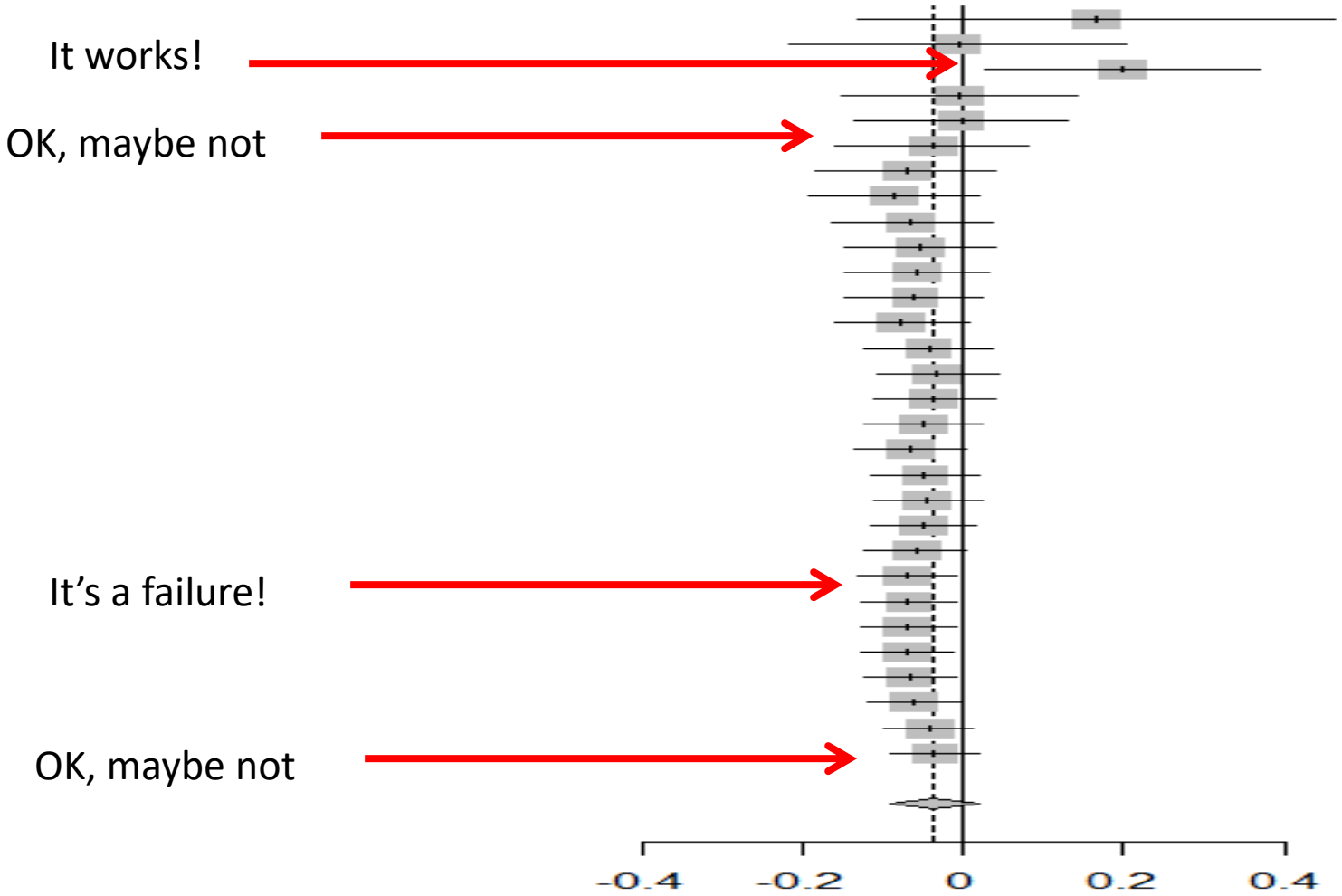
Some issues

- When can we stop updating?
- Which meta-analyses should have
- priority for updating?
- Conclusions can change over time
  - Risk of error if we stop too soon

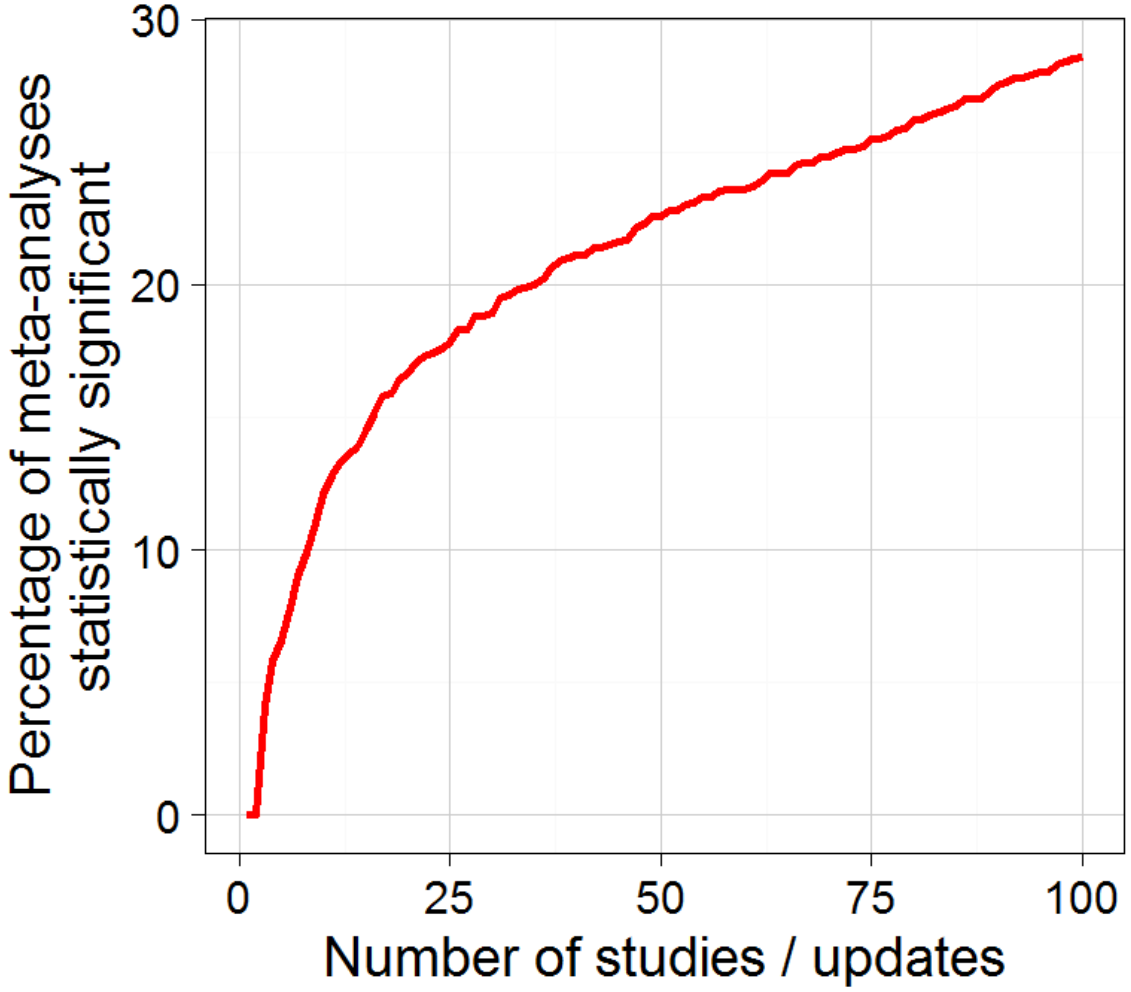
Type I error

- Assuming an intervention is effective when it isn't
- Usually set at 5%
- Increases the more updates we perform
- Can we accept a conventionally "statistically significant" meta-analysis?

Cumulative meta-analysis



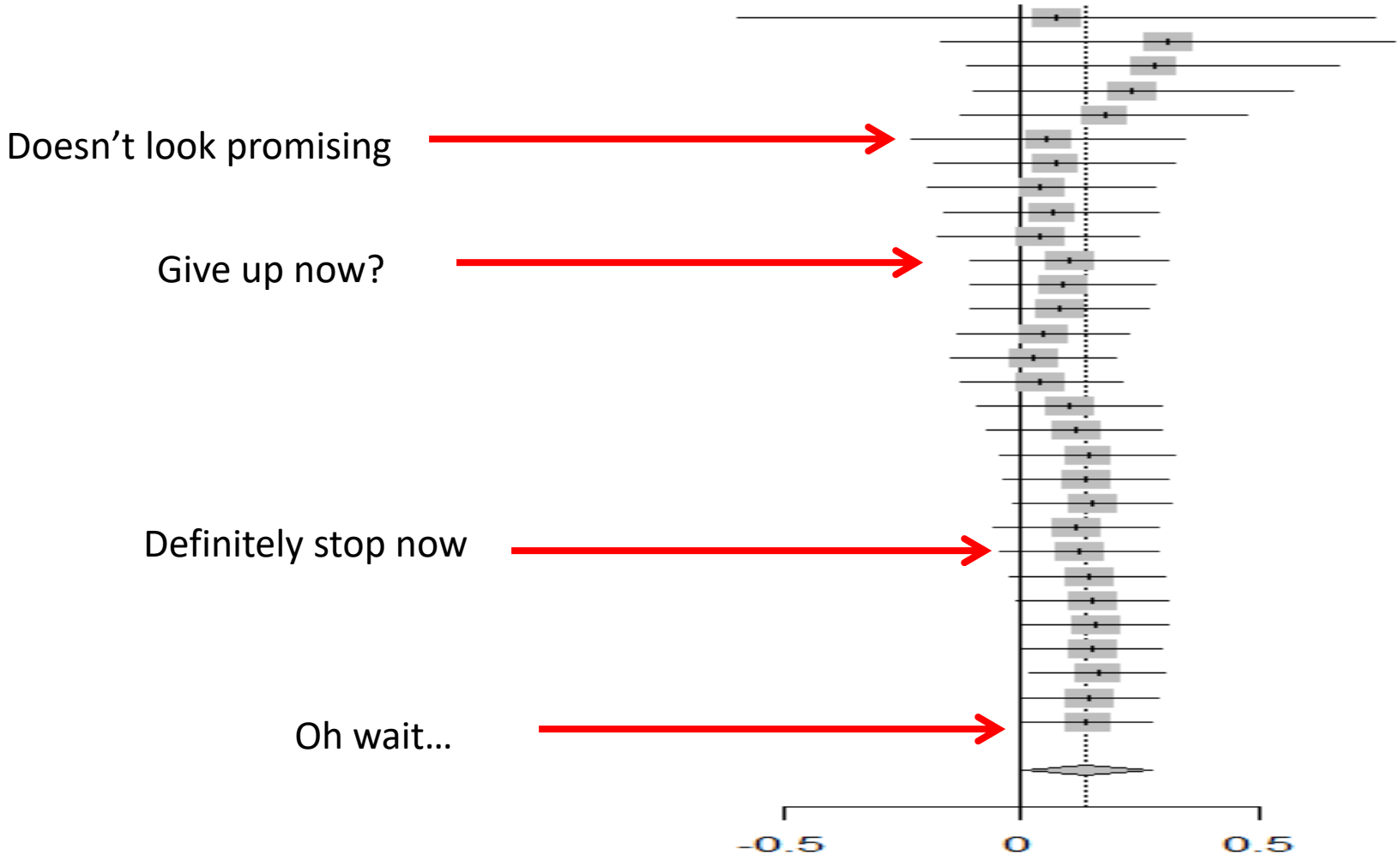
Type I error in meta-analyses



Type II error

- Assuming an intervention isn't effective when it is
- Not controlled in a meta-analysis
- When can we stop updating non-significant meta-analyses?

Cumulative meta-analysis



A caveat

- The summary ***effect estimates*** (and confidence intervals?) are valid at each update

- ***Decisions*** made on the basis of the results may not be
  - Particularly decisions about whether to update

Parallels with sequential trial design

- Aim to stop a ~~trial~~ as soon as possible
  review
- Select a desired Type I and II error rate
- And desired clinical effect

- Perform ~~interim~~ analyses throughout ~~trial~~
  Meta- review

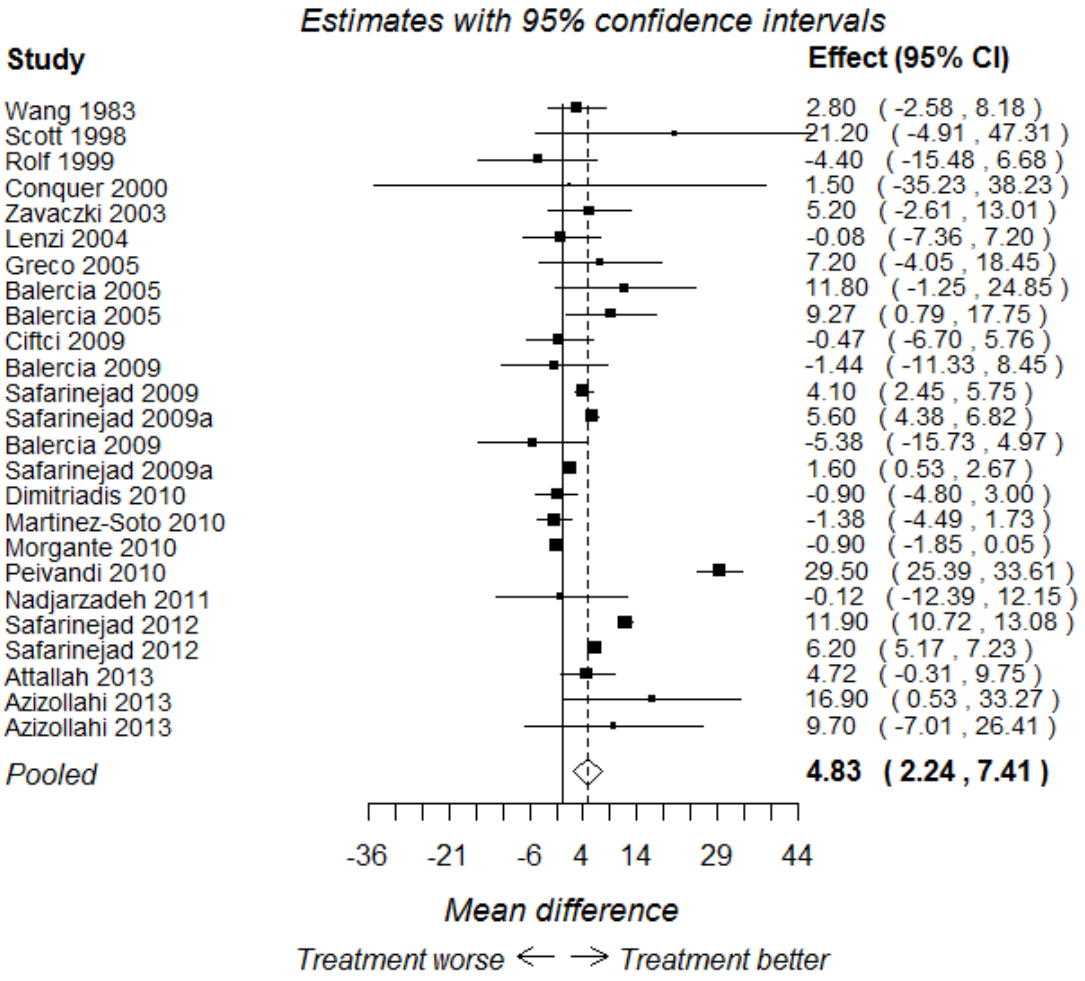
Key differences

- Meta-analysis is not controlled
  - No control over timing of studies
  - Size of studies
- Heterogeneity
  - Studies have different protocols
  - Estimated effects may not be consistent

Controlling error

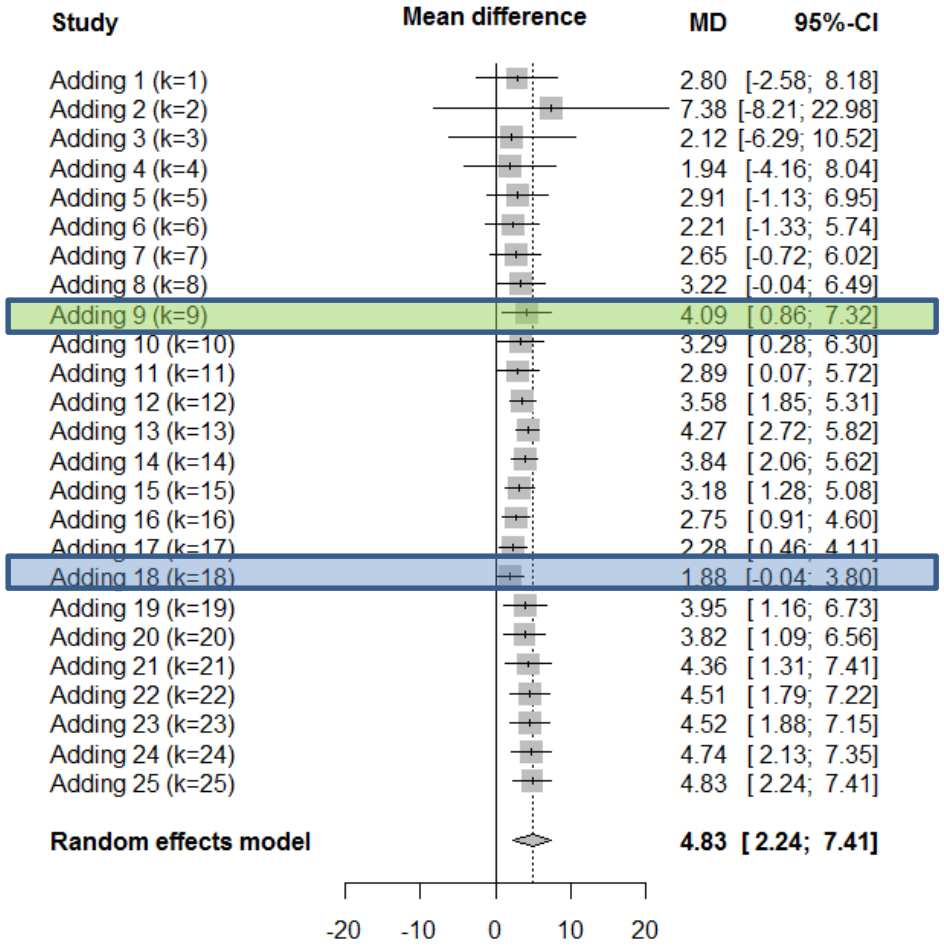
- Control Type I and Type II error
  - Sequential meta-analysis
  - Trial sequential analysis
- Control Type I error
  - Law of Iterated Logarithm
  - “Shuster-Pocock” method
- Other methods
  - Fully Bayesian analysis
  - Robustness or stability of analysis
  - Consequences of adding new studies
  - Power gains from adding new studies

Example from Cochrane



$I^2 = 95\%$

Cumulative meta-analysis



Sequential meta-analysis (SMA)

Higgins, Simmonds, Whitehead 2010

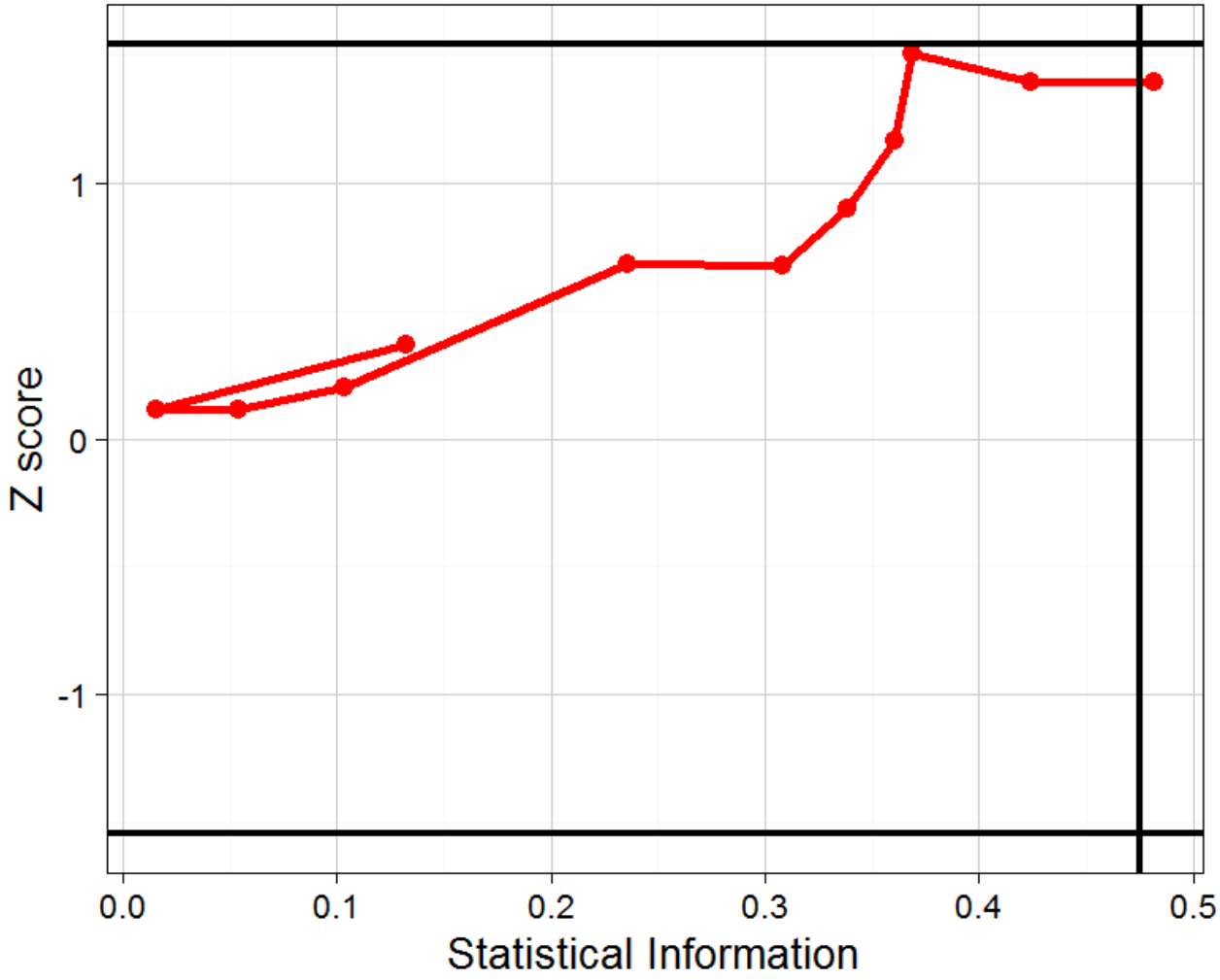
- Calculate cumulative Z score and cumulative Information for each updated meta-analysis

- Stop when a pre-specified boundary is crossed
- Boundary designed to control type I and II error

Accounting for heterogeneity

- Select a prior estimate of heterogeneity
  - Generally assuming high heterogeneity
- Use Bayesian methods to calculate posterior heterogeneity estimate at each update
- Use this Bayesian estimate in the updated meta-analysis

Sequential meta-analysis



Trial sequential analysis (TSA)

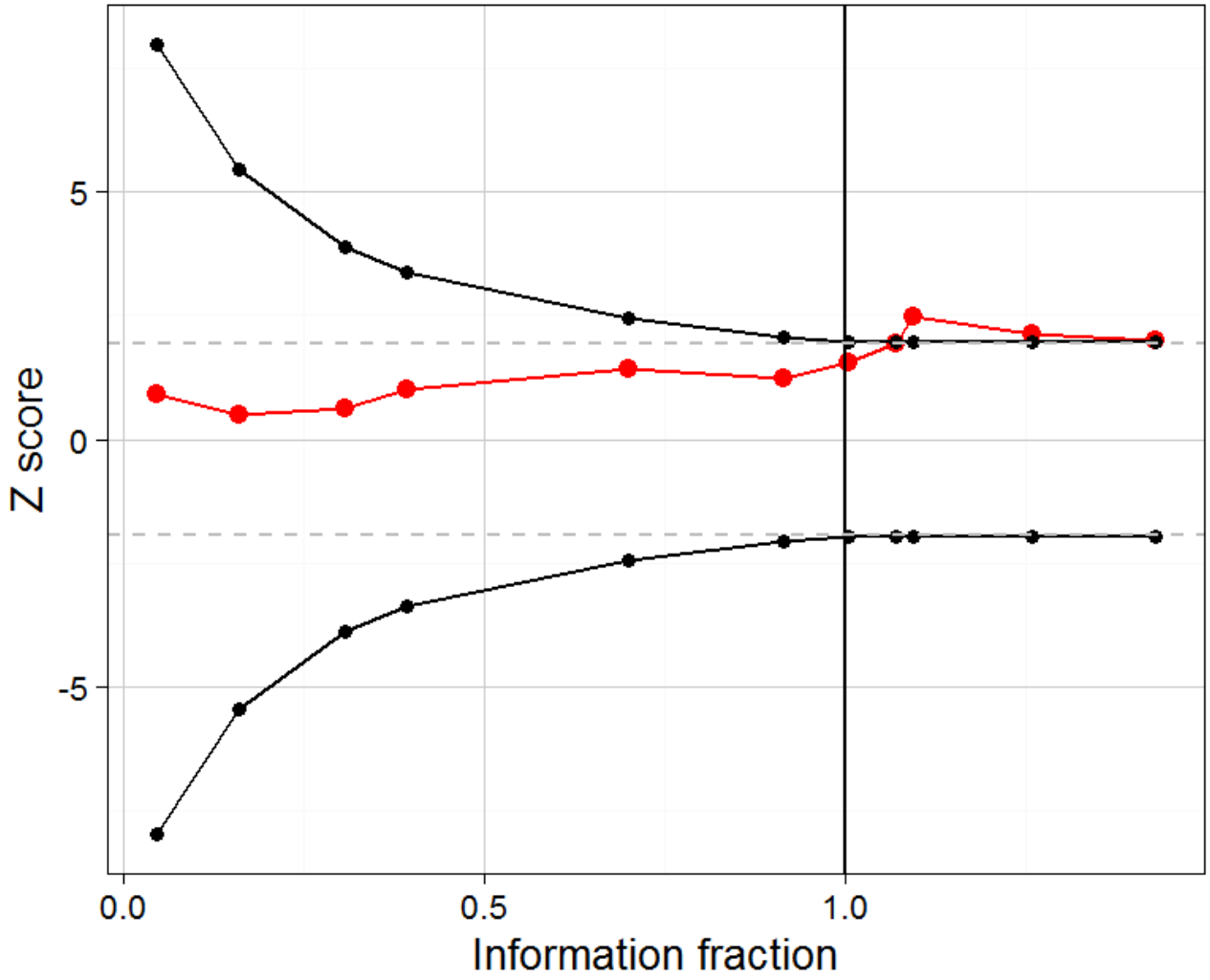
Wetterslev, Thorlund, Brok, Gluud 2008

- Select a 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



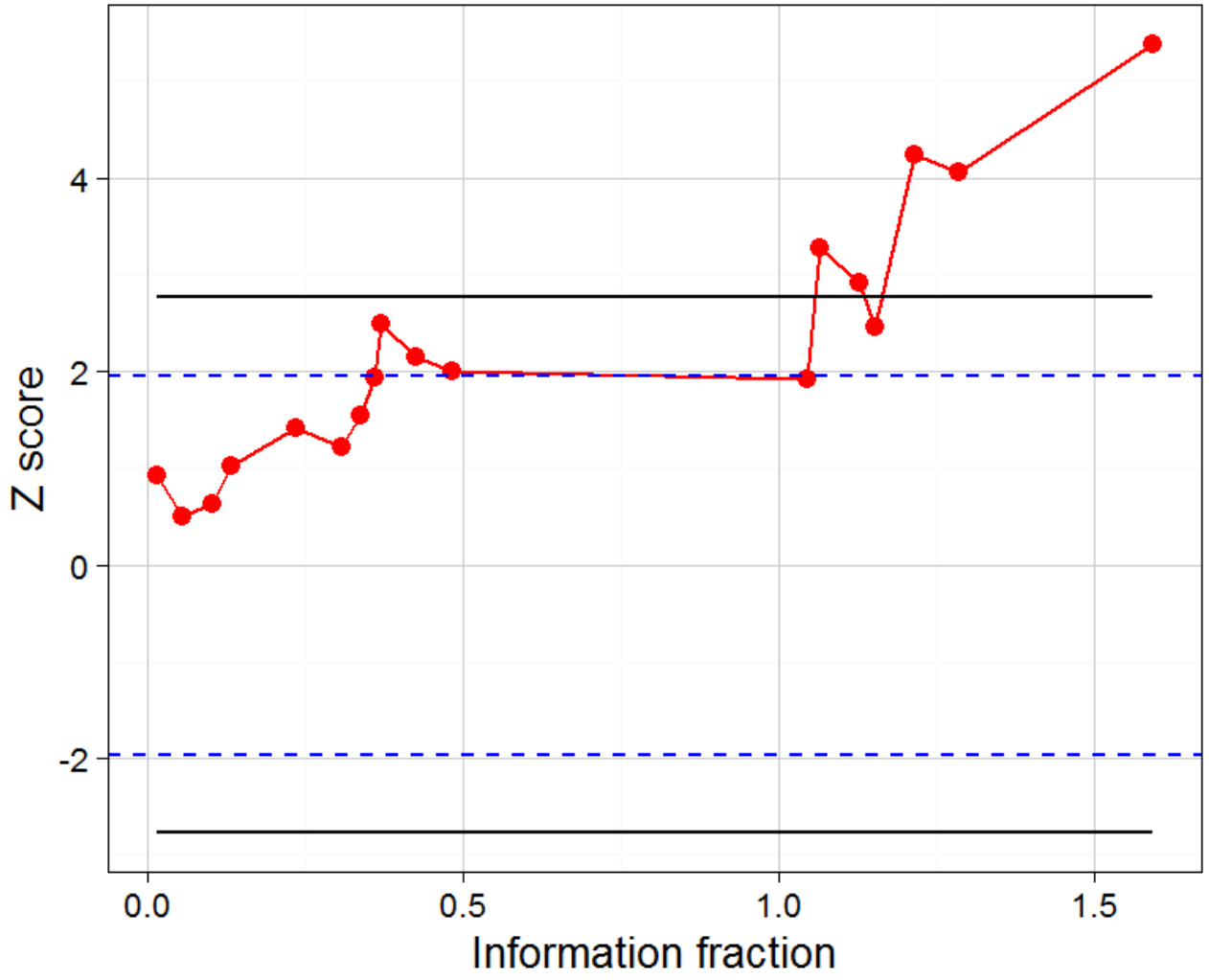
Law of Iterated Logarithm (LIL)

Lan, Hu, Cappelleri 2007

- Uses an adjusted Z statistic
- $Z^* = \frac{Z}{\sqrt{\lambda \log(\log(N))}}$
- This is bounded as $N \to \infty$
- So controls Type I error

- Commonly sets $\lambda = 2$

Example

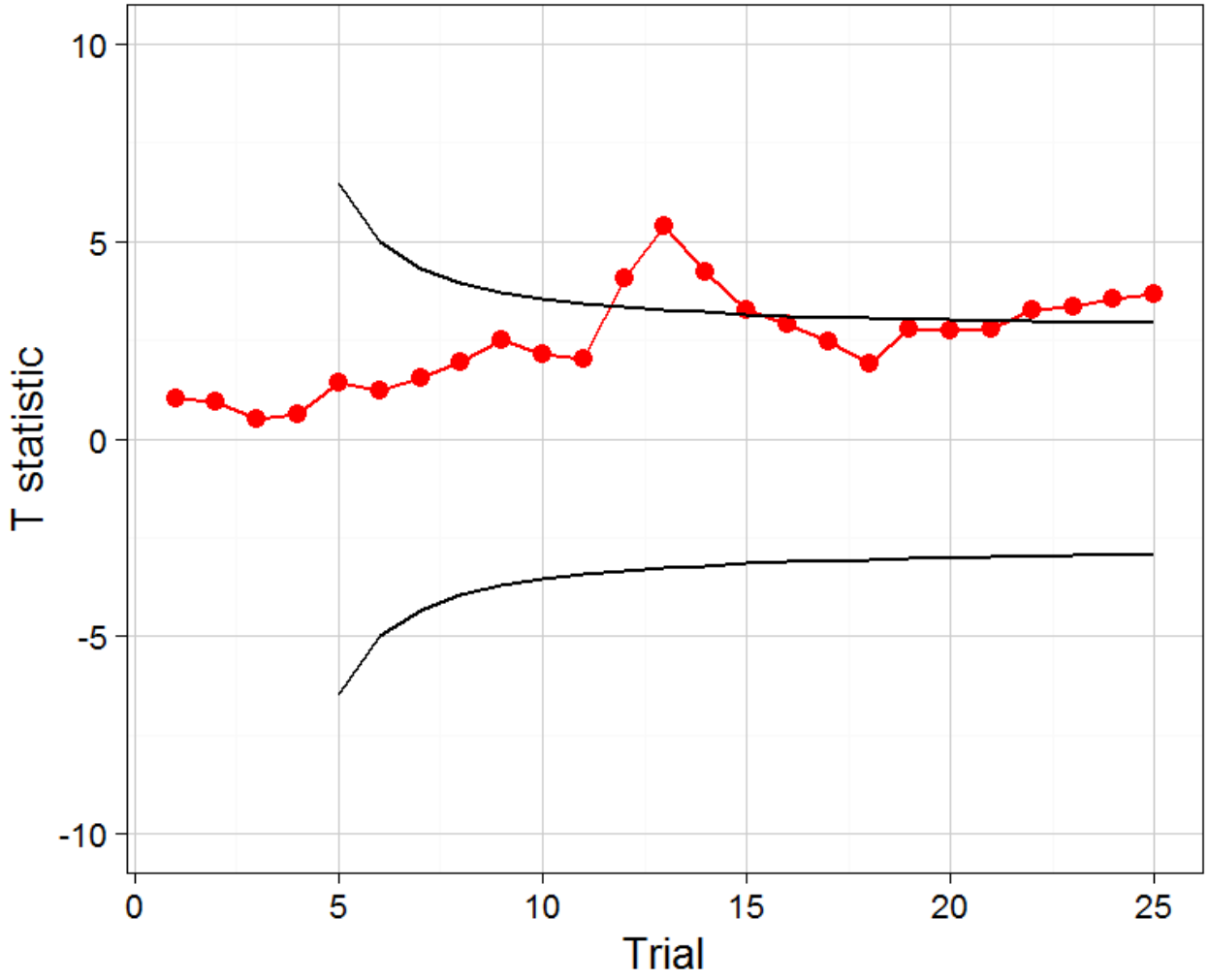


Shuster-Pocock method

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

Example



76 Cochrane Reviews

- 76 Reviews: 286 meta-analyses

Binary outcome	**194 (68%)**	**Continuous outcome**	**92**
Stat. sig.	178 (62%)	**Not stat. sig.**	108
Trials per MA	Median 9	IQR: 6 to 14	Max: 200
Effect size *	Median 0.47	If stat sig. 0.69	If not 0.25
I^2		$I^2 = 0$: 32%	$I^2 > 90\%$: 7.0%
		If stat sig. 46%	If not: 13%

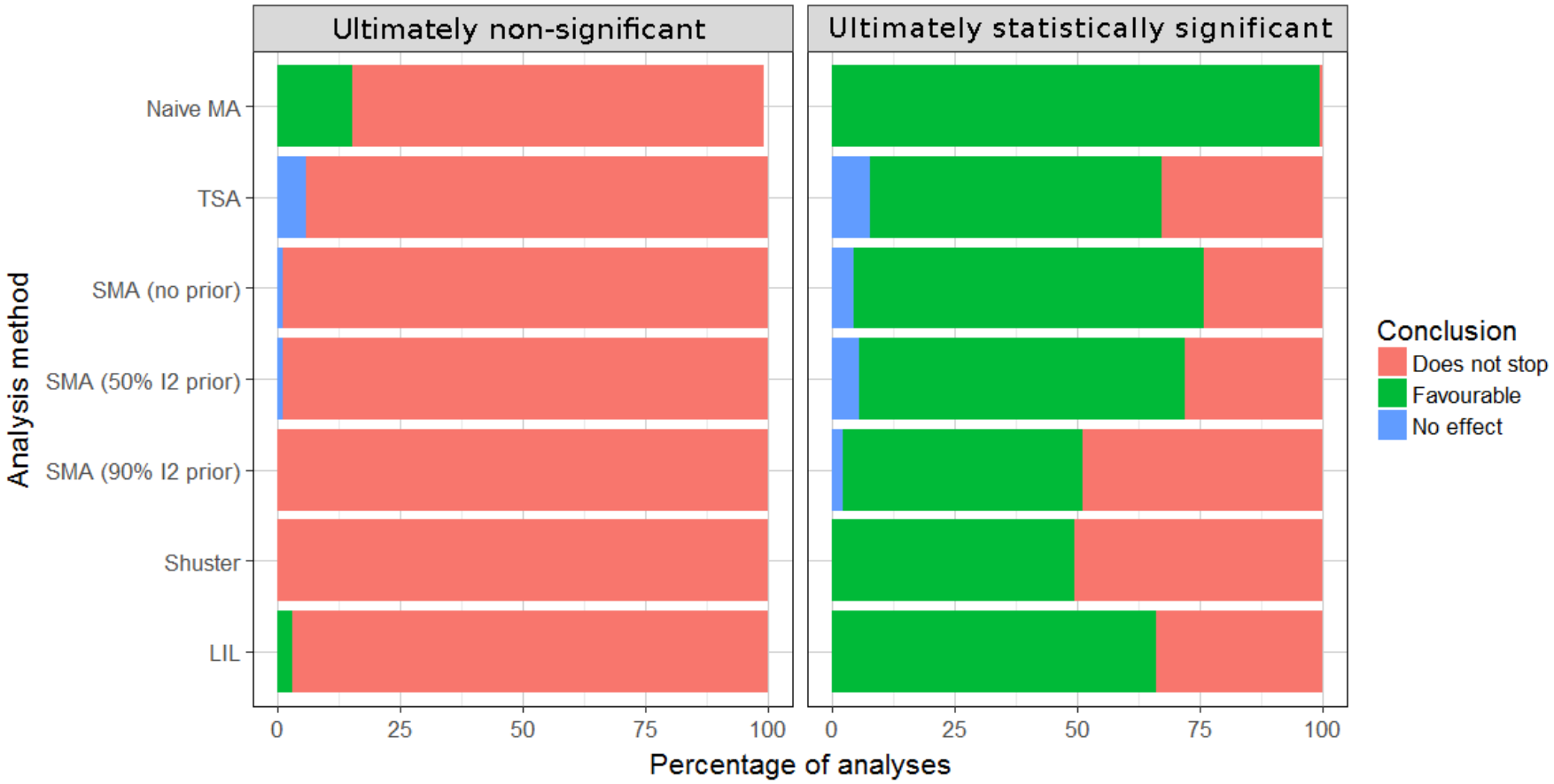
** Log odds ratio or standardised mean difference*

Applying meta-analysis updating methods

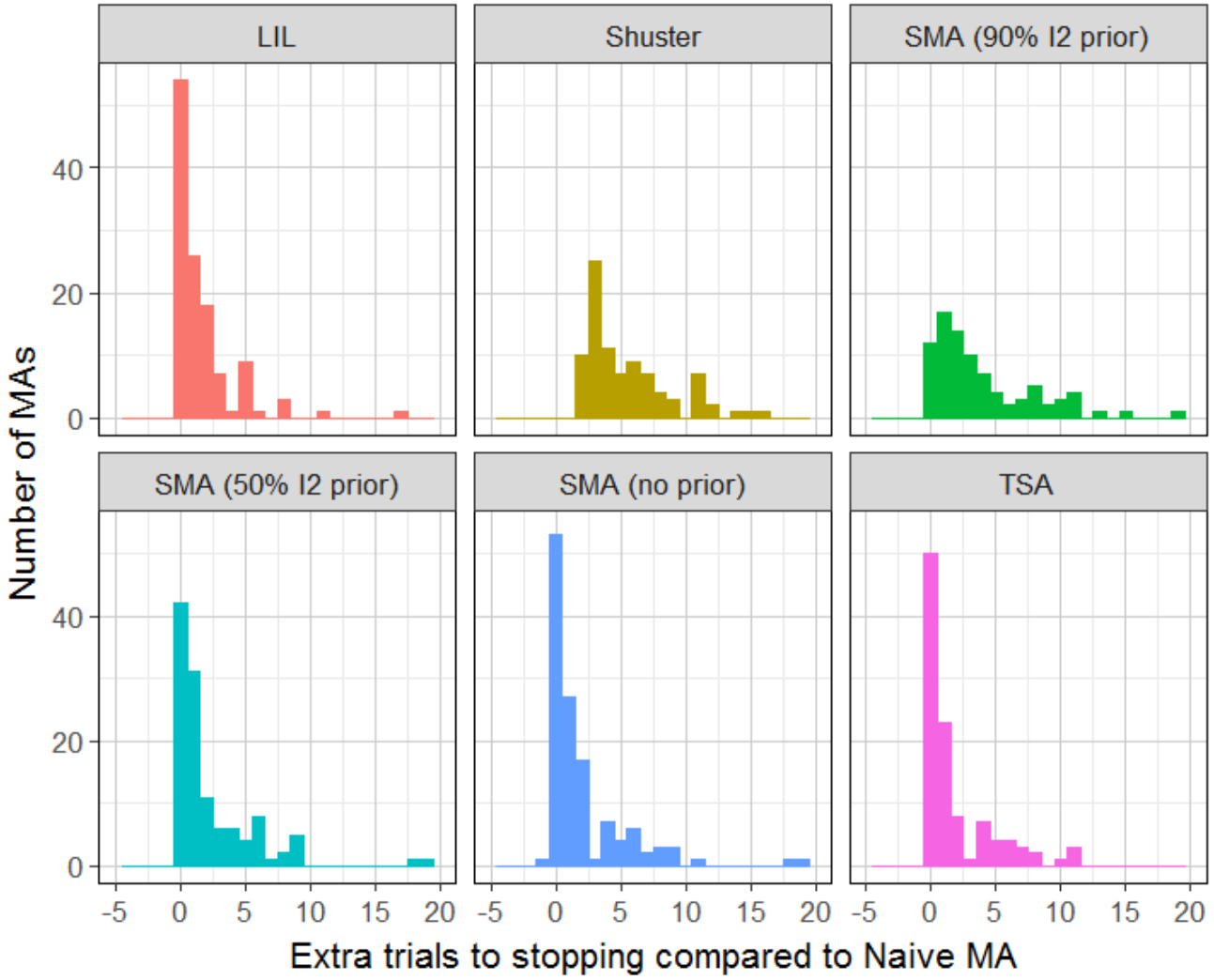
- Apply to all 286 meta-analyses:

- "Naïve" cumulative meta-analysis
- Trial sequential analysis
  - (heterogeneity adjusted)
- Sequential meta-analysis
  - With no prior, 50% I^2 and 90% I^2 priors
- Law of iterated logarithm
- Shuster-Pocock

Conclusions of analyses



Extra trials required to reach a conclusion



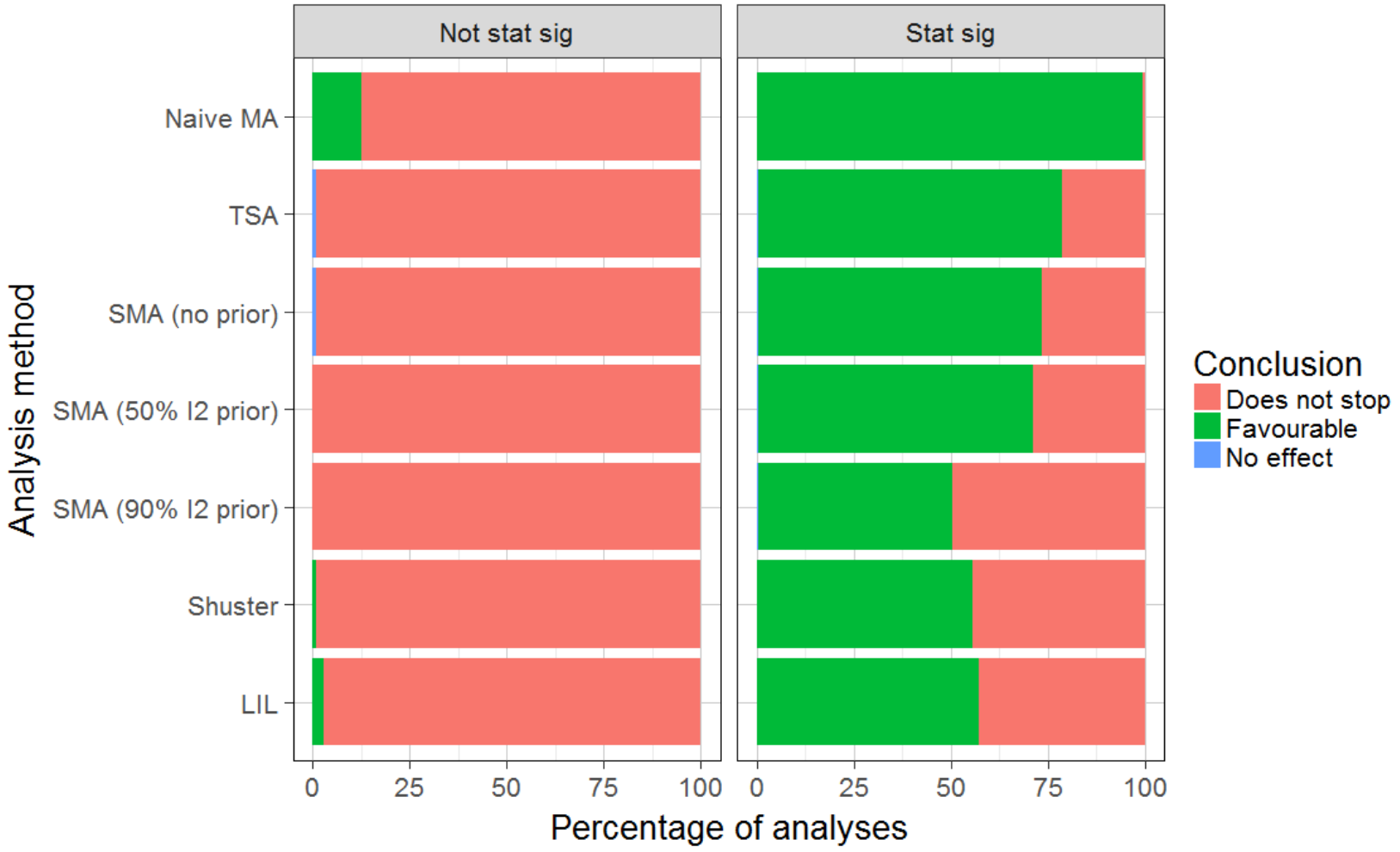
Realistic review updating

- Have assumed a new meta-analysis after each new trial

- In reality updates are less frequent
- First analysis will have good proportion of total trials

- Re-analyse assuming updates once 50%, 70%, 90% and 100% of trials are available

Conclusions using realistic updating



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%

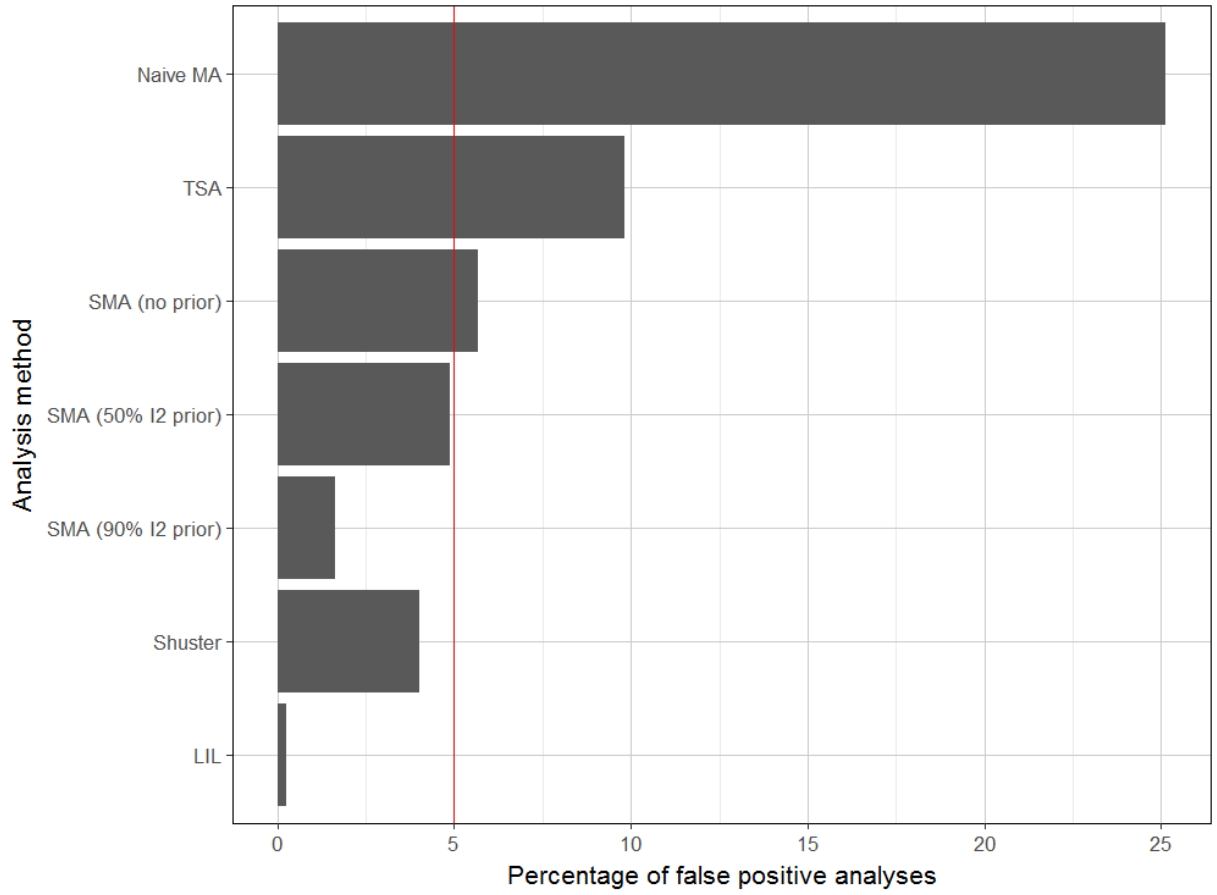
- Fixed total sample size of 9000
  - 90% power to detect effect of 0.1 if $I^2 = 50\%$

Methods applied

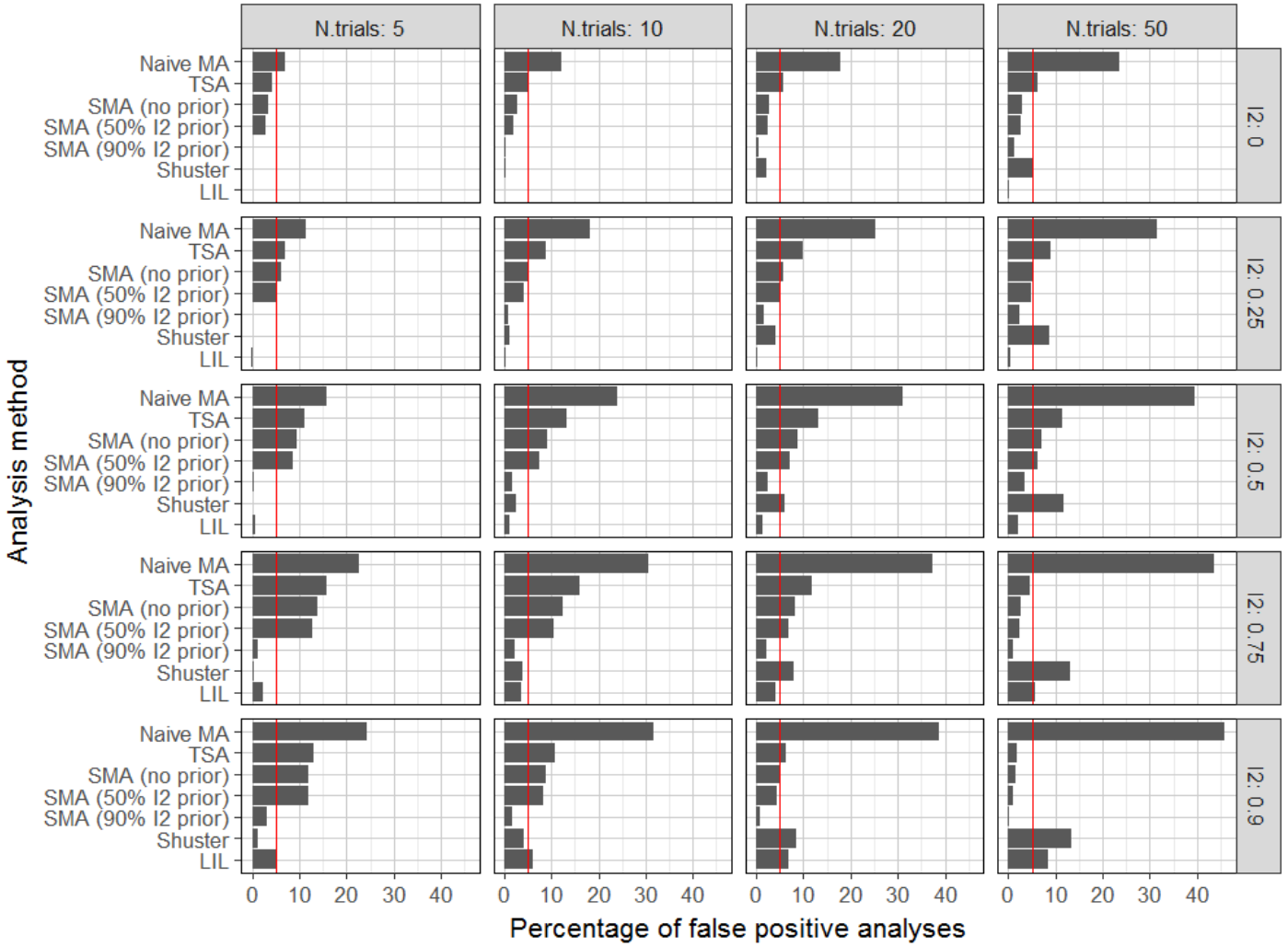
- Naïve analysis (standard cumulative MA)
- Trial Sequential Analysis (TSA)
- Sequential Meta-Analysis (SMA)
  - No prior heterogeneity
  - Prior I^2 of 50% or 90%
- Law of Iterated Logarithm (LIL)
- Shuster method

False positive rates – Type I error

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

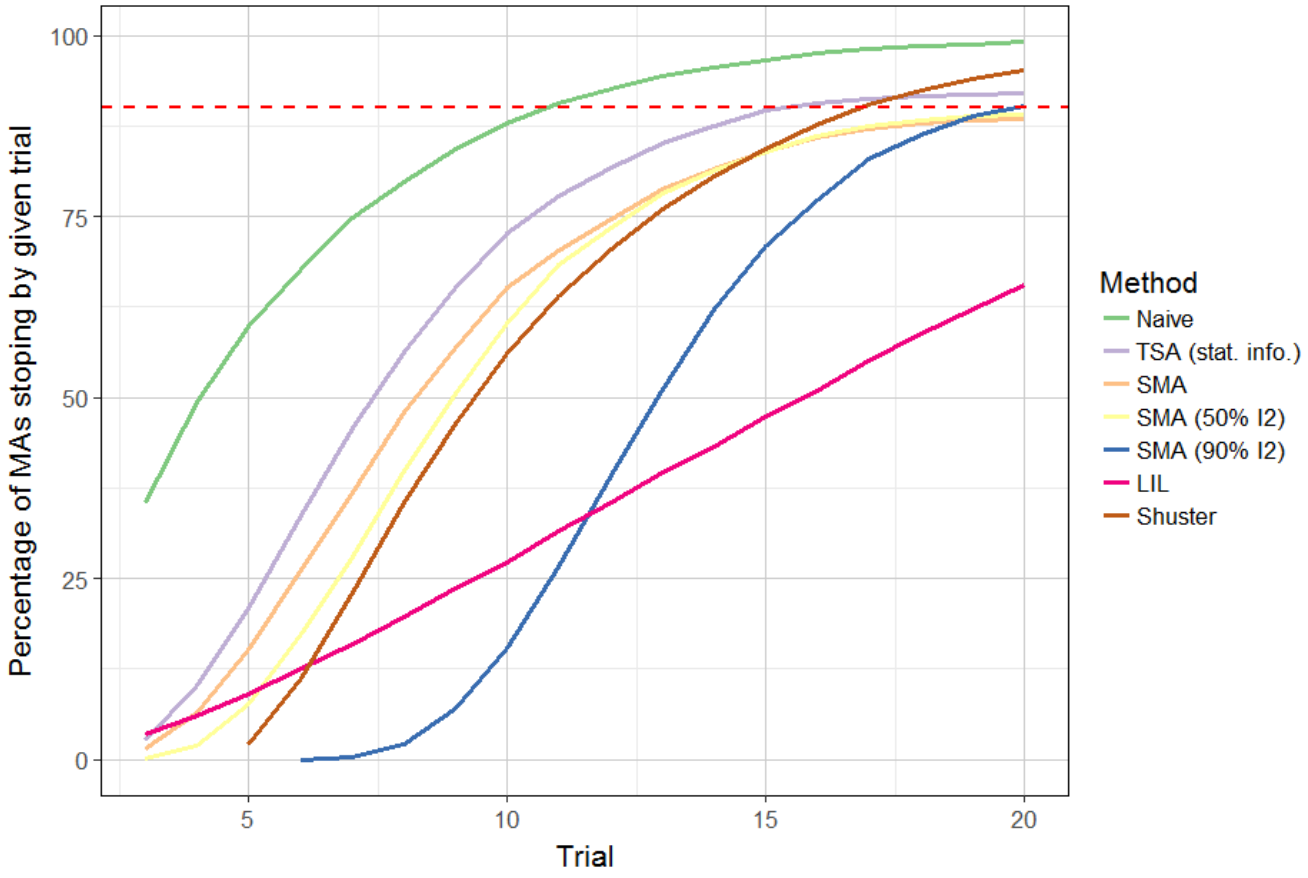


False positive rates – Type I error

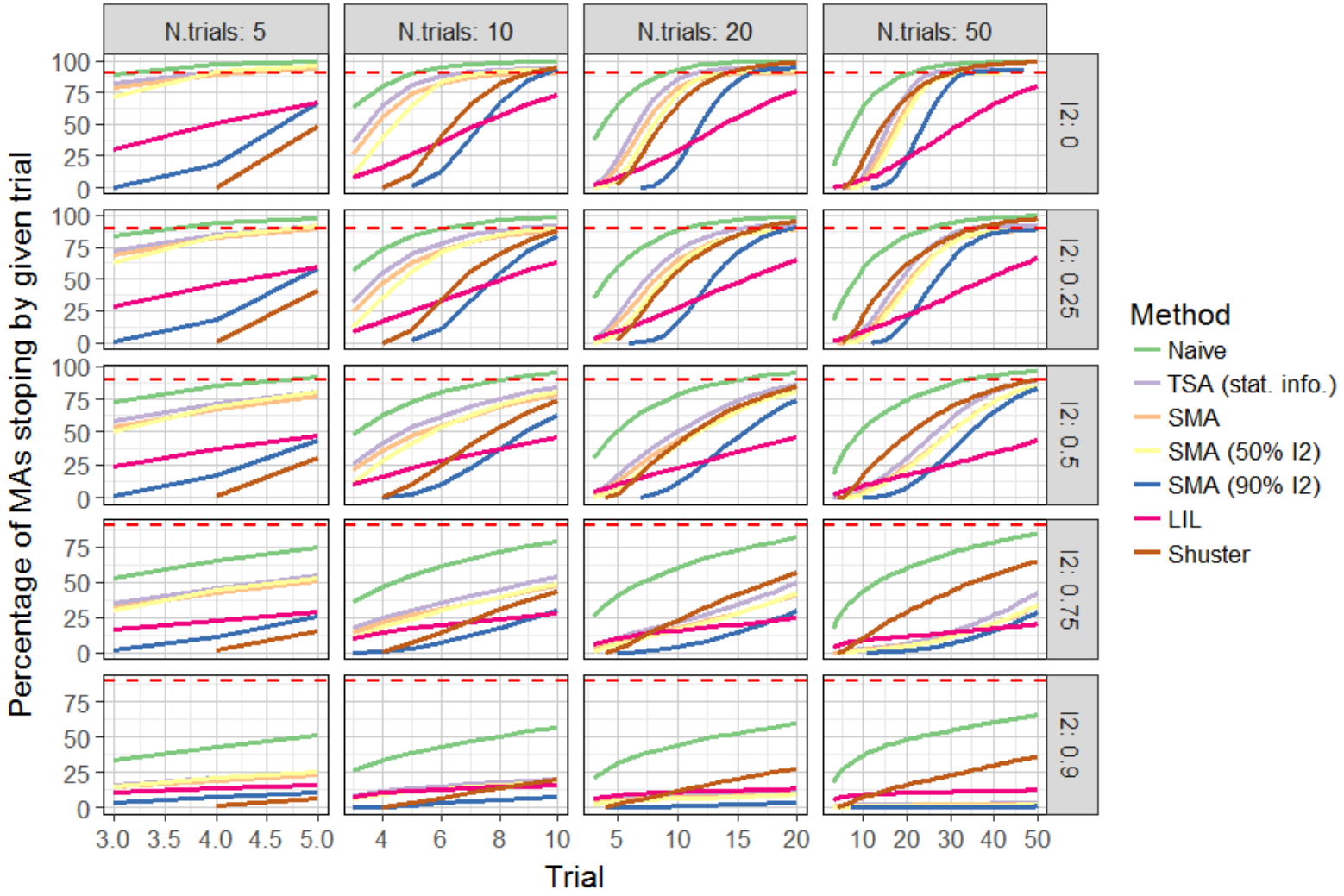


Cumulative power

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



Cumulative power



Conventional “Naïve” analysis

- Too many inappropriate positive conclusions
  - Elevated Type I error rate
  - But not vastly elevated for most updated reviews?
- Biased estimates of effect
- Half of all analyses showing significant results are based on too little evidence?

Trial Sequential Analysis

- Controls for Type I and II error

- Need to set desired effect
- Complex to run
- Required sample size varies with time
  - Can lead to inconsistent updates

Sequential Meta-Analysis

- Controls for Type I and II error

- Need to set desired effect
- Complex to run
- Statistical information not intuitive
- Limited choice of boundaries

- Bayesian heterogeneity too conservative?
- Not needed in practice?

Law of Iterated Logarithm

- Controls for Type I error
- Easy to implement

- Biased estimates of effect at stopping?
- Over-conservative: low-power
- Uncertainty over λ parameter

Shuster-Pocock

- Controls for Type I error
- Fairly easy to implement

- Needs more trials before stopping
- Need to pre-specify number of updates?
- Needs many studies to have adequate power

Do we need these methods?

- Is the problem with “naïve” analysis serious enough in real Cochrane reviews?
- Do the methods needlessly delay a statistically significant result?
- Too much focus on decision making over estimation?
- More complex than necessary?

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

What are Cochrane reviews for?

- To present the best evidence at the current time?
- To aid in making medical decisions or guiding future trials?