INTRODUCTION TO META-ANALYSIS 3

dealing with heterogeneity

Anna Chaimani

Research Center of Epidemiology & Statistics, Sorbonne Paris Cité (CRESS-UMR1153), Paris Descartes University, France

Mark Simmonds

Centre for Reviews and Dissemination, University of York, UK

25th Cochrane Colloquium, Edinburgh, UK
18 September 2018

Acknowledgements: Julian Higgins, Georgia Salanti

We have no actual or potential conflict of interest in relation to this presentation

General principles of meta-analysis

- Participants of one study are not compared directly with the participants in another study
  - each study is analyzed separately
  - in each study we estimate the intervention effect preserving the randomization (e.g. RR, OR)
- In each study we assign a weight depending on the information it provides
  - in a way that large studies have greater influence in the summary effect
- The study-specific intervention effects are synthesized to obtain the summary effect of the meta-analysis

Why performing a meta-analysis?

- To increase the power of the analysis and get more precise results
  - obtaining narrower confidence intervals
  - detecting statistically significant effects

Why performing a meta-analysis?

- To increase the power of the analysis and get more precise results
  - obtaining narrower confidence intervals
  - detecting statistically significant effects

- To investigate the intervention effect under different conditions
  - exploration of heterogeneity

What is heterogeneity?

- The differences observed between the studies of a systematic review.

- Types of heterogeneity – diversity:
  1. Clinical
  2. Methodological
  3. Statistical

Clinical heterogeneity

- Participants
  - Age
  - Severity of condition
  - Geographical variation
- Interventions
  - Intensity / dose / duration
  - Sub-type of drug
  - Mode of administration,
  - Nature of the control (placebo/none/standard care)

Methodological heterogeneity

- Design
  - Randomised vs non-randomised
  - Cross-over vs parallel group vs cluster randomised
  - Follow-up duration
- Conduct
  - Allocation concealment
  - Blinding
  - Analysis method
- Outcomes
  - Definition of an event
  - Choice of measurement scale

Statistical heterogeneity

- Effect estimates will vary across studies

- Some variation is chance variation:
  - Studies are small
  - All results come with uncertainty
  - Effect estimates will vary by chance

- Some variation is genuine differences in the effect across studies
  - Clinical / methodological heterogeneity

- Statistical heterogeneity is the observed variation in effect estimates that cannot be explained by chance alone

Outcome data required from each study

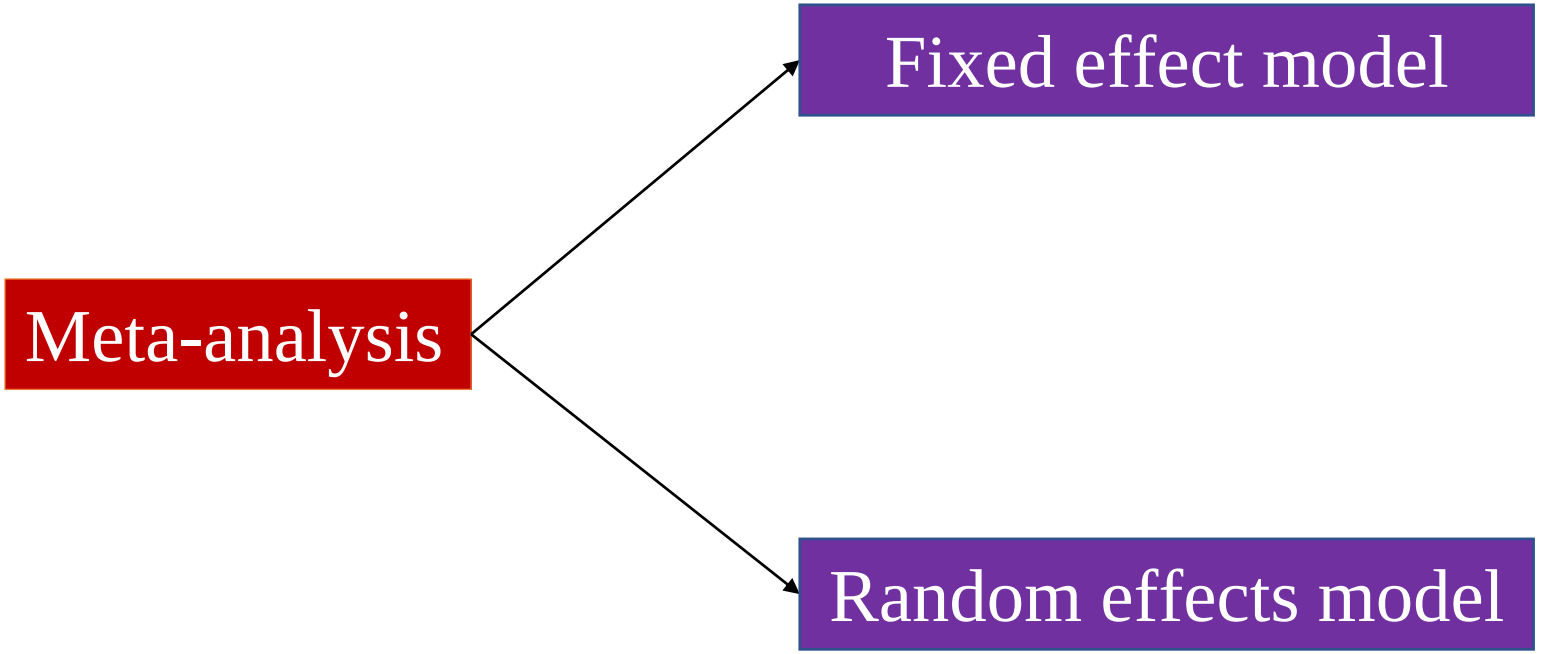
- Extract from each study an effect size and its uncertainty (standard error)

- Usually we present the effect sizes from all studies in a forest plot

How to synthesize these studies?

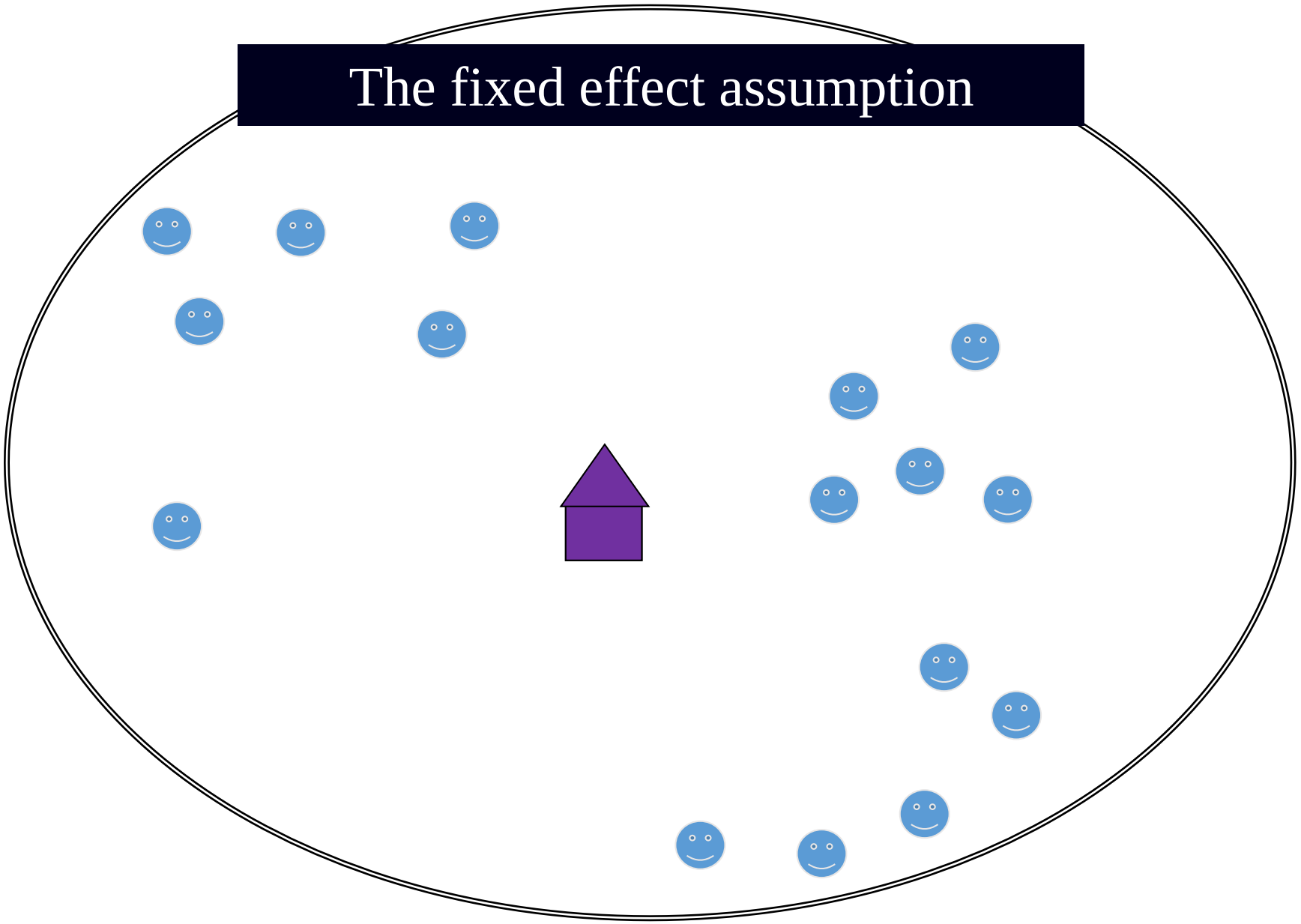
- By obtaining an average effect
  - Differences in level of uncertainty across the studies are ignored
- By pooling the different intervention arms across all studies
  - this approach breaks the randomization of the studies – comparison between treatment and control valid within studies but potentially invalid across studies
- By obtaining a weighted average
  - Randomization is preserved and larger (more precise) studies have larger weight in the analysis

Meta-analysis models

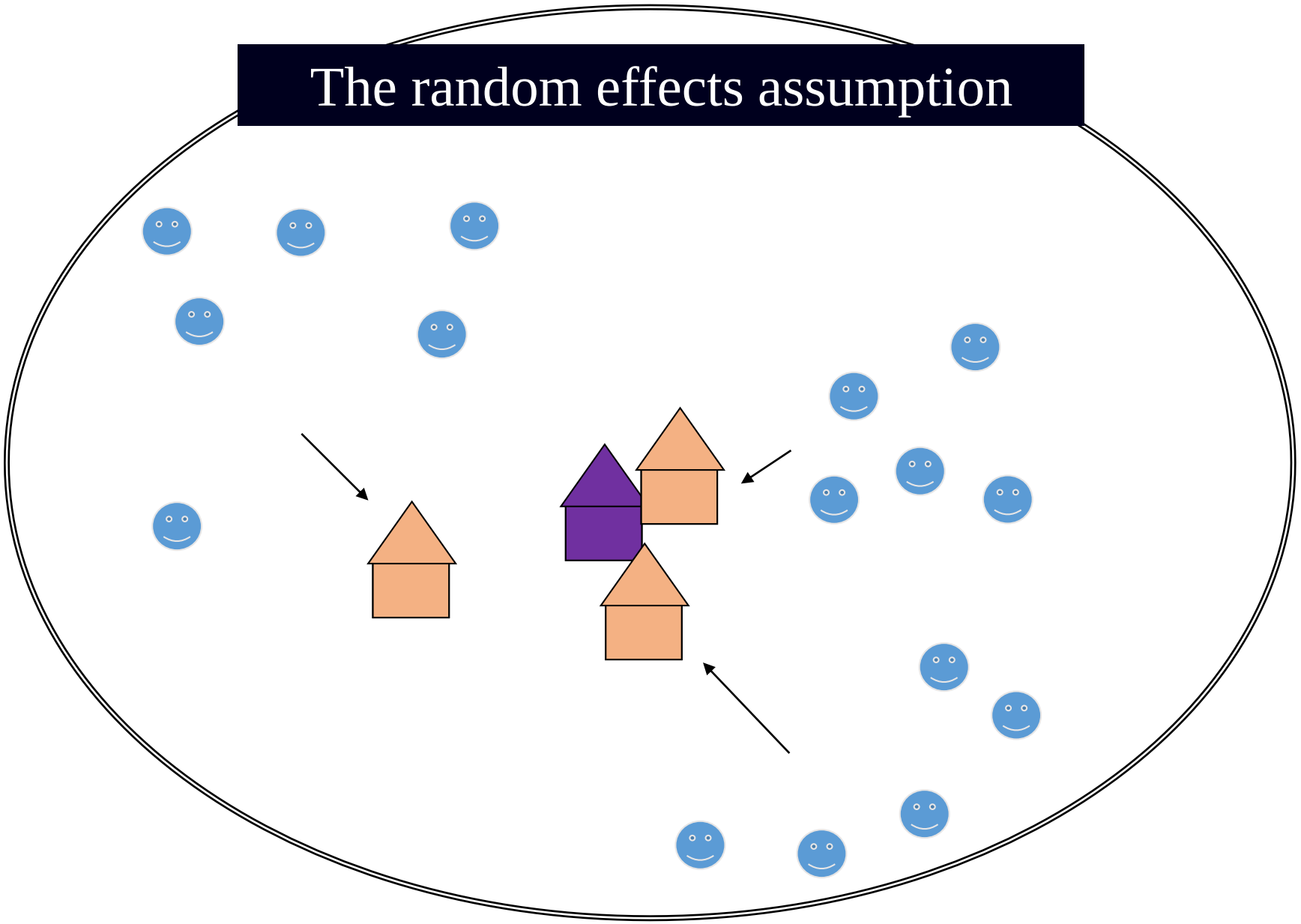


What are these models?

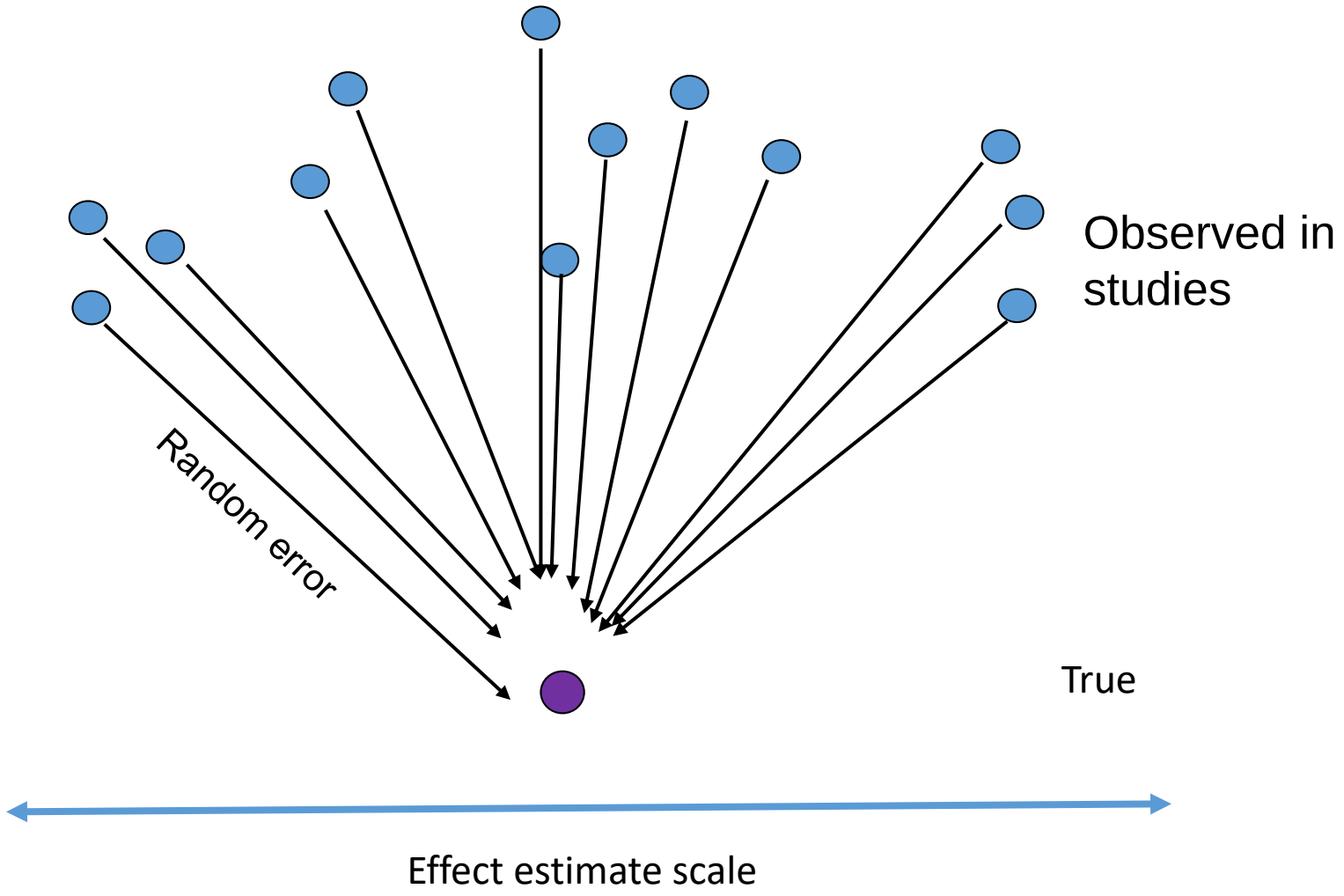
The fixed effect assumption



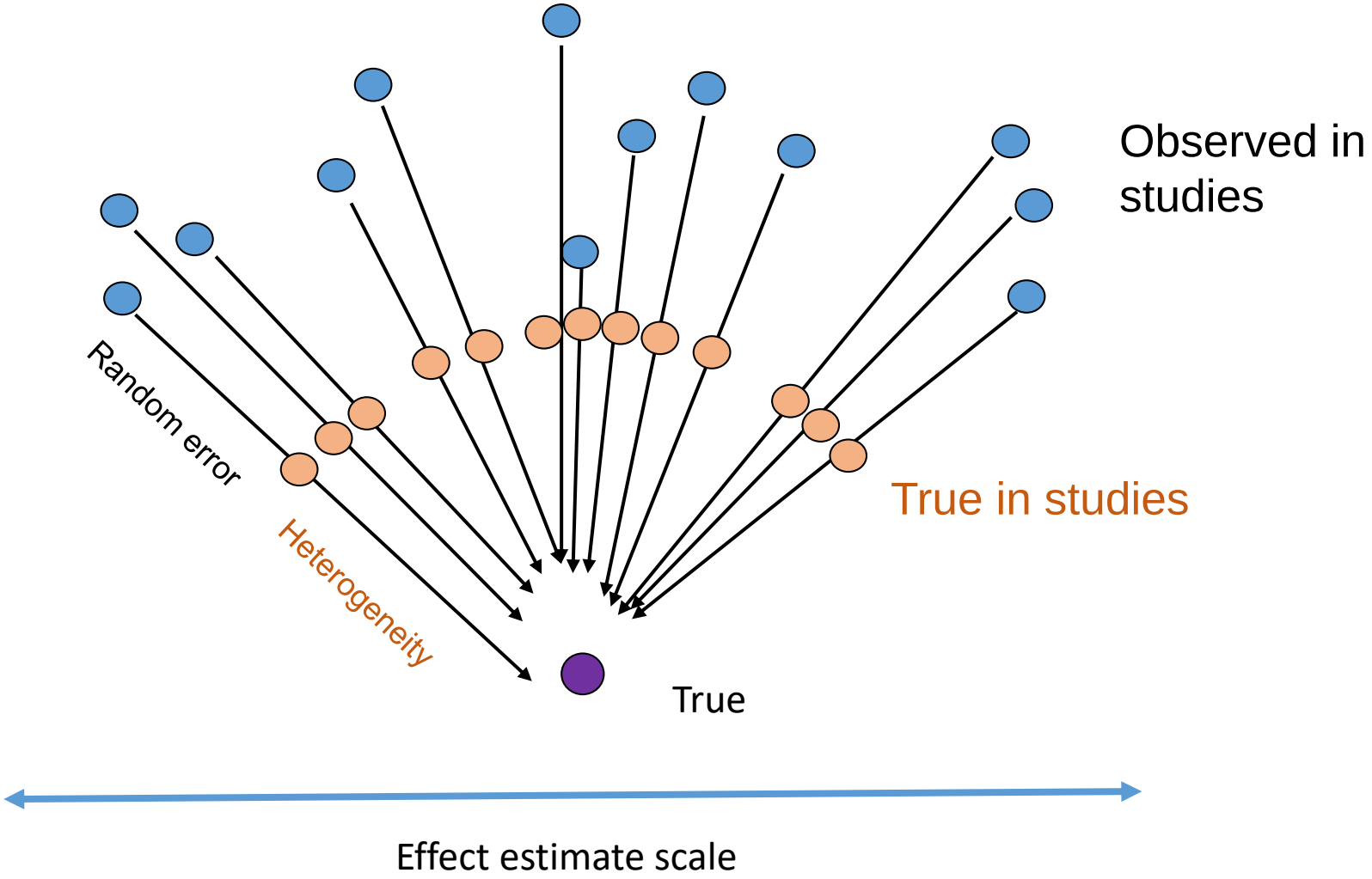
The random effects assumption



The fixed effect assumption



The random effects assumption



How to assign weights to the studies?

- Inverse variance method
  - any type of data, both fixed and random effects
  - in fact this is the maximum likelihood estimator!

- Mantel-Haenszel method
  - only binary data, only fixed effect (but there are ways to account for the heterogeneity)

- Peto method
  - only binary data, only odds ratio, only fixed effect

Fixed effect model

- Inverse variance method

- Weight is 1 ÷ variance

$$w_i = \frac{1}{var(\hat{y}_i)} \qquad \text{for each study } i$$

$$\hat{\theta}_{FE} = \frac{\sum w_i \hat{y}_i}{\sum w_i} \qquad var\left(\hat{\theta}_{FE}\right) = \frac{1}{\sum w_i}$$

Random effects model

- Inverse variance method
- Uncertainty in each trial is now BOTH the random variation AND the heterogeneity
- Weight is 1 ÷ (variance + heterogeneity)

$$w_i^* = \frac{1}{var(\hat{y}_i) + \tau^2} \quad \text{for each study } i$$

$$\hat{\theta}_{FE} = \frac{\sum w_i^* \hat{y}_i}{\sum w_i^*} \qquad var(\hat{\theta}_{FE}) = \frac{1}{\sum w_i^*}$$

The weights are smaller than before

Example: Organized inpatient rehabilitation

	OR	ln (OR)	var	weight FE		weight RE	
Study		y_i	v_i	w_i	$w_i y_i$	w_i^*	$w_i^* y_i$
Cameron 1993	0.98	-0.02	0.10	10.0	-0.2	7.6	-0.2
Fordham 1986	1.36	0.31	0.26	3.8	1.2	3.4	1.1
Galvard 1995	1.28	0.25	0.06	16.6	4.2	10.9	2.7
Gilchrist 1988	0.75	-0.29	0.14	7.1	-2.1	5.8	-1.7
Kennie 1988	0.45	-0.79	0.21	4.8	-3.8	4.1	-3.3
Total				**42.3**	**-0.65**	**31.8**	**-1.3**

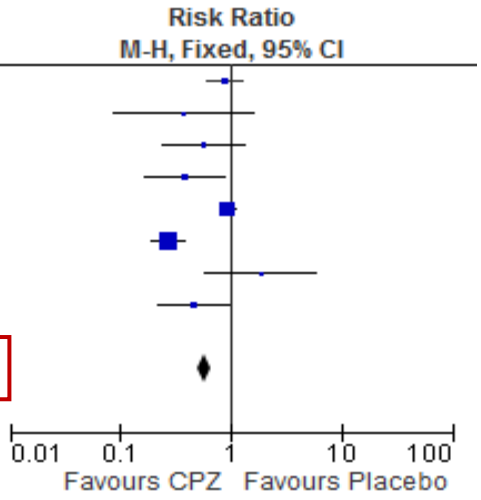
- Random effects meta-analysis
  - pooled odds ratio = exp{– 0.045} = 0.96
  - 95% confidence interval from 0.68 to 1.35
- Fixed effect analysis
  - pooled odds ratio = exp{– 0.02} = 0.98
  - 95% confidence interval from 0.72 to 1.32

Random effects model gives wider confidence intervals!

Example: Behaviour Deteriorated/Disturbed/Unco-operative

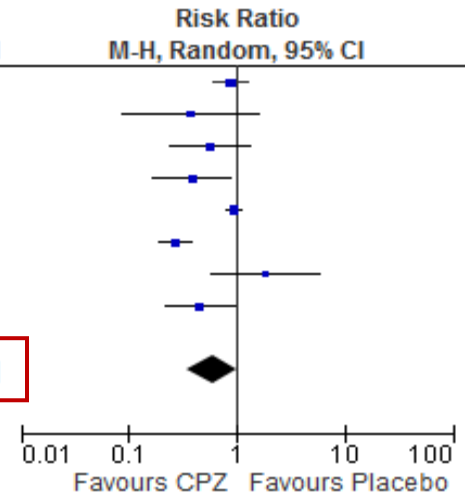
Study or Subgroup	Chlorpromazine Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Chouinard 1990	14	21	16	21	7.2%	0.88 [0.60, 1.29]
Clark 1970a	2	15	5	14	2.3%	0.37 [0.09, 1.62]
Clark 1970b	10	53	6	18	4.0%	0.57 [0.24, 1.34]
Fleming 1959	5	21	13	21	5.8%	0.38 [0.17, 0.89]
Hall 1955	65	87	70	88	31.2%	0.94 [0.80, 1.10]
Prien 1968	37	416	70	212	41.5%	0.27 [0.19, 0.39]
Serafetinides 1962	6	14	3	13	1.4%	1.86 [0.58, 5.94]
Somerville 1960	5	15	22	30	6.6%	0.45 [0.22, 0.96]
Total (95% CI)		**642**		**417**	**100.0%**	**0.58 [0.50, 0.67]**
Total events	144		205			

Heterogeneity: Chi² = 61.84, df = 7 (P < 0.00001); I² = 89%
Test for overall effect: Z = 7.25 (P < 0.00001)



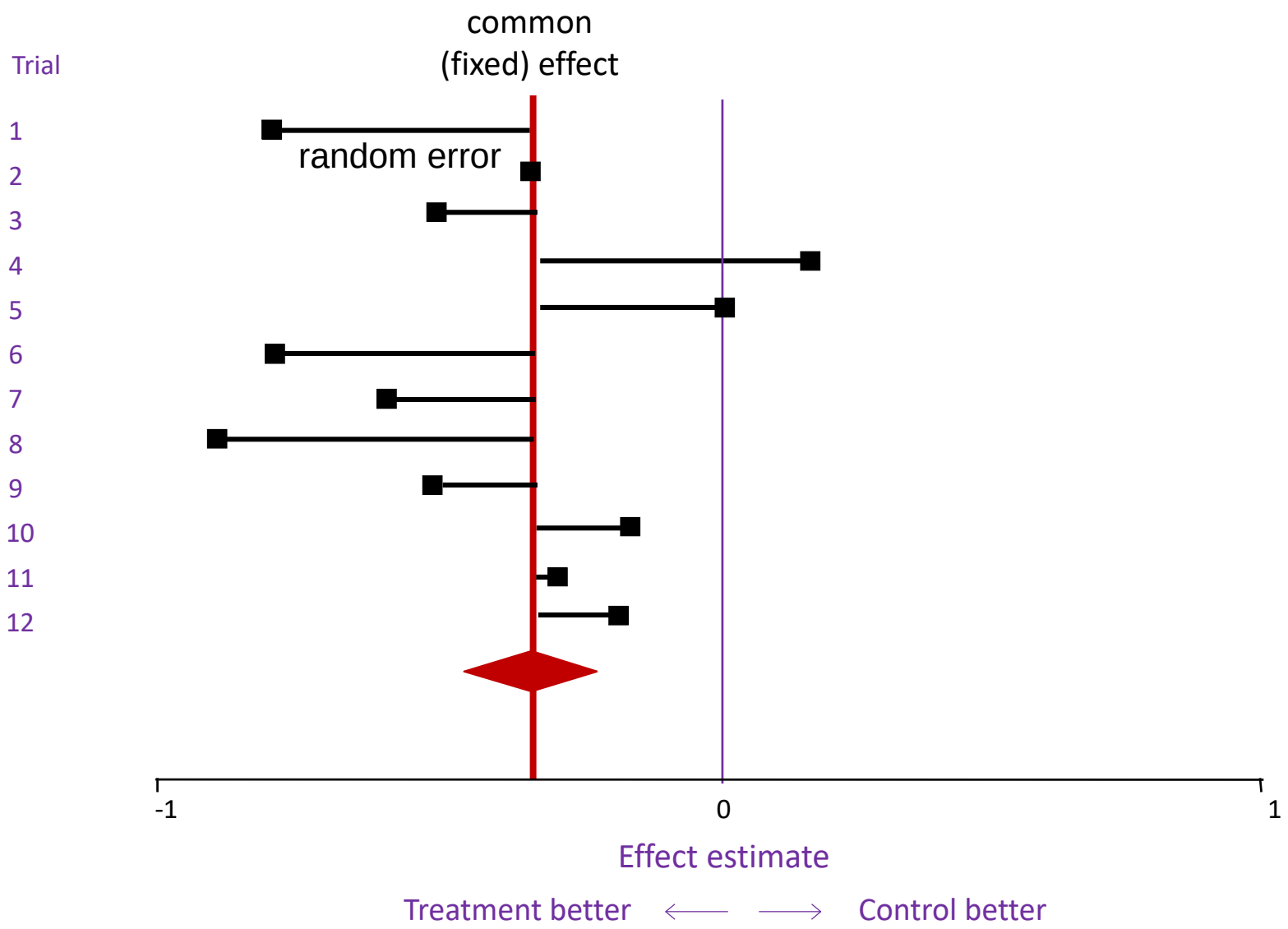
Study or Subgroup	Chlorpromazine Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Chouinard 1990	14	21	16	21	15.3%	0.88 [0.60, 1.29]
Clark 1970a	2	15	5	14	7.4%	0.37 [0.09, 1.62]
Clark 1970b	10	53	6	18	11.7%	0.57 [0.24, 1.34]
Fleming 1959	5	21	13	21	11.9%	0.38 [0.17, 0.89]
Hall 1955	65	87	70	88	16.4%	0.94 [0.80, 1.10]
Prien 1968	37	416	70	212	15.4%	0.27 [0.19, 0.39]
Serafetinides 1962	6	14	3	13	9.4%	1.86 [0.58, 5.94]
Somerville 1960	5	15	22	30	12.6%	0.45 [0.22, 0.96]
Total (95% CI)		**642**		**417**	**100.0%**	**0.59 [0.34, 1.01]**
Total events	144		205			

Heterogeneity: Tau² = 0.45; Chi² = 61.84, df = 7 (P < 0.00001); I² = 89%
Test for overall effect: Z = 1.92 (P = 0.05)

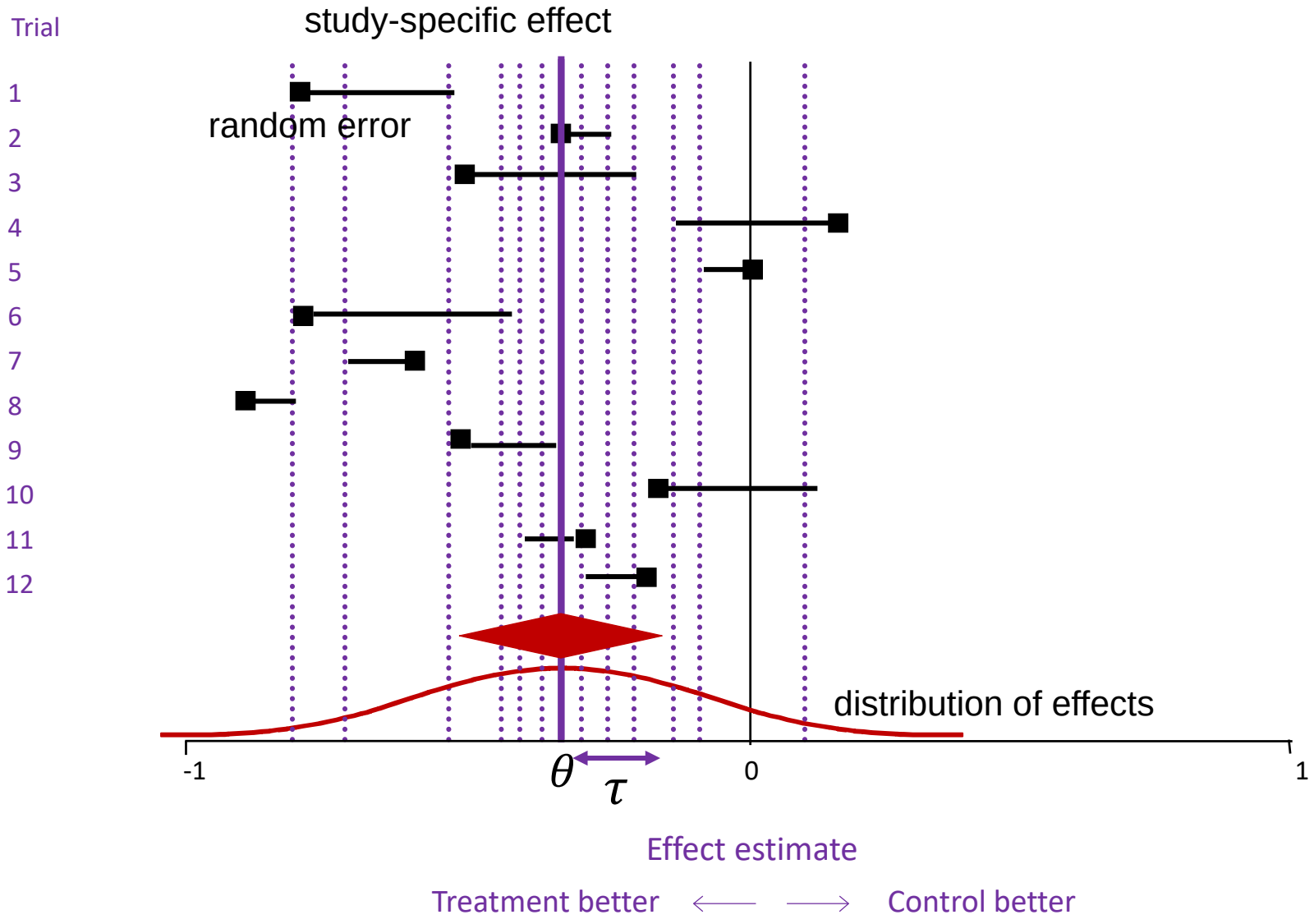


RE gives more conservative results

Fixed effect meta-analysis



Random effects meta-analysis



Random effects meta-analysis

- Heterogeneity suggests that the studies have important underlying differences.
- We can allow the true effects underlying the studies to differ.
- We assume the true effects underlying the studies follow a distribution.
  - conventionally a normal distribution
- We use a simple adaptation of the inverse-variance weighted average.

DerSimonian and Laird (1986)

Identifying heterogeneity

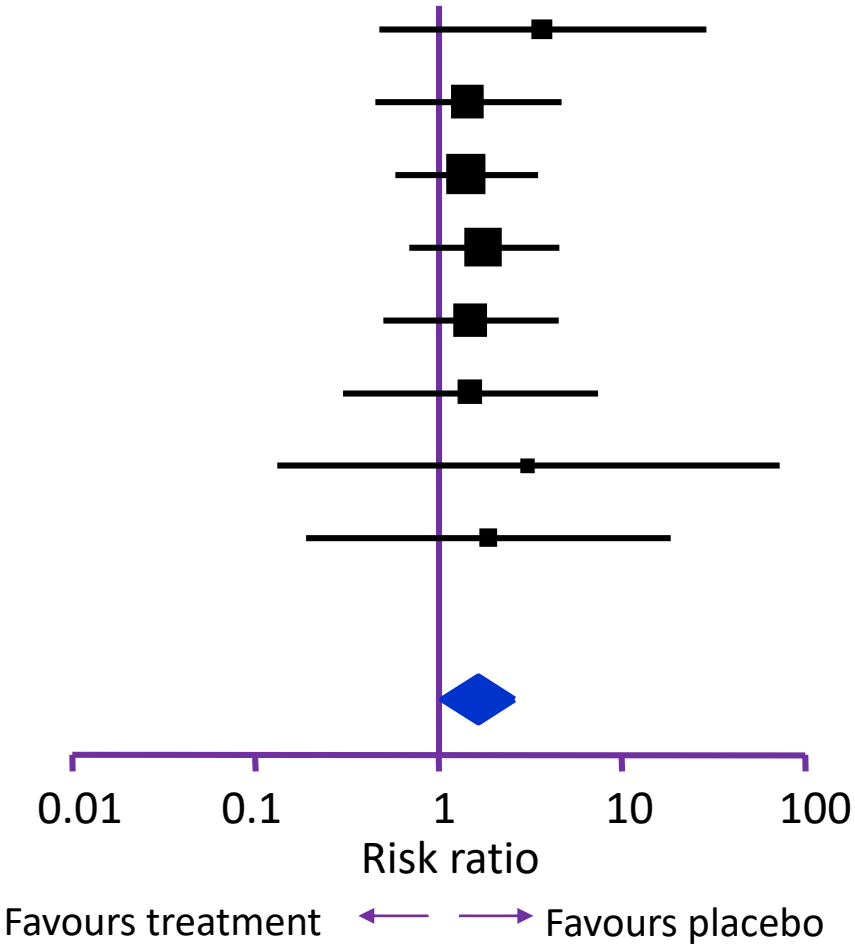
1. Visual inspection of the forest plots

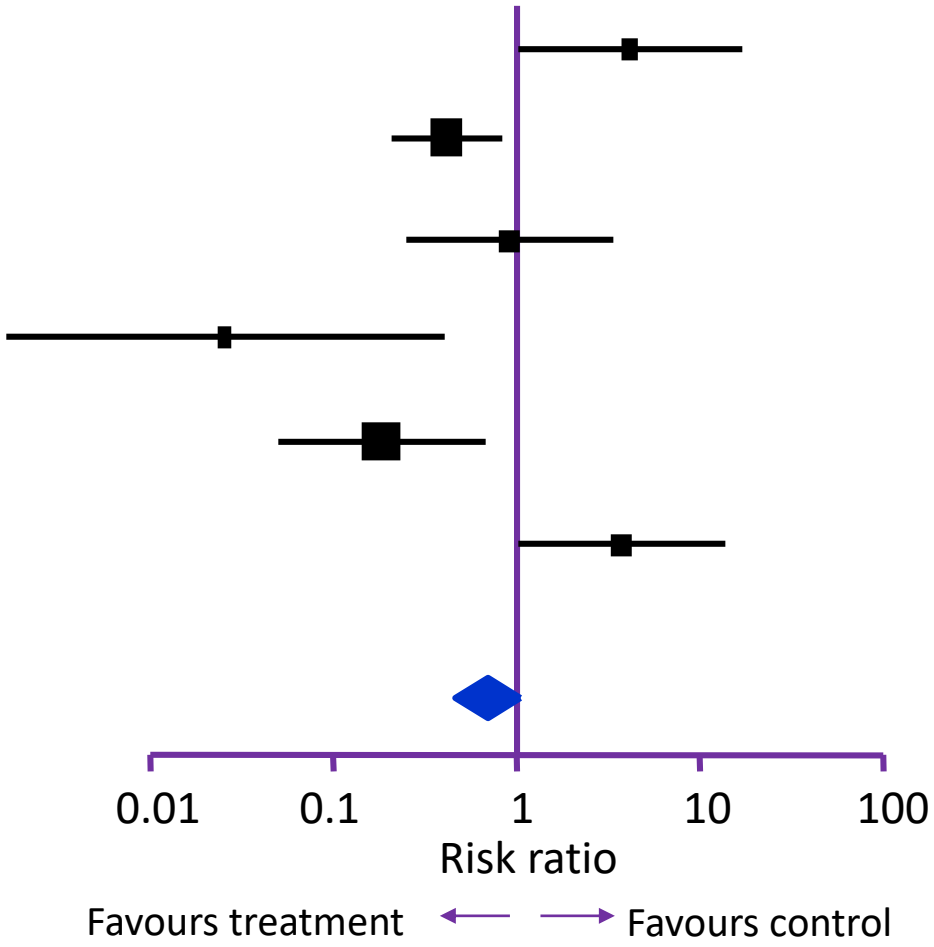
2. Q test for the presence of heterogeneity

3. I^2 statistic that quantifies heterogeneity as a proportion

Visual inspection of the forest plot

- A graphical inspection of the results is usually the first step
- A lack of overlap in confidence intervals indicates heterogeneity





Q-test

- chi-squared (χ^2) test

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

- has χ^2 distribution with $k - 1$ d.f. under null hypothesis of an identical effect in every study
- k is the number of studies in the meta-analysis
- rejection of H_0 suggests heterogeneity

Q-test drawbacks

- Has low power since there are usually very few studies
  - i.e. test is not very good at detecting heterogeneity as statistically significant when it exists

- But, has excessive power to detect clinically unimportant heterogeneity when there are many studies

I-square statistic

Higgins and Thompson (2002)

- Q-test is not asking a useful question if heterogeneity is inevitable
- Quantify heterogeneity
  - based on the χ^2 statistic Q and its degrees of freedom

$$I^2 = \frac{Heterogenity}{Heterogeneity + Average\ study\ variance}$$

$$I^2 = \frac{Q-k+1}{Q} * 100\%$$

describes the proportion of total variability that is due to heterogeneity

Estimation of tau-square

- Estimate the heterogeneity variance τ^2 from the Q-test (method of moments/DL estimator) :

$$\tau^2 = \frac{Q - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$$

- We set $\tau^2 = 0$ if $Q < (k-1)$

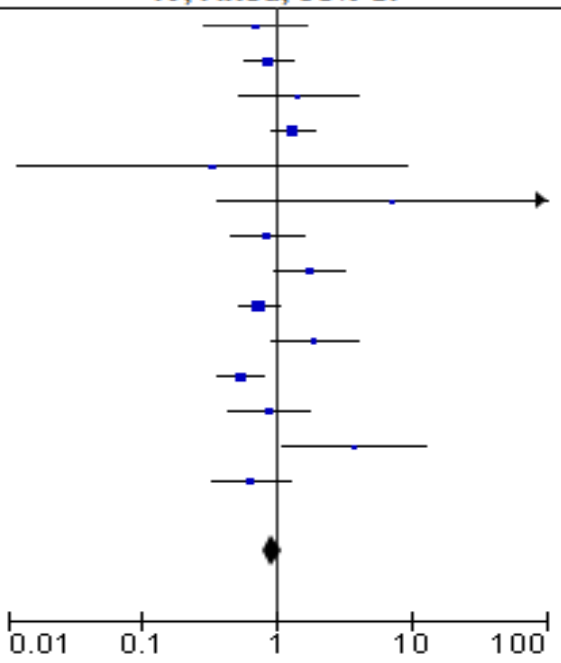
- Many other ways to estimate the heterogeneity variance exist (e.g. restricted maximum likelihood)
  - under certain conditions perform better than the DL estimator

Example: Bleeding

Study or Subgroup	Vitamine K Events	Vitamine K Total	Placebo Events	Placebo Total	Weight	Odds Ratio IV, Fixed, 95% CI
Crowther	12	56	15	53	3.1%	0.69 [0.29, 1.66]
Duley	48	412	56	421	13.9%	0.86 [0.57, 1.30]
Gates	8	32	12	64	2.3%	1.44 [0.52, 3.99]
Gyte	67	612	53	617	16.4%	1.31 [0.90, 1.91]
Hampson	0	8	1	9	0.2%	0.33 [0.01, 9.40]
Henderson	3	63	0	62	0.3%	7.23 [0.37, 142.97]
Hodnett	28	97	31	96	6.3%	0.85 [0.46, 1.57]
Hofmeyr	34	143	22	145	6.7%	1.74 [0.96, 3.16]
Horey	82	342	102	341	20.4%	0.74 [0.53, 1.04]
McKnight	25	76	15	73	4.3%	1.90 [0.90, 3.98]
Mugford	43	764	65	654	14.7%	0.54 [0.36, 0.81]
Neilson	20	80	22	80	4.7%	0.88 [0.43, 1.78]
Sakala	12	44	4	44	1.6%	3.75 [1.10, 12.74]
Winterbottom	18	102	26	103	5.2%	0.63 [0.32, 1.25]
Total (95% CI)		**2831**		**2762**	**100.0%**	**0.92 [0.79, 1.07]**
Total events	400		424			

Heterogeneity: Chi² = 29.54, df = 13 (P = 0.005) I² = 56%

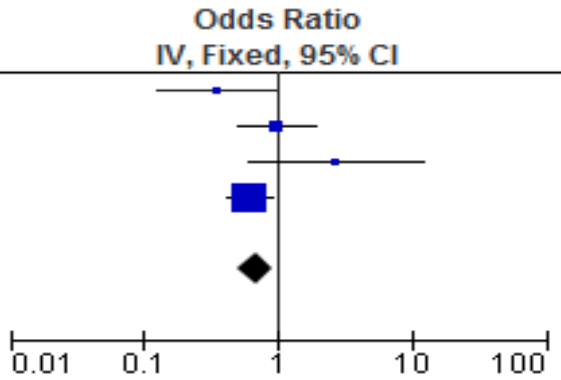
Test for overall effect: Z = 1.07 (P = 0.28)



Study or Subgroup	Vitamine K Events	Vitamine K Total	No Treatment Events	No Treatment Total	Weight	Odds Ratio IV, Fixed, 95% CI
Ashby	7	42	15	41	8.8%	0.35 [0.12, 0.97]
Enkin	23	80	24	82	20.3%	0.98 [0.49, 1.92]
Keirse	8	14	5	15	4.1%	2.67 [0.59, 12.04]
Renfrew	74	243	100	241	66.7%	0.62 [0.42, 0.90]
Total (95% CI)		**379**		**379**	**100.0%**	**0.68 [0.50, 0.93]**
Total events	112		144			

Heterogeneity: Chi² = 6.13, df = 3 (P = 0.11) I² = 51%

Test for overall effect: Z = 2.43 (P = 0.01)



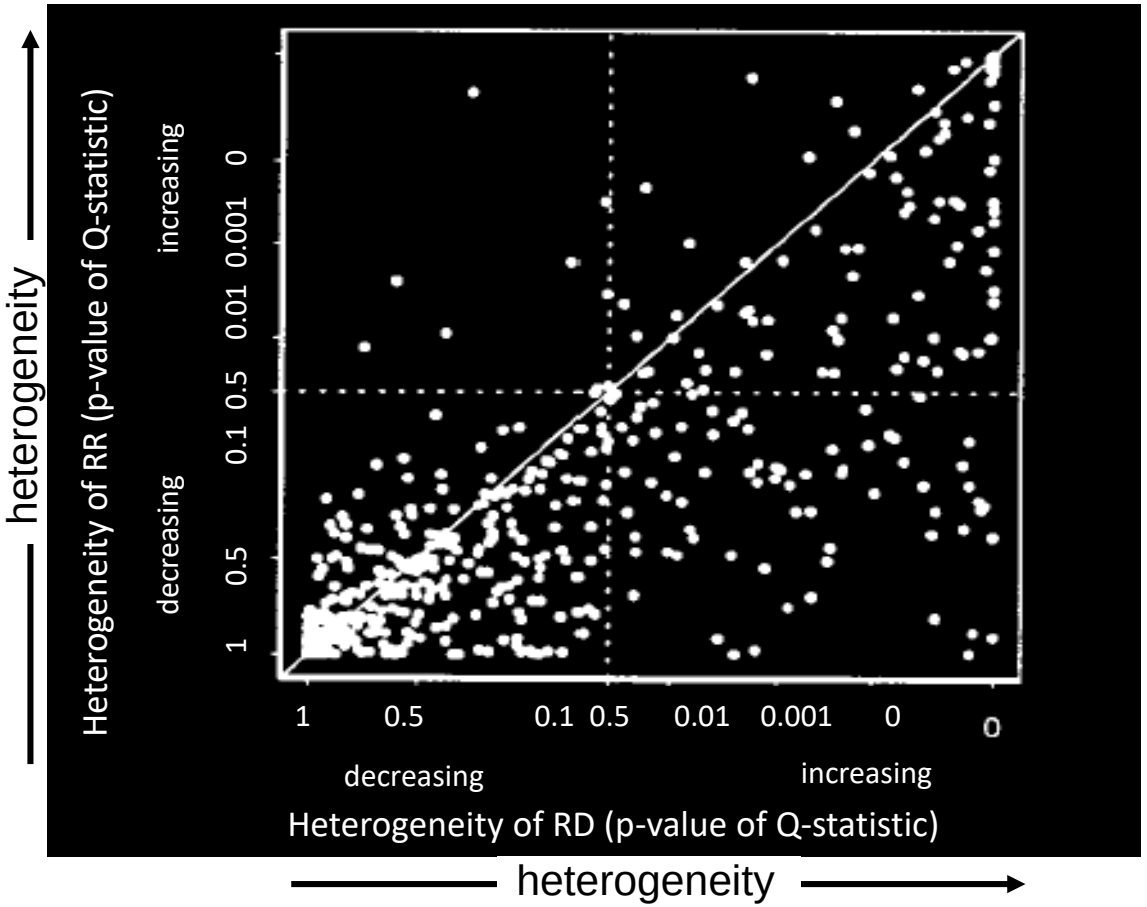
What can we do with heterogeneity?

- Check the data
  - Incorrect data extraction; unit of analysis errors (e.g. with crossover trials, cluster randomized trials, counts)
- Try to bypass it
  - Change effect measure
- Encompass it
  - Random effects meta-analysis
- Explore it
  - Subgroup analysis
    Meta-regression
- Resign to it
  - Do no meta-analysis
- Ignore it
  - Don’t do that!

Heterogeneity of effect measures

Empirical evidence

- Ratio measures (RR and OR) considerably less heterogeneous than difference measures (RD)



Deeks et al. 2002

What can we do with heterogeneity?

• Check the data	• Incorrect data extraction; unit of analysis errors (e.g. with crossover trials, cluster randomized trials, counts)
• Try to bypass it	• Change effect measure
• Encompass it	• Random effects meta-analysis
• Explore it	• Subgroup analysis Meta-regression
• Resign to it	• Do no meta-analysis
• Ignore it	• Don’t do that!

What not to do!

- Fixed or random effects meta-analysis should be specified *a priori* if possible and not on the basis of the Q test

What to do:

Think about the question you asked, the included studies etc: do you expect them to be very diverse?

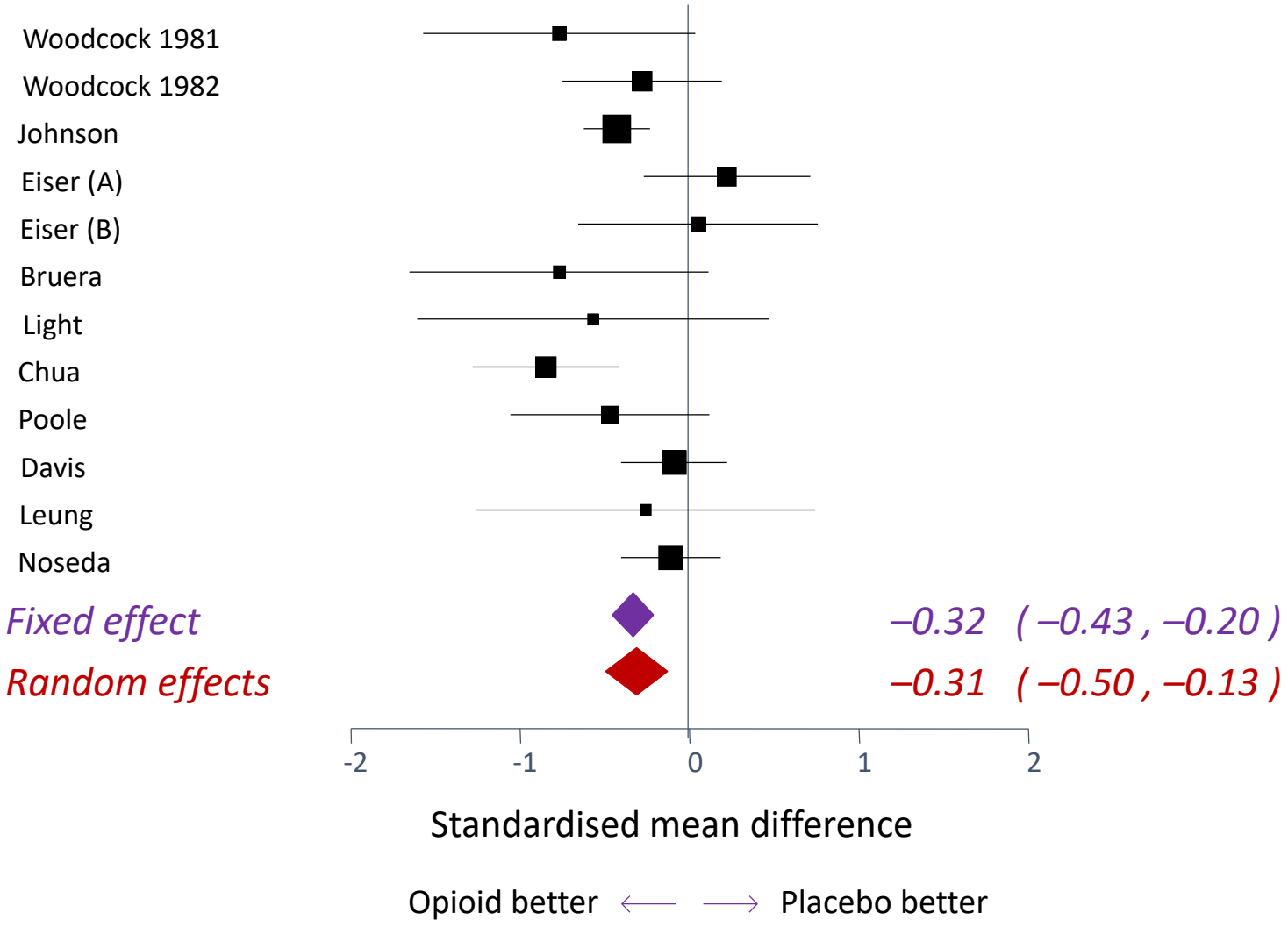
You can apply and present both fixed and random effects

Fixed vs. random effects

- Fixed effect model is often unrealistic
- Random effects model difficult to interpret

- Fixed and random effects inverse-variance meta-analyses may
  - be identical (when $\tau^2 = 0$)
  - give similar point estimate, different confidence intervals
    (the 95% CI from FE should fall within the 95% CI from the RE)

Example: Opioids for breathlessness



Fixed vs. random effects

- Fixed effect model is often unrealistic
- Random effects model difficult to interpret

- Fixed and random effects inverse-variance meta-analyses may
  - be identical (when $\tau^2 = 0$)
  - give similar point estimate, different confidence intervals
    (the 95% CI from FE should fall within the 95% CI from the RE)

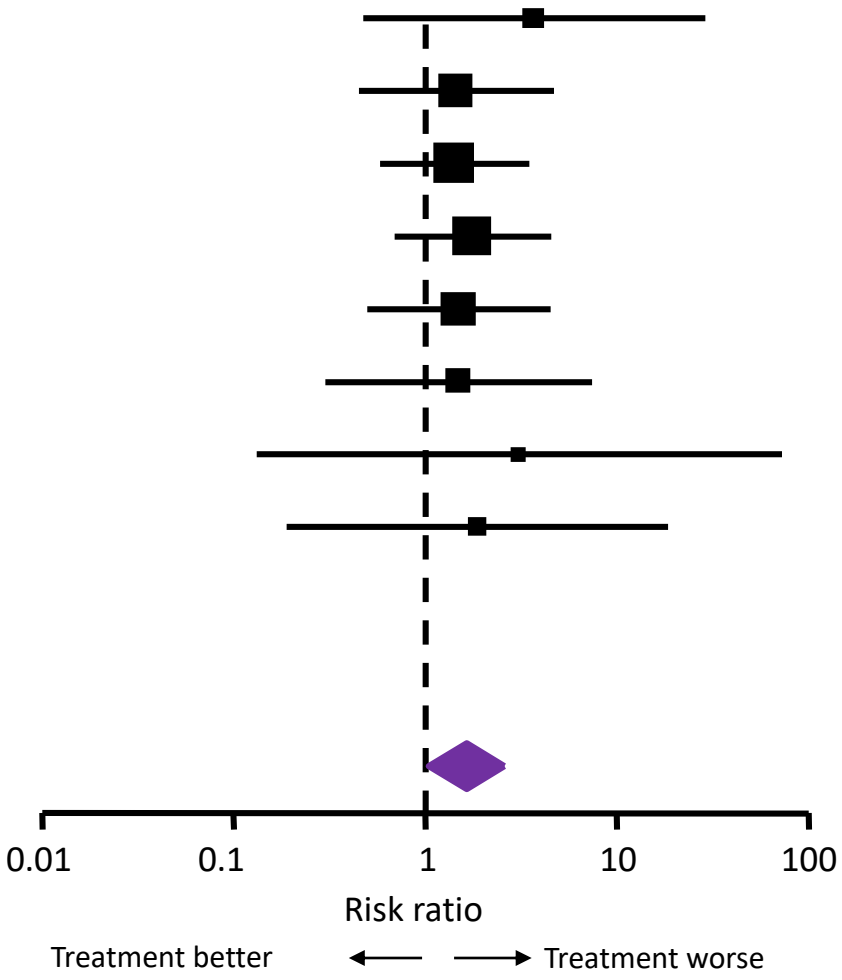
- Random effects analysis may give spurious results when effect size depends on precision
  - gives relatively more weight to smaller studies
  - important because
    - smaller studies may be of lower quality (hence biased)
    - publication bias may result in missing smaller studies

Interpreting random effects meta-analysis

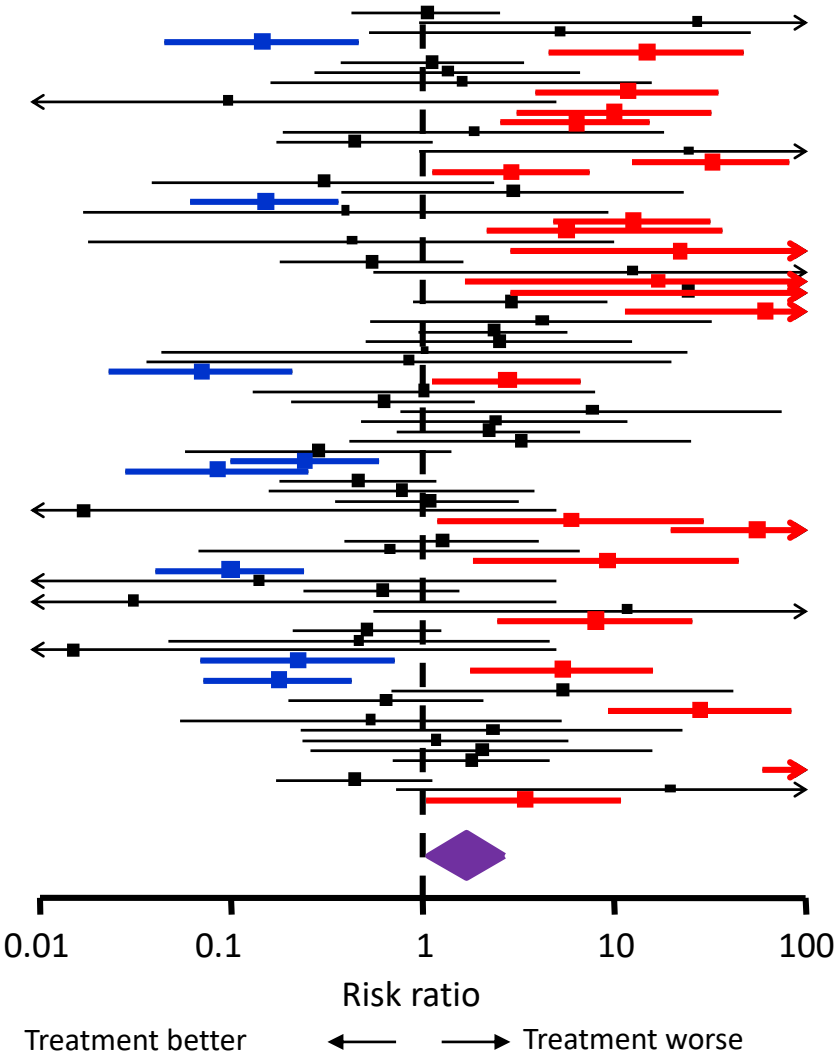
- Random-effects meta-analysis suitable for unexplained heterogeneity
  - Random effects may not explain all the heterogeneity of the data if covariates are responsible

- Conventionally, inference is focused on the mean of the distribution ($\hat{\theta}$)
  - i.e. we report mean and 95% from a meta-analysis
  - This may be misleading...

Example



Fixed effect meta-analysis:
1.64 (1.04 , 2.58) P = 0.03



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

Interpreting random effects meta-analysis

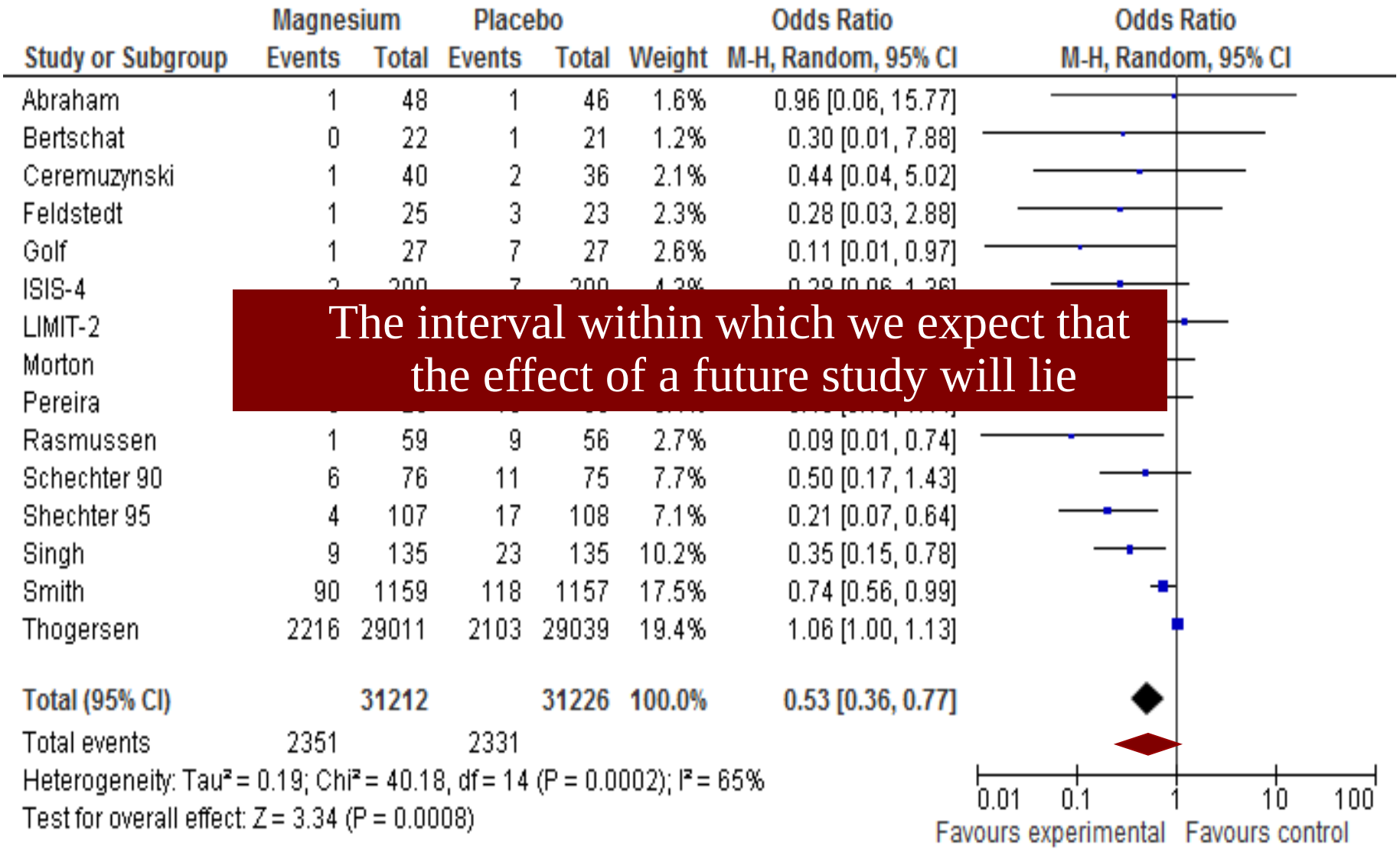
- Random-effects meta-analysis suitable for unexplained heterogeneity
  - Random effects may not explain all the heterogeneity of the data if covariates are responsible

- Conventionally, inference is focused on the mean of the distribution ($\hat{\theta}$)
  - i.e. we report mean and 95% from a meta-analysis
  - This may be misleading...

- Look also at the prediction interval

$$\theta \pm 1.96\sqrt{se(\theta)^2 + \tau^2}$$
$$\theta \pm t_{0.025,n-1}\sqrt{se(\theta)^2 + \tau^2}$$

Example


Magnesium
Placebo
Odds Ratio
Odds Ratio
Study or Subgroup
Events
Total
Events
Total
Weight
M-H, Random, 95% CI
M-H, Random, 95% CI
Abraham 1 48 1 46 1.6% 0.96 [0.06, 15.77]
Bertschat 0 22 1 21 1.2% 0.30 [0.01, 7.88]
Ceremuzynski 1 40 2 36 2.1% 0.44 [0.04, 5.02]
Feldstedt 1 25 3 23 2.3% 0.28 [0.03, 2.88]
Golf 1 27 7 27 2.6% 0.11 [0.01, 0.97]
ISIS-4
LIMIT-2
Morton
Pereira
Rasmussen 1 59 9 56 2.7% 0.09 [0.01, 0.74]
Schechter 90 6 76 11 75 7.7% 0.50 [0.17, 1.43]
Shechter 95 4 107 17 108 7.1% 0.21 [0.07, 0.64]
Singh 9 135 23 135 10.2% 0.35 [0.15, 0.78]
Smith 90 1159 118 1157 17.5% 0.74 [0.56, 0.99]
Thogersen 2216 29011 2103 29039 19.4% 1.06 [1.00, 1.13]
Total (95% CI) 31212 31226 100.0% 0.53 [0.36, 0.77]
Total events 2351 2331
Heterogeneity: Tau² = 0.19; Chi² = 40.18, df = 14 (P = 0.0002); I² = 65%
Test for overall effect: Z = 3.34 (P = 0.0008)
The interval within which we expect that the effect of a future study will lie
0.01
0.1
1
10
100
Favours experimental
Favours control

Interpreting the diamond

- Conventional Interpretations
  1. Statistical significance and direction
  2. Magnitude of the pooled estimate
  3. Width of the confidence interval

- Heterogeneity
  - Too much heterogeneity challenges the meaning of the diamond

- Quality of the included studies

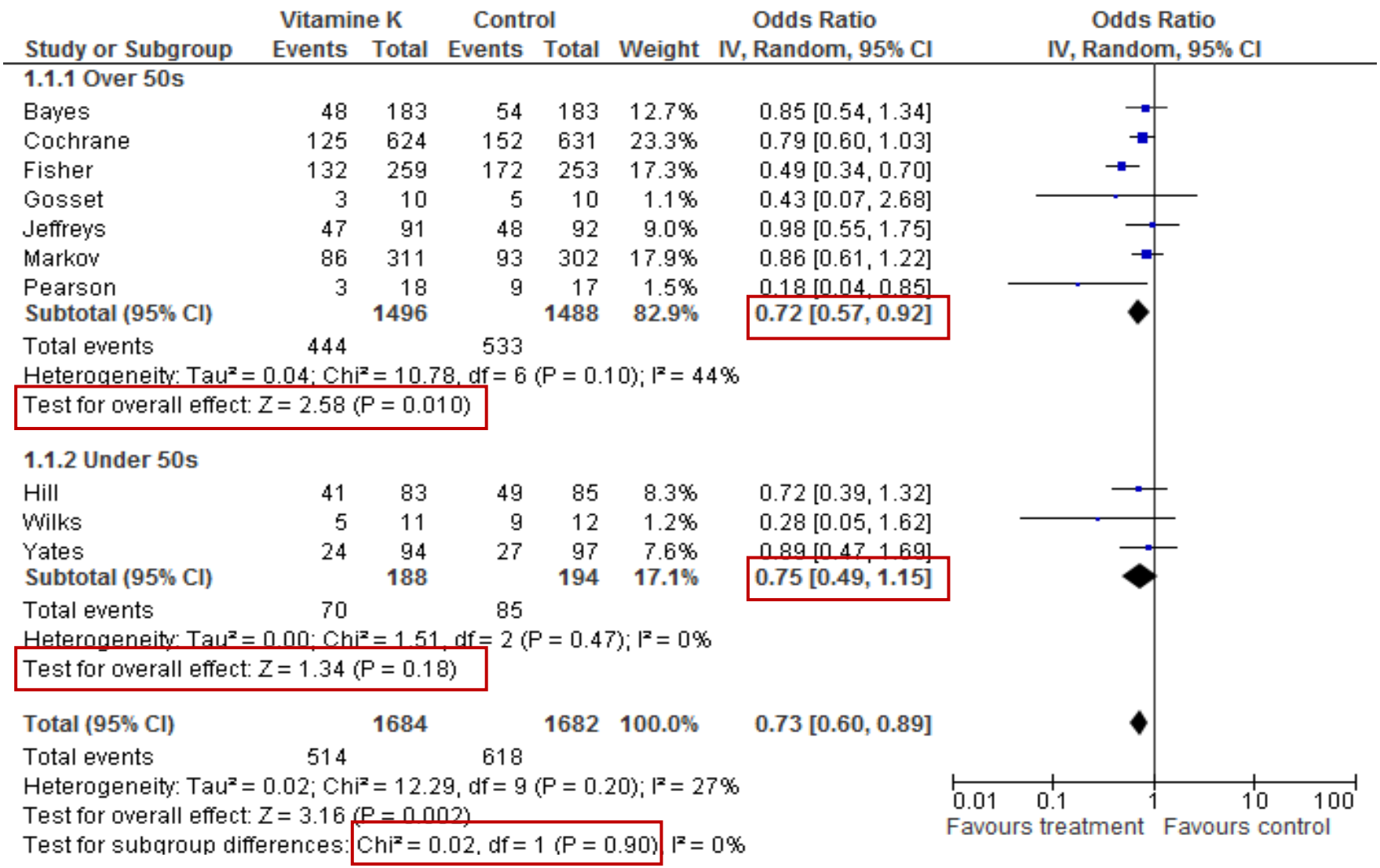
What can we do with heterogeneity?

→ • Check the data	• Incorrect data extraction; unit of analysis errors (e.g. with crossover trials, cluster randomized trials, counts)
→ • Try to bypass it	• Change effect measure
→ • Encompass it	• Random effects meta-analysis
→ • Explore it	• Subgroup analysis Meta-regression
• Resign to it	• Do no meta-analysis
• Ignore it	• Don’t do that!

Exploring heterogeneity

- Characteristics of studies may be associated with the size of treatment effect
- For example,
  - adequacy of allocation concealment
  - average age of patients
  - setting of study
  - dose of drug
- For discrete characteristics, we can use subgroup analyses
- For discrete or continuous characteristics, we can use meta-regression

Subgroup analysis (example: bleeding)



Study or Subgroup	Vitamine K Events	Vitamine K Total	Control Events	Control Total	Weight	Odds Ratio IV, Random, 95% CI
1.1.1 Over 50s						
Bayes	48	183	54	183	12.7%	0.85 [0.54, 1.34]
Cochrane	125	624	152	631	23.3%	0.79 [0.60, 1.03]
Fisher	132	259	172	253	17.3%	0.49 [0.34, 0.70]
Gosset	3	10	5	10	1.1%	0.43 [0.07, 2.68]
Jeffreys	47	91	48	92	9.0%	0.98 [0.55, 1.75]
Markov	86	311	93	302	17.9%	0.86 [0.61, 1.22]
Pearson	3	18	9	17	1.5%	0.18 [0.04, 0.85]
Subtotal (95% CI)		**1496**		**1488**	**82.9%**	**0.72 [0.57, 0.92]**
Total events	444		533			
Heterogeneity: Tau² = 0.04; Chi² = 10.78, df = 6 (P = 0.10); I² = 44%						
Test for overall effect: Z = 2.58 (P = 0.010)						
1.1.2 Under 50s						
Hill	41	83	49	85	8.3%	0.72 [0.39, 1.32]
Wilks	5	11	9	12	1.2%	0.28 [0.05, 1.62]
Yates	24	94	27	97	7.6%	0.89 [0.47, 1.69]
Subtotal (95% CI)		**188**		**194**	**17.1%**	**0.75 [0.49, 1.15]**
Total events	70		85			
Heterogeneity: Tau² = 0.00; Chi² = 1.51, df = 2 (P = 0.47); I² = 0%						
Test for overall effect: Z = 1.34 (P = 0.18)						
Total (95% CI)		**1684**		**1682**	**100.0%**	**0.73 [0.60, 0.89]**
Total events	514		618			
Heterogeneity: Tau² = 0.02; Chi² = 12.29, df = 9 (P = 0.20); I² = 27%						
Test for overall effect: Z = 3.16 (P = 0.002)						
Test for subgroup differences: Chi² = 0.02, df = 1 (P = 0.90), I² = 0%						

Test for differences between subgroups

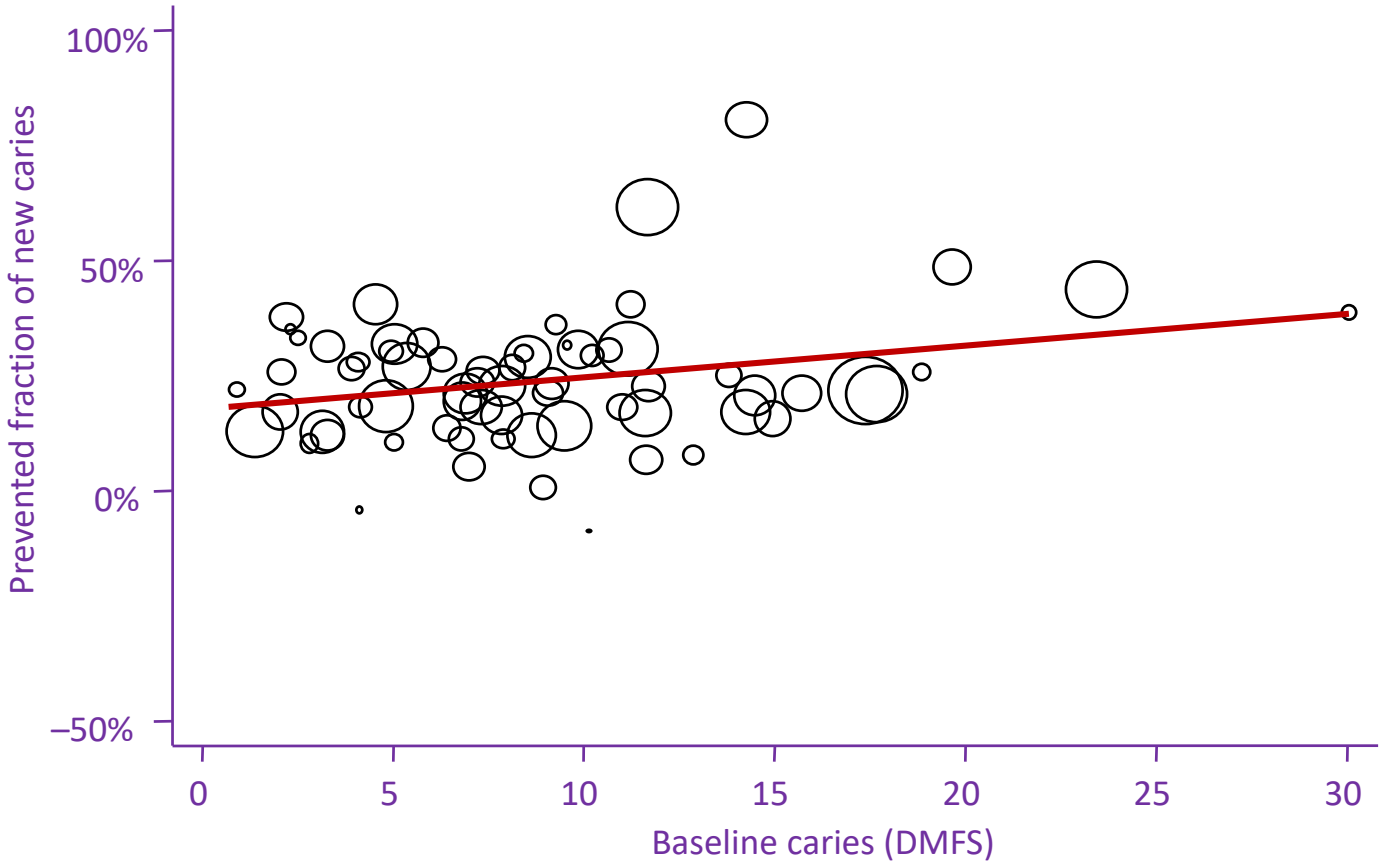
- H_0: No differences across the K subgroups
- H_1: There are differences across the K subgroups

$$Q = Q_{tot} - (Q_1 + Q_2 + \cdots + Q_K) \sim \chi^2_{K-1}$$

Meta-regression

Does effectiveness of toothpaste depend on baseline population levels of caries?

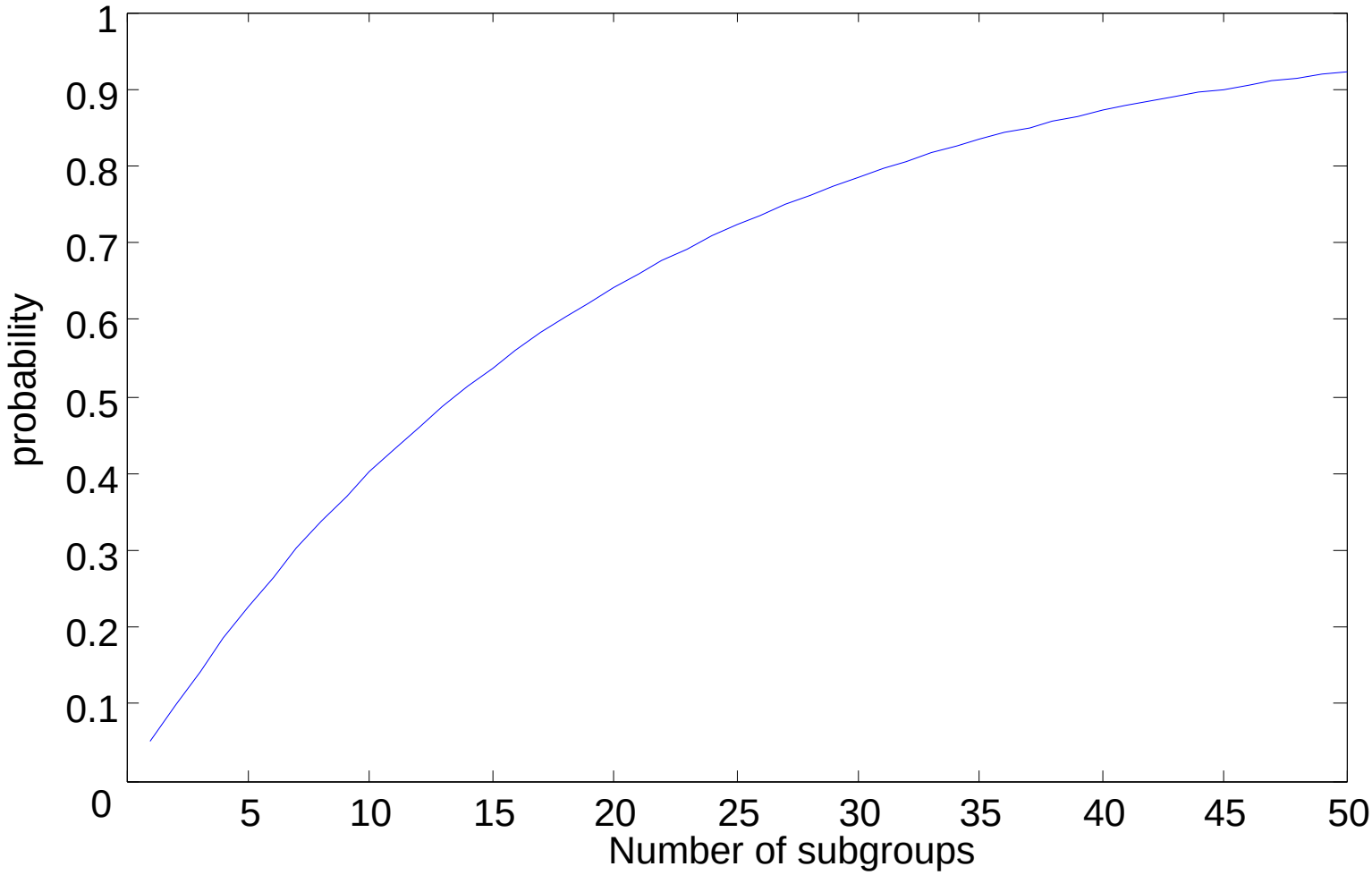
Marinho et al (2003)



Selecting variables for subgroup analysis and meta-regression

- Specify characteristics in advance
- Select a small number of characteristics
- Ensure there is scientific rationale for investigating the characteristics
  - beware ‘prognostic factors’
- Make sure the effect of a characteristic can be identified
  - does it differentiate studies?
- Think about whether the characteristic is closely related to another characteristic

Probability of false positive findings



Problems using published results

- Limited to what is reported
  - Subgroups are rarely reported in all trials
- Limited to “trial-level” characteristics
  - Things that vary between studies; constant within studies
    - Drug dose
    - Treatment duration
- Hard to analyze “participant-level” characteristics
  - Varying between patients in a trial
    - Age
    - Disease severity
  - Rarely reported
  - Using averages (average age, proportion of men) is biased

Using Individual Participant Data

- Obtain all the “raw” data for all participants of all trials
- Gives full data on all characteristics of interest for every participant
  - Age
  - Sex
  - Drug / dose received
  - Exact nature of condition
- Permits analysis of all characteristics of interest
- Usually analyzed using regression modelling
  - Linear / logistic regression
  - NOT subgroup analysis or meta-regression

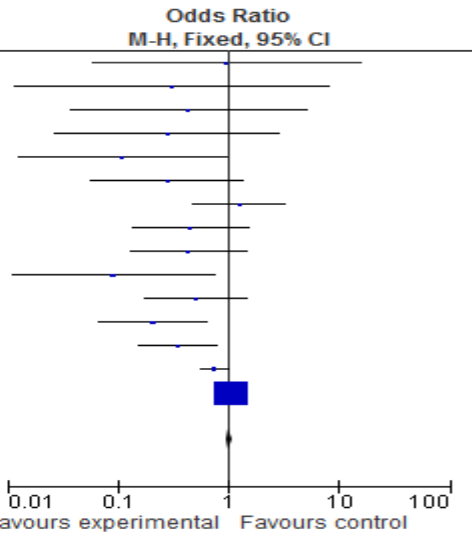
Small-study effects and random effects

Magnesium for acute myocardial infraction

Outcome: Mortality

Study or Subgroup	Magnesium Events	Magnesium Total	Placebo Events	Placebo Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Abraham	1	48	1	46	0.0%	0.96 [0.06, 15.77]
Bertschat	0	22	1	21	0.1%	0.30 [0.01, 7.88]
Ceremuzynski	1	40	2	36	0.1%	0.44 [0.04, 5.02]
Feldstedt	1	25	3	23	0.1%	0.28 [0.03, 2.88]
Golf	1	27	7	27	0.3%	0.11 [0.01, 0.97]
ISIS-4	2	200	7	200	0.3%	0.28 [0.06, 1.36]
LIMIT-2	10	150	8	148	0.3%	1.25 [0.48, 3.26]
Morton	4	130	8	122	0.4%	0.45 [0.13, 1.54]
Pereira	5	23	13	33	0.4%	0.43 [0.13, 1.44]
Rasmussen	1	59	9	56	0.4%	0.09 [0.01, 0.74]
Schechter 90	6	76	11	75	0.5%	0.50 [0.17, 1.43]
Shechter 95	4	107	17	108	0.8%	0.21 [0.07, 0.64]
Singh	9	135	23	135	1.0%	0.35 [0.15, 0.78]
Smith	90	1159	118	1157	5.1%	0.74 [0.56, 0.99]
Thogersen	2216	29011	2103	29039	90.2%	1.06 [1.00, 1.13]
Total (95% CI)		**31212**		**31226**	**100.0%**	**1.01 [0.95, 1.07]**
Total events	2351		2331			

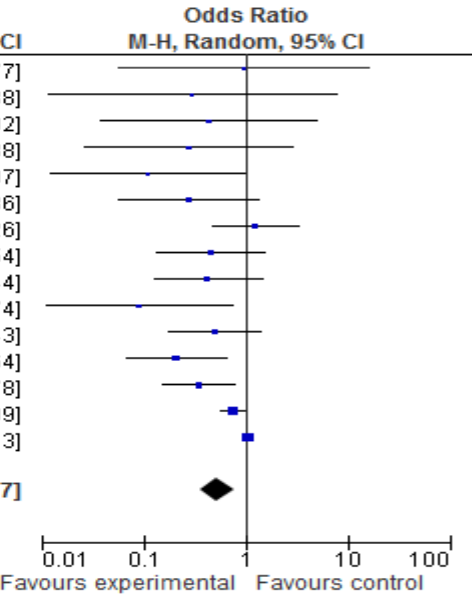
Heterogeneity: Chi² = 40.18, df = 14 (P = 0.0002); I² = 65%
Test for overall effect: Z = 0.36 (P = 0.72)



RE gives less 'contrasted' weights between big and small studies

Study or Subgroup	Magnesium Events	Magnesium Total	Placebo Events	Placebo Total	Weight	Odds Ratio M-H, Random, 95% CI
Abraham	1	48	1	46	1.6%	0.96 [0.06, 15.77]
Bertschat	0	22	1	21	1.2%	0.30 [0.01, 7.88]
Ceremuzynski	1	40	2	36	2.1%	0.44 [0.04, 5.02]
Feldstedt	1	25	3	23	2.3%	0.28 [0.03, 2.88]
Golf	1	27	7	27	2.6%	0.11 [0.01, 0.97]
ISIS-4	2	200	7	200	4.3%	0.28 [0.06, 1.36]
LIMIT-2	10	150	8	148	8.6%	1.25 [0.48, 3.26]
Morton	4	130	8	122	6.3%	0.45 [0.13, 1.54]
Pereira	5	23	13	33	6.4%	0.43 [0.13, 1.44]
Rasmussen	1	59	9	56	2.7%	0.09 [0.01, 0.74]
Schechter 90	6	76	11	75	7.7%	0.50 [0.17, 1.43]
Shechter 95	4	107	17	108	7.1%	0.21 [0.07, 0.64]
Singh	9	135	23	135	10.2%	0.35 [0.15, 0.78]
Smith	90	1159	118	1157	17.5%	0.74 [0.56, 0.99]
Thogersen	2216	29011	2103	29039	19.4%	1.06 [1.00, 1.13]
Total (95% CI)		**31212**		**31226**	**100.0%**	**0.53 [0.36, 0.77]**
Total events	2351		2331			

Heterogeneity: Tau² = 0.19; Chi² = 40.18, df = 14 (P = 0.0002); I² = 65%
Test for overall effect: Z = 3.34 (P = 0.0008)



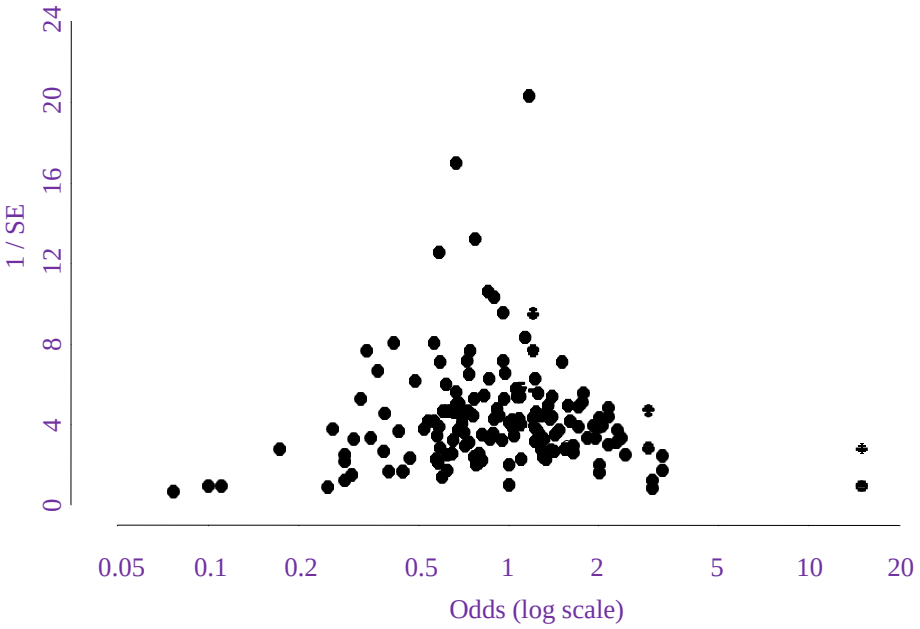
Small-study effects as a source of heterogeneity

- When the results of your review are related to the size of the study
  - For example smaller studies may give larger treatment effects

Selective outcome reporting as source of heterogeneity

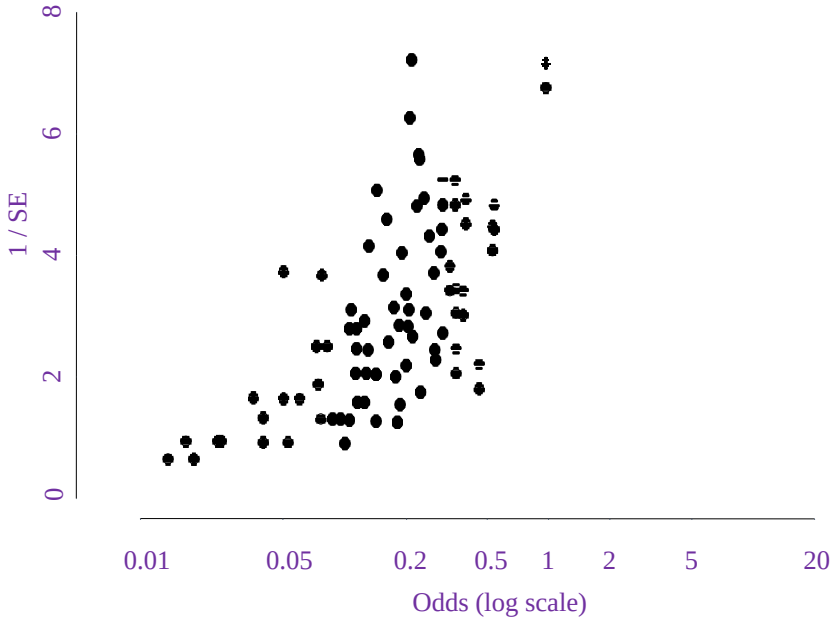
Bown et al. (2002)

Suggests absence of bias



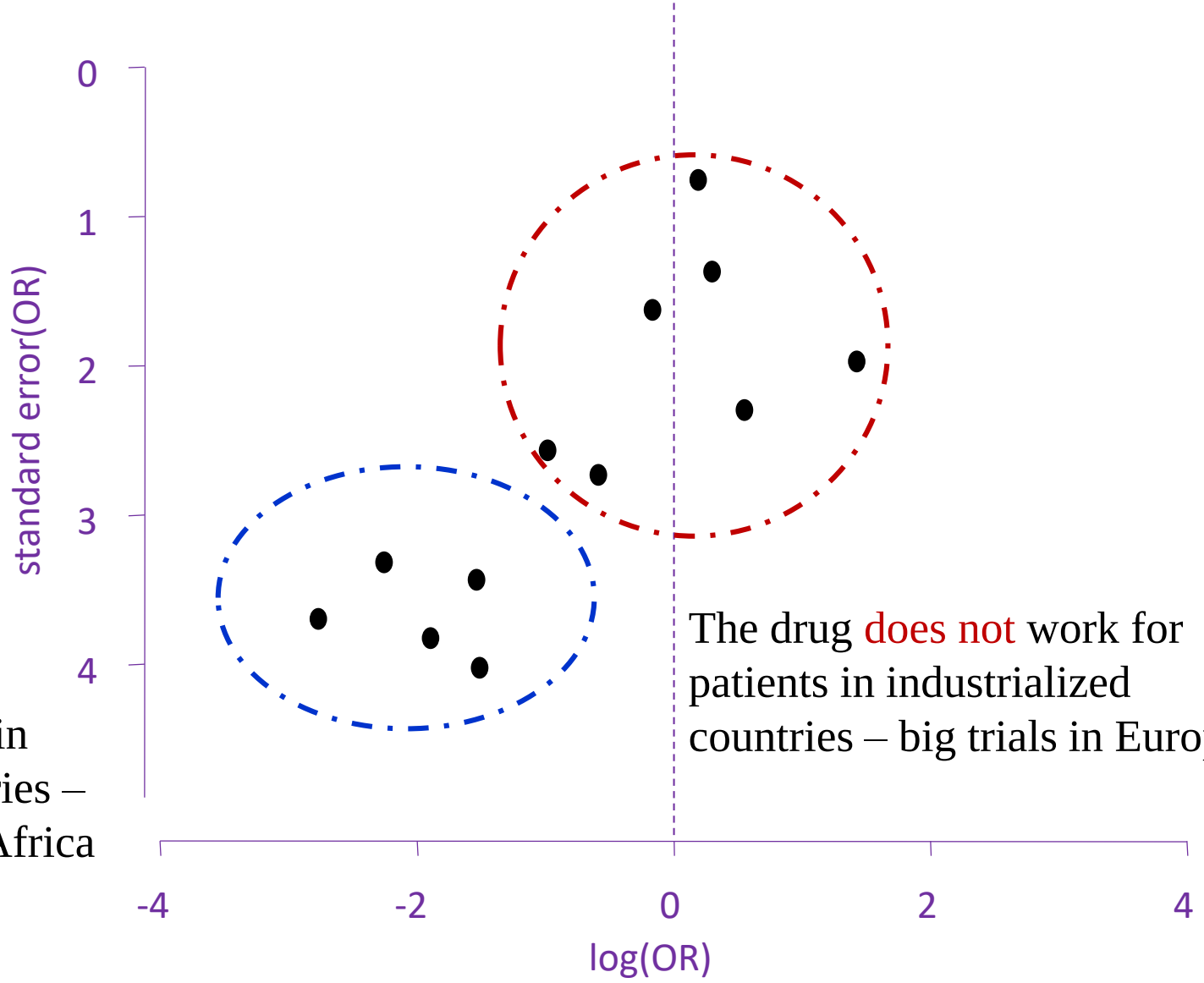
Overall mortality following emergency aneurysm repair (n=171)

Suggests presence of bias



Mortality in operating theatre following emergency aneurysm repair (n=77)

Baseline risk as source of heterogeneity



The drug does work for patients in developing countries – small trials from Africa

The drug does not work for patients in industrialized countries – big trials in Europe

What can we do with heterogeneity?

• Check the data	• Incorrect data extraction; unit of analysis errors (e.g. with crossover trials, cluster randomized trials, counts)
• Try to bypass it	• Change effect measure
• Encompass it	• Random effects meta-analysis
• Explore it	• Subgroup analysis Meta-regression
• Resign to it	• Do no meta-analysis
• **Ignore it**	• **Don't do that!**

Methods available in RevMan

- Estimate of overall effect with CI (fixed effect model)
- Estimate of mean effect with CI (random effects model)
- Test for heterogeneity, with P value
- I^2 measure of heterogeneity
- τ^2 heterogeneity variance
- Test for subgroup differences

Methods not available in RevMan

- Meta-regression
- Random-effects methods that account for the fact that tau-square is estimated
- Prediction intervals
- Individual participant data methods

References

DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clin Trials 1986; 7: 177-188

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21: 1539-1558

Higgins JPT, Thompson SG. Controlling the risk of spurious results from meta-regression. Stat Med 2004 ; 23(11),1663-1682.

Higgins, JPT., Thompson, SG, Spiegelhalter,DJ (2009). A re-evaluation of random effect meta-analysis. JRSC series A 2009; 172(1),137-159.

Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011

Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. Br Med J 1994; 309: 1351-1355

Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002; 21: 1559-1574