Explaining heterogeneity

Julian Higgins

School of Social and Community Medicine, University of Bristol, UK

Outline

1. Revision and remarks on fixed-effect and random-effects meta-analysis methods (and interpretation under heterogeneity)

Explaining heterogeneity:

2. Subgroup analysis
3. Meta-regression
4. Problems

5. Closing remarks

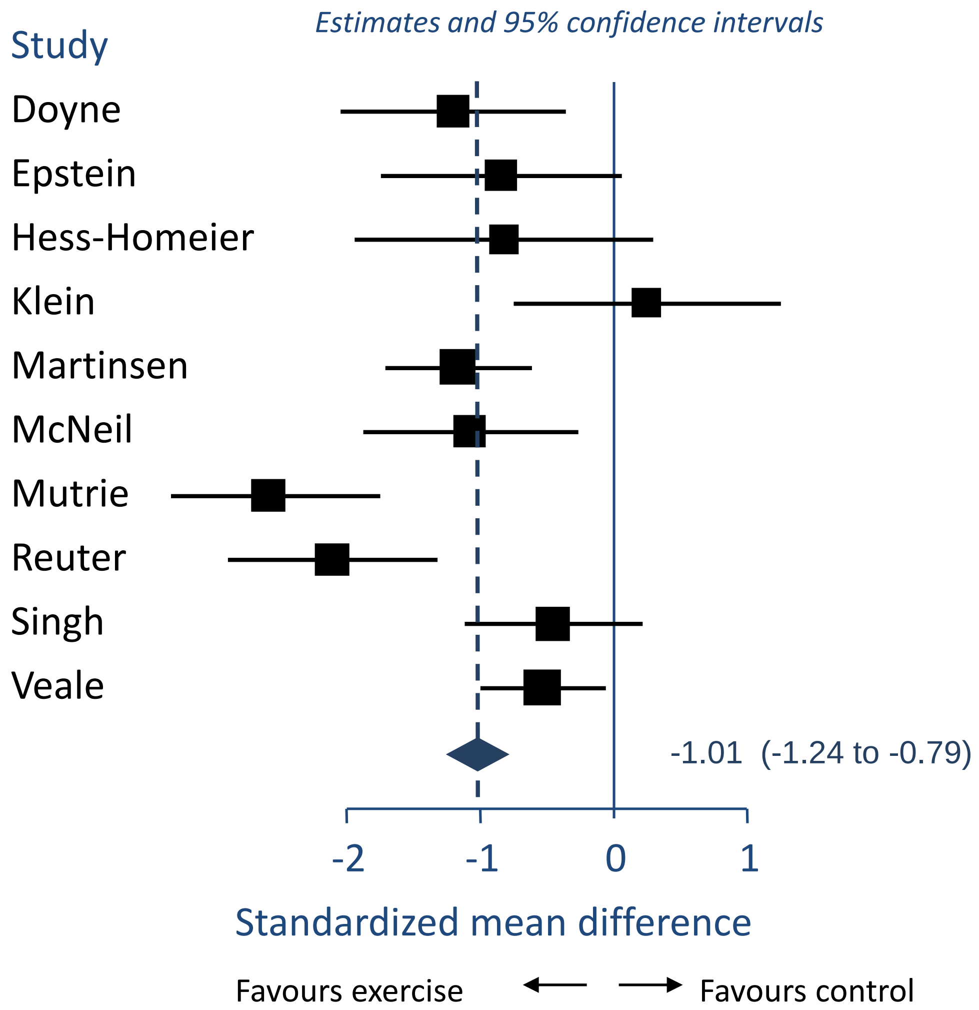
- Example 1: Trials of exercise for treatment of depression

Part 1: Revision and remarks on fixed-effect and random-effects meta-analysis methods

University of BRISTOL

Revision (1)

- Fixed-effect(s) meta-analysis
  - weights $w_i = \frac{1}{v_i}$
  - (minimise variability of the summary estimate)



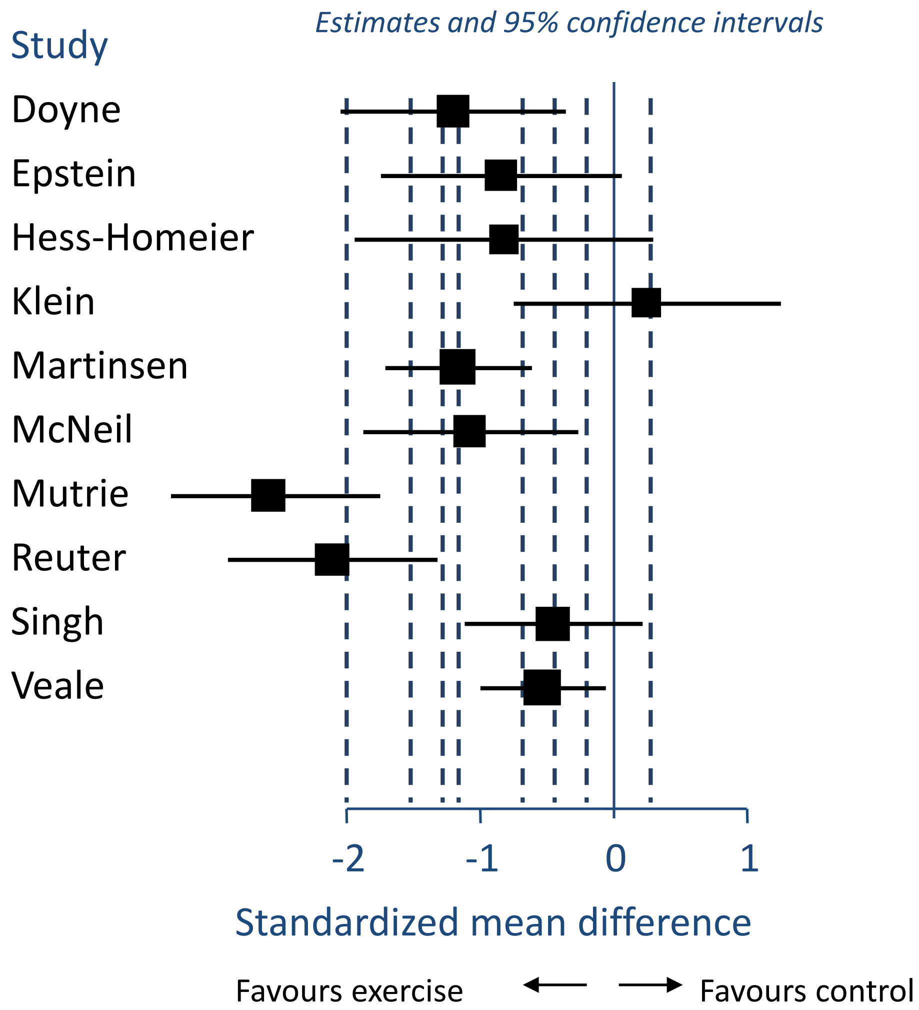
Revision (2)

- examine whether effect vary across studies

$Q = 35.4$ (9 d.f.)
($p = 0.00005$)

$\tau_{DL} = 0.64$ (for example)

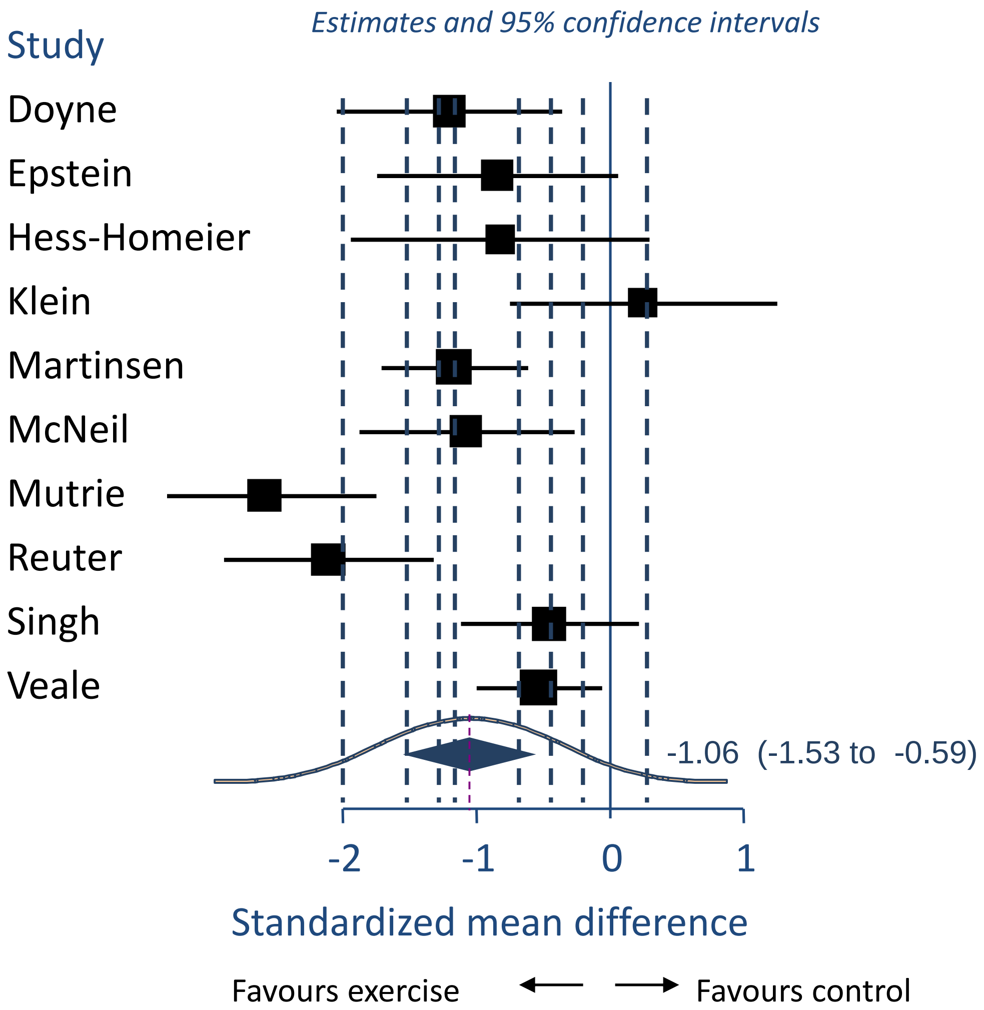
$I^2 = 75\%$



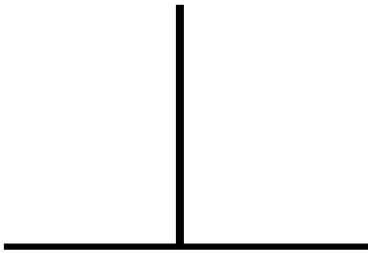
Revision (3)

- Meta-analysis under random-effects assumption
  - effect size varies across studies
  - between-study variance= τ^2
  - weights

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



Three simple models for relationships of effect size parameters across studies



Effect size

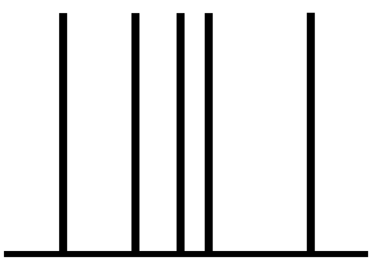
The underlying effect sizes are identical

Common-effect model
(or fixed-effect model)

Effect size

The underlying effect size are different and related through some distribution

Random-effects model

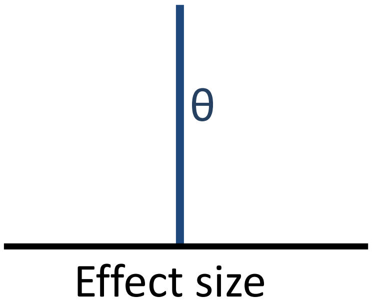


Effect size

The underlying effect sizes are different and unrelated (independent)

Fixed-effects model

Two simple *methods* for combining effect size *estimates* across studies

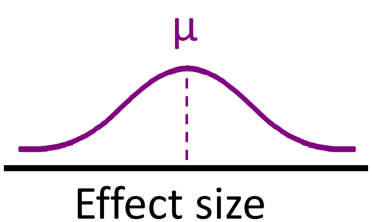


Common-effect model

$$\hat{\theta} = \frac{\sum w_i y_i}{\sum w_i}$$

Estimate of: θ

Variance: $\dfrac{1}{\sum w_i}$ $\quad w_i = \dfrac{1}{s_i^2}$

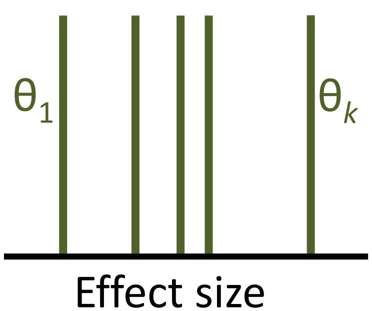


Random-effects model

$$\hat{\mu} = \frac{\sum w_i{}^* y_i}{\sum w_i{}^*}$$

Estimate of: μ

Variance: $\dfrac{1}{\sum w_i{}^*}$ $\quad w_i^* = \dfrac{1}{s_i^2 + \hat{\tau}^2}$



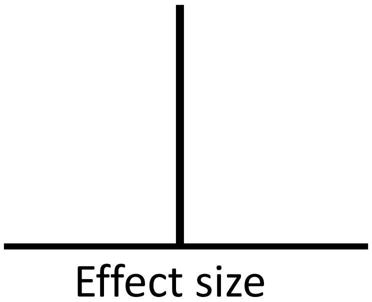
Fixed-effects model

$$\hat{\phi} = \frac{\sum w_i y_i}{\sum w_i}$$

Estimate of: $\dfrac{\sum w_i \theta_i}{\sum w_i}$

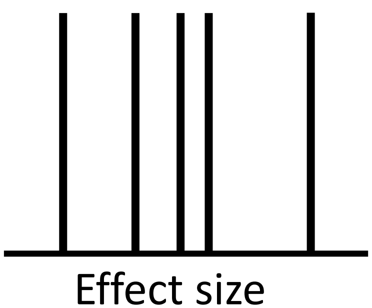
Variance: $\dfrac{1}{\sum w_i}$

Three simple models for relationships of effect size parameters across studies



Common-effect model
(or fixed-effect model)

Complication is that the fixed-effect meta-analysis **method** can be interpreted under either of these **models**



Fixed-effects model

Aside: an interpretation of the weighted average under fixed-effects model

$$\mu = \frac{\sum p_i r_i \theta_i}{\sum p_i r_i}$$

- A weighted average of the study-specific effects seen in the individual study populations, where the weight for study population *i* is proportional to both the number of participants selected (p_i) and to how much information is contributed per subject (r_i)
- No assumption of homogeneity is required
- But heterogeneity is ignored in the analysis so inference on this is required
  - this inference is statistically independent
- The empirical weights are poor estimates of $p_i r_i$ unless the studies are large

Rice, Higgins and Lumley (submitted)

What is heterogeneity?

- Variation beyond that expected by chance alone
  - Observed result = **true effect** + **bias** + **random error**
- True effects vary due to 'clinical diversity'
  - variants in PICO
- Biases vary due to 'methodological diversity'
  - design, conduct, problems

- In practice, true effects and biases inseparable, so group them together
- Variation in true effect + bias measured by the between-study variance τ^2

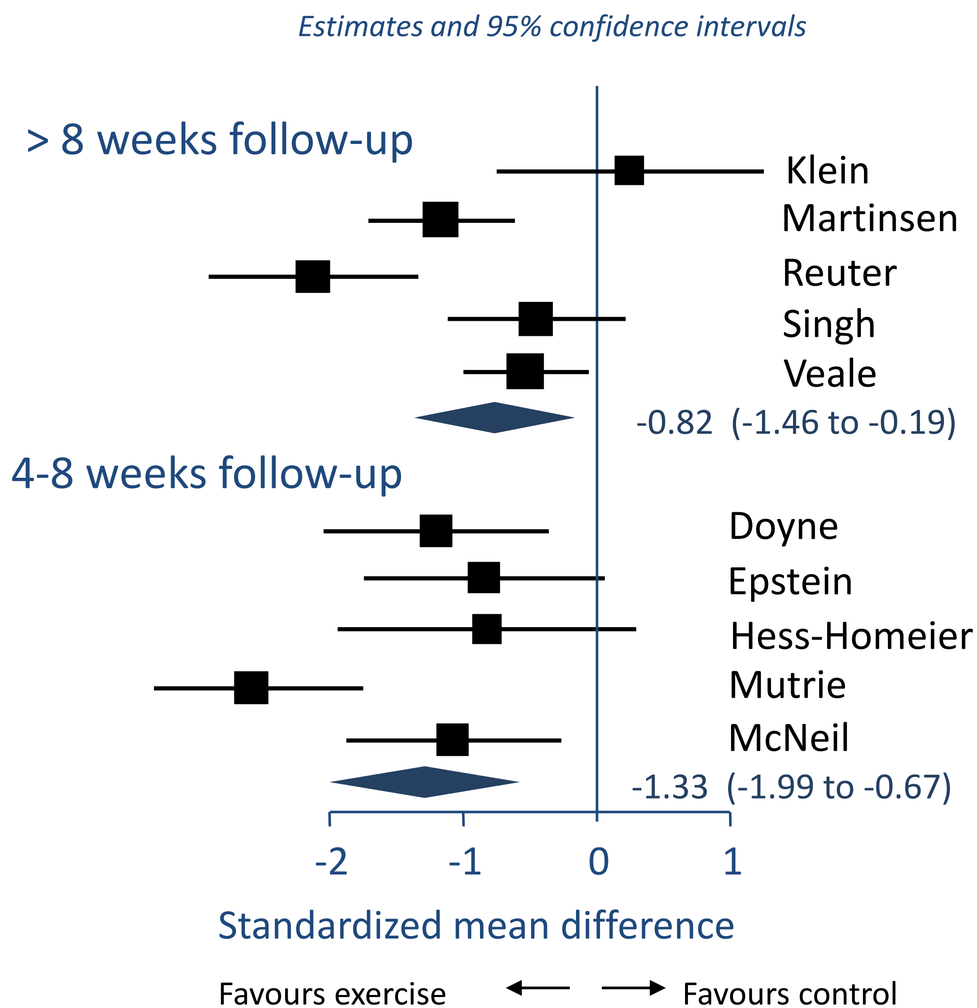
Exploring heterogeneity

In which types of trials does the intervention work best?

- Characteristics of studies may be associated with the size of treatment effect
  - Does the intervention work better if given for longer?
  - Are smaller odds ratios observed in high-risk populations?
  - Is there a relationship between sample size and effect size (e.g. due to publication bias)?
  - Is inadequate allocation concealment associated with a larger effect estimate?
  - Is A better than B, when they've each only been compared with C? For discrete characteristics, can use subgroup analyses
- We can use subgroup analyses or meta-regression to answer questions like these

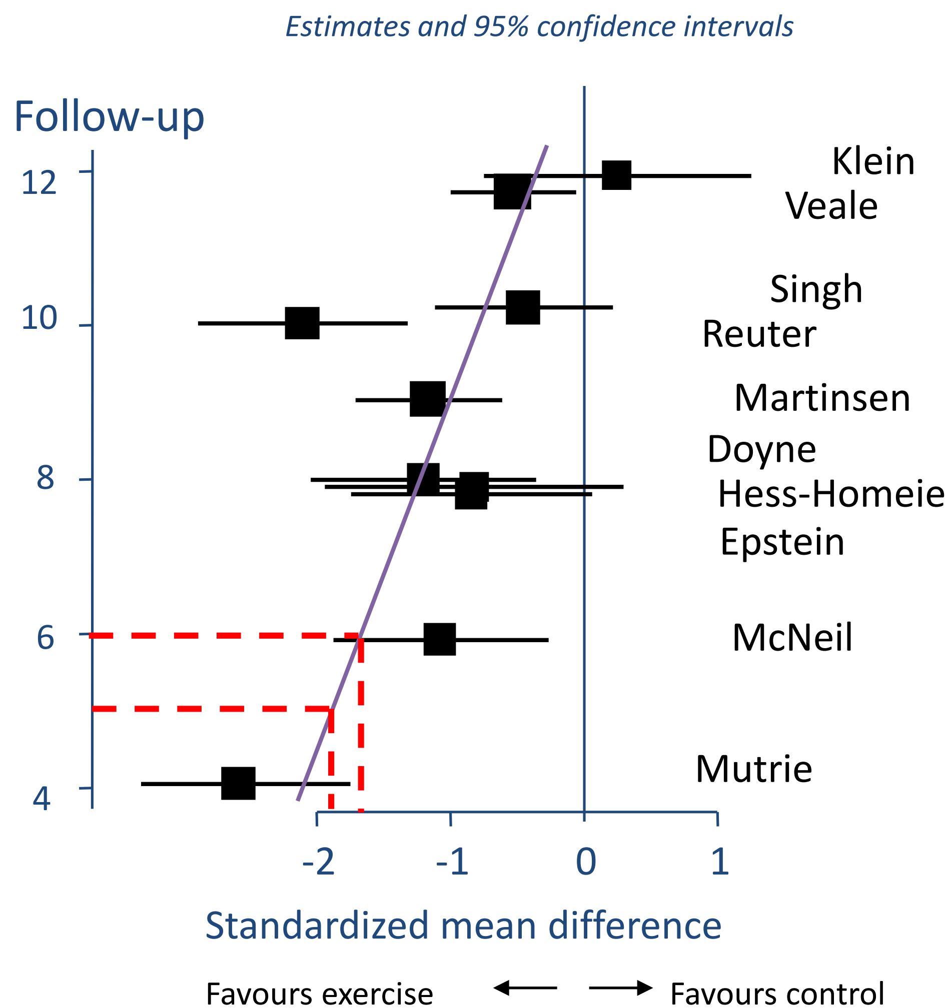
Subgroup analysis

- 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
- *Difference in SMDs* = 0.18 (95% CI 0.02 to 0.34): i.e. SMD decreases by 0.2 for each extra week of follow up



Part 2: Subgroup analysis

Subgroup analyses

- Split data into subgroups:
  - Subsets of patients
  - Subsets of studies
- Do separate meta-analyses for each

- Subgroup analyses done for three reasons
  - To answer multiple questions (one for each subset of studies)
  - To examine whether the effect is different in different subgroups (comparing subgroups)
  - To assess whether restricting analysis to one subgroup changes the conclusion (sensitivity analysis)

Subgroup analyses

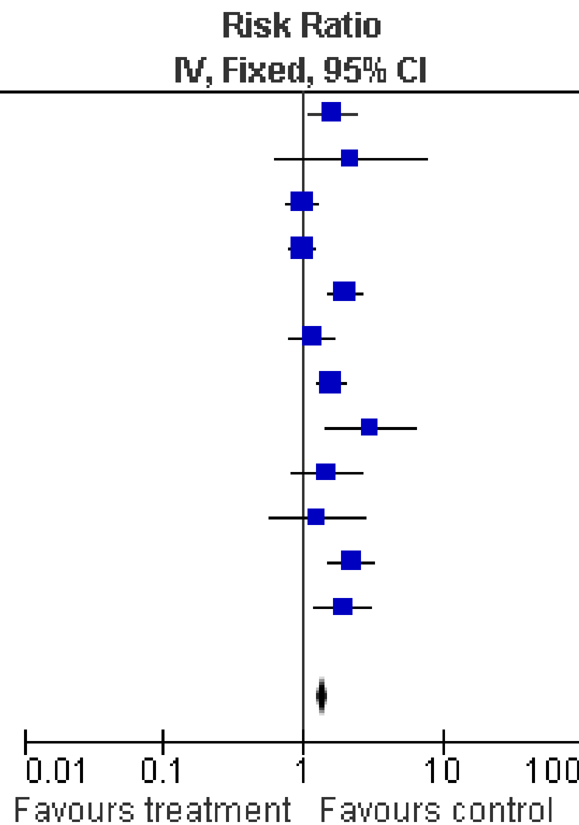
Behavioural interventions for smokeless tobacco use cessation

Abstinence from all tobacco use (where reported) at 6 months or more

	Treatment		Control			Risk Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**IV, Fixed, 95% CI**
Boyle 2004	44	109	28	112	6.3%	1.61 [1.09, 2.39]
Cigrang 2002	7	31	3	29	0.6%	2.18 [0.62, 7.65]
Cummings 1995	76	316	102	417	14.6%	0.98 [0.76, 1.27]
Gansky 2005	103	285	130	352	23.0%	0.98 [0.80, 1.20]
Seversen 2006	119	393	60	392	12.8%	1.98 [1.50, 2.61]
Seversen 2007a	60	535	52	534	7.9%	1.15 [0.81, 1.64]
Seversen 2007b	159	1260	100	1263	17.3%	1.59 [1.26, 2.02]
Severson 1998	40	394	8	239	1.8%	3.03 [1.44, 6.37]
Stevens 1995	25	245	19	273	3.0%	1.47 [0.83, 2.60]
Stotts 2003	19	198	8	105	1.6%	1.26 [0.57, 2.78]
Walsh 1999	60	171	30	189	6.6%	2.21 [1.50, 3.25]
Walsh 2003	38	141	23	166	4.5%	1.95 [1.22, 3.10]
Total (95% CI)	**750**	**4078**	**563**	**4071**	**100.0%**	**1.38 [1.25, 1.53]**

Heterogeneity: $\text{Chi}^2 = 39.56$, $df = 11$ ($P < 0.0001$); $I^2 = 72\%$

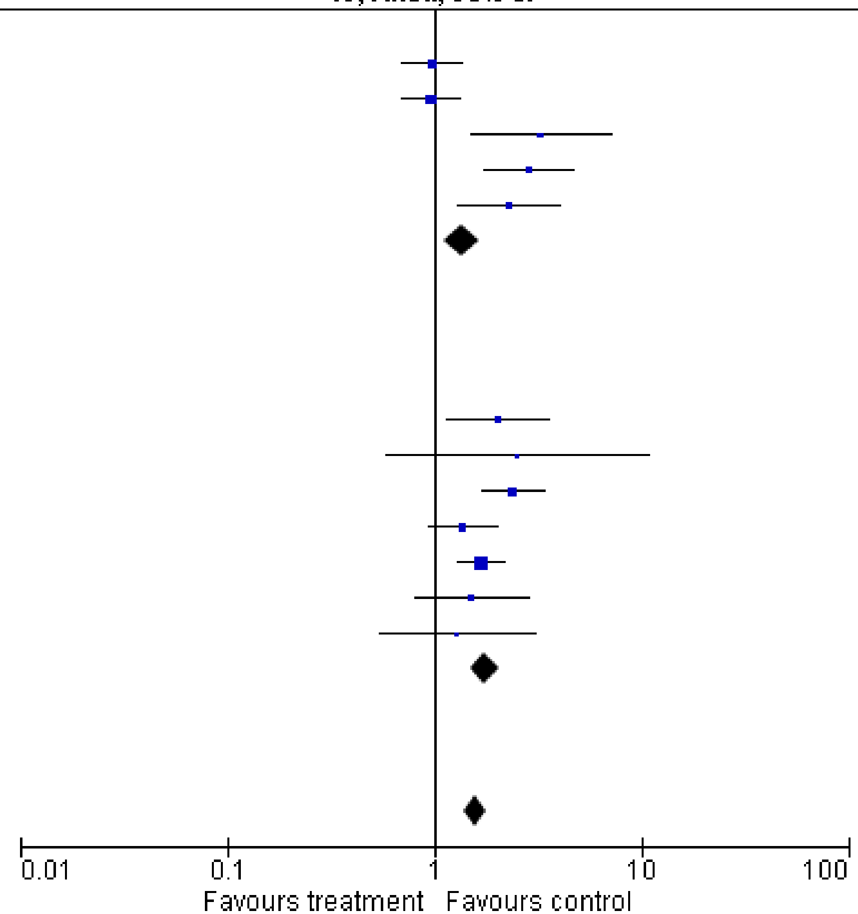
Test for overall effect: $Z = 6.41$ ($P < 0.00001$)



Subgroup analyses

Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio IV, Fixed, 95% CI
1.1.1 Randomisation by organisation						
Cummings 1995	76	316	102	417	13.5%	0.98 [0.70, 1.38]
Gansky 2005	103	285	130	352	14.9%	0.97 [0.70, 1.34]
Severson 1998	40	394	8	239	2.6%	3.26 [1.50, 7.10]
Walsh 1999	60	171	30	189	6.3%	2.86 [1.74, 4.73]
Walsh 2003	38	141	23	166	4.7%	2.29 [1.29, 4.08]
Subtotal (95% CI)		**1307**		**1363**	**42.0%**	**1.36 [1.12, 1.64]**
Total events	317		293			
Heterogeneity: Chi² = 24.38, df = 4 (P < 0.0001); I² = 84%						
Test for overall effect: Z = 3.08 (P = 0.002)						
1.1.2 Individual randomisation						
Boyle 2004	44	109	28	112	4.8%	2.03 [1.14, 3.60]
Cigrang 2002	7	31	3	29	0.7%	2.53 [0.59, 10.90]
Seversen 2006	119	393	60	392	12.9%	2.40 [1.69, 3.41]
Seversen 2007a	69	535	52	534	10.8%	1.37 [0.94, 2.01]
Seversen 2007b	159	1260	100	1263	22.6%	1.68 [1.29, 2.19]
Stevens 1995	25	245	19	273	4.0%	1.52 [0.81, 2.83]
Stotts 2003	19	198	8	105	2.1%	1.29 [0.54, 3.05]
Subtotal (95% CI)	**442**	**2771**	**270**	**2708**	**58.0%**	**1.76 [1.49, 2.07]**
Heterogeneity: Chi² = 6.00, df = 6 (P = 0.42); I² = 0%						
Test for overall effect: Z = 6.72 (P < 0.00001)						
Total (95% CI)	**759**	**4078**	**563**	**4071**	**100.0%**	**1.58 [1.39, 1.79]**
Heterogeneity: Chi² = 34.42, df = 11 (P = 0.0003); I² = 68%						
Test for overall effect: Z = 7.12 (P < 0.00001)						
Test for subgroup differences: Chi² = 4.04, df = 1 (P = 0.04), I² = 75.3%						

Odds Ratio
IV, Fixed, 95% CI



Comparing two subgroups

- Simple statistics: take the difference in point estimates and add their variances
  - all on the log scale for ratio measures

Subgroup	**OR**	**OR_CI_low**	**OR_CI_upp**	**lnOR**	**lnOR_CI_low**	**lnOR_CI_upp**	**lnOR_var**
Randomisation by organisation	1.36	1.12	1.64	0.307	0.113	0.495	0.009
Individual randomisation	1.76	1.49	2.07	0.565	0.399	0.728	0.007
	RATIO			DIFF			SUM
Comparison of subgroups	0.773	0.601	0.994	-0.258	-0.510	-0.006	0.016

Fixed-effects test for subgroup differences

- Simple method

Q_{all} : chi-squared statistic for all the studies

$Q_1, \ldots, Q_m$: chi-squared statistic for *m* different subgroups

- Heterogeneity explained by differences between subgroups (Q_{bet}):

$$Q_{bet} = Q_{all} - (Q_1 + \ldots + Q_m)$$

- Has degrees of freedom

$$df = m - 1$$

Fixed-effects test for subgroup differences

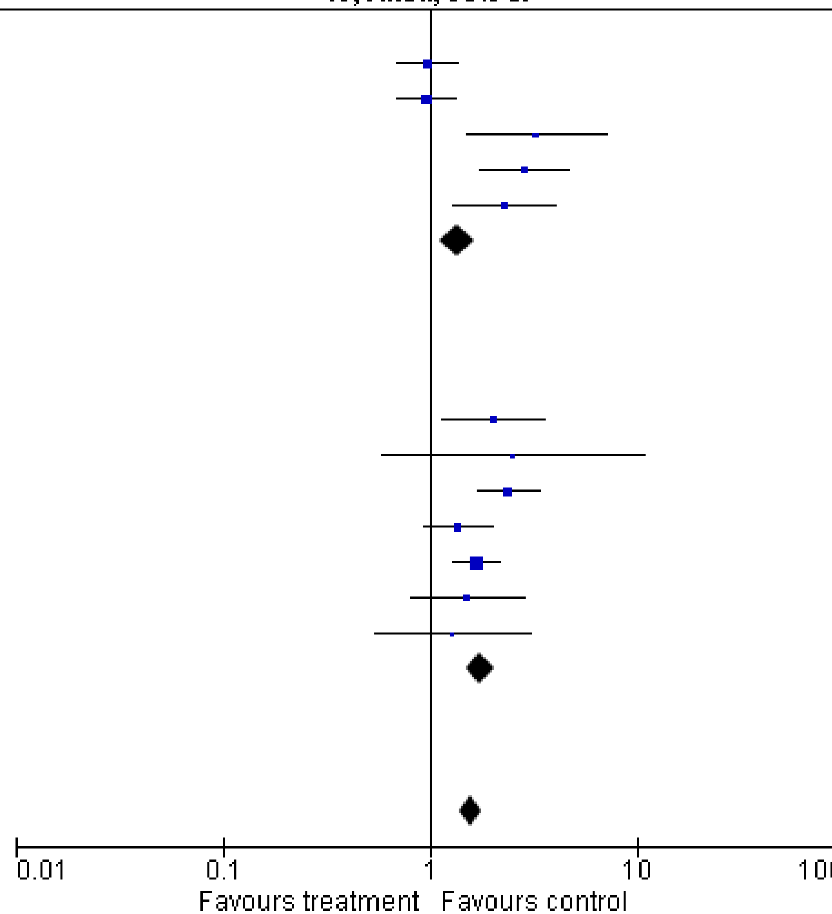
Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio IV, Fixed, 95% CI
1.1.1 Randomisation by organisation						
Cummings 1995	76	316	102	417	13.5%	0.98 [0.70, 1.38]
Gansky 2005	103	285	130	[illegible]	[illegible]4.9%	0.97 [0.70, 1.34]
Severson 1998	40	394	8	[illegible]	[illegible]2.6%	3.26 [1.50, 7.10]
Walsh 1999	60	171	30	[illegible]	[illegible]6.3%	2.86 [1.74, 4.73]
Walsh 2003	38	141	23	[illegible]	[illegible]4.7%	2.29 [1.29, 4.08]
Subtotal (95% CI)	**317**	**1307**	**293**	**1363**	**42.0%**	**1.36 [1.12, 1.64]**
Heterogeneity: Chi² = 24.38, df = 4 (P < 0.0001); I² = 84%						
Test for overall effect: Z = 3.08 (P = 0.002)						
1.1.2 Individual randomisation						
Boyle 2004	44	109	28	112	4.8%	2.03 [1.14, 3.60]
Cigrang 2002	7	31	3	29	0.7%	2.53 [0.59, 10.90]
Seversen 2006	119	393	60	392	12.9%	2.40 [1.69, 3.41]
Seversen 2007a	69	535	5[illegible]	[illegible]	10.8%	1.37 [0.94, 2.01]
Seversen 2007b	159	1260	10[illegible]	[illegible]	22.6%	1.68 [1.29, 2.19]
Stevens 1995	25	245	1[illegible]	[illegible]	4.0%	1.52 [0.81, 2.83]
Stotts 2003	19	198	8	[illegible]	2.1%	1.29 [0.54, 3.05]
Subtotal (95% CI)	**442**	**2771**	**270**	**2708**	**58.0%**	**1.76 [1.49, 2.07]**
Heterogeneity: Chi² = 6.00, df = 6 (P = 0.42); I² = 0%						
Test for overall effect: Z = 6.72 (P < 0.00001)						
Total (95% CI)	**759**	**4078**	**563**	**4071**	**100.0%**	**1.58 [1.39, 1.79]**
Heterogeneity: Chi² = 34.42, df = 11 (P = 0.0003); I² = 68%						
Test for overall effect: Z = 7.12 (P < 0.00001)						
Test for subgroup differences: Chi² = 4.04, df = 1 (P = 0.04), I² = 75.3%						

Q_1

Q_2

Q_{all}

Odds Ratio
IV, Fixed, 95% CI



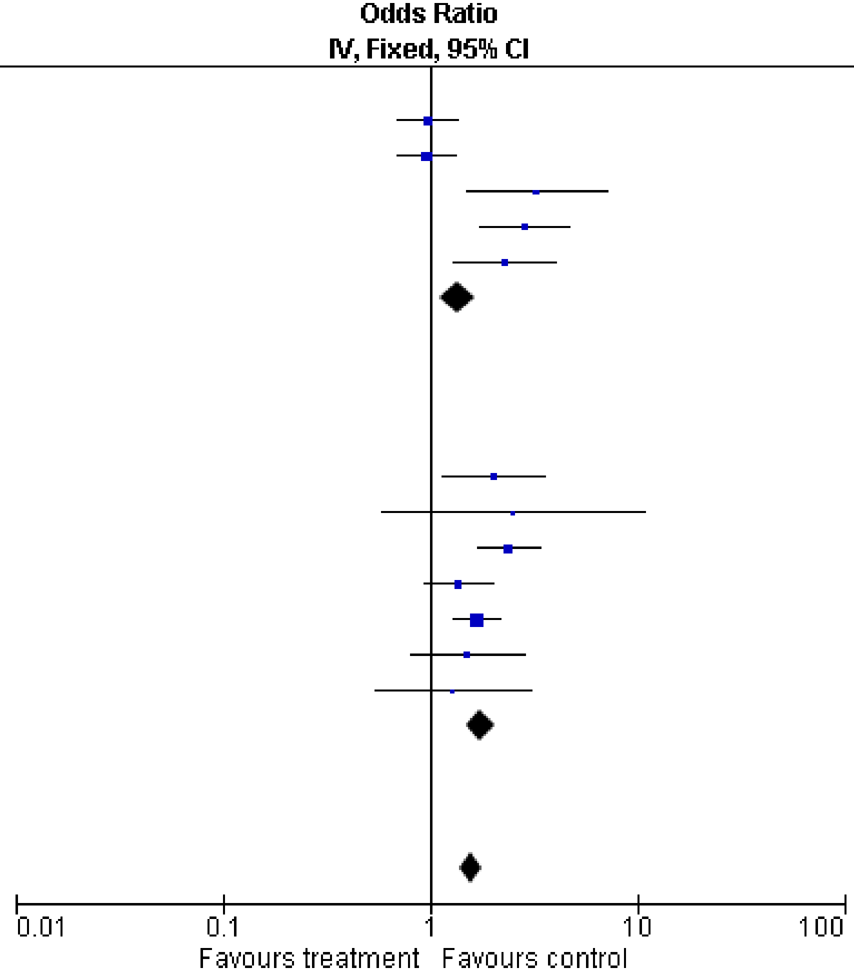
Fixed-effects test for subgroup differences

$$Q_{bet} = 34.42 - (24.38 + 6.00)$$
$$= 4.04$$

- Two subgroups, so 2 – 1 = 1 degree of freedom
- Chi-squared value of 4.04 with 1 degree of freedom:
  P = 0.044

Test for subgroup differences

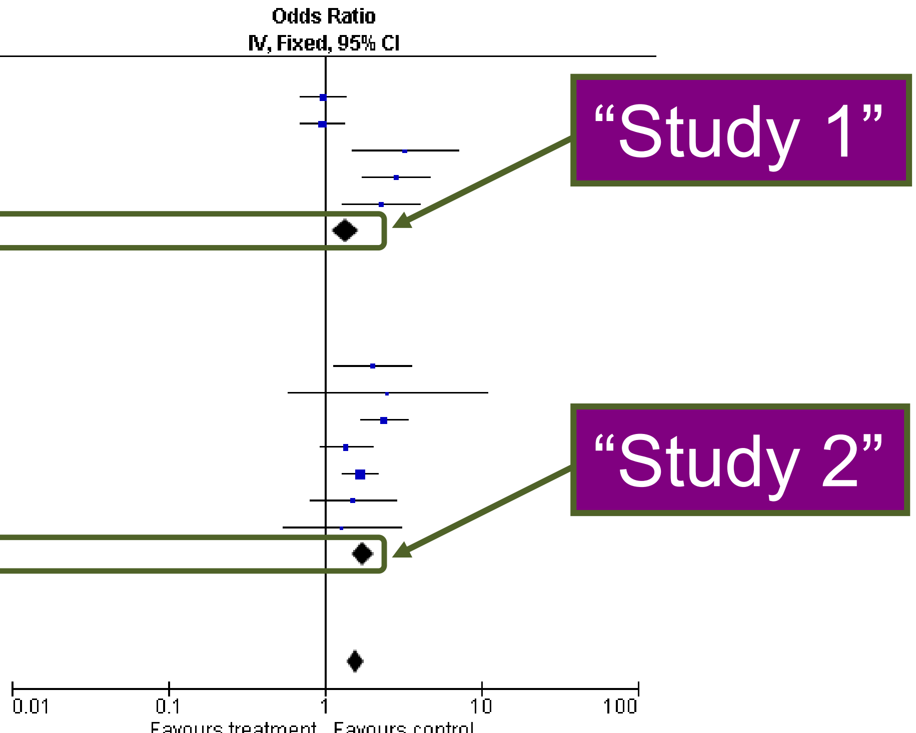
Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio IV, Fixed, 95% CI
1.1.1 Randomisation by organisation						
Cummings 1995	76	316	102	417	13.5%	0.98 [0.70, 1.38]
Gansky 2005	103	285	130	352	14.9%	0.97 [0.70, 1.34]
Severson 1998	40	394	8	239	2.6%	3.26 [1.50, 7.10]
Walsh 1999	60	171	30	189	6.3%	2.86 [1.74, 4.73]
Walsh 2003	38	141	23	166	4.7%	2.29 [1.29, 4.08]
Subtotal (95% CI)		**1307**		**1363**	**42.0%**	**1.36 [1.12, 1.64]**
Total events	**317**		**293**			
Heterogeneity: Chi² = 24.38, df = 4 (P < 0.0001); I² = 84%						
Test for overall effect: Z = 3.08 (P = 0.002)						
1.1.2 Individual randomisation						
Boyle 2004	44	109	28	112	4.8%	2.03 [1.14, 3.60]
Cigrang 2002	7	31	3	29	0.7%	2.53 [0.59, 10.90]
Seversen 2006	119	393	60	392	12.9%	2.40 [1.69, 3.41]
Seversen 2007a	69	535	52	534	10.8%	1.37 [0.94, 2.01]
Seversen 2007b	159	1260	100	1263	22.6%	1.68 [1.29, 2.19]
Stevens 1995	25	245	19	273	4.0%	1.52 [0.81, 2.83]
Stotts 2003	19	198	8	105	2.1%	1.29 [0.54, 3.05]
Subtotal (95% CI)		**2771**		**2708**	**58.0%**	**1.76 [1.49, 2.07]**
Total events	**442**		**270**			
Heterogeneity: Chi² = 6.00, df = 6 (P = 0.42); I² = 0%						
Test for overall effect: Z = 6.72 (P < 0.00001)						
Total (95% CI)		**4078**		**4071**	**100.0%**	**1.58 [1.39, 1.79]**
Total events	**759**		**563**			
Heterogeneity: Chi² = 34.42, df = 11 (P = 0.0003); I² = 68%						
Test for overall effect: Z = 7.12 (P < 0.00001)						
Test for subgroup differences: Chi² = 4.04, df = 1 (P = 0.04), I² = 75.3%						



General test for subgroup differences

- Just do a test for heterogeneity across the results of all the subgroups
- Fine for any number of subgroups, and any type of analysis within the subgroups

Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Weight	Odds Ratio IV, Fixed, 95% CI
1.1.1 Randomisation by organisation						
Cummings 1995	76	316	102	417	13.5%	0.98 [0.70, 1.38]
Gansky 2005	103	285	130	352	14.9%	0.97 [0.70, 1.34]
Severson 1998	40	394	8	239	2.6%	3.26 [1.50, 7.10]
Walsh 1999	60	171	30	189	6.3%	2.86 [1.74, 4.73]
Walsh 2003	38	141	23	166	4.7%	2.29 [1.29, 4.08]
Subtotal (95% CI)	**317**	**1307**	**293**	**1363**	**42.0%**	**1.36 [1.12, 1.64]**
Heterogeneity: Chi² = 24.38, df = 4 (P < 0.0001); I² = 84%						
Test for overall effect: Z = 3.08 (P = 0.002)						
1.1.2 Individual randomisation						
Boyle 2004	44	109	28	112	4.8%	2.03 [1.14, 3.60]
Cigrang 2002	7	31	3	29	0.7%	2.53 [0.59, 10.90]
Seversen 2006	119	393	60	392	12.9%	2.40 [1.69, 3.41]
Seversen 2007a	69	535	52	534	10.8%	1.37 [0.94, 2.01]
Seversen 2007b	159	1260	100	1263	22.6%	1.68 [1.29, 2.19]
Stevens 1995	25	245	19	273	4.0%	1.52 [0.81, 2.83]
Stotts 2003	19	198	8	105	2.1%	1.29 [0.54, 3.05]
Subtotal (95% CI)	**442**	**2771**	**270**	**2708**	**58.0%**	**1.76 [1.49, 2.07]**
Heterogeneity: Chi² = 6.00, df = 6 (P = 0.42); I² = 0%						
Test for overall effect: Z = 6.72 (P < 0.00001)						
Total (95% CI)	**759**	**4078**	**563**	**4071**	**100.0%**	**1.58 [1.39, 1.79]**
Heterogeneity: Chi² = 34.42, df = 11 (P = 0.0003); I² = 68%						
Test for overall effect: Z = 7.12 (P < 0.00001)						
Test for subgroup differences: Chi² = 4.04, df = 1 (P = 0.04), I² = 75.3%						



Part 3: Meta-regression

Meta-regression

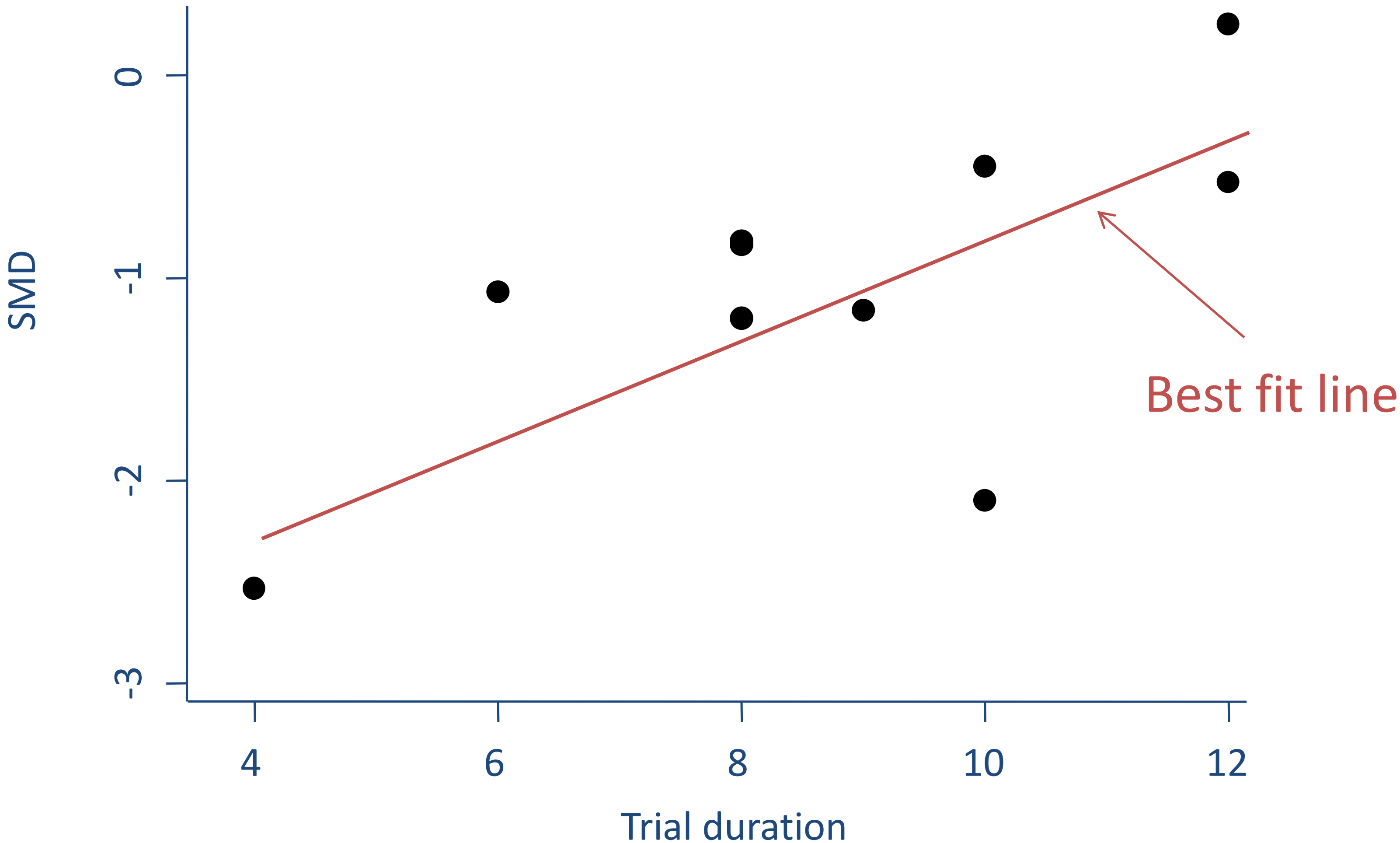
- A general framework for looking at possible explanations for differences in effect sizes across studies

- A test for subgroup differences can be done using meta-regression

- But meta-regression also good for continuous characteristics

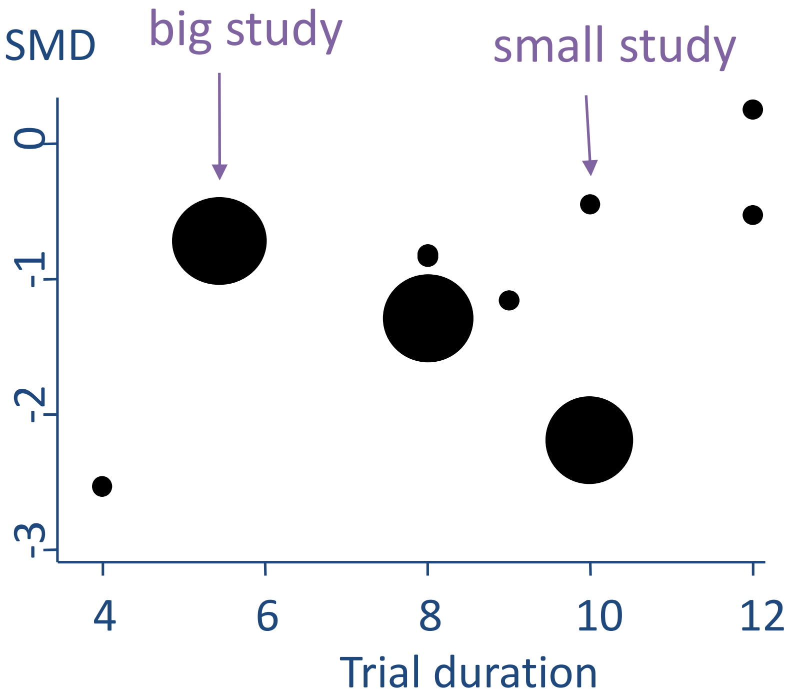
- Linear regression
  - outcome variable = effect size (e.g. SMD or logRR)
  - explanatory variable = summary description of study (e.g. duration or location)

Simple linear regression

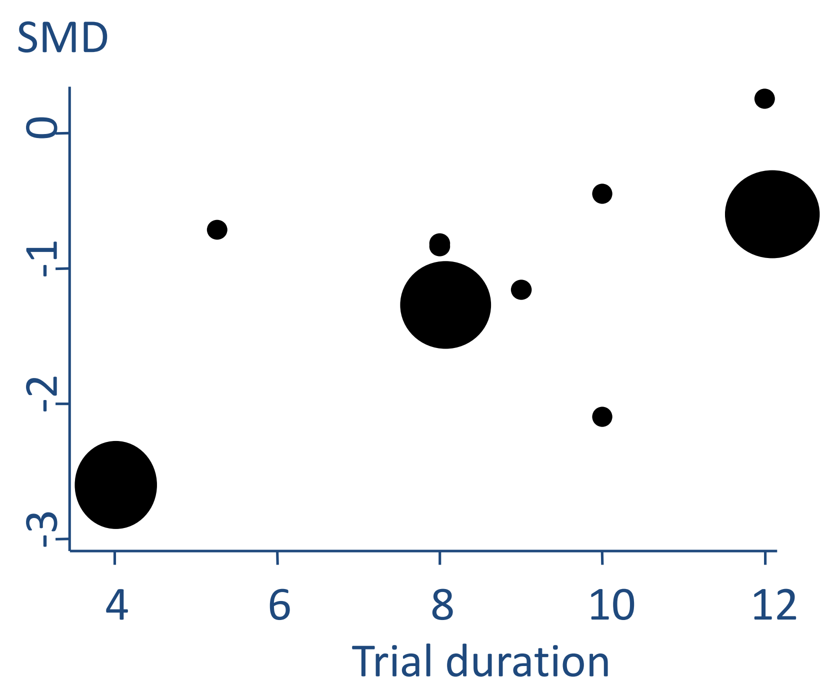


Why we don't use simple linear regression

- Just as in meta-analysis, the studies are different sizes, and should have different influences on the analysis
- There's a difference between

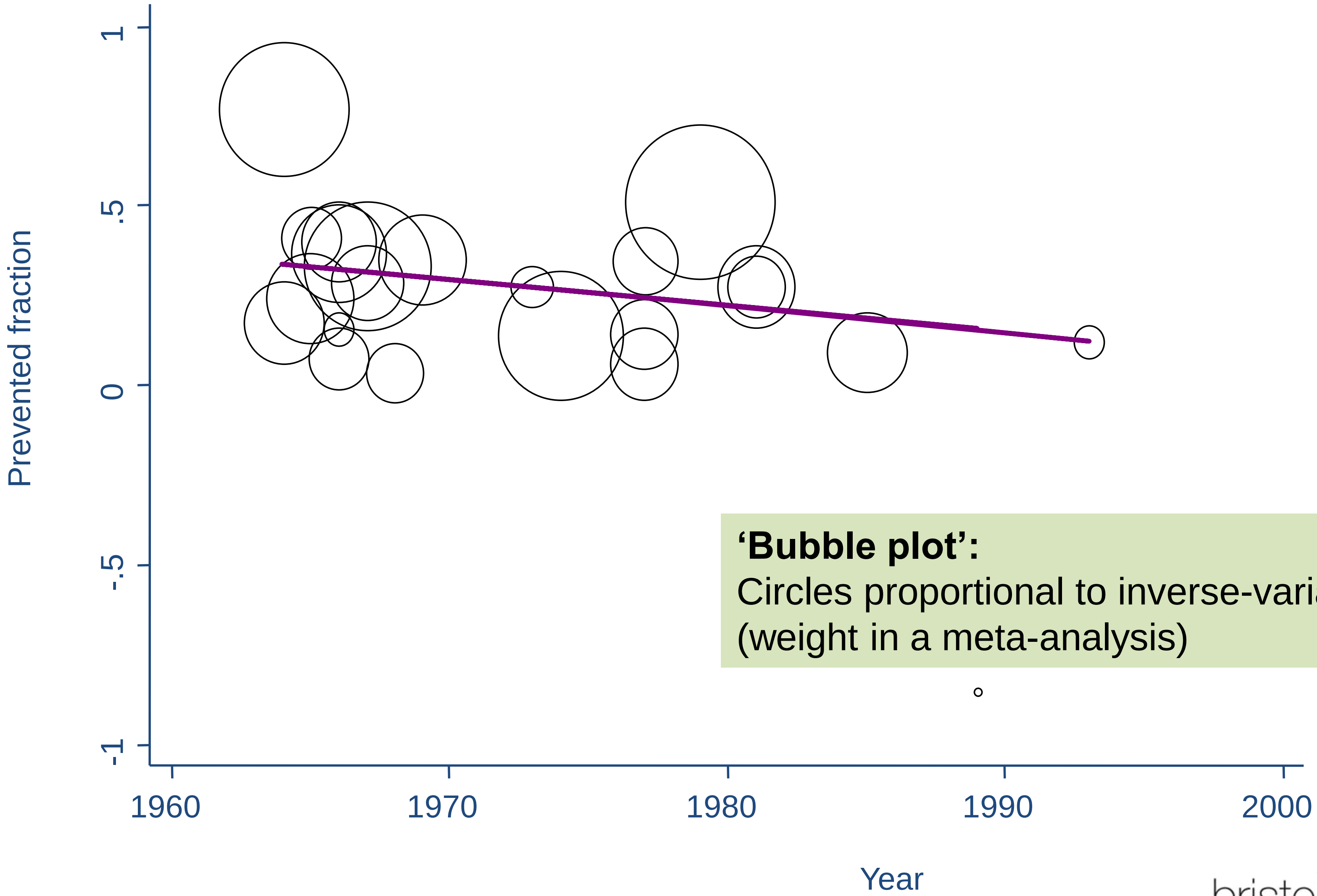


and



Has effectiveness of fluoride gels changed over time?

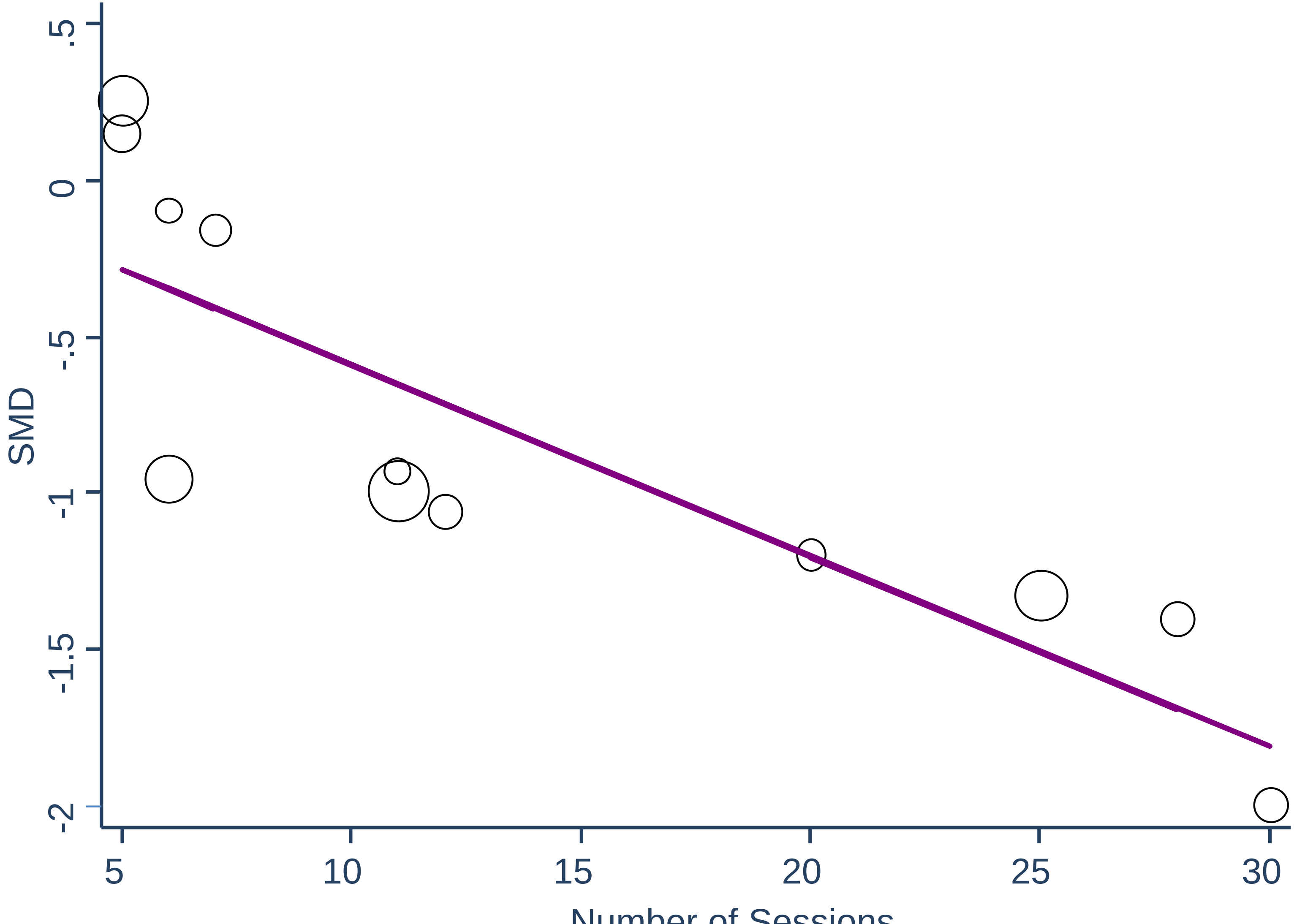
Marinho et al (2004)



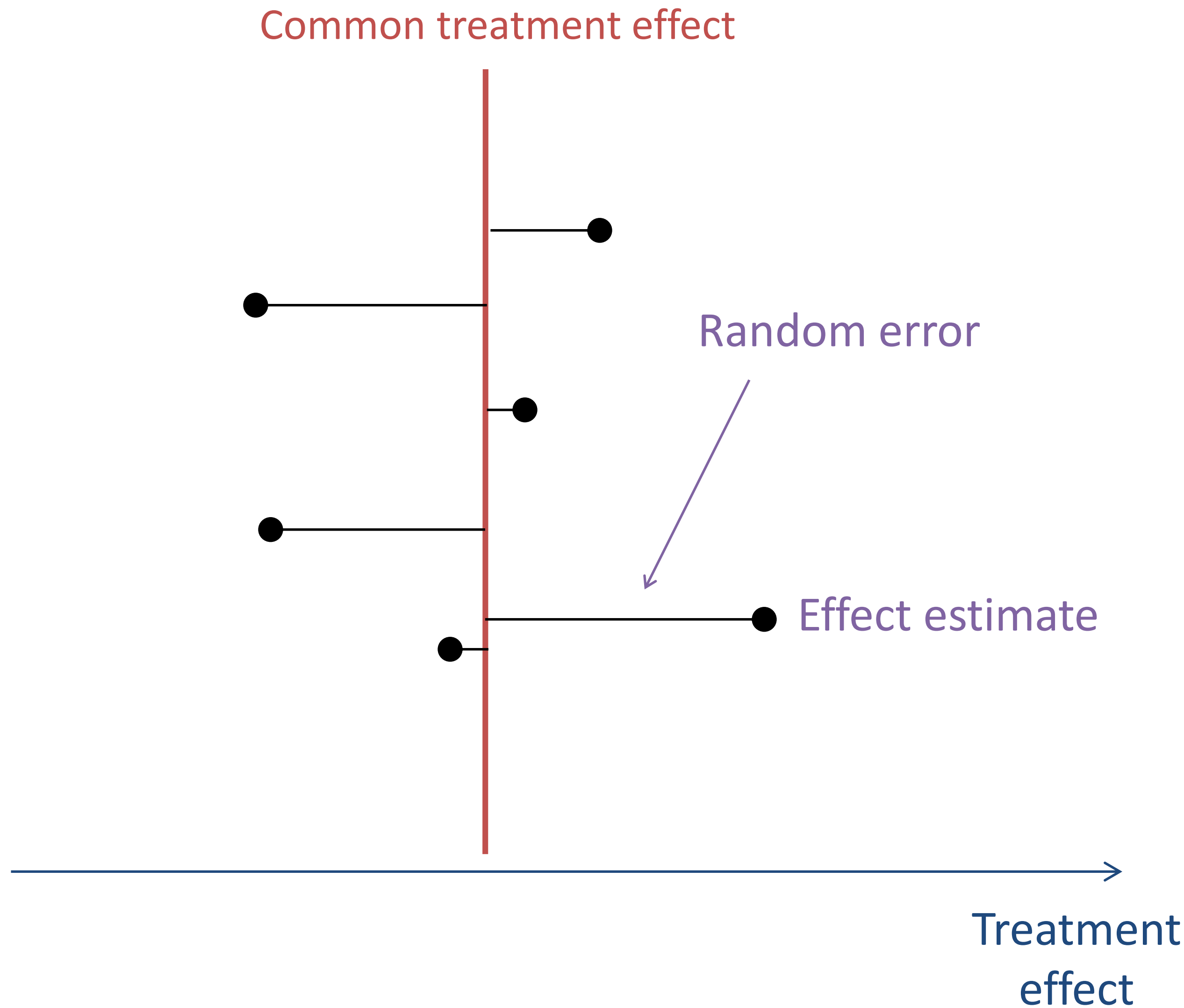
‘Bubble plot’:
Circles proportional to inverse-variance (weight in a meta-analysis)

University of BRISTOL

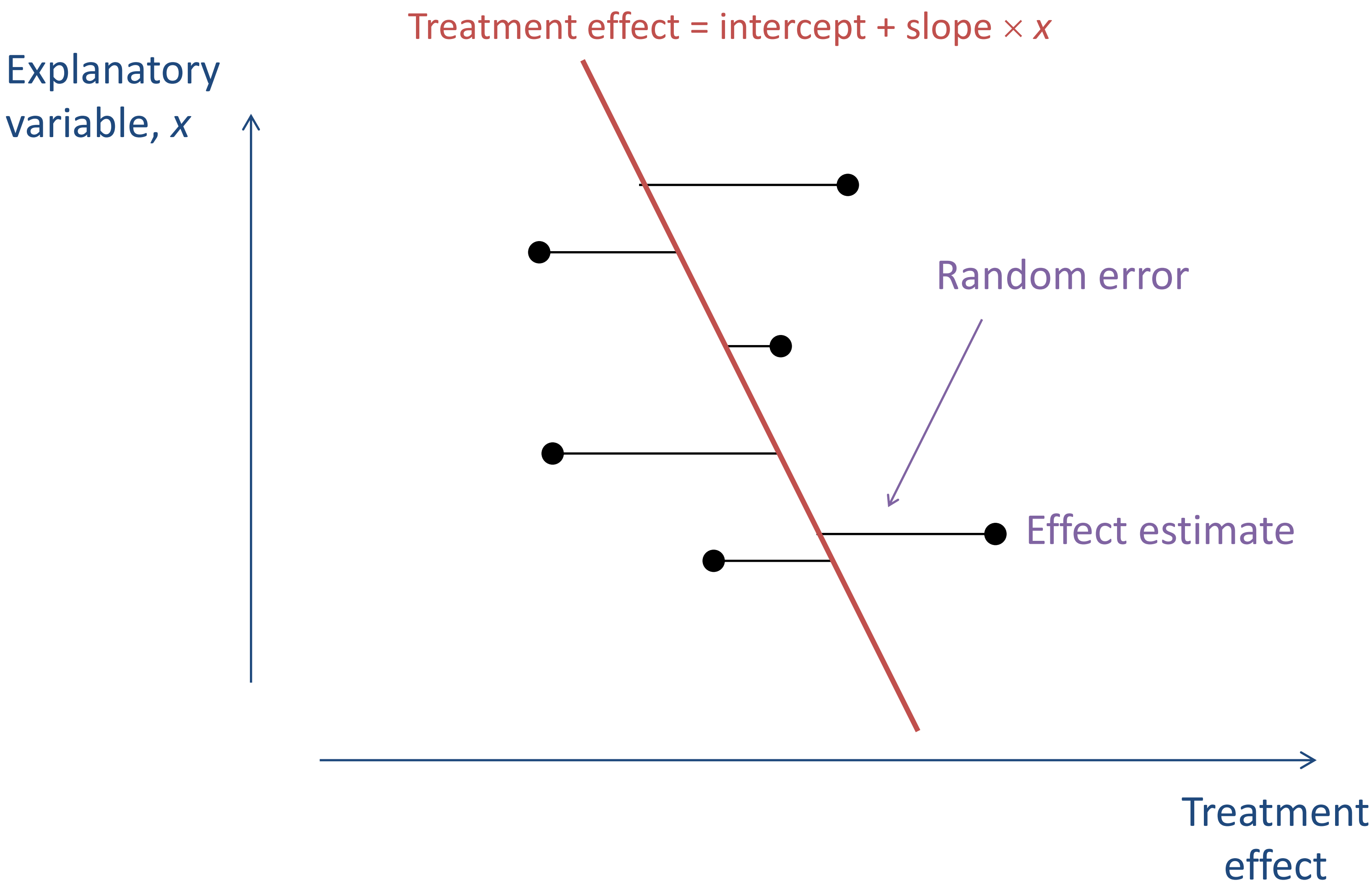
Number of sessions of CBT



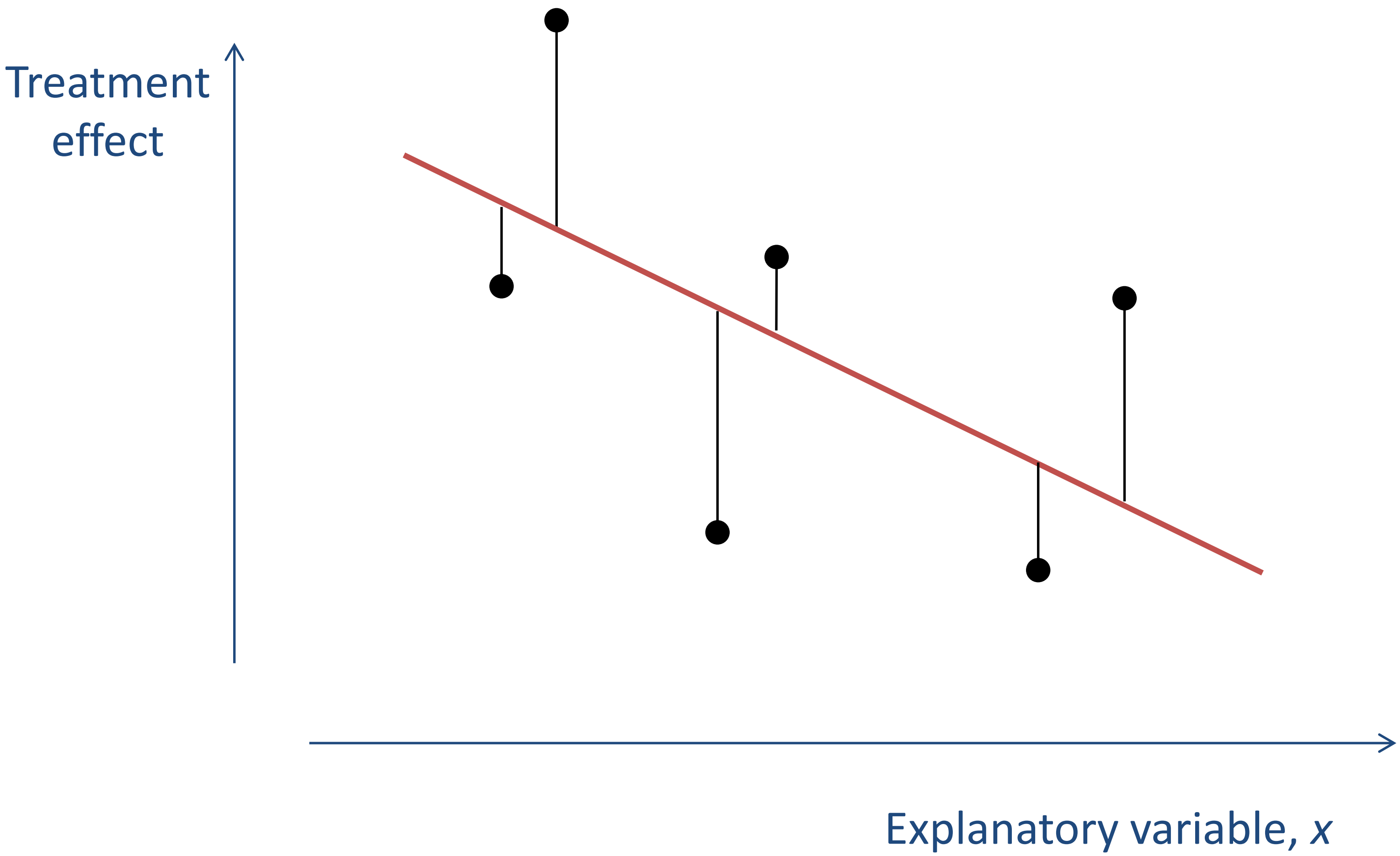
Fixed-effect meta-analysis



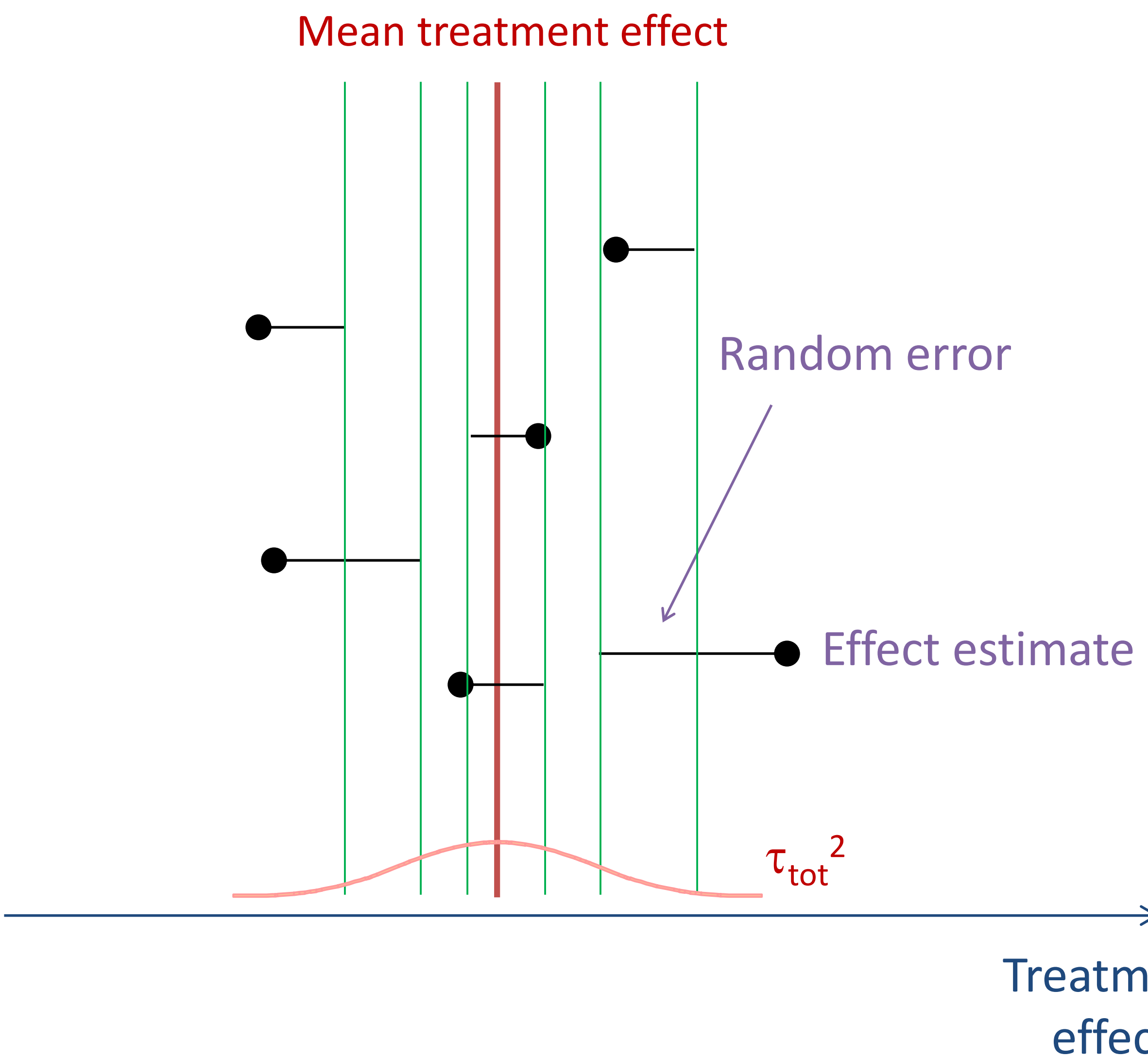
Fixed-effect meta-regression



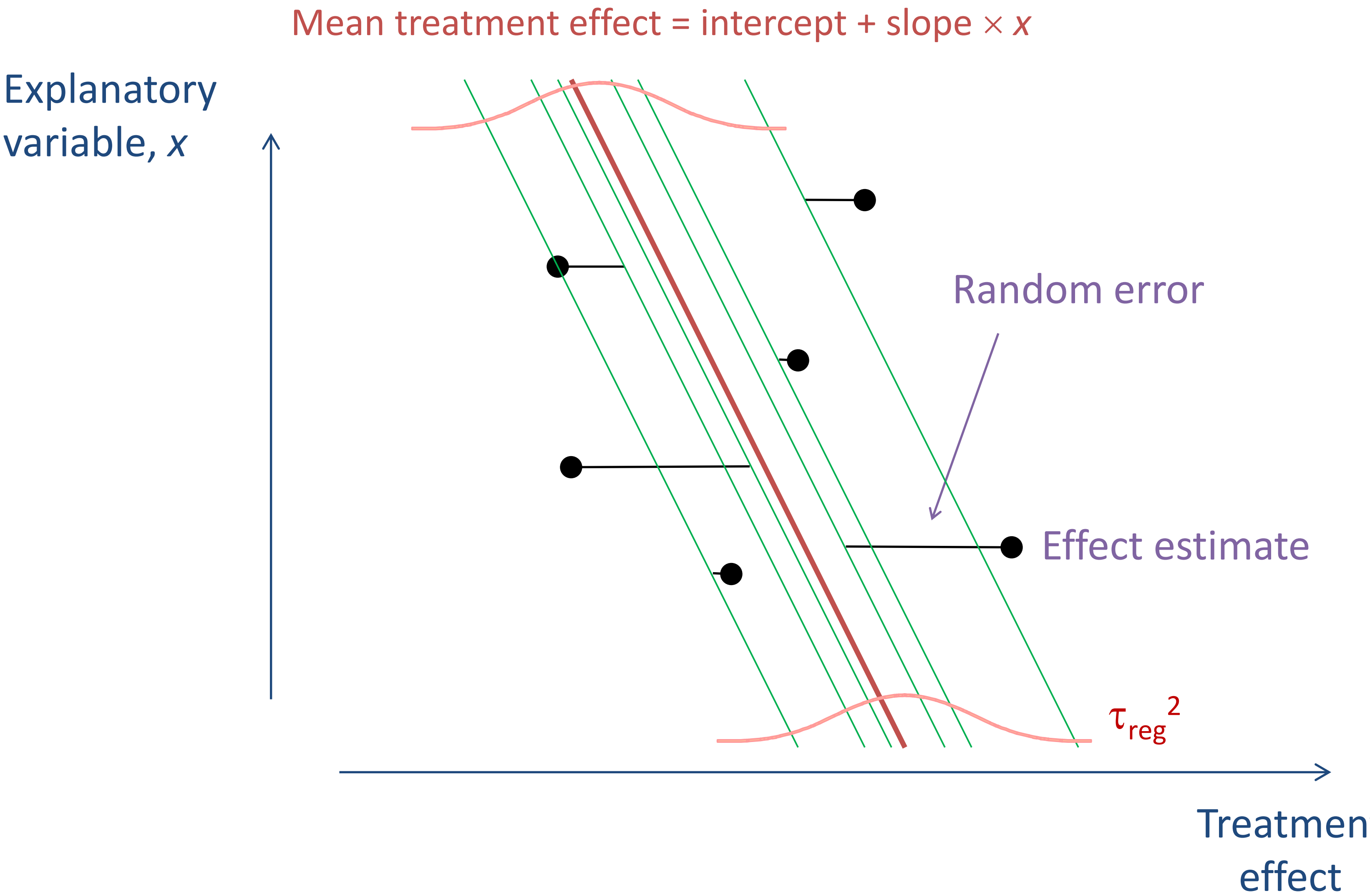
The other way round
(the convention for plotting)



Random-effects meta-analysis



Random-effects meta-regression



Proportion of heterogeneity explained

- Compare

heterogeneity variance from random-effects meta-analysis (τ^2_{tot})

with

heterogeneity variance from random-effects meta-regression (τ^2_{reg})

- % variance explained = $100\% \times \dfrac{\tau^2_{tot} - \tau^2_{reg}}{\tau^2_{tot}}$

- A useful measure of the explanatory ability of a (set of) covariate(s)

Example - BCG vaccination

It has been recognised for many years that the protection given by BCG against tuberculosis varied between trials

Risk ratios:

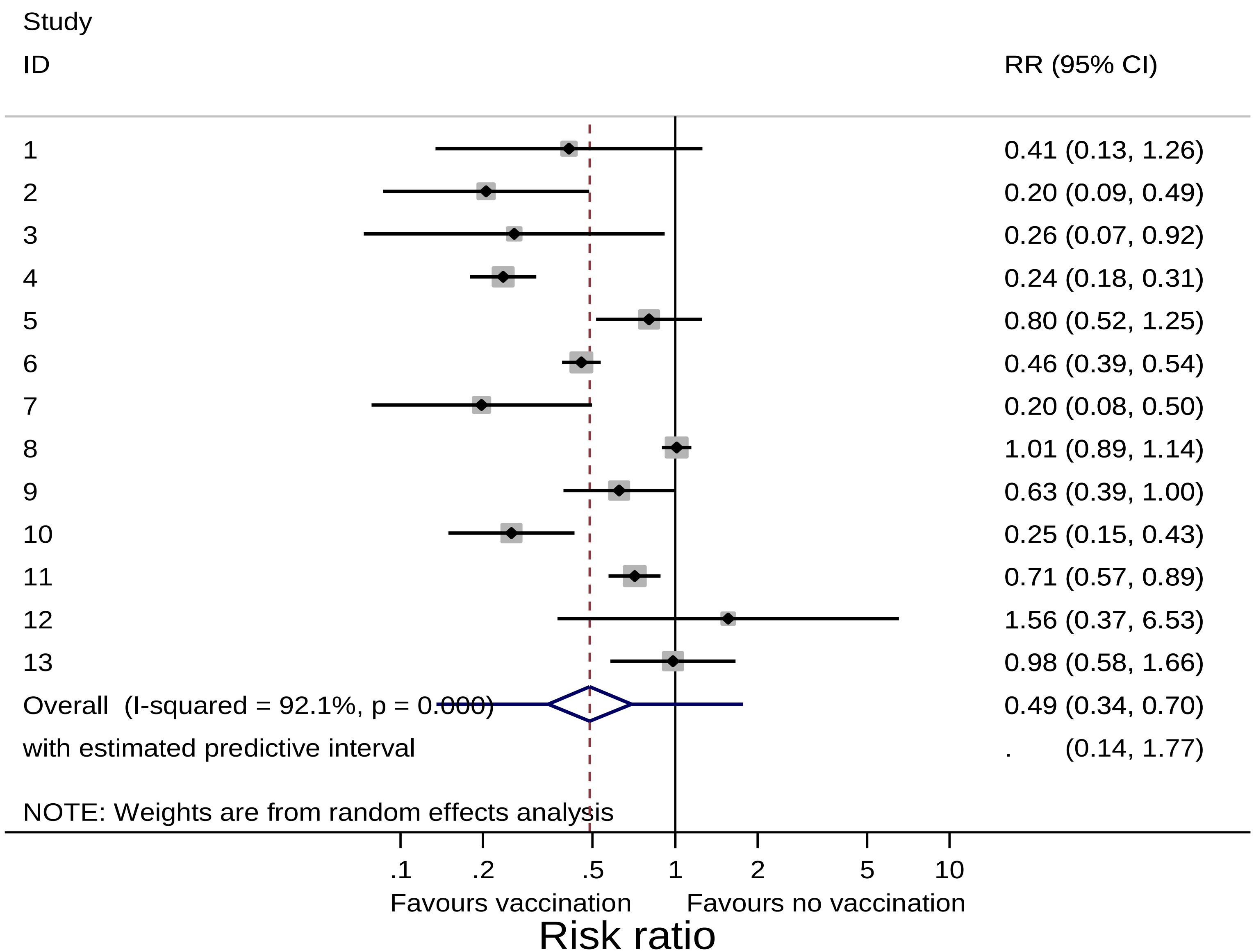
MRC trial (UK)	**0.24 (95% CI 0.18 to 0.31)**
Madras, (South India)	**1.01 (95% CI 0.89 to 1.14)**

Random-effects meta-analysis (Colditz et al. 1994)

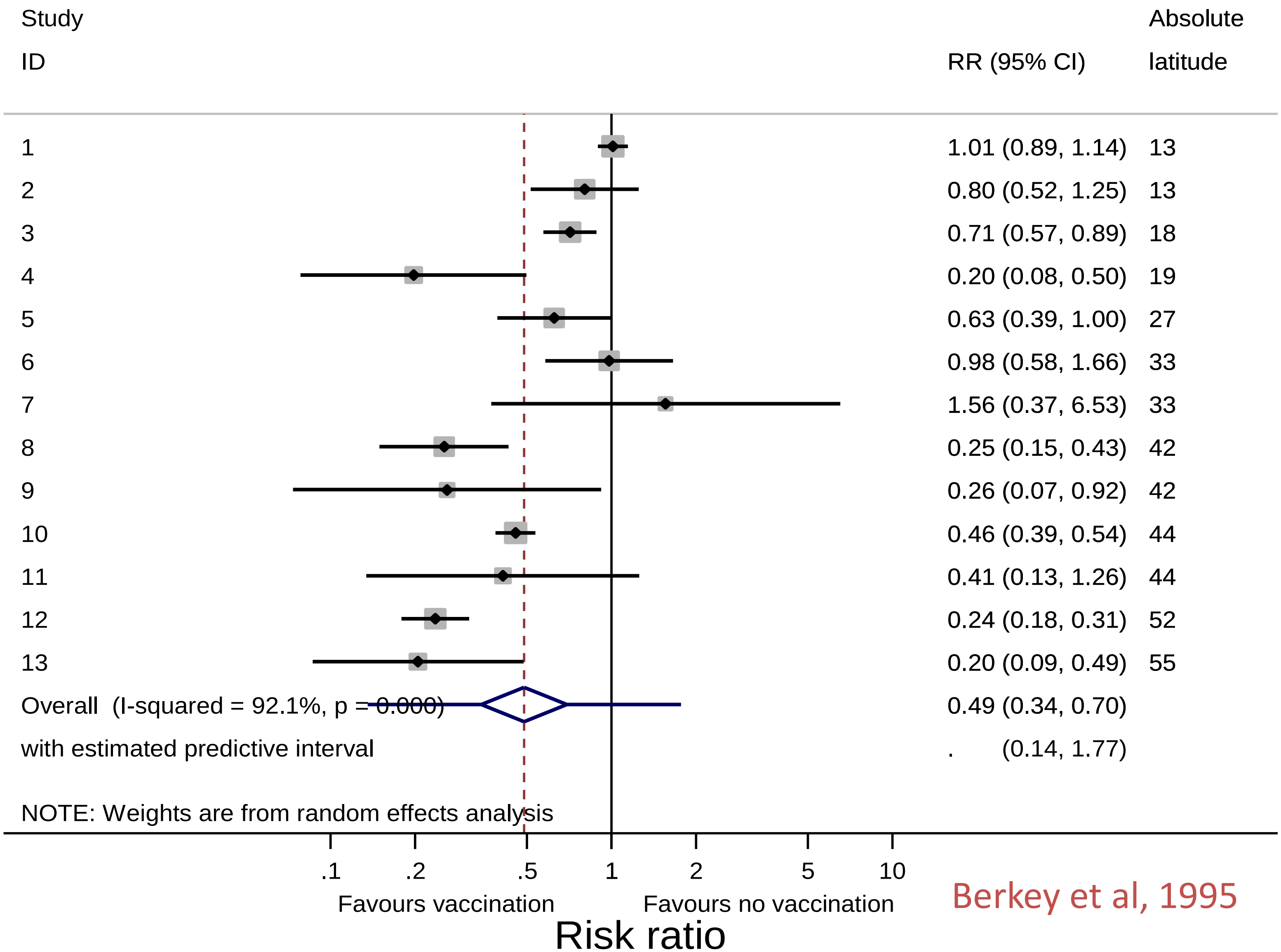
Summary (random effects) RR: 0.49 (95% CI 0.34 to 0.70)

“the results of this meta-analysis lend added weight and confidence to arguments favouring the use of BCG vaccine”

BCG vaccination to prevent tuberculosis

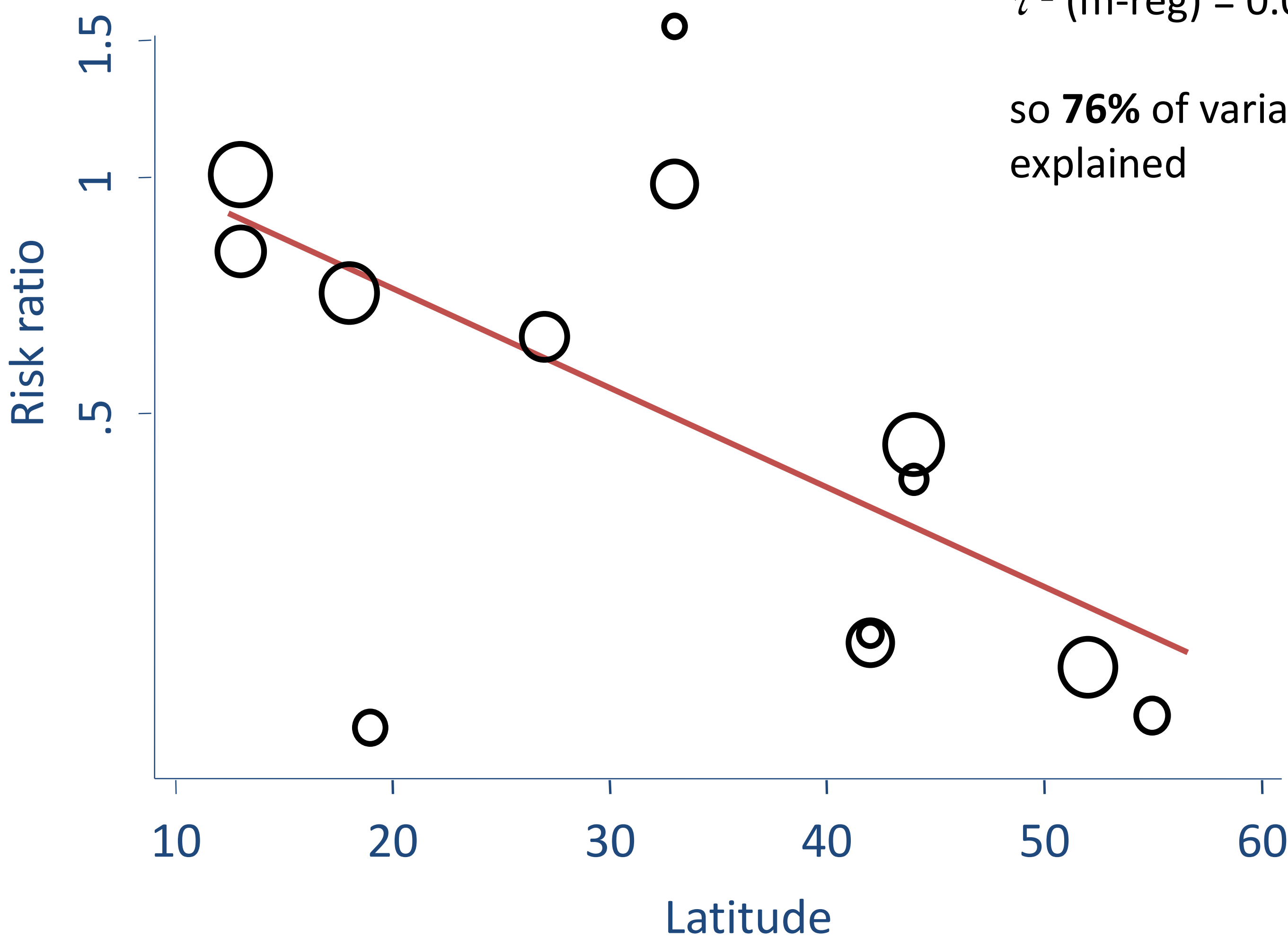


BCG vaccination to prevent tuberculosis



Berkey et al, 1995

Meta-regression: Dependence of BCG vaccine efficacy on study latitude



τ^2 (m-a) = 0.313
τ^2 (m-reg) = 0.076

so **76%** of variance is explained

Hartung-Knapp

- Simulation study:

"Of the different methods evaluated, only the Knapp and Hartung method and the permutation test provide adequate control of the Type I error rate across all conditions. Due to its computational simplicity, the Knapp and Hartung method is recommended as a suitable option for most meta-analyses."

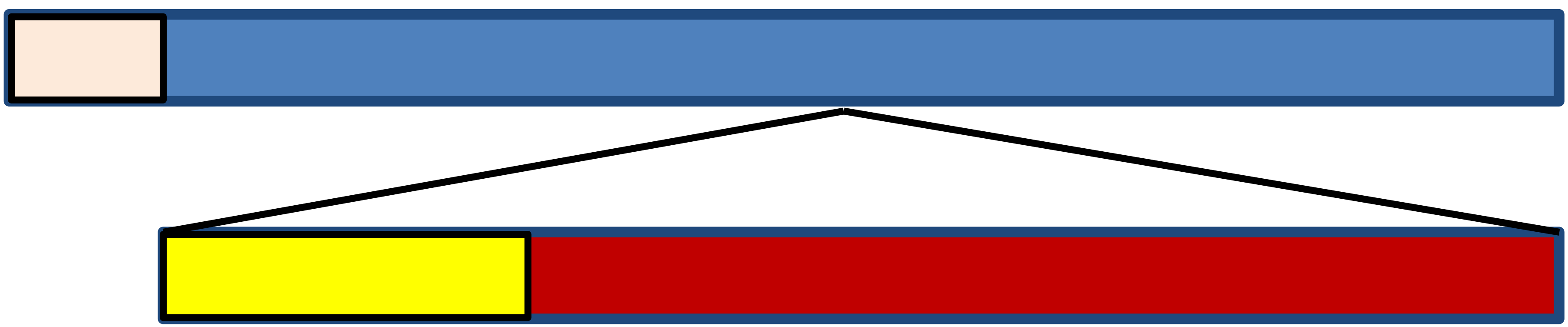
Viechtbauer et al 2015

- More on permutation tests later

Variance explained

Within studies:
8%

Between studies (I^2):
92%



Unexplained:
24%

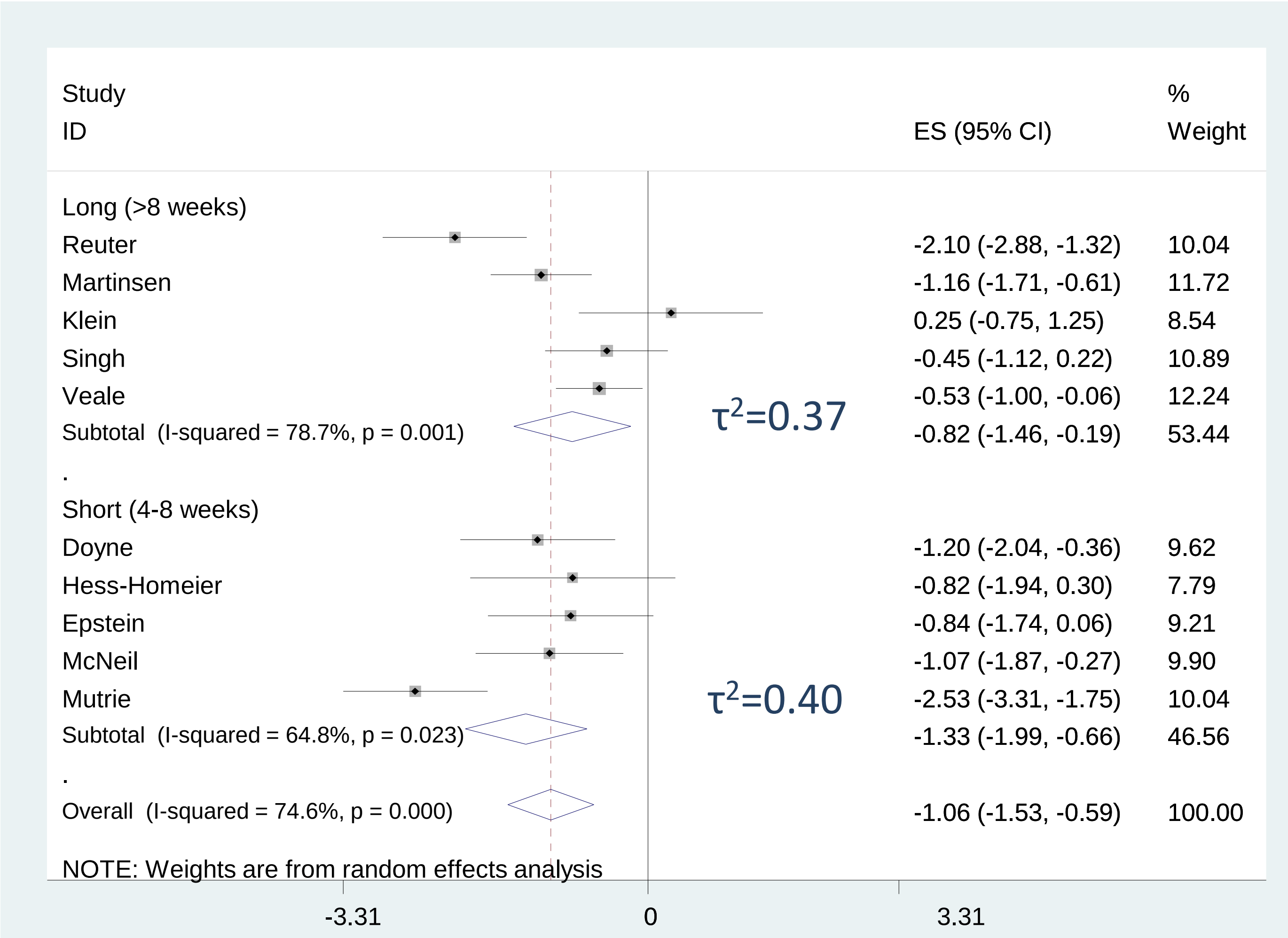
Explained by latitude (R^2):
76%

Subgroup analysis vs meta-regression

- The key difference is the assumption made about heterogeneity in random-effects analyses
  - estimated separately for each subgroup in subgroup analyses
    - (as usually implemented)
  - assumed equal across subgroups in meta-regression
    - (as usually implemented)

- Of course, there is flexibility to do things in different ways
  - some software assumes equal heterogeneity when performing subgroup analyses

```
. metan smd sesmd, random by(duration) label(namevar=study)
```



Tests for subgroup differences

1. Q test from subgroup results: P = 0.28 (RevMan/metan)

2. Meta-regression: P = 0.34 (metareg)

```
. metareg smd duration, wsse(sesmd) mm
Meta-regression                          Number of studies = 10
Method of moments estimate of between-study variance: tau2 =.3893
 smd       |  Coef.     Std. Err.  t     P>|t|  [95% Conf. Interval]
duration |-.5026523 .4990484 -1.01  0.343  -1.65346     .6481553
```

Interpretation of meta-regression coefficients

	Continuous outcome (MD/SMD)	**Dichotomous outcome (OR/RR)**
Continuous explanatory variable	Difference in mean differences or in SMDs *associated with 1 point increase in covariate*	Ratio of ORs or of RRs *associated with 1 point increase in covariate*
Dichotomous explanatory variable	Difference in mean differences or in SMDs *between 2 groups*	Ratio of ORs or of RRs *between 2 groups*

Meta-regression with dichotomous variable gives formal comparison between subgroups

Part 4: Problems

Problems with exploration of heterogeneity

- Observational relationships
  - Confounding (correlation between characteristics)

Treatment effect ↑ Year of randomization

Quality

Confounder

(associated with treatment effect and year)

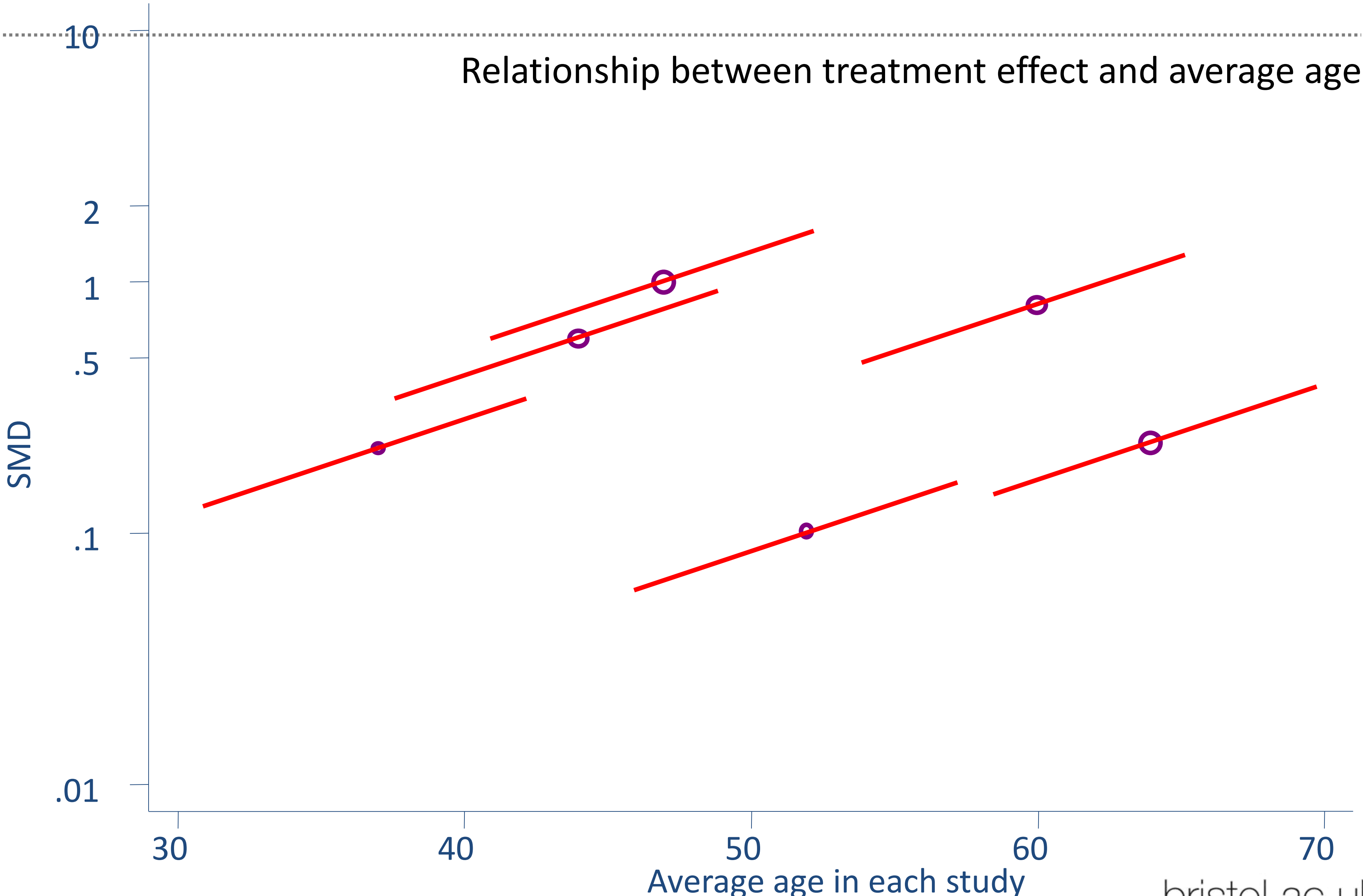
  - Aggregation bias
- Low power with few studies: false-negative findings
- Multiple characteristics: false-positive findings

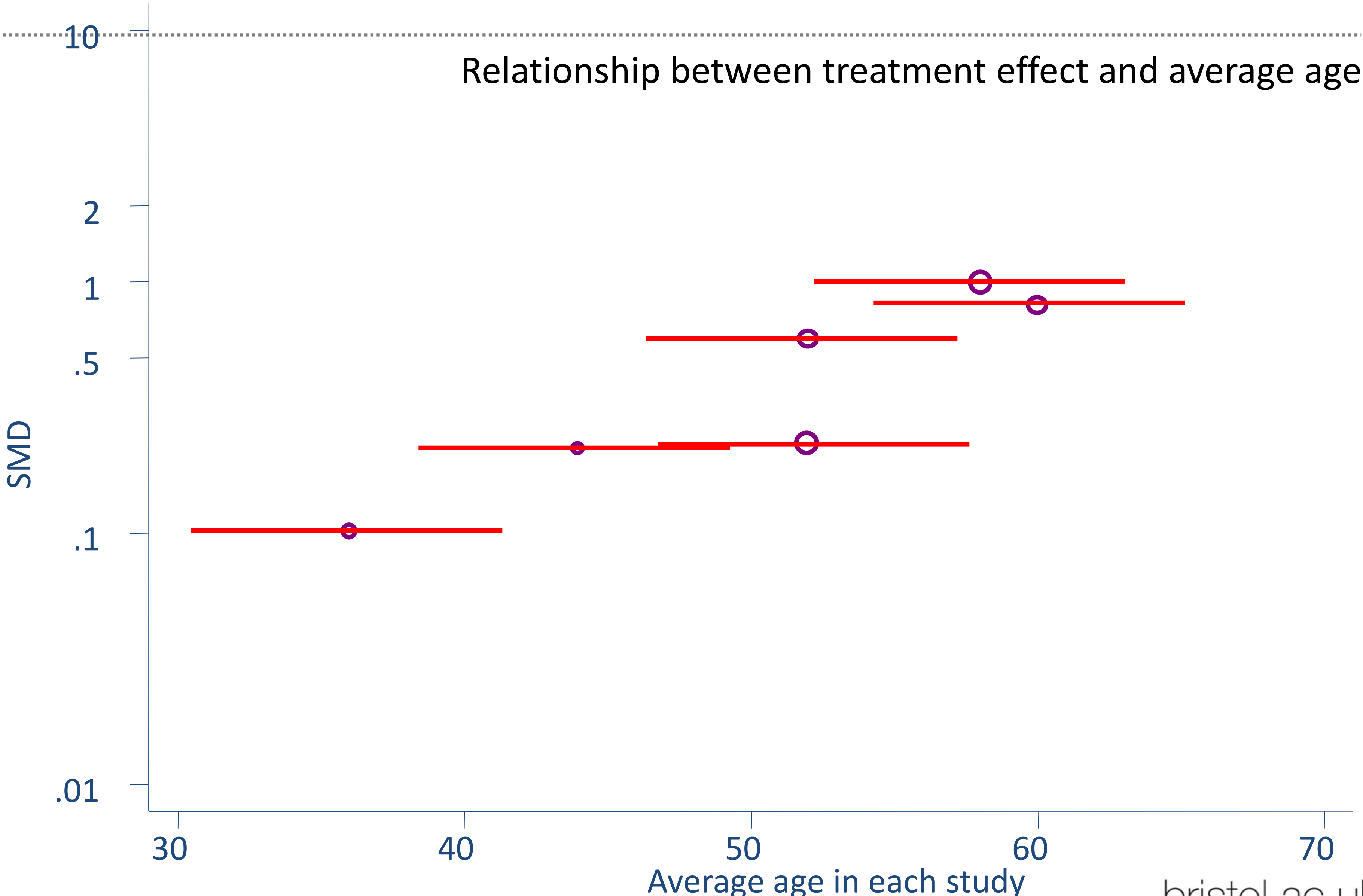
Thompson & Higgins (2002)

Aggregation bias

- Beware study characteristics that summarize participants within a study, e.g.
  - average age
  - % females
  - average duration of follow-up
  - % drop out

University of BRISTOL




Relationship between treatment effect and average age
SMD
10
2
1
.5
.1
.01
30
40
50
60
70
Average age in each study

Aggregation bias

- Beware study characteristics that summarize participants within a study, e.g.
  - average age
  - % females
  - average duration of follow-up
  - % drop out

- Relationships **across** studies may not reflect relationships **within** studies

- The relationship between treatment effect and age, sex, etc is best measured within a study
  - Needs individual participant data

University of BRISTOL

Lack of power

- Unfortunately most meta-analyses in Cochrane reviews do not have many studies

- Meta-regression typically has low power to detect relationships

- Model diagnostics / adequacy difficult to assess

University of BRISTOL

Selecting explanatory variables

- There are typically many, many explanatory variables to choose from
  - Heterogeneity can always be explained if you look at enough of them
  - Great risk of spurious findings

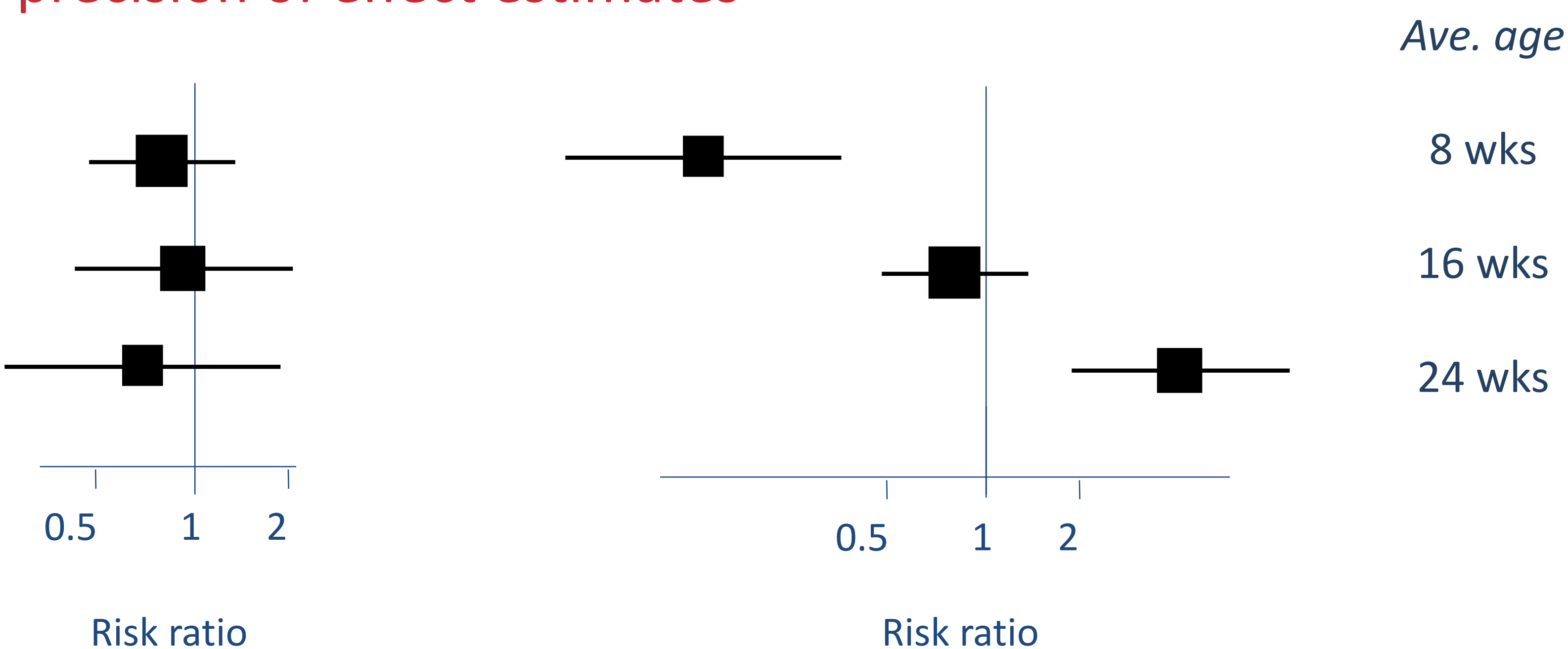
False-positive findings

- To guard against false-positives, meta-analysts are advised to
  - Pre-specify characteristics in a protocol
  - Limit to a small number
  - Have a scientific rationale

  - Beware ‘prognostic factors’
    - Variables that predict clinical outcome don’t necessarily affect treatment effects
    - e.g. age may be strongly prognostic, but risk ratios may well be the same irrespective of age

Risk of false-positive findings

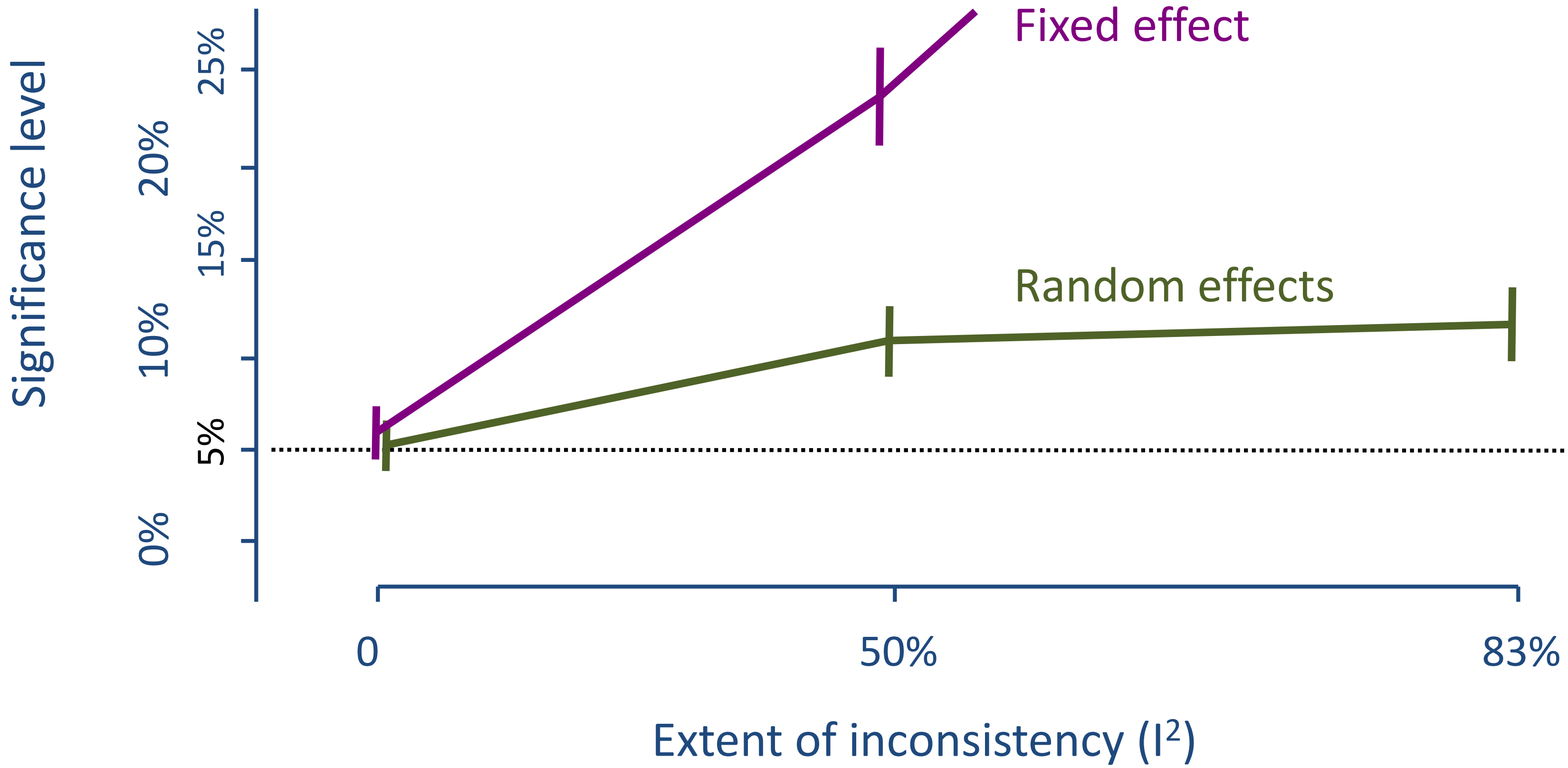
- Depends on
  - number of studies
  - extent of heterogeneity
  - precision of effect estimates



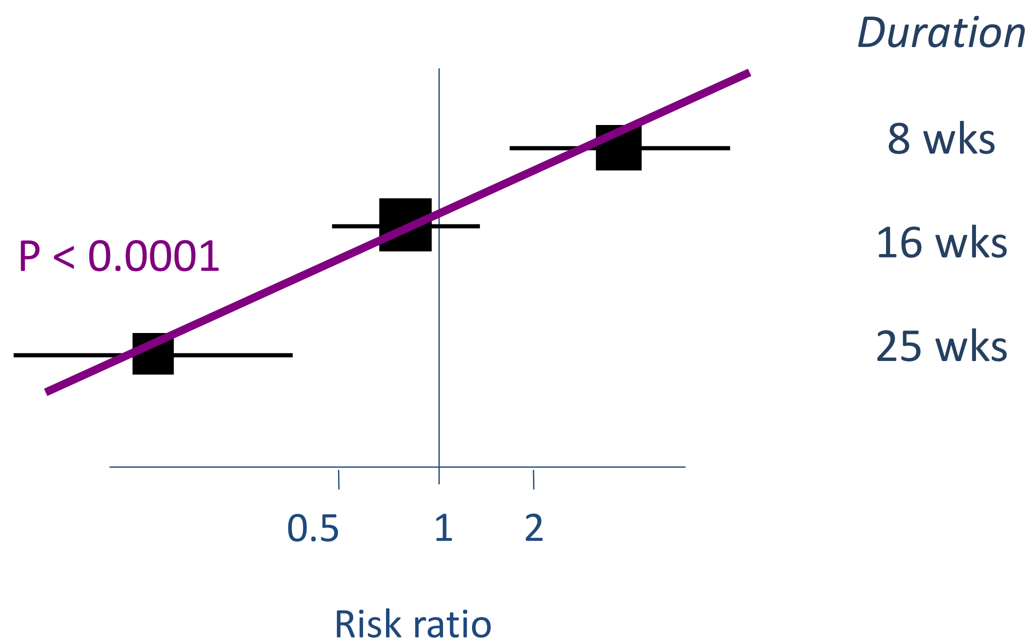
University of BRISTOL

Simulation study of false-positive rates

Example results: 10 studies, typical study weights, 1 covariate



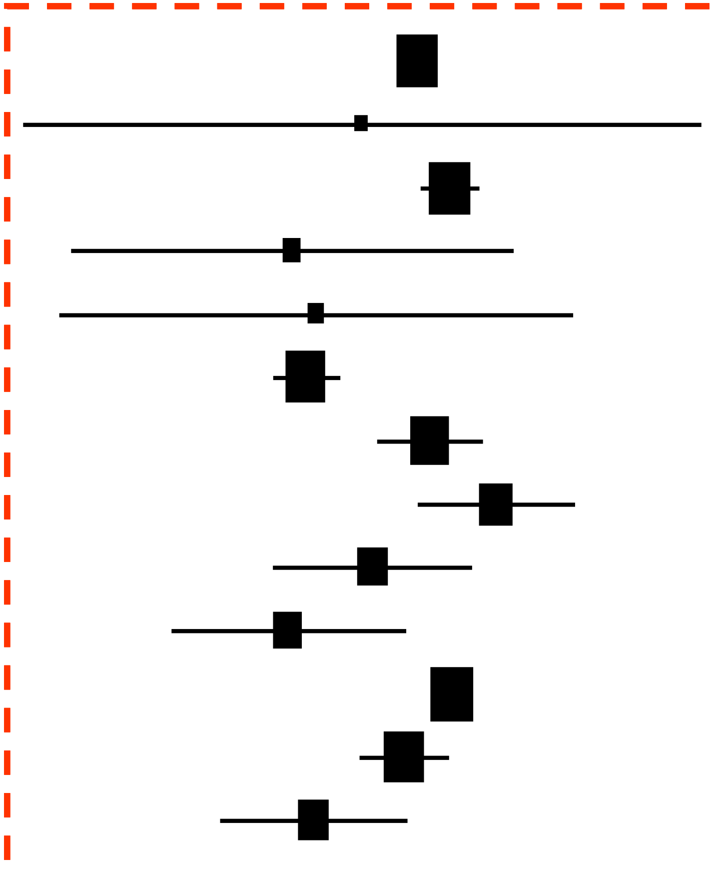
Solution? A permutation test



- But 1/3 of possible permutations give a relationship *at least as strong* as this
- So ‘real’ p = 0.33

University of BRISTOL

Illustration



Study	Absolute latitude	Shuffle 1		Shuffle 1000
1	18	55		33
2	44	44		27
3	33	42		52
4	55	13		42
5	42	33		44
6	52	52	*1000 times…*	18
7	13	42		19
8	33	18		13
9	44	44		33
10	19	19		44
11	13	13		55
12	27	33		42
13	42	27		13
Meta-regression test statistic	T^{obs}	$T^{(1)}$		$T^{(1000)}$

Permutation test *p*-value

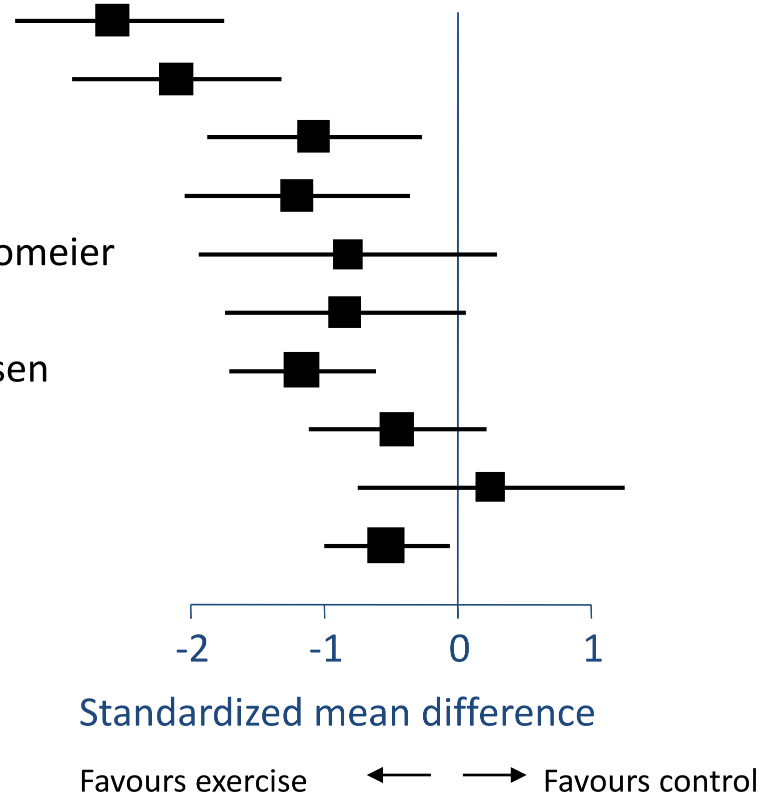
- Compare $\boxed{T^{obs}}$ with the values $\boxed{T^{(1)}}$ … , $\boxed{T^{(1000)}}$
- Count the number of $\boxed{T}$'s that exceed $\boxed{T^{obs}}$

- e.g. for BCG example $\boxed{T^{obs}}$ $= 2.44$ (= slope/SE);
- 54/1000 $\boxed{T}$'s exceed $\boxed{T^{obs}}$: $p = 0.054$

- c.f.
  - FE: $p < 10^{-10}$
  - RE (standard normal): $p = 0.012$

Multiple testing example: Exercise for depression

Study	abstract	duration	ITT	allocation	PhD
Mutrie	1	4	0	0	0
Reuter	1	10	0	0	0
McNeil	0	6	0	0	0
Doyne	0	8	0	0	0
Hess-Homeier	0	8	0	0	1
Epstein	0	8	0	0	1
Martinsen	0	9	0	1	0
Singh	0	10	1	1	0
Klein	0	12	0	0	0
Veale	0	12	0	1	0



	abstract	duration	ITT	allocation	PhD
RE meta-regression (STATA)	p=0.0006	p=0.01	p=0.24	p=0.25	p=0.98
Perm. test for "n^{th}" most significant	p=0.11	p=0.09	p=0.25	p=0.10	p=0.88

Part 5: Closing remarks

Selecting explanatory variables

Meta-regression in Cochrane reviews

Extensions

Summary

Selecting explanatory variables

- Specify a **small number** of characteristics **in advance**
- Ensure there is scientific rationale for investigating each characteristic
  - effect modifier or prognostic factor?
- Make sure the effect of a characteristic can be identified
  - does it differentiate studies?
  - aggregation bias
- Think about whether the characteristic is closely related to another characteristic
  - confounding

How many studies / characteristics?

- Typical guidance is to have 10 studies for each characteristic examined
- Some say 5 studies is enough

Including meta-regression in a Cochrane review

- Authors are encouraged to use meta-regression if appropriate
- ‘Bubble’ plots may be included as an ‘Other figure’
- Results should be presented in an ‘Additional table’
- Consider presenting something like:

Explanatory variable	Slope or Exp(slope)	95% confidence interval	Proportion of variation explained	Interpretation
Duration	OR = 1.3	0.9 to 1.8	14%	Odds ratio increases with duration (not statistically significant)

University of BRISTOL

Extensions

- Baseline risk of the studied population (as measured in the control group) might be considered as an explanatory variable
  - Beware! It is inherently correlated with treatment effects
  - Special methods are needed (see Thompson et al 1997)

- Precision of the estimate (e.g. its standard error) might be considered as an explanatory variable to look at small study effects (may be suggestive of publication bias)
  - Beware! It is often inherently correlated with treatment effects
  - Special methods are needed (see Sterne et al 2011)

Key messages

- Subgroup analysis and meta-regression examine the relationship between treatment effects and one or more study-level characteristics

- Meta-regression is like usual linear regression, but
  - study weights incorporated
  - variation not explained by the characteristics should be accounted for
  - random effects meta-regression does this;
    fixed effect meta-regression does not

Key messages

- Subgroup analyses are meta-regression are great in theory, and easy to perform in STATA or R (and other software)

- They are fraught with dangers and should be undertaken and interpreted with caution
  - observational relationships
  - too few studies
  - too many sources of diversity and bias
  - confounding and aggregation bias

References

- Berkey CS, Hoaglin DC, Mosteller F and Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995 **14**, 395-411
- Borenstein et al, *Introduction to Meta-analysis*, chapters 19, 20, 21
- Higgins JPT, Thompson SG. Controlling the risk of spurious results from meta-regression. *Statistics in Medicine* 2004; **23**: 1663-1682
- Higgins J, Thompson S, Altman D and Deeks J. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *Journal of Health Services Research and Policy* 2002; **7**: 51-61
- Hughes JR, Stead LF, Lancaster T. Antidepressants for smoking cessation (Cochrane Review). In: *The Cochrane Library*, Issue 4, 2004. Chichester, UK: John Wiley & Sons, Ltd
- Marinho VCC, Higgins JPT, Logan S, Sheiham A. Fluoride gels for preventing dental caries in children and adolescents (Cochrane Review). In: *The Cochrane Library*, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd
- Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; **343**: d4002

References

- Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; **2**: 173-192
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994; **309**: 1351-1355
- Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 2002; **21**: 1559-1574
- Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**: 2693-2708
- Thompson SG, Smith TC and Sharp SJ. Investigation underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**, 2741-2758
- van Houwelingen HC, Arends LR, Stijnen T. Tutorial in Biostatistics: Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**: 589–624
- Viechtbauer W, López-López JA, Sánchez-Meca J, Marín-Martínez F. A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods* 2015; **20**: 360-74.