Meta-analysis of individual participant data from observational studies

Simon Thompson

University of Cambridge, UK

1. Introduction

Freiburg, March 2013

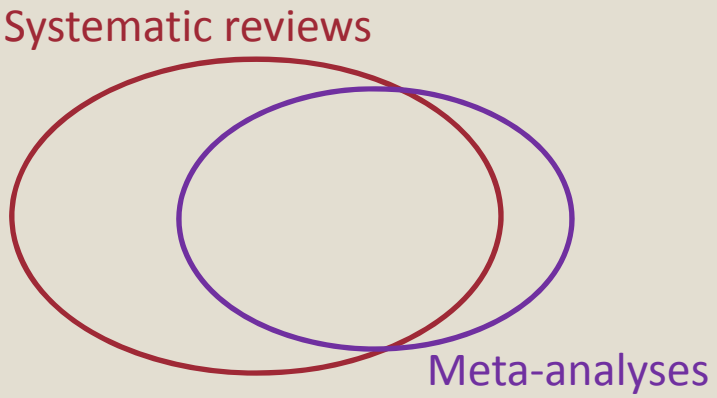
Systematic review vs. meta-analysis

Systematic review

- A systematic identification and evaluation of all the available relevant evidence

Meta-analysis

- Statistical combination of the numerical results of several studies



Rationale for systematic reviews and meta-analysis

- To minimize bias
  - of the reviewer, and in the research studies themselves
- To enhance precision
  - by including all the relevant evidence
- To put results into context
  - examine conflicts and understand differences
- To help prioritize research
  - by knowing exactly what has been done, how well, and with what findings

Key ideas of meta-analysis

- Characteristics of participants likely to vary across studies, so participants should only be compared with others in the same study
- Each study is summarized by an estimate of effect
- **The overall measure of effect is a weighted average of the results of the individual studies**
- The weights reflect the precision of each study

Focus of today

1. Meta-analysis
   - i.e. statistical techniques and issues
2. Individual participant data
   - not published results / summary data
3. Observational studies
   - rather than randomized trials

Outline of today

1. The basics
2. Epidemiological associations
3. Risk prediction
4. Causal associations

Four 40-minute sessions:

30 minutes + 10 minutes discussion

Acknowledgements

- Thanks to
  - Julian Higgins, University of Bristol, UK
  - Stephen Kaptoge, University of Cambridge, UK
  - Stephen Burgess, University of Cambridge, UK

- ... for some borrowed slides

Fixed-effect meta-analysis

- Require from each study, i=1…k :
  - estimate of effect, y_i
  - estimate of variance of effect, s_i^2

  When using ratio measures, natural log of the ratio is used

- Combine the estimates using a weighted average

- Take weight = inverse variance: $w_i = 1 / s_i^2$

- Gives more weight to the more precise studies

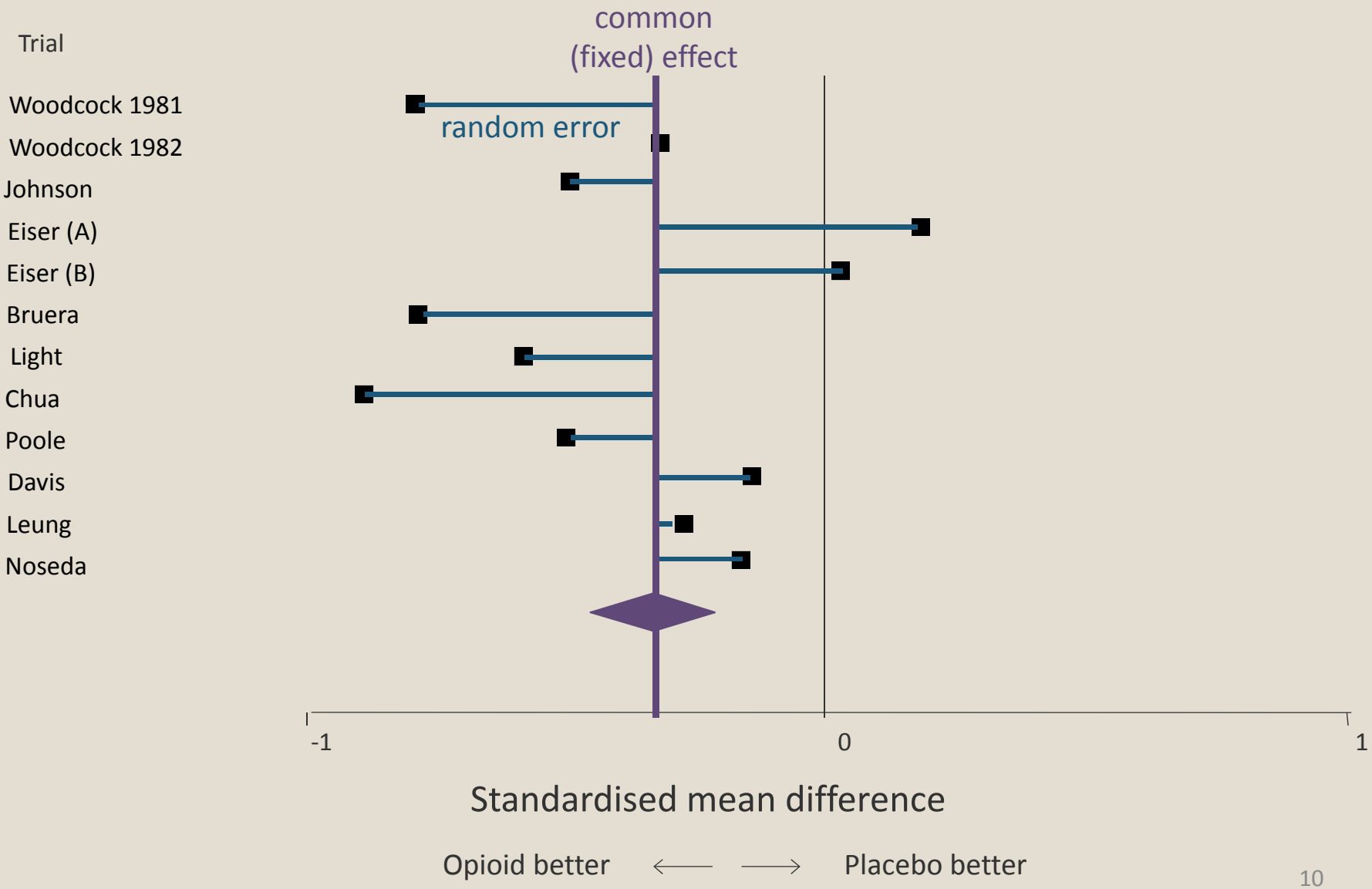
Fixed-effect inverse-variance weighted average

- Summary estimate $\hat{\theta} = \frac{\sum w_i y_i}{\sum w_i}$
- If we additionally assume $y_i \sim N\left(\theta, s_i^2\right)$ then $\hat{\theta} \sim N\left(\theta, \frac{1}{\sum w_i}\right)$
- Approximate 95% confidence interval for θ :

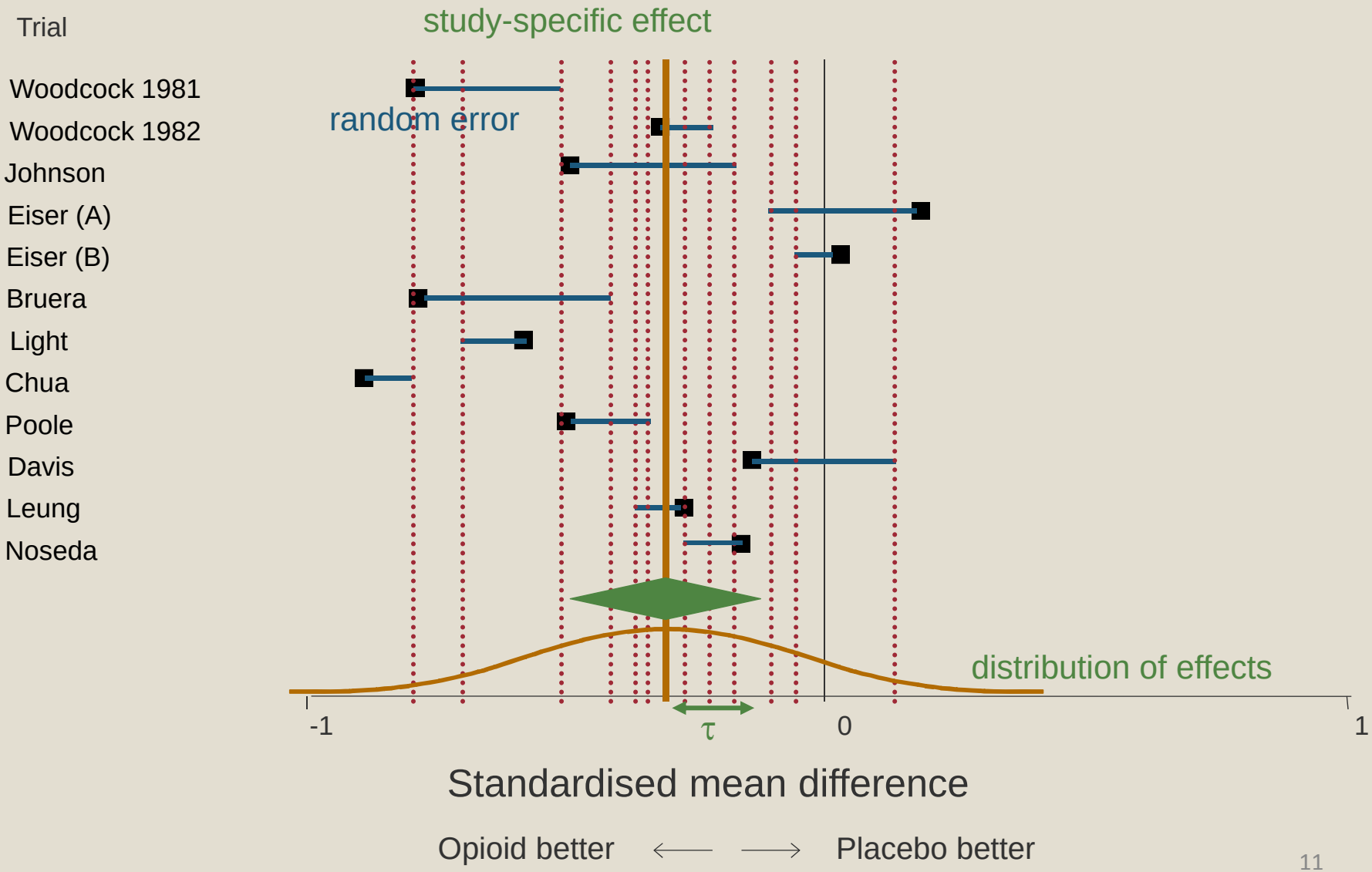
$$\hat{\theta} - 1.96\sqrt{1/\sum w_i} \quad to \quad \hat{\theta} + 1.96\sqrt{1/\sum w_i}$$

- Inverse-variance weights give most precise estimate of θ

Fixed-effect meta-analysis

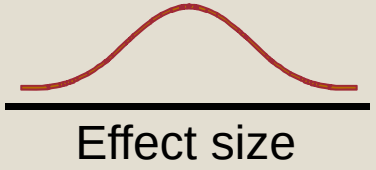


Random-effects meta-analysis



Random-effects meta-analysis

- We suppose the *true* treatment effect in each study is randomly, normally distributed across studies
  - with variance τ^2



- Estimate the between-study variance τ^2, and use this to modify the weights used to calculate the summary estimate

- The usual estimate of τ^2 is called the *DerSimonian and Laird (DL)* estimate, or *method of moments* estimate, calculated from the test statistic for heterogeneity:

$$Q = \sum w_i \left(y_i - \hat{\theta} \right)^2$$

Random-effects meta-analysis

- We revise weights to incorporate an estimate of the heterogeneity variance:

$$w_i^* = \frac{1}{s_i^2 + \hat{\tau}^2}$$

- To estimate the mean of the random-effects distribution:

$$\hat{\mu} = \frac{\sum w_i^* y_i}{\sum w_i^*} \qquad \text{var}(\hat{\mu}) = \frac{1}{\sum w_i^*}$$

- These simple methods ignore uncertainty in estimate of τ^2

Technicalities

- Assume study-specific effect θ_i in study *i*

$$\mathrm{E}\left[y_i \mid \theta_i\right] = \theta_i \qquad \mathrm{var}\left(y_i \mid \theta_i\right) = s_i^2$$

- Allow underlying effects to vary:

$$\mathrm{E}\left[\theta_i\right] = \mu \qquad \mathrm{var}\left(\theta_i\right) = \tau^2$$

- Unconditional mean and variance of estimates are then

$$\mathrm{E}\left[y_i\right] = \mu \qquad \mathrm{var}\left(y_i\right) = s_i^2 + \tau^2$$

- Confidence interval follows from assuming normality within and between studies

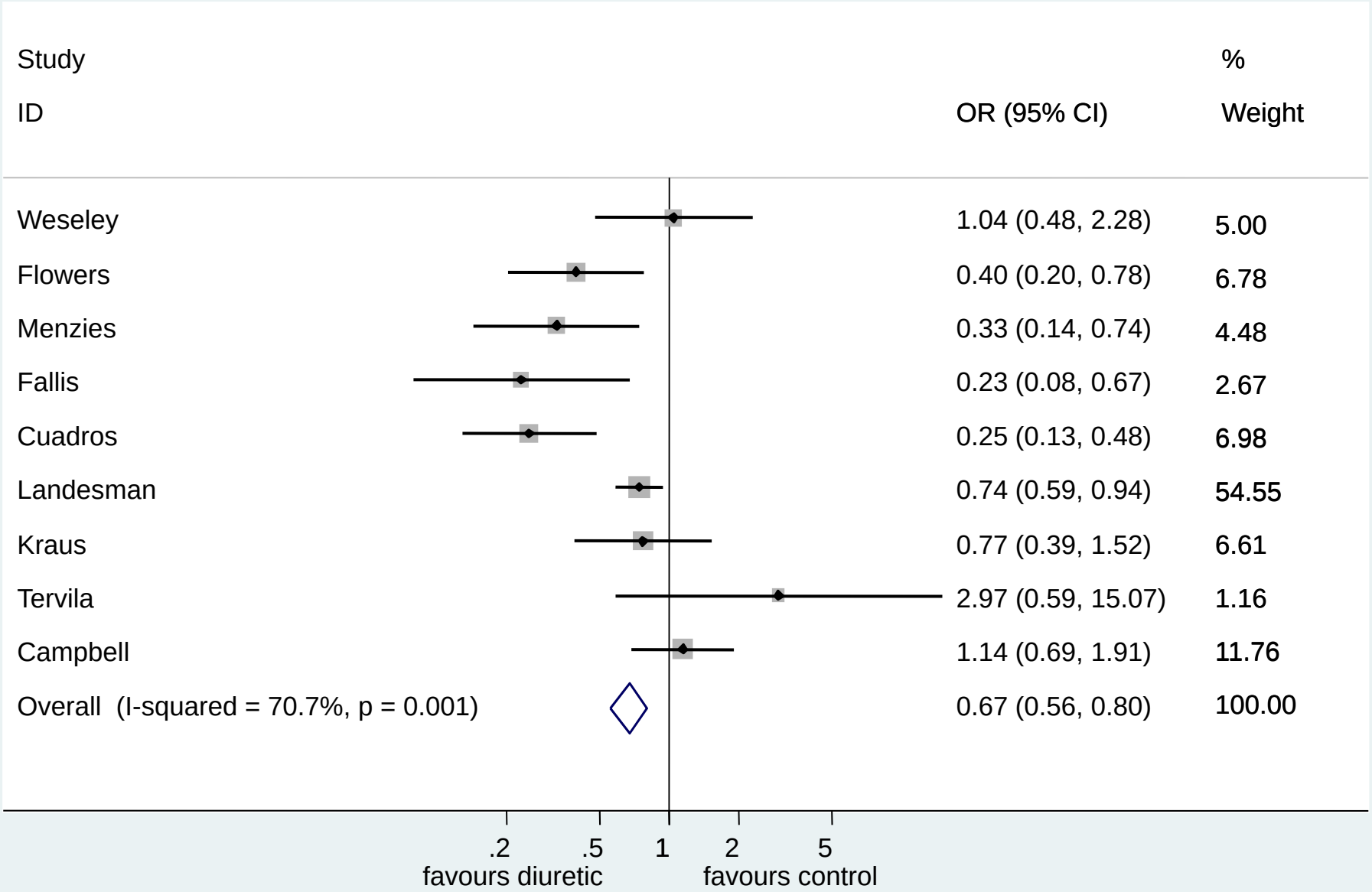
Example

Randomised trials to examine the effect of diuretics on pre-eclampsia (very high blood pressure) in pregnancy

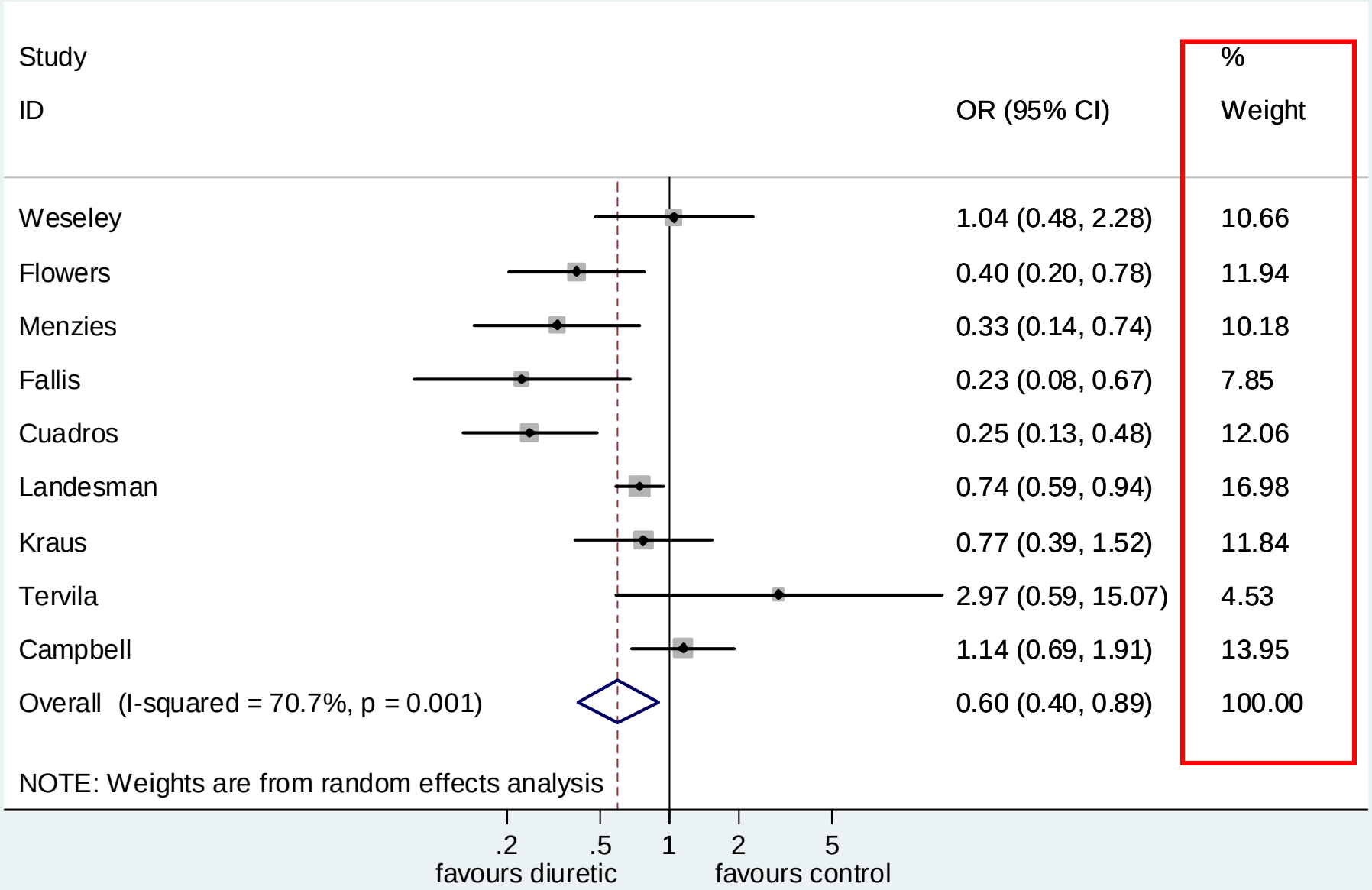
First author	P.E./total (treated)	P.E./total (control)	Odds ratio (95% CI)
Weseley	14/131	14/136	1.043 (0.477 to 2.28)
Flowers	21/385	17/134	0.397 (0.203 to 0.778)
Menzies	14/57	24/48	0.326 (0.142 to 0.744)
Fallis	6/38	18/40	0.229 (0.078 to 0.669)
Cuadros	12/1011	35/760	0.249 (0.128 to 0.483)
Landesman	138/1370	175/1336	0.743 (0.586 to 0.942)
Kraus	15/506	20/524	0.770 (0.390 to 1.52)
Tervila	6/108	2/103	2.971 (0.586 to 15.1)
Campbell	65/153	40/102	1.145 (0.687 to 1.91)

Meta-analysis on log odds ratio scale

Forest plot with fixed-effect summary estimate

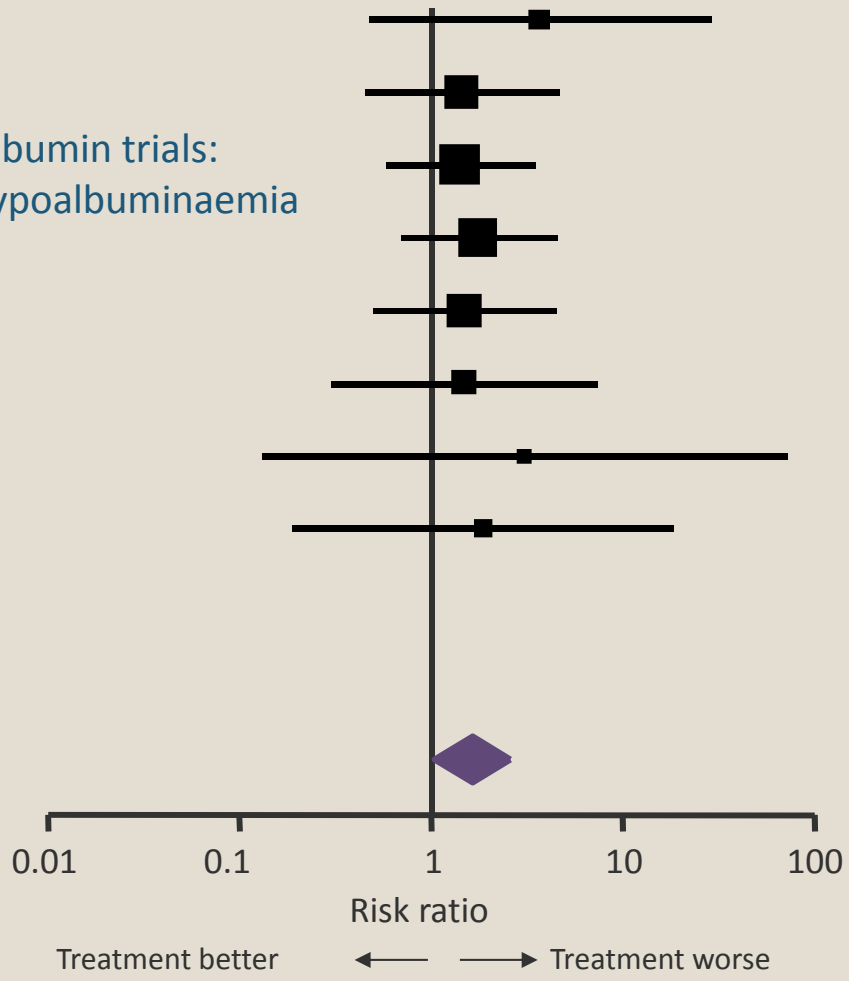


Forest plot with random-effects summary estimate



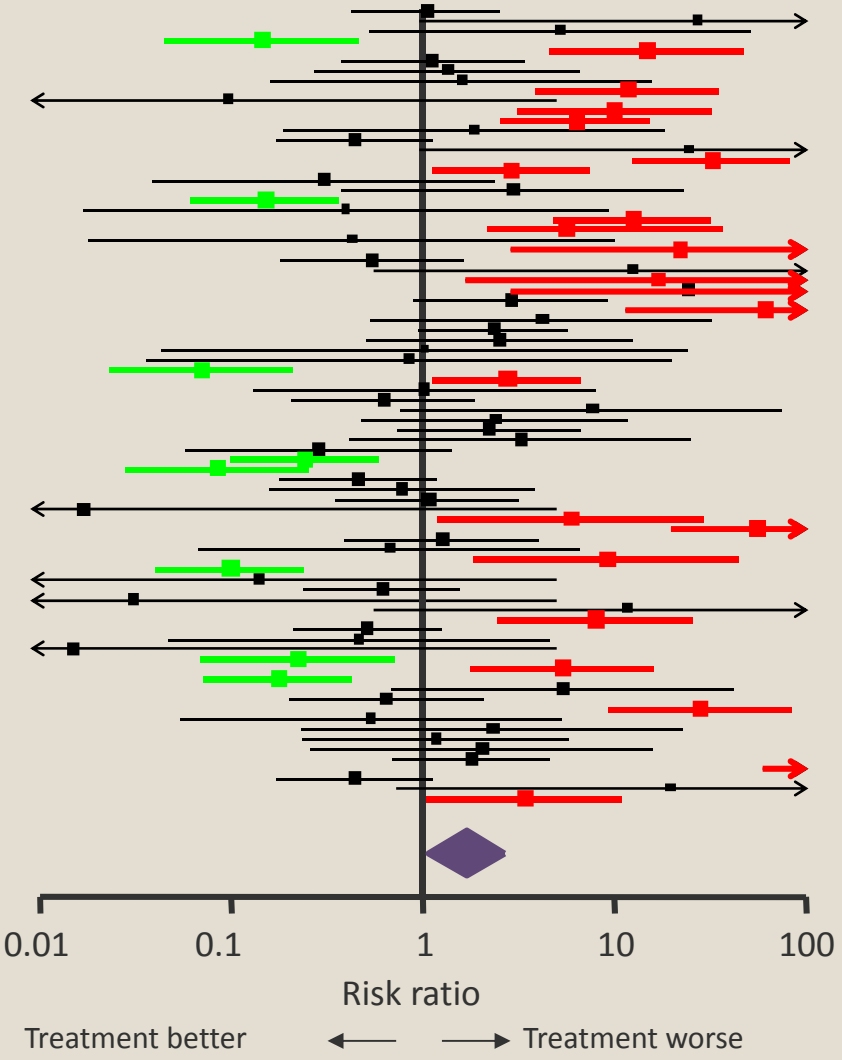
Estimates with 95% confidence intervals

Albumin trials:
hypoalbuminaemia



Random effects meta-analysis:
1.64 (1.04 , 2.58) P = 0.03

Estimates with 95% confidence intervals

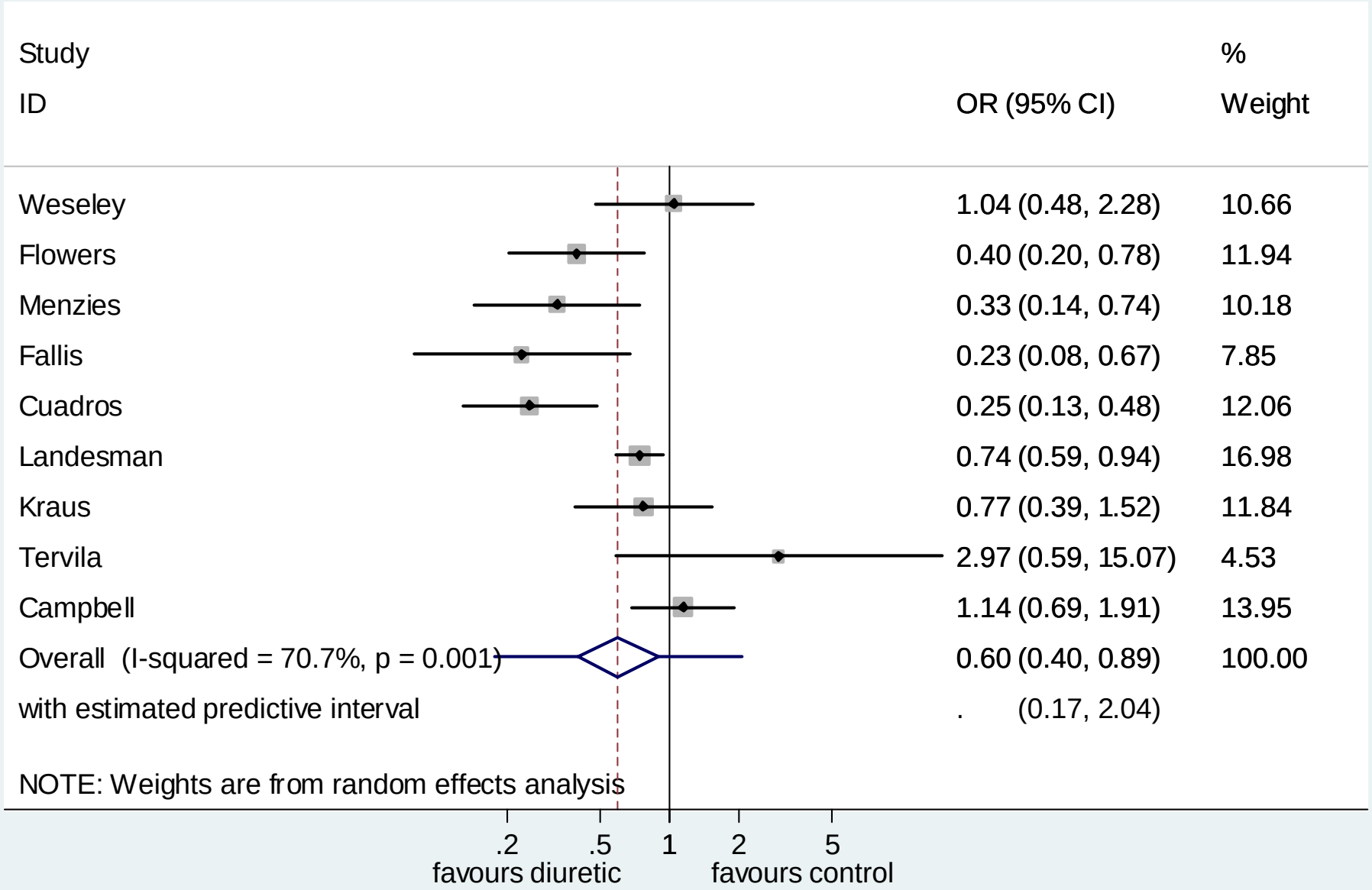


Random effects meta-analysis:
1.64 (1.04 , 2.58) P = 0.03

Interpreting random-effects meta-analysis

- Conventionally, inference is focused on the mean of the distribution
- This may be misleading … the amount of between-study variation should be taken into account
- A predictive distribution describes the uncertainty about the true effect in a (hypothetical) future study
  - e.g. we expect the true treatment effect in 95% of future studies to lie in a certain interval

Pre-eclampsia trials: random-effects predictive interval



Statistical heterogeneity

Statistical heterogeneity is:

Variation in the true effects underlying the studies

... which may manifest itself in more observed variation than expected by chance

Is heterogeneity inevitable?

Bias (design, conduct, attrition)

Diversity (participants, interventions / exposures, outcomes)

Testing for heterogeneity

- To test the null hypothesis that the true treatment effect is the same in all studies we can calculate a *heterogeneity statistic*:

$$Q = \sum w_i \left(y_i - \hat{\theta} \right)^2$$

- To calculate a P value, Q is compared with the χ^2 distribution on (k – 1) degrees of freedom (k is number of studies)

- The heterogeneity statistic Q and associated test assess the evidence for heterogeneity, not the amount of heterogeneity:
  - The size of Q depends on the number of studies
  - The test has low power when there are few studies, high power when there are many

Impact of heterogeneity: I^2

- Alternative measure, to quantify inconsistency
  - based on χ^2 statistic, Q, and its degrees of freedom.

$$I^2 = \frac{Q - \text{d.f.}}{Q} \times 100\%$$

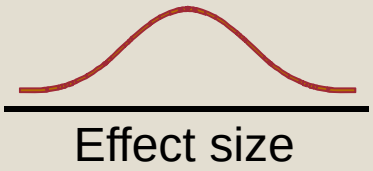
I^2 can be interpreted as the proportion of total variability explained by heterogeneity, rather than chance

- If I^2 is computed to be < 0, it is set to 0%

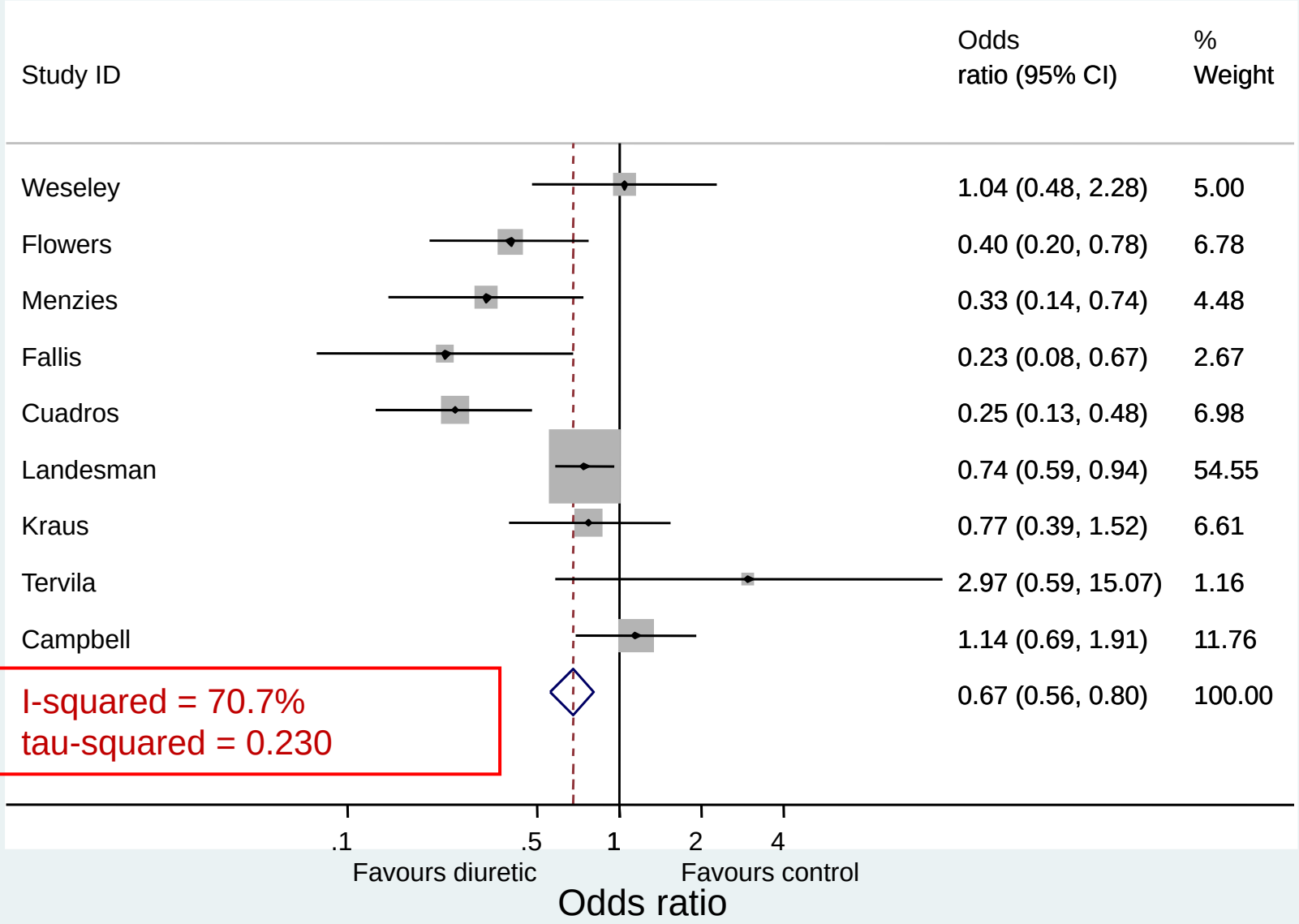
Quantifying heterogeneity: τ^2

- The between-studies variance, τ^2 is estimated as part of a random-effects meta-analysis

- It provides a useful measure of the true extent of heterogeneity across studies, but the scale can be awkward to interpret (e.g. log odds ratio)



Pre-eclampsia trials: using Stata

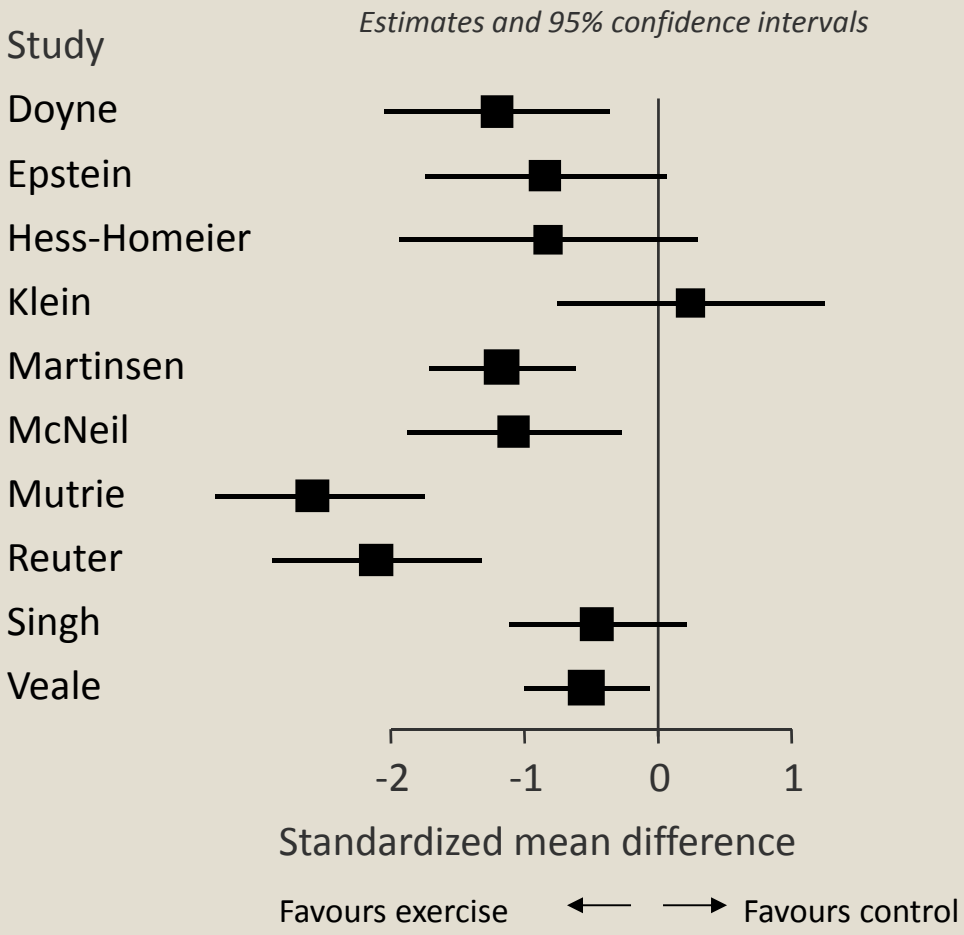


Heterogeneity chi-squared = 27.26 (d.f. = 8), p = 0.001

Exploring heterogeneity

- Characteristics of studies may be associated with the size of treatment effect
- For example,
  - age group of patients
  - dose of drug
  - adequate allocation concealment in a randomized trial
  - sample size
- For discrete characteristics, can use subgroup analyses
- For discrete or continuous characteristics, can use meta-regression

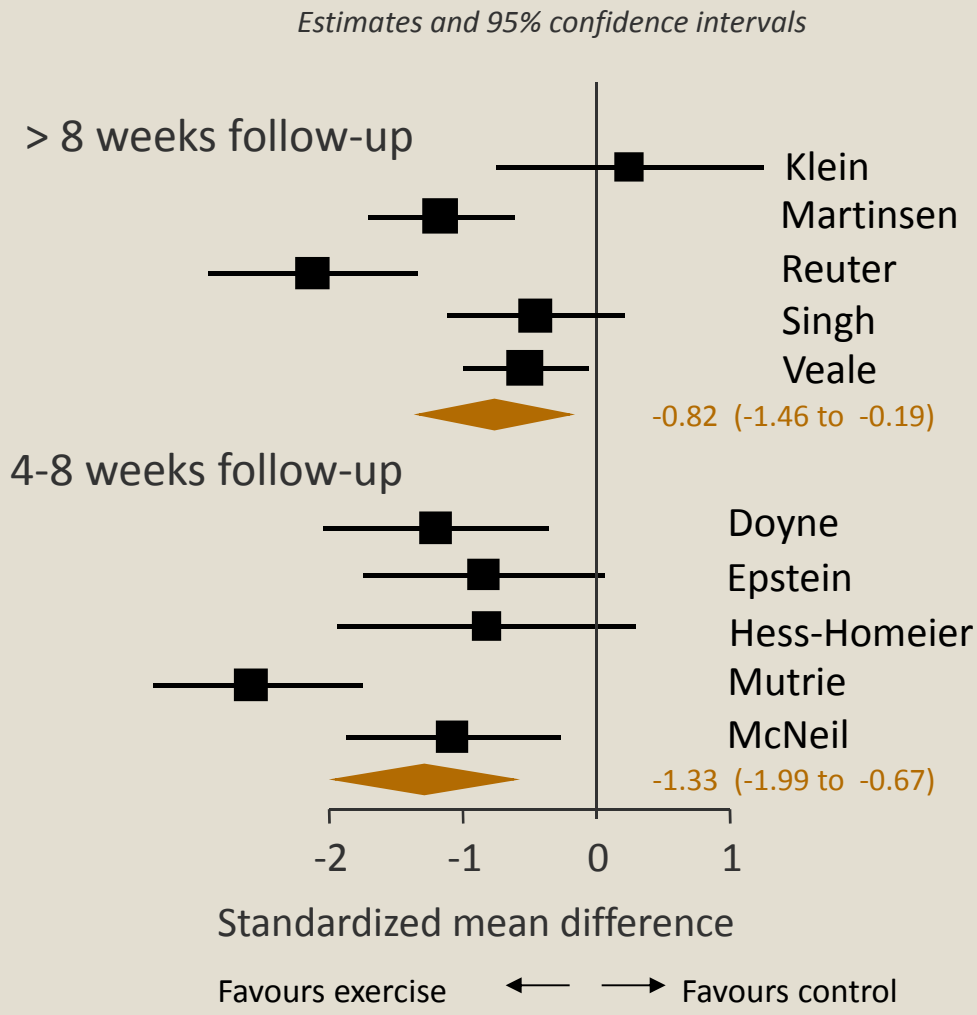
Example: exercise for depression



Lawlor & Hopker, *BMJ* 2001

Subgroup analysis

- Divide up the studies
  - e.g. by duration of trial
- Test for subgroup differences:
  - can apply Q test to subgroup results
  - here, *P* = 0.28

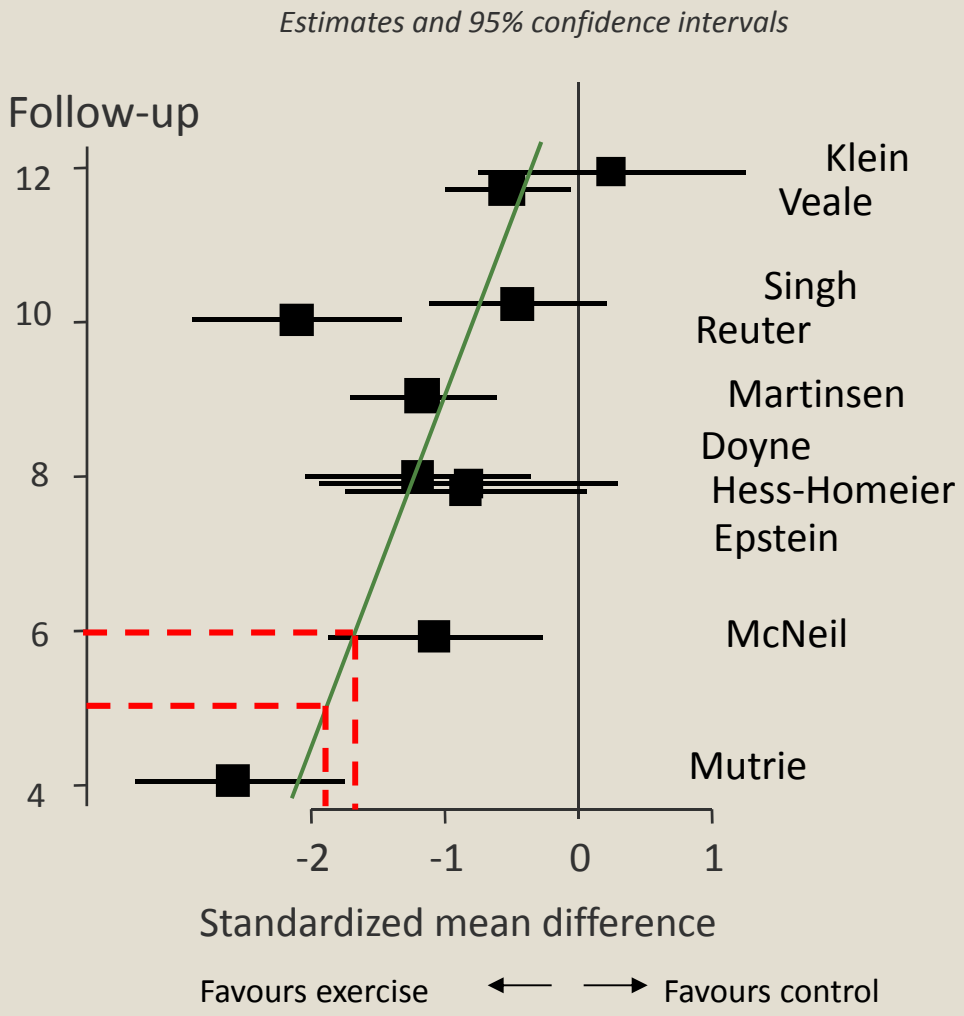


Meta-regression

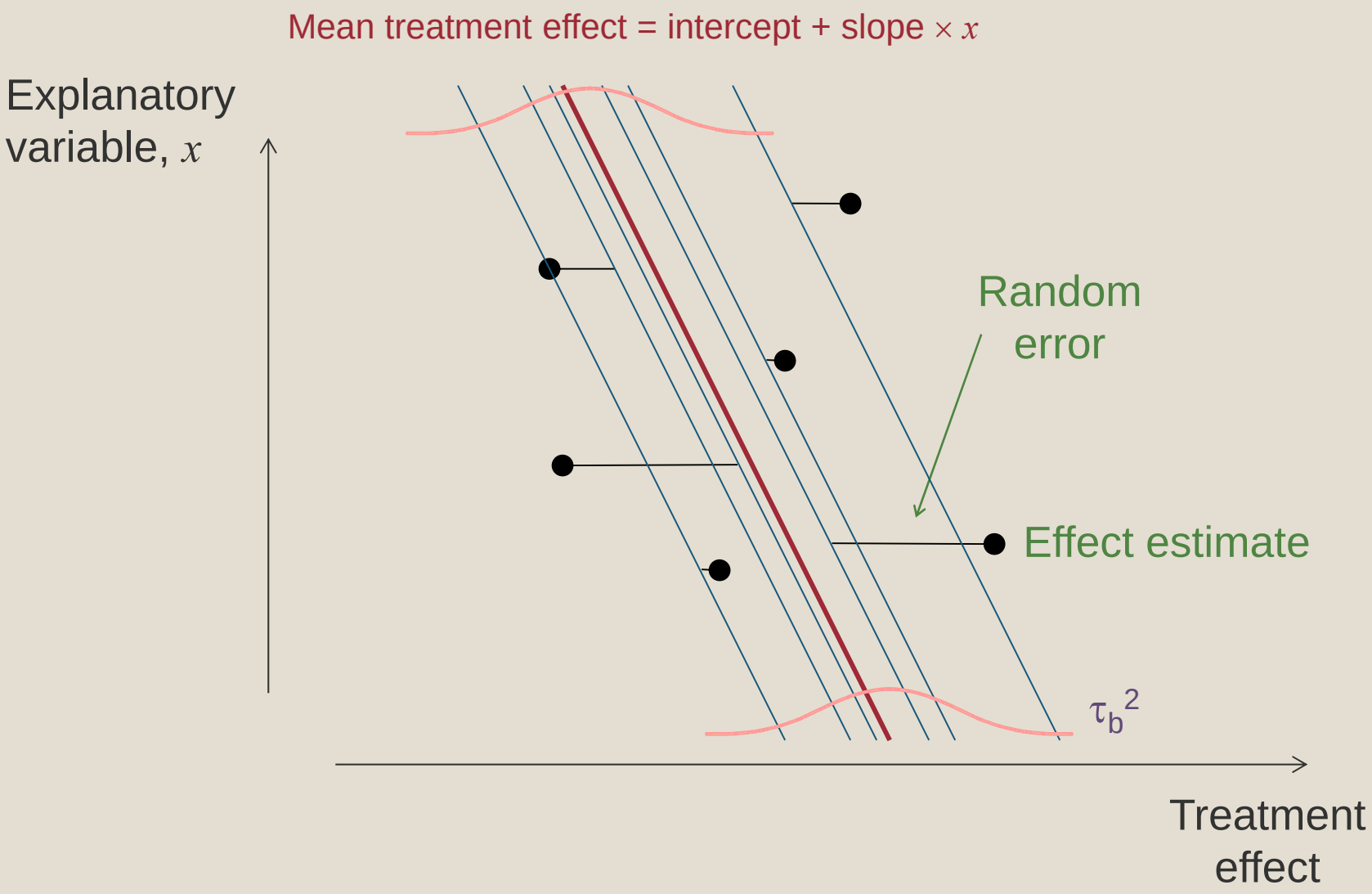
- Examine heterogeneity

- predict effect according to length of follow-up

  - SMD decreases by 0.2 for each extra week ($P = 0.008$)



Random-effects meta-regression



Conclusions regarding heterogeneity

- Heterogeneity is inevitable
- It is preferable to quantify heterogeneity, than just test for it
- Random effects meta-analysis incorporates the heterogeneity, but predictive interval may be more relevant than CI

- Exploring heterogeneity is fraught with dangers
  - observational relationships / too few studies / too many sources of clinical and methodological heterogeneity
- Avoid over-interpretation of findings

Some references

- Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011; 342: d549.
- Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009 Jan;172(1):137-159.
- Whitehead A. *Meta-Analysis of Controlled Clinical Trials.* Wiley 2002
- *Cochrane Handbook for Systematic Reviews of Interventions.* Higgins and Green (eds); Wiley 2008, updated online