Reflections on quantifying heterogeneity (and more)

Ian White <ian.white@ucl.ac.uk>

MRC Clinical Trials Unit at UCL

A Day with the Cochrane SMG

11 May 2021

Heterogeneity is probably always there

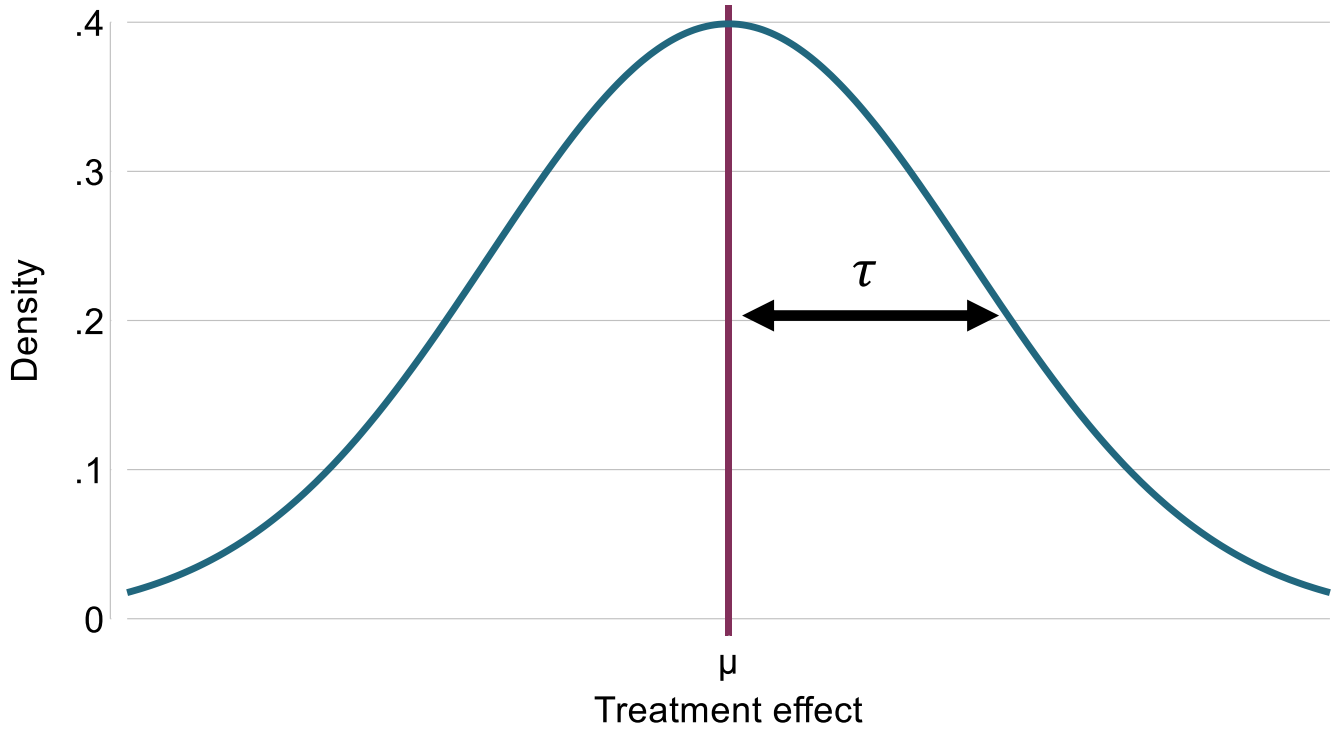
- Heterogeneity was chillingly obvious in Celina's review of meta-analyses (MAs) of prevalence
- It is less obvious in MAs of treatment effects, but still sure to be present
  - why would studies all estimate exactly the same thing?
- Specifically, $\hat{\tau}$ may be 0, but τ isn't

Can we explain heterogeneity?

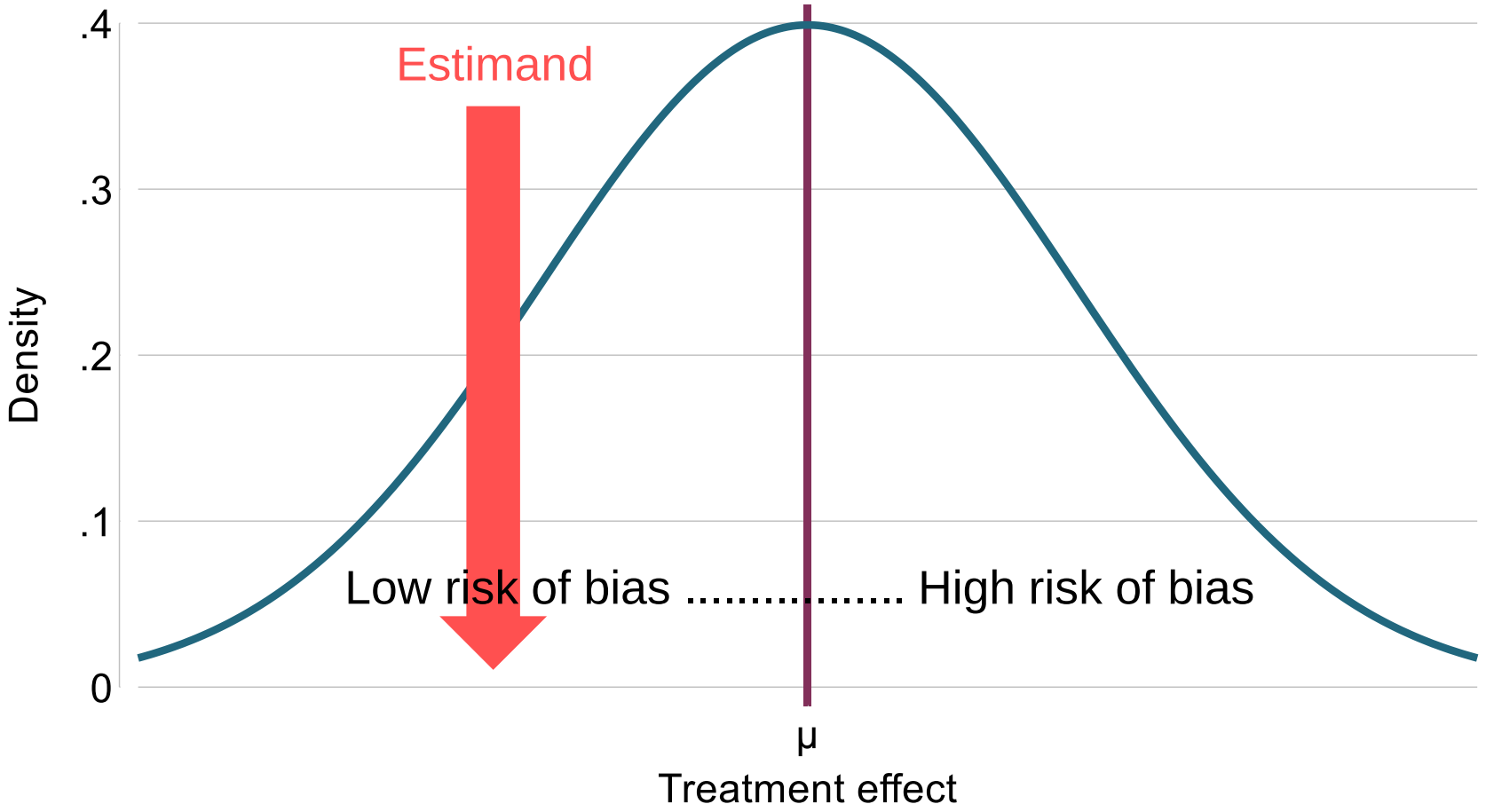
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351–5.
  - "Although meta-analysis is now well established as a method of reviewing evidence, an uncritical use of the technique can be very misleading. One common problem is the failure to investigate appropriately the sources of heterogeneity, in particular the clinical differences between the studies included."
- Interesting that 61% of Celina's SRs tried to explain heterogeneity
- Important but difficult
- Most credible if pre-specified

What's the estimand in the presence of heterogeneity?

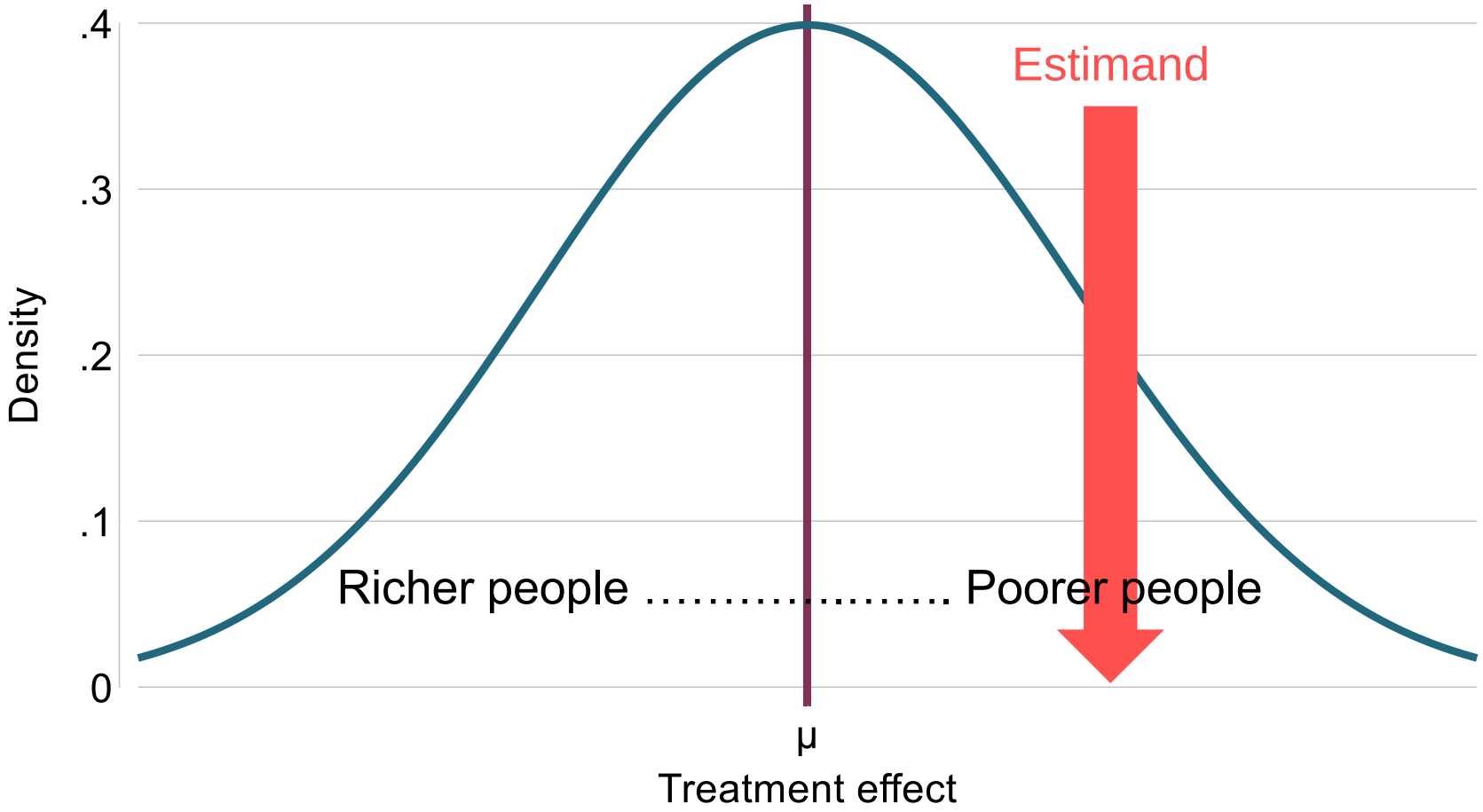
- Estimand – the thing we are estimating
- Move beyond focus on parameters in statistical models
- For example in random-effects MA:



Explained by study risk of bias?

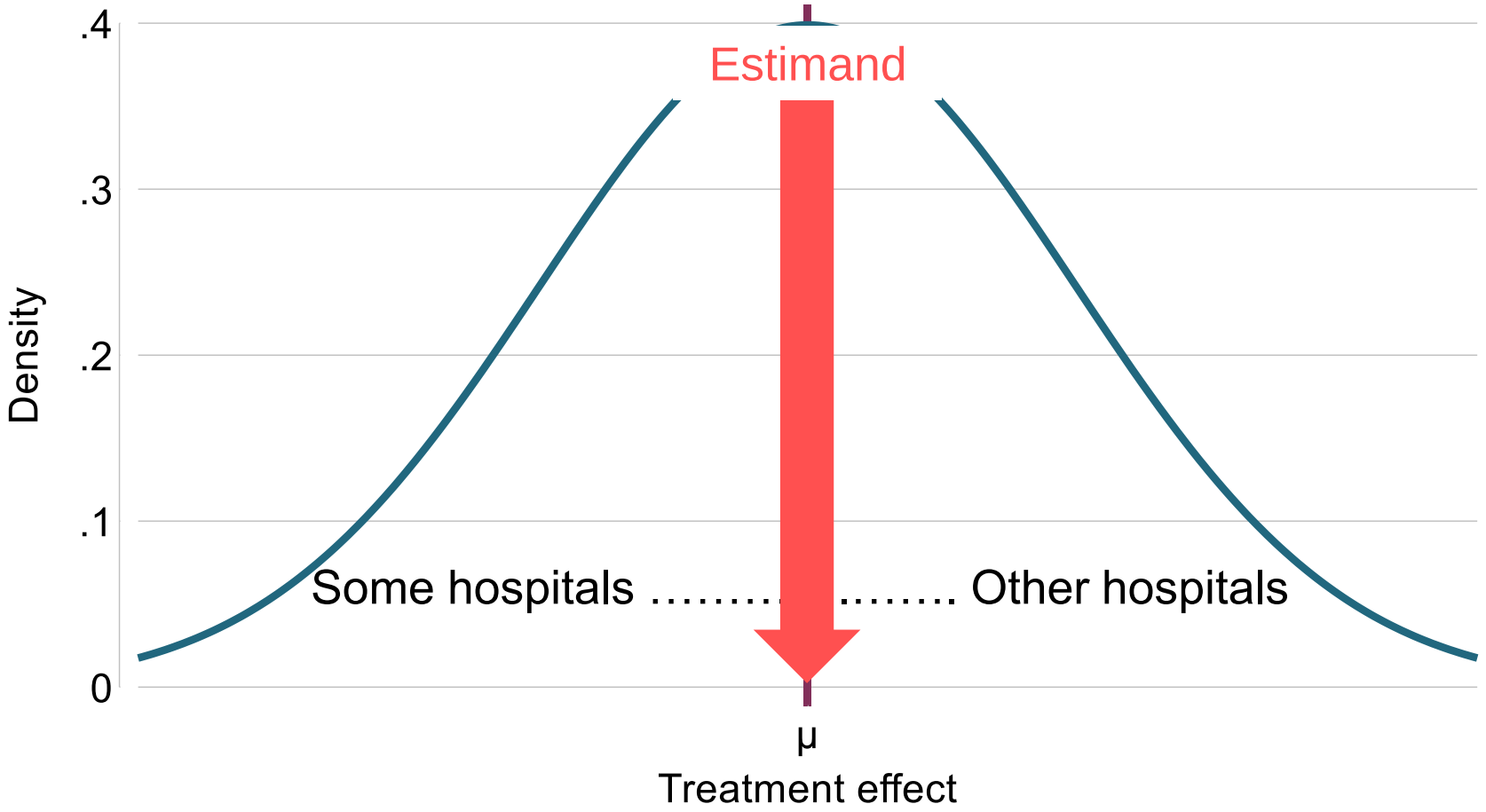


Explained by socio-economic status?



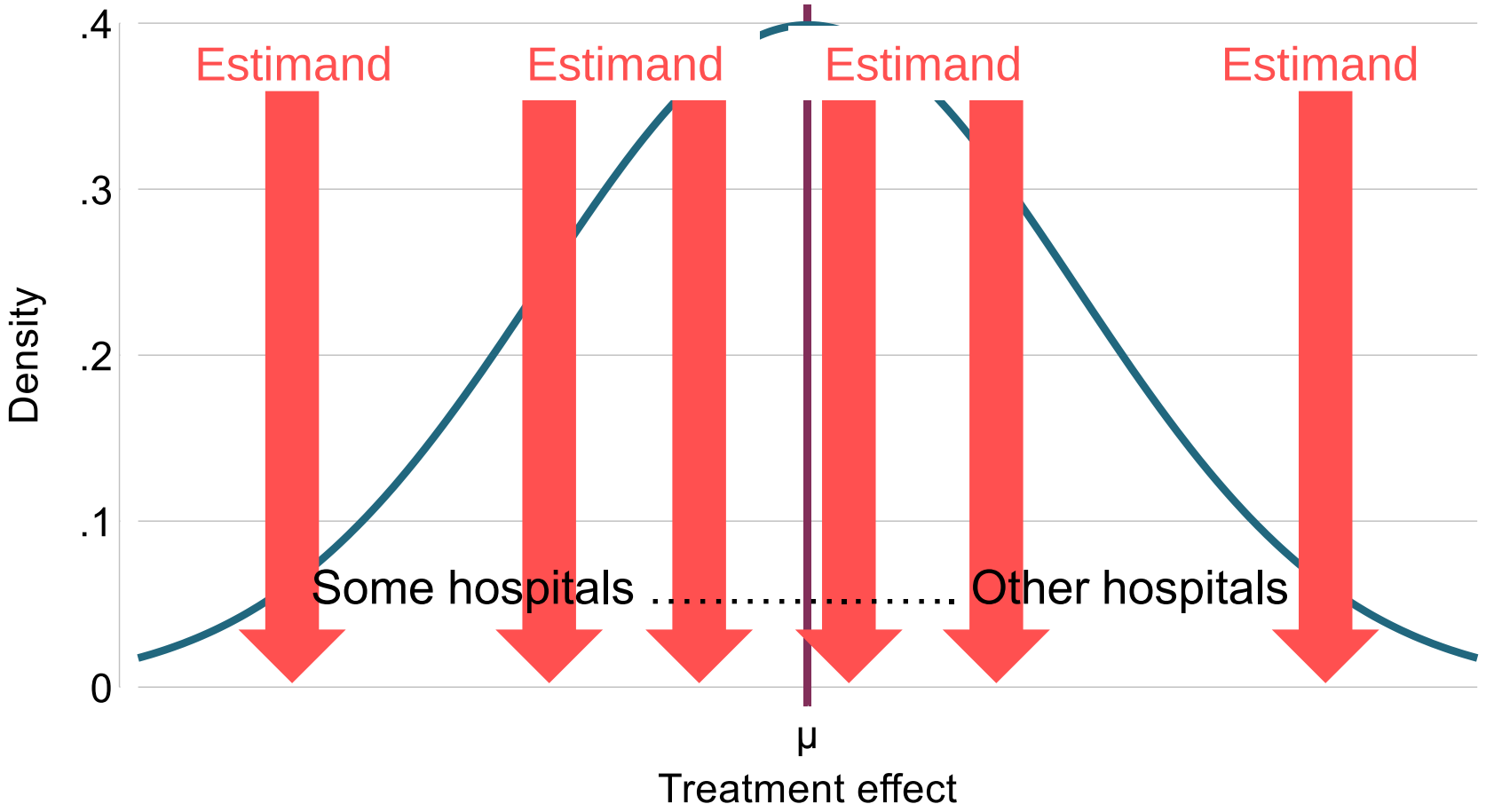
Not explained?

National perspective?



Not explained?

Local perspective?



How should we account for heterogeneity in estimation?

- Ralf's approach – apply a prior to τ
- Excellent compromise on uncertainty in $\hat{\tau}$ between DL (no uncertainty) and HK (from t-distribution)
- Challenge is to produce a prior that can't be disputed
  - hence Ralf's empirical approach is valuable
- Second challenge is to account for estimation error in $\hat{\tau}$: difference between
  - distribution of τ (wanted)
  - distribution of $\hat{\tau}$ conditional on $\hat{\tau} > 0,\ \hat{\tau}$ n.s. (estimated)

  cf Turner et al's empirical approach did account for estimation error

How should we report heterogeneity?

- I^2 seems to be a statistic that methodologists criticise but can't live without (e.g. Karla's talk)
- Part of the appeal for non-statisticians is the classification of I^2, e.g. Cochrane Handbook:
  - 0 - 40%: might not be important
  - 30 - 60%: may represent moderate heterogeneity
  - 50 - 90%: may represent substantial heterogeneity
  - 75 - 100%: considerable heterogeneity

  cf Cohen's small, medium, large effect size = .2, .5, .8
- Predictive intervals are more closely tied to the review question: what is the treatment effect (in a new study)?

Things to ask questions about

- How should we develop Bayesian priors in order for policy-makers to believe them?
- Should we state the estimand in a systematic review?
- Is there still a need for I^2?
- Anything else!