Cochrane Training

Exploring heterogeneity

Trusted evidence.
Informed decisions.
Better health.

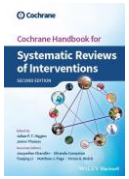
Cochrane Training

Steps of a Cochrane Review

1. define the question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. **analyse and present results**
9. **interpret results and draw conclusions**
10. improve and update review

Session outline

- **what is heterogeneity?**
- assumptions about heterogeneity
- identifying heterogeneity
- exploring your results



See Chapter 10 of the Handbook

What is heterogeneity?

Variation or differences

- three broad types:
  - clinical
  - methodological
  - statistical

Clinical diversity

- participants
  - e.g. condition, age, gender, location, study eligibility criteria
- interventions
  - intensity/dose, duration, delivery, additional components, experience of practitioners, control (placebo, none, standard care)
- outcomes
  - follow-up duration, ways of measuring, definition of an event, cut-off points

Methodological diversity

- design
  - e.g. randomised vs non-randomised, crossover vs parallel, individual vs cluster randomised
- conduct
  - e.g. risk of bias, approach to analysis

Statistical heterogeneity

- there will always be random (sampling) variation between the results of different studies
- heterogeneity is variation between the effects being evaluated in the different studies
  - caused by clinical and methodological diversity
  - alternative to homogeneity (identical true effects underlying every study)
  - study results will be more different from each other than if random variation is the only reason for the differences between the estimated intervention effects

Cochrane Training

Session outline

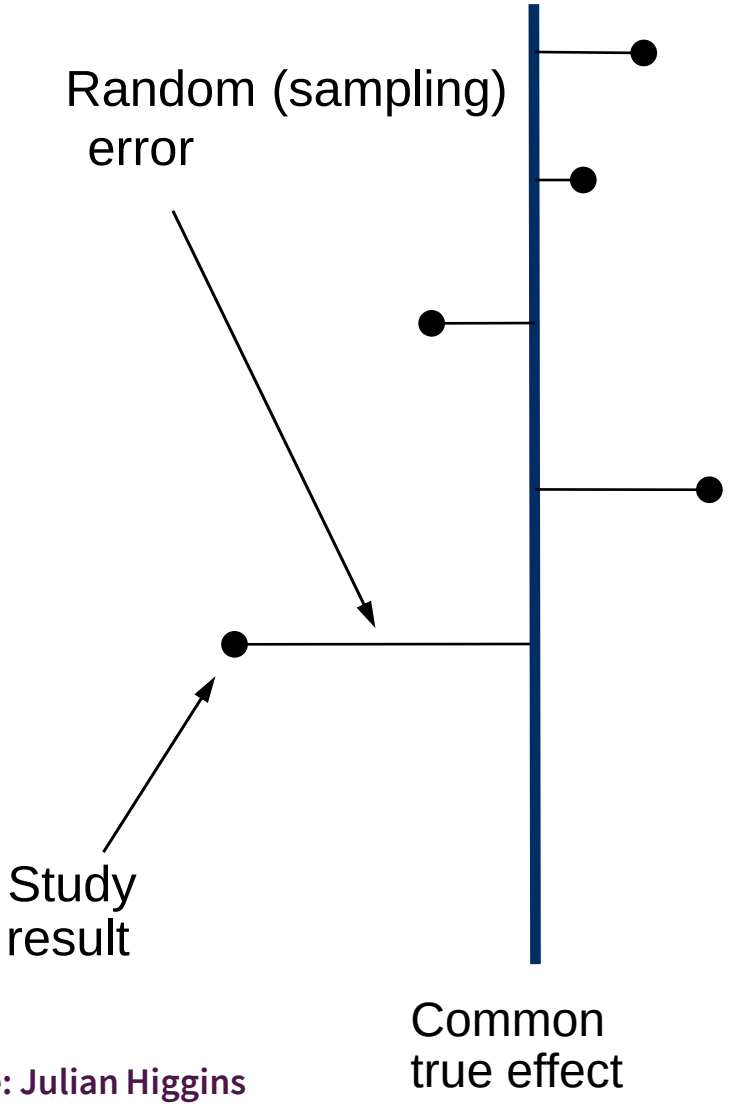
- what is heterogeneity?
- **assumptions about heterogeneity**
- identifying heterogeneity
- exploring your results

Cochrane Training

Fixed-effect vs random-effects

- two models for meta-analysis available in RevMan
- make different assumptions about heterogeneity
- pre-specify your planned approach in your protocol

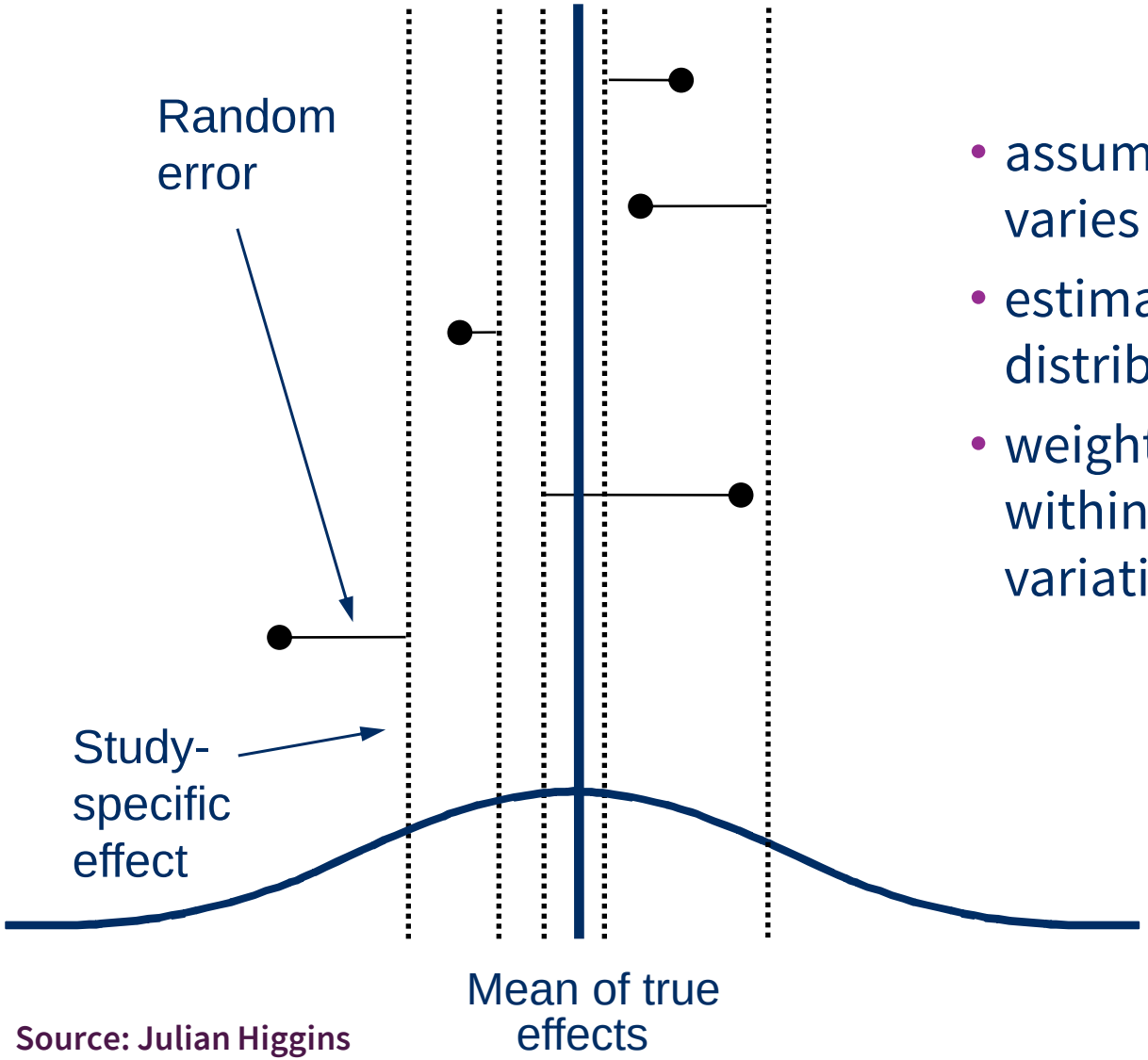
Fixed-effect model



- assumes all studies are measuring the same treatment effect
- estimates that one effect
- if not for random (sampling) error, all results would be identical

Source: Julian Higgins

Random-effects model



Source: Julian Higgins

- assumes the treatment effect varies between studies
- estimates the **mean** of the distribution of effects
- weighted for both within-study and between-study variation (tau^2, τ^2)

What's the difference?

- random-effects (RE) meta-analyses are:
  - almost identical to fixed-effect when there is no heterogeneity
  - similar to fixed-effect but with wider confidence intervals when there is heterogeneity of the sort assumed by RE model
  - different from fixed-effect meta-analyses when results are related to study size
    - RE model gives relatively more weight to smaller studies

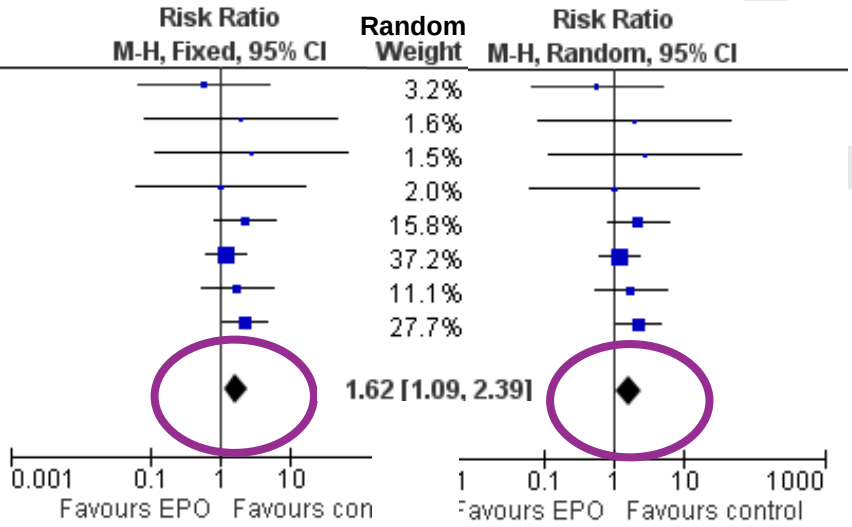
$$\text{weight} = \frac{1}{\text{variance within} + \text{variance between}} = \frac{1}{SE^2 + tau^2}$$

No heterogeneity

Study or Subgroup	Treatment Events	Treatment Total	Control Events	Control Total	Fixed Weight	Risk Ratio M-H, Fixed, 95% CI	Random Weight	Risk Ratio M-H, Random, 95% CI
Bierer 2006	1	7	2	8	5.1%	0.57 [0.06, 5.03]	3.2%	
Fauchère 2008	1	24	0	15	1.7%	1.92 [0.08, 44.29]	1.6%	
Haiden 2005	1	21	0	19	1.4%	2.73 [0.12, 63.19]	1.5%	
Maier 1994	1	120	1	121	2.7%	1.01 [0.06, 15.94]	2.0%	
Maier 2002	12	67	5	62	14.3%	2.22 [0.83, 5.94]	15.8%	
Ohls 2001A	17	87	14	85	39.0%	1.19 [0.62, 2.25]	37.2%	
Ohls 2001B	7	59	4	59	11.0%	1.75 [0.54, 5.66]	11.1%	
Romagnoli 2000	20	115	9	115	24.8%	2.22 [1.06, 4.67]	27.7%	
Total (95% CI)		**500**		**484**	**100.0%**	**1.65 [1.12, 2.43]**		**1.62 [1.09, 2.39]**
Total events	60		35					

Heterogeneity: Tau² = 0.00; Chi² = 3.14, df = 7 (P = 0.87); I² = 0%

Test for overall effect: Z = 2.41 (P = 0.02)



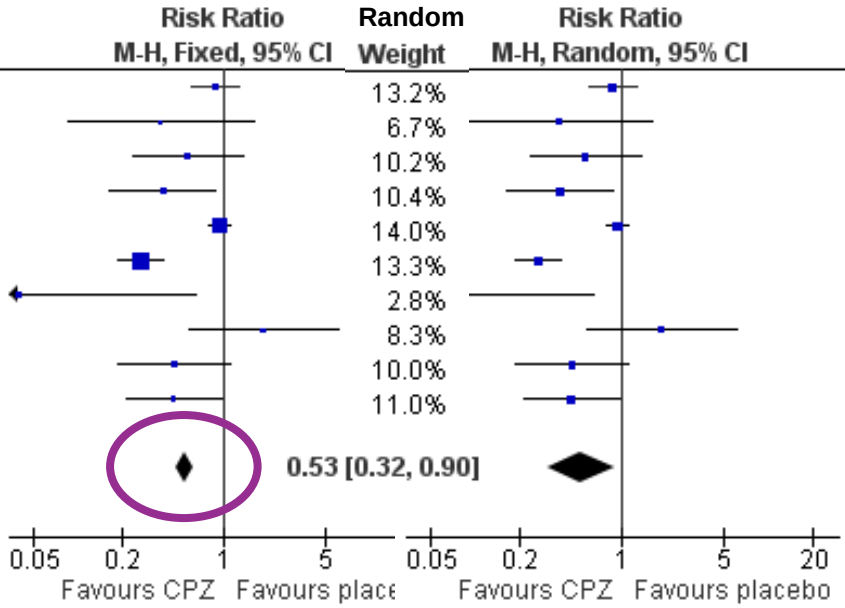
Adapted from Ohlsson A, Aher SM. Early erythropoietin for preventing red blood cell transfusion in preterm and/or low birth weight infants. *Cochrane Database of Systematic Reviews* 2006, Issue 3.

Some heterogeneity

Study or Subgroup	Chlorpromazine Events	Total	Placebo Events	Total	Fixed Weight	Risk Ratio M-H, Fixed, 95% CI	Random Weight	Risk Ratio M-H, Random, 95% CI
Chouinard 1990	14	21	16	21	6.5%	0.88 [0.60, 1.29]	13.2%	
Clark 1970a	2	15	5	14	2.1%	0.37 [0.09, 1.62]	6.7%	
Clark 1970b	10	53	6	18	3.7%	0.57 [0.24, 1.34]	10.2%	
Fleming 1959	5	21	13	21	5.3%	0.38 [0.17, 0.89]	10.4%	
Hall 1955	65	87	70	88	28.4%	0.94 [0.80, 1.10]	14.0%	
Prien 1968	37	416	70	212	37.8%	0.27 [0.19, 0.39]	13.3%	
Schiele 1961	0	20	12	20	5.1%	0.04 [0.00, 0.63]	2.8%	
Serafetinides 1972	6	14	3	13	1.3%	1.86 [0.58, 5.94]	8.3%	
Smith 1961	4	13	10	15	3.8%	0.46 [0.19, 1.12]	10.0%	
Somerville 1960	5	15	22	30	6.0%	0.45 [0.22, 0.96]	11.0%	
Total (95% CI)		**675**		**452**	**100.0%**	**0.55 [0.47, 0.63]**		**0.53 [0.32, 0.90]**
Total events	148		227					

Heterogeneity: Tau² = 0.50; Chi² = 72.76, df = 9 (P < 0.00001); I² = 88%

Test for overall effect: Z = 2.37 (P = 0.02)



Adapted from Adams CE, Awad G, Rathbone J, Thornley B. Chlorpromazine versus placebo for schizophrenia. *Cochrane Database of Systematic Reviews* 2007, Issue 2.

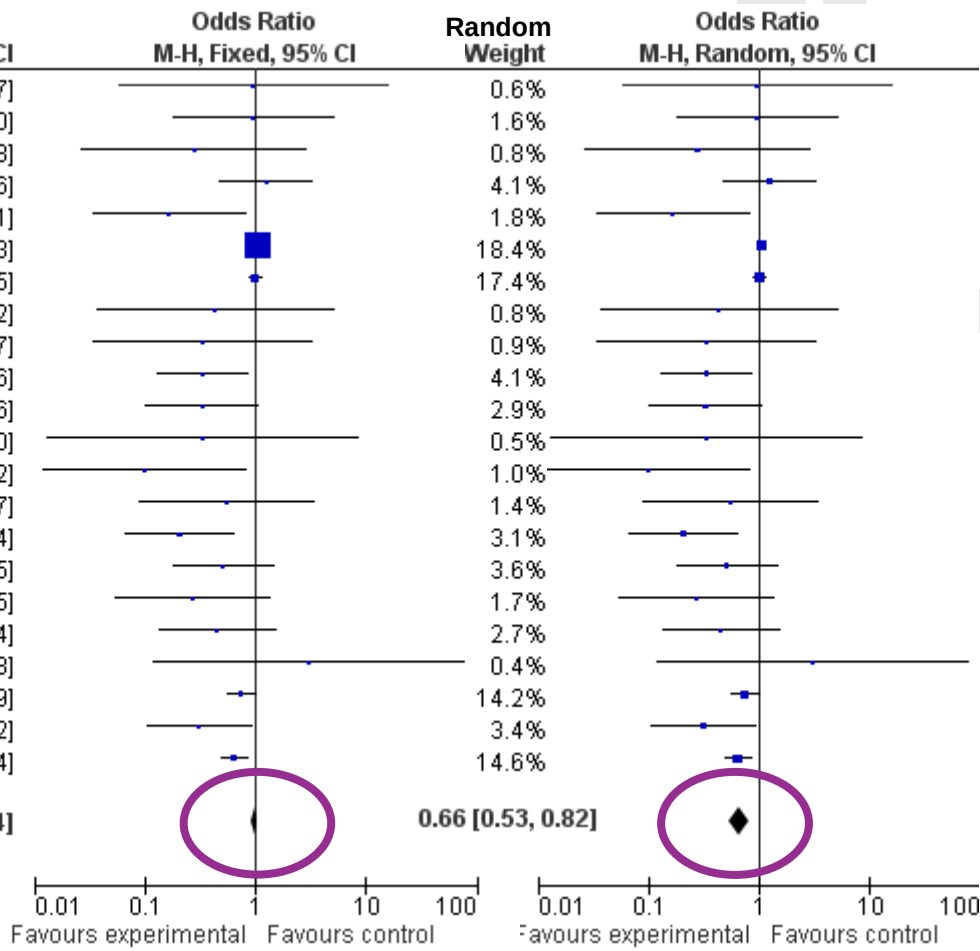
Cochrane Training

Small study effects

Study or Subgroup	Magnesium Events	Magnesium Total	Placebo Events	Placebo Total	Fixed Weight	Odds Ratio M-H, Fixed, 95% CI	Random Weight	Odds Ratio M-H, Random, 95% CI
Abraham 1987	1	48	1	46	0.0%	0.96 [0.06, 15.77]	0.6%	
Bhargava 1995	3	40	3	38	0.1%	0.95 [0.18, 5.00]	1.6%	
Ceremuzynski 1989	1	25	3	23	0.1%	0.28 [0.03, 2.88]	0.8%	
Feldstedt 1991	10	150	8	148	0.3%	1.25 [0.48, 3.26]	4.1%	
Gyamlani 2000	2	50	10	50	0.4%	0.17 [0.03, 0.81]	1.8%	
ISIS-4 1995	2216	29008	2103	29038	71.6%	1.06 [1.00, 1.13]	18.4%	
MAGIC 2000	475	3113	472	3100	14.8%	1.00 [0.87, 1.15]	17.4%	
Morton 1984	1	40	2	36	0.1%	0.44 [0.04, 5.02]	0.8%	
Nakashima 2004	1	89	3	91	0.1%	0.33 [0.03, 3.27]	0.9%	
Raghu 1999	6	169	18	181	0.6%	0.33 [0.13, 0.86]	4.1%	
Rasmussen 1986	4	56	14	74	0.4%	0.33 [0.10, 1.06]	2.9%	
Santoro 2000	0	75	1	75	0.1%	0.33 [0.01, 8.20]	0.5%	
Shechter 1990	1	50	9	53	0.3%	0.10 [0.01, 0.82]	1.0%	
Shechter 1991	2	21	4	25	0.1%	0.55 [0.09, 3.37]	1.4%	
Shechter 1995	4	96	17	98	0.6%	0.21 [0.07, 0.64]	3.1%	
Singh 1990	6	81	11	81	0.4%	0.51 [0.18, 1.45]	3.6%	
Smith 1986	2	92	7	93	0.3%	0.27 [0.06, 1.35]	1.7%	
Thogersen 1995	4	130	8	122	0.3%	0.45 [0.13, 1.54]	2.7%	
Urek 1996	1	31	0	30	0.0%	3.00 [0.12, 76.58]	0.4%	
Woods 1992	90	1150	118	1150	4.0%	0.74 [0.56, 0.99]	14.2%	
Wu 1992	5	125	12	102	0.5%	0.31 [0.11, 0.92]	3.4%	
Zhu 2002	101	1691	134	1488	4.9%	0.64 [0.49, 0.84]	14.6%	
Total (95% CI)		**36330**		**36142**	**100.0%**	**0.99 [0.94, 1.04]**	**0.66 [0.53, 0.82]**	
Total events	2936		2958					

Heterogeneity: Tau² = 0.07; Chi² = 57.78, df = 21 (P < 0.0001); I² = 64%

Test for overall effect: Z = 3.69 (P = 0.0002)



Adapted from Li J, Zhang Q, Zhang M, Egger M. Intravenous magnesium for acute myocardial infarction. Cochrane Database of Systematic Reviews 2007, Issue 2.

Which to choose?

- plan your approach at the protocol stage
- do you expect your results to be very diverse?
- consider the underlying assumptions of the model
  - fixed-effect
    - may be unrealistic – ignores heterogeneity
  - random-effects
    - allows for heterogeneity
    - estimate of distribution of studies may not be accurate if biases are present, few studies or few events

Cochrane Training

Session outline

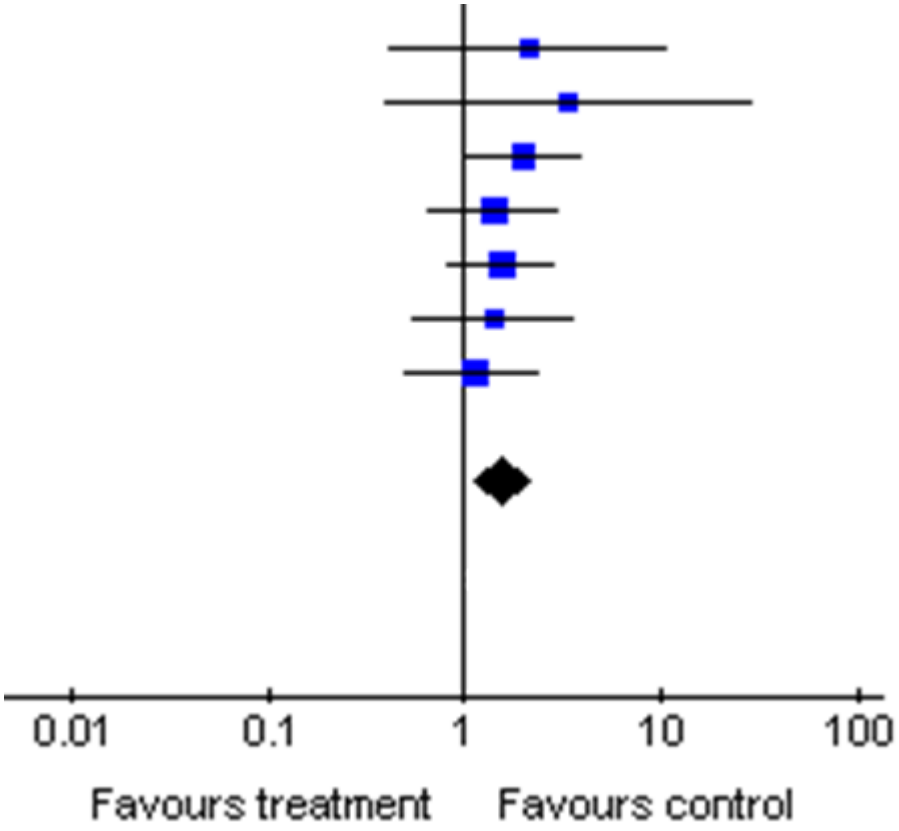
- what is heterogeneity?
- assumptions about heterogeneity
- **identifying heterogeneity**
- exploring your results

Identifying heterogeneity

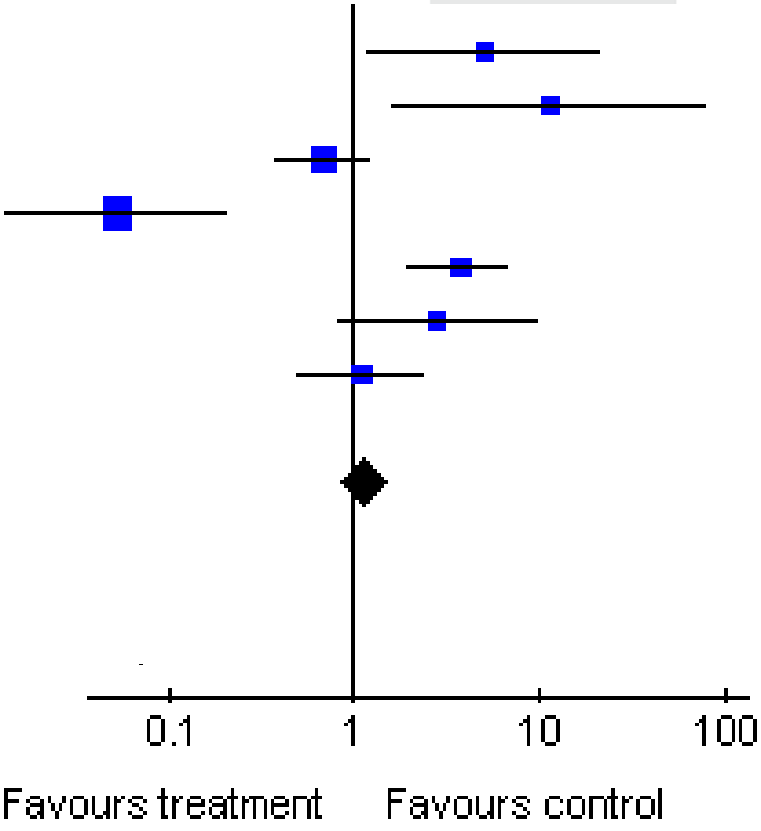
- visual inspection of the forest plots
- Chi-squared (χ^2) test (Q test)
- I^2 statistic to quantify heterogeneity

Visual inspection

Forest plot A



Forest plot B



The chi-squared (χ^2) test

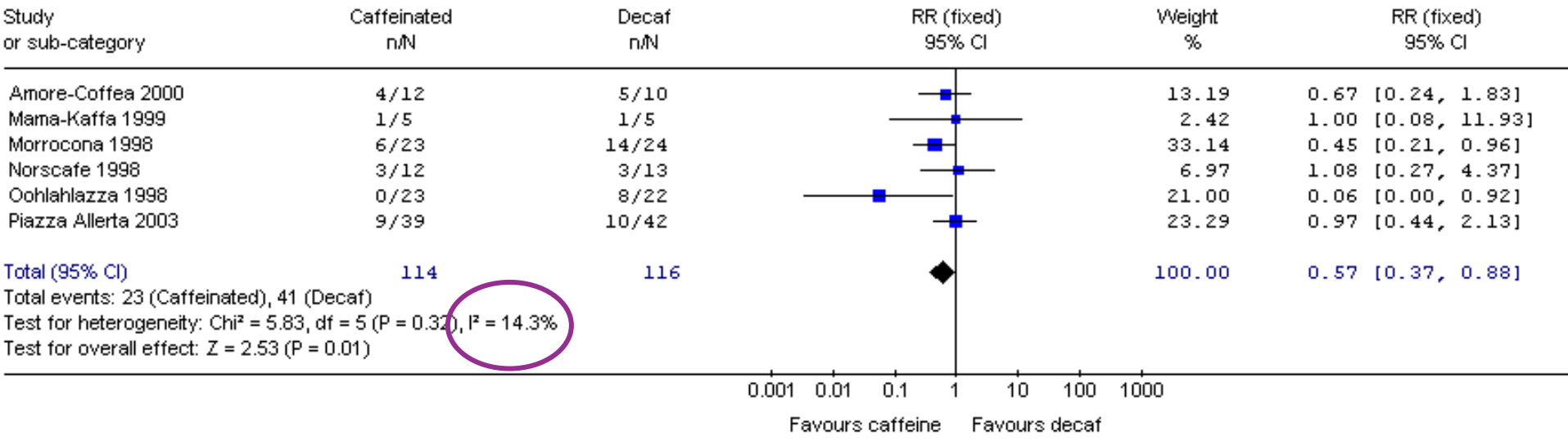
- tests the null hypothesis of homogeneity
  - low power with few studies
  - may detect clinically unimportant differences with many studies
  - narrow question (yes/no) not useful if heterogeneity is inevitable
- calculated automatically by RevMan

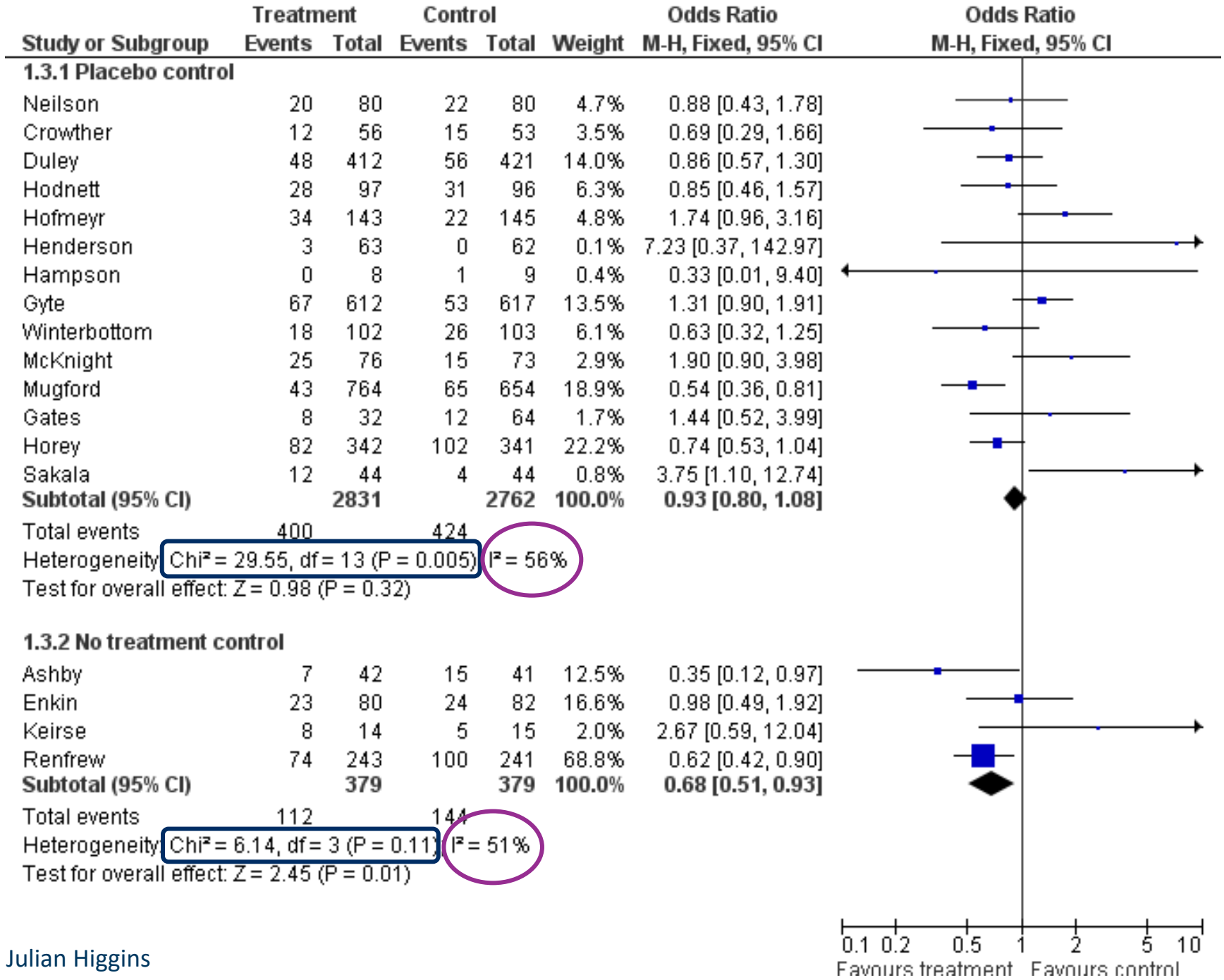
The I^2 statistic

- I^2 statistic describes the percentage of variability due to heterogeneity rather than chance (0% to 100%)
  - low values indicate no, or little, heterogeneity
  - high values indicate a lot of heterogeneity
- calculated automatically by RevMan
- be cautious in interpreting

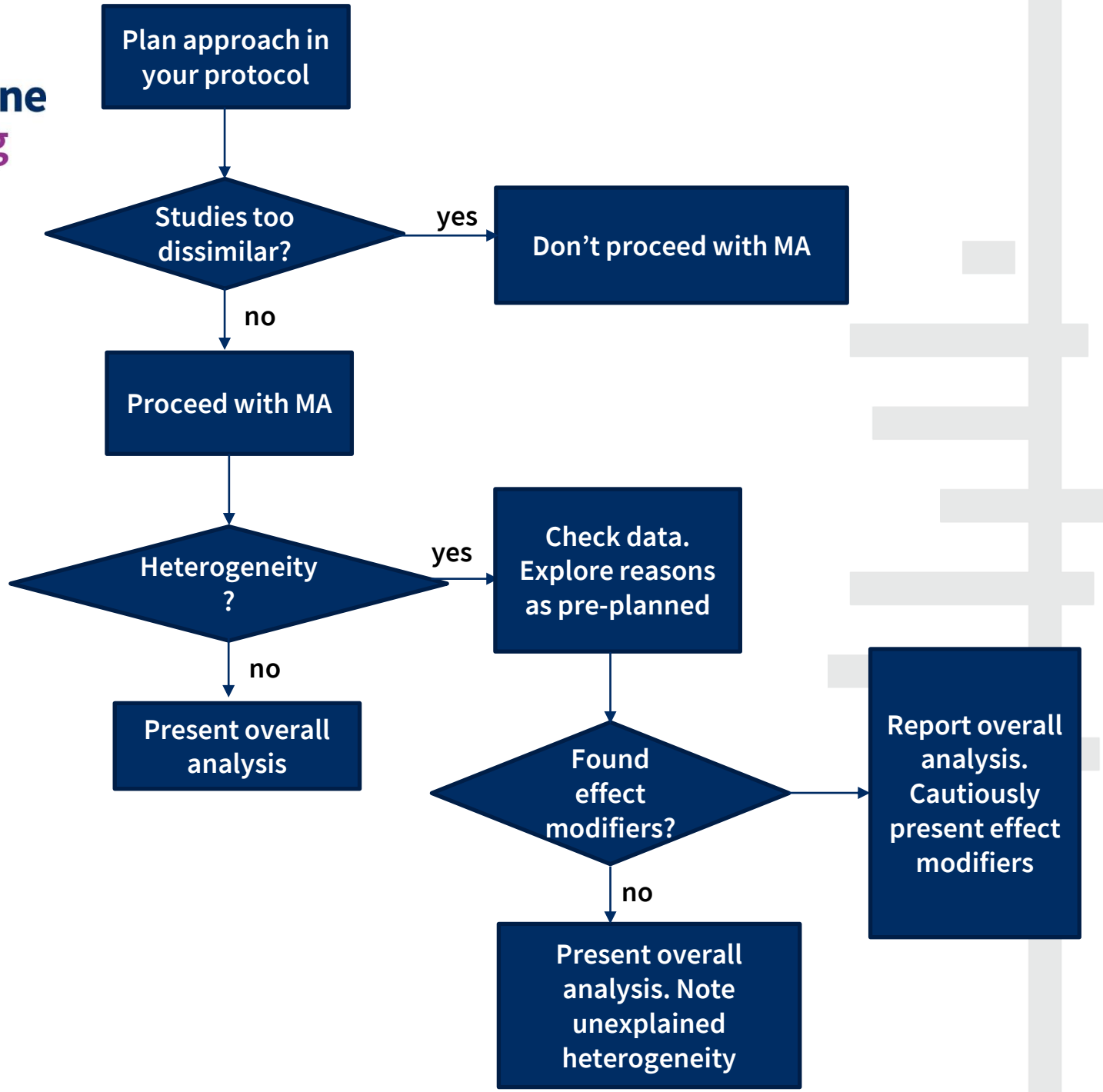
The I^2 statistic

Review: Caffeine for daytime drowsiness (version with data)
Comparison: 01 Caffeinated Coffee versus Decaffeinated Coffee
Outcome: 07 Asleep





Source: Julian Higgins


Plan approach in your protocol
Studies too dissimilar?
yes
Don't proceed with MA
no
Proceed with MA
Heterogeneity ?
yes
Check data. Explore reasons as pre-planned
no
Present overall analysis
Found effect modifiers?
Report overall analysis. Cautiously present effect modifiers
no
Present overall analysis. Note unexplained heterogeneity

Cochrane Training

Session outline

- what is heterogeneity?
- assumptions about heterogeneity
- identifying heterogeneity
- **exploring your results**

Cochrane Training

Exploring your results

- what factors appear to modify the effect?
  - clinical diversity (population, interventions, outcomes)
  - methodological diversity (study design, risk of bias)
- plan your strategy in your protocol
  - identify a limited number of important factors to investigate
  - have a scientific rationale for each factor chosen
  - declare any post-hoc investigations

Two methods available

- subgroup analysis
  - group studies by pre-specified factors
  - look for differences in results and heterogeneity
- meta-regression
  - examine interaction with categorical and continuous variables
  - not available in RevMan

Cochrane Training

Proceed with caution

- results are observational, not randomised
- be wary of multiple and post hoc comparisons
- may not be useful with few studies
- may not be able to investigate aggregate data
- look for confounding factors
- follow the plan specified in the protocol without over-emphasising particular findings

Results are rarely definitive

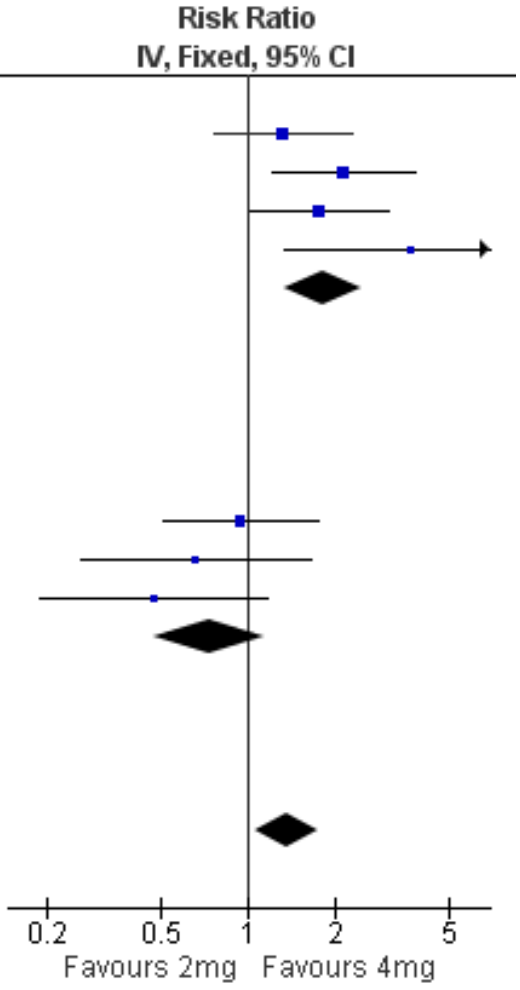
Cochrane Training

Interpreting subgroup analyses

- look at results and heterogeneity within subgroups
- are the subgroups genuinely different?
  - if only 2 subgroups – do the confidence intervals overlap?
  - statistical tests for subgroup difference
- can be more confident about:
  - pre-specified analyses
  - within-study analyses
  - effect is clinically plausible and supported by indirect evidence
  - effect is clinically important and will alter recommendations

Participant subgroups

Study or Subgroup	4 mg dose Events	Total	2 mg dose Events	Total	Weight	Risk Ratio IV, Fixed, 95% CI
5.1.1 High dependency smokers						
Garvey 2000	24	116	18	115	20.9%	1.32 [0.76, 2.30]
Herrera 1995	30	87	13	81	19.3%	2.15 [1.21, 3.82]
Kornitzer 1987	24	73	16	86	21.2%	1.77 [1.02, 3.06]
Tonnesen 1988	12	27	4	33	6.3%	3.67 [1.33, 10.08]
Subtotal (95% CI)		**303**		**315**	**67.8%**	**1.83 [1.34, 2.49]**
Total events	90		51			
Heterogeneity: Chi² = 3.45, df = 3 (P = 0.33); I² = 13%						
Test for overall effect: Z = 3.84 (P = 0.0001)						
5.1.2 Low Dependency Smokers						
Garvey 2000	16	87	17	87	17.0%	0.94 [0.51, 1.74]
Hughes 1990	5	19	8	20	7.5%	0.66 [0.26, 1.66]
Kornitzer 1987	5	17	5	8	7.7%	0.47 [0.19, 1.17]
Subtotal (95% CI)		**123**		**115**	**32.2%**	**0.73 [0.47, 1.15]**
Total events	26		30			
Heterogeneity: Chi² = 1.60, df = 2 (P = 0.45); I² = 0%						
Test for overall effect: Z = 1.36 (P = 0.17)						
Total (95% CI)		**426**		**430**	**100.0%**	**1.36 [1.06, 1.75]**
Total events	116		81			
Heterogeneity: Chi² = 15.96, df = 6 (P = 0.01); I² = 62%						
Test for overall effect: Z = 2.39 (P = 0.02)						
Test for subgroup differences: Chi² = 10.91, df = 1 (P = 0.0010), I² = 90.8%						

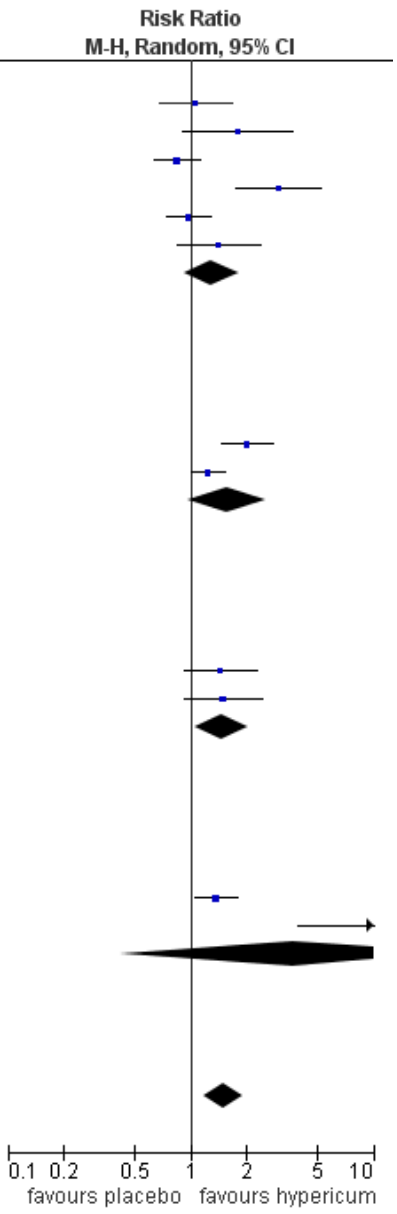


Based on Stead LF, Perera R, Bullen C, Mant D, Lancaster T. Nicotine replacement therapy for smoking cessation. *Cochrane Database of Systematic Reviews* 2008, Issue 1. Art. No.: CD000146. DOI: 10.1002/14651858.CD000146.pub3.

Intervention subgroups

Database of .pub3.

Study or Subgroup	hypericum Events	Total	placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
7.4.1 extract LI 160						
Bjerkenstedt 2005	22	54	21	55	8.2%	1.07 [0.67, 1.70]
Fava 2005	17	45	9	43	6.0%	1.80 [0.90, 3.60]
HDTSG 2002	46	113	56	116	10.0%	0.84 [0.63, 1.13]
Hänsgen 1996	35	53	12	55	7.4%	3.03 [1.77, 5.17]
Montgomery 2000	55	123	57	124	10.2%	0.97 [0.74, 1.28]
Shelton 2001	26	98	19	102	7.6%	1.42 [0.84, 2.40]
Subtotal (95% CI)		**486**		**495**	**49.3%**	**1.31 [0.92, 1.86]**
Total events	201		174			
Heterogeneity: Tau² = 0.14; Chi² = 20.94, df = 5 (P = 0.0008); I² = 76%						
Test for overall effect: Z = 1.47 (P = 0.14)						
7.4.2 extract WS 5570						
Kasper 2006	159	243	26	81	9.6%	2.04 [1.47, 2.83]
Lecrubier 2002	98	186	80	189	10.7%	1.24 [1.00, 1.54]
Subtotal (95% CI)		**429**		**270**	**20.3%**	**1.57 [0.96, 2.56]**
Total events	257		106			
Heterogeneity: Tau² = 0.11; Chi² = 6.26, df = 1 (P = 0.01); I² = 84%						
Test for overall effect: Z = 1.79 (P = 0.07)						
7.4.3 extract WS 5572						
Kalb 2001	23	37	15	35	8.2%	1.45 [0.92, 2.29]
Laakmann 1998	24	49	16	49	7.9%	1.50 [0.92, 2.46]
Subtotal (95% CI)		**86**		**84**	**16.1%**	**1.47 [1.05, 2.06]**
Total events	47		31			
Heterogeneity: Tau² = 0.00; Chi² = 0.01, df = 1 (P = 0.92); I² = 0%						
Test for overall effect: Z = 2.26 (P = 0.02)						
7.4.4 extract STW3-VI						
Gastpar 2006	71	131	51	130	10.3%	1.38 [1.06, 1.80]
Uebelhack 2004	41	70	4	70	4.0%	10.25 [3.88, 27.09]
Subtotal (95% CI)		**201**		**200**	**14.3%**	**3.59 [0.41, 31.56]**
Total events	112		55			
Heterogeneity: Tau² = 2.33; Chi² = 18.65, df = 1 (P < 0.0001); I² = 95%						
Test for overall effect: Z = 1.15 (P = 0.25)						
Total (95% CI)		**1202**		**1049**	**100.0%**	**1.51 [1.19, 1.92]**
Total events	617		366			
Heterogeneity: Tau² = 0.13; Chi² = 51.55, df = 11 (P < 0.00001); I² = 79%						
Test for overall effect: Z = 3.34 (P = 0.0008)						
Test for subgroup differences: Chi² = 1.11, df = 3 (P = 0.77), I² = 0%						



Sensitivity analysis

- not the same as subgroup analysis
- testing the impact of decisions made during the review
  - inclusion of studies in the review
  - inclusion of results at high risk of bias
  - choice of effect measure
  - assumptions about missing data
  - cut-off points for dichotomised ordinal scales
  - correlation coefficients
- repeat analysis using an alternative method or assumption
  - don’t present multiple forest plots – just report the results
  - if difference is minimal, can be more confident of conclusions
  - if difference is large, interpret results with caution

Cochrane Training

What to include in your protocol

- assessment of heterogeneity
  - assessment of comparability of studies before meta-analysis
  - visual inspection and use of statistics such as I2
- data synthesis
  - fixed-effect or random-effects model (or both)
- subgroup analyses and investigation of heterogeneity
  - planned subgroup analyses
  - any other strategies for investigating heterogeneity

Cochrane Training

Take home message

- statistical heterogeneity is the presence of differences between estimated intervention effects greater than expected because of random (sampling) variation alone
- it can be caused by clinical and methodological diversity
- fixed and random-effects models make different assumptions about heterogeneity
- explore any heterogeneity you find

References

- Deeks JJ, Higgins JPT, Altman DG. **Chapter 10: Analysing data and undertaking meta-analyses.** In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook.

Acknowledgements

- Compiled by Miranda Cumpston
- Based on materials by Georgia Salanti, Julian Higgins, Steff Lewis, Cochrane Statistical Methods Group, Cochrane Australia and Cochrane Netherlands

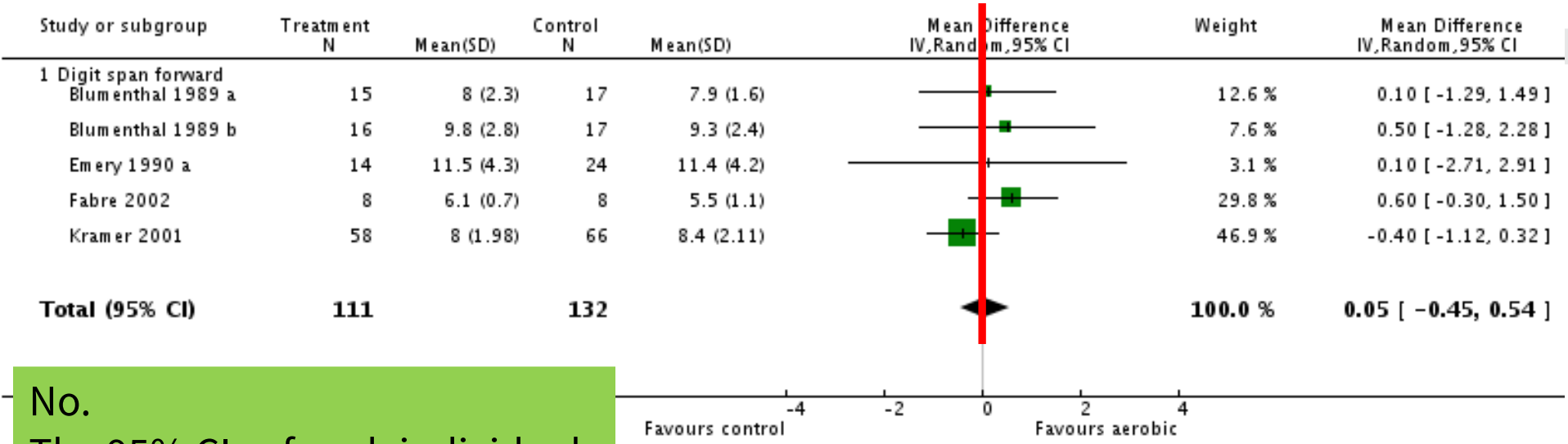
Cochrane Training

Examples: Assessing heterogeneity

Trusted evidence.
Informed decisions.
Better health.

Does visual inspection show heterogeneity?

Review: Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment
Comparison: 1 Aerobic exercise vs. any intervention
Outcome: 10 Auditory attention

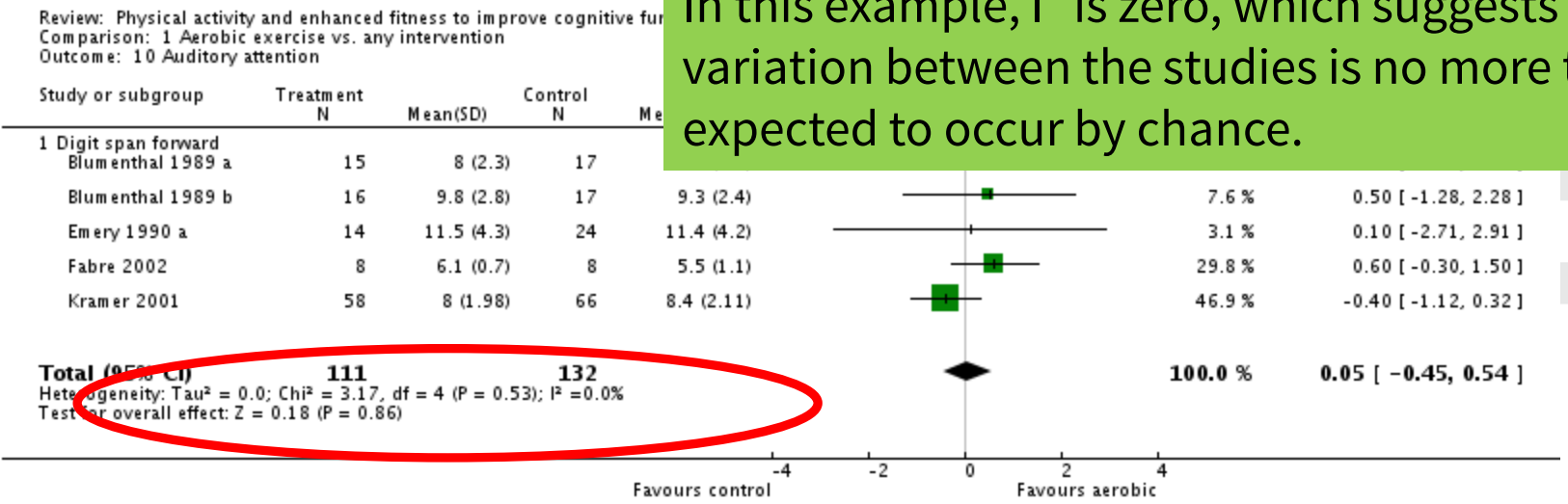


No.
The 95% CIs of each individual study overlap

Source: Angevaren M, Aufdemkampe G, Verhaar HJJ, Aleman A, Vanhees L. Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment. *Cochrane Database of Systematic Reviews* 2008, Issue 3.

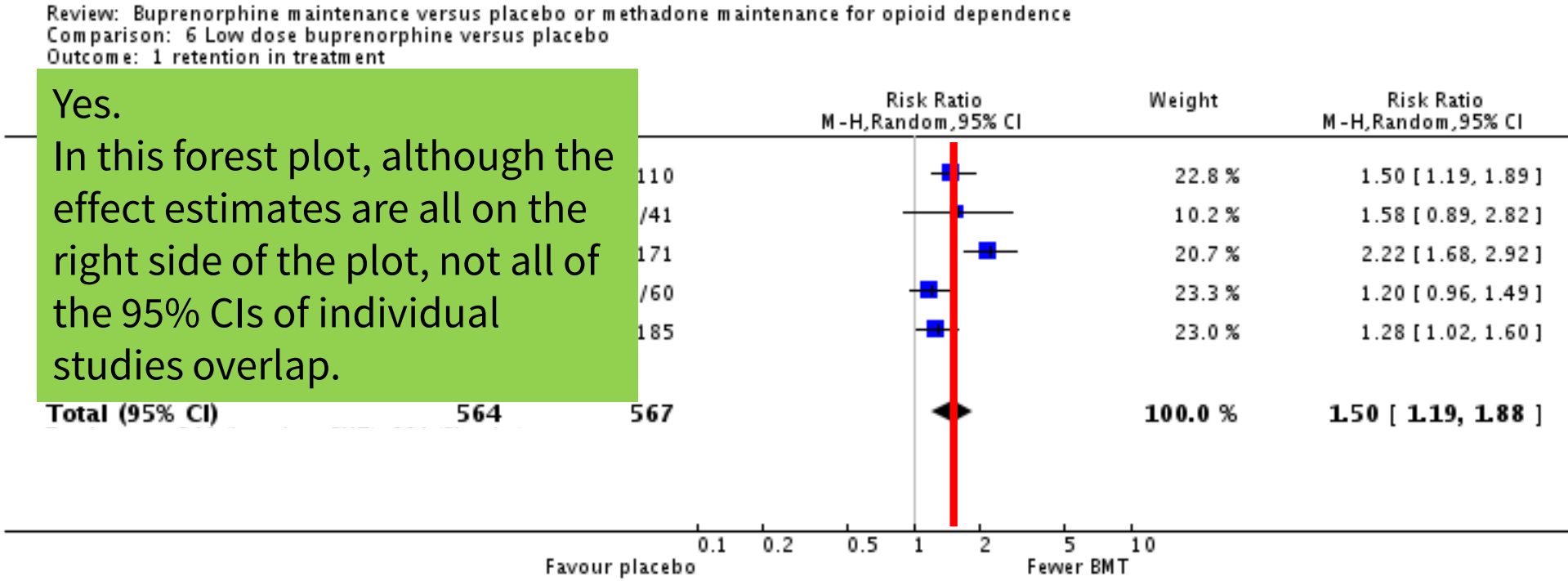
Do the statistics show heterogeneity?

No.
In this example, I^2 is zero, which suggests that the variation between the studies is no more than that expected to occur by chance.



Source: Angevaren M, Aufdemkampe G, Verhaar HJJ, Aleman A, Vanhees L. Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment. *Cochrane Database of Systematic Reviews* 2008, Issue 3.

Does visual inspection show heterogeneity?



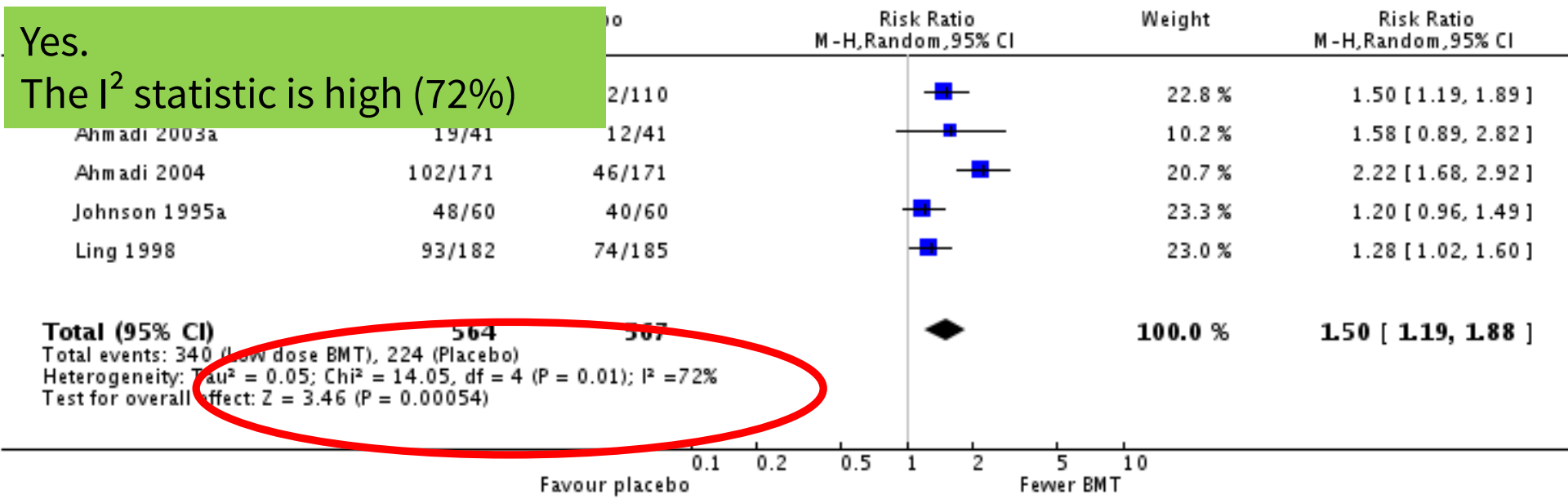
Yes.
In this forest plot, although the effect estimates are all on the right side of the plot, not all of the 95% CIs of individual studies overlap.

Source: Mattick RP, Kimber J, Breen C,Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database of Systematic Reviews* 2008, Issue 2.

Do the statistics show heterogeneity?

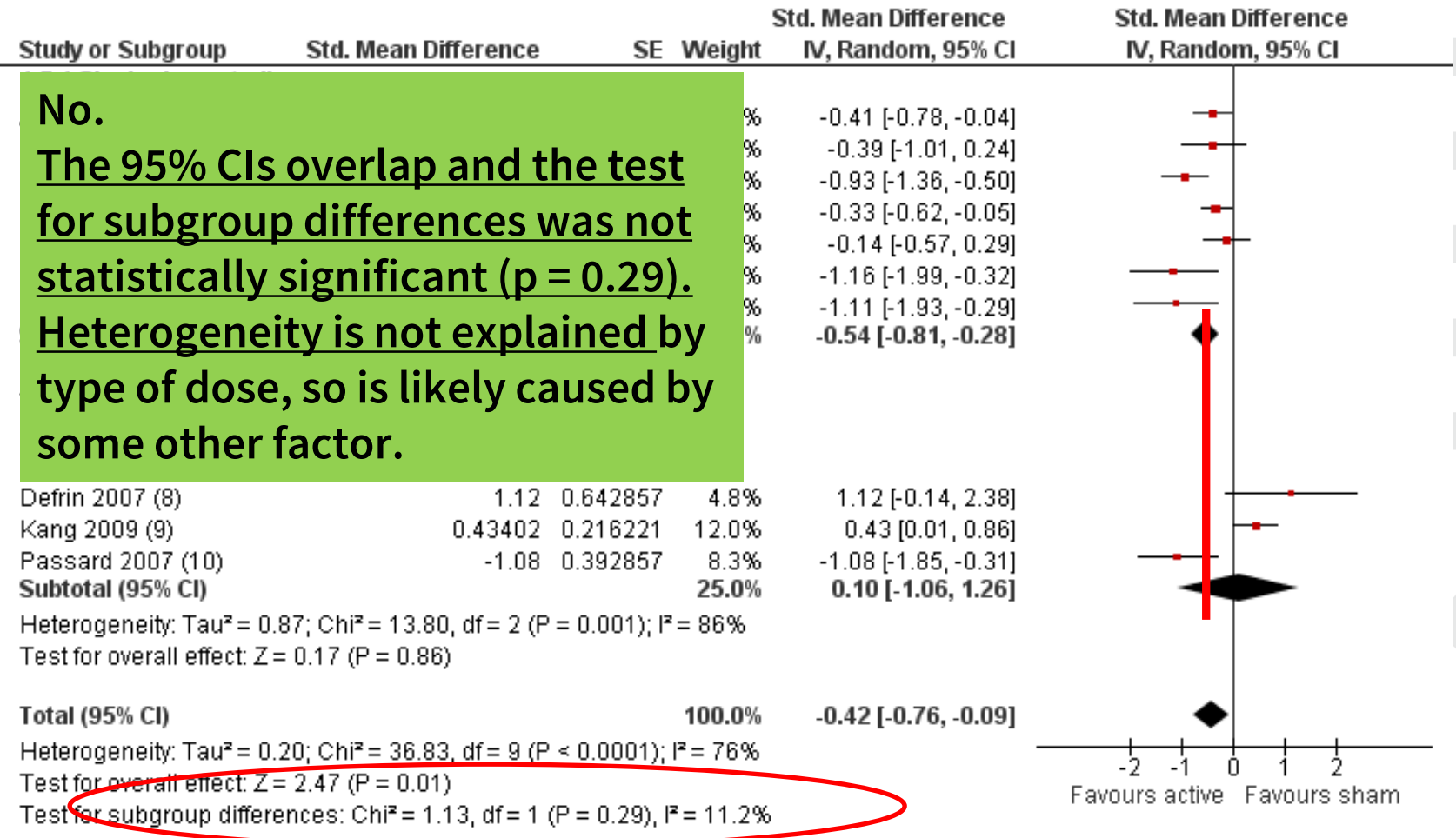
Review: Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence
Comparison: 6 Low dose buprenorphine versus placebo
Outcome: 1 retention in treatment

Yes.
The I^2 statistic is high (72%)



Source: Mattick RP, Kimber J, Breen C,Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database of Systematic Reviews* 2008, Issue 2.

Do these subgroups explain the observed heterogeneity?



No.

The 95% CIs overlap and the test for subgroup differences was not statistically significant (p = 0.29). Heterogeneity is not explained by type of dose, so is likely caused by some other factor.

Based on: O'Connell NE, Wand BM, Marston L, Spencer S, DeSouza LH. Non-invasive brain stimulation techniques for chronic pain. *Cochrane Database of Systematic Reviews* 2010, Issue 9.

Cochrane Training

References

- Deeks JJ, Higgins JPT, Altman DG. **Chapter 10**: **Analysing data and undertaking meta-analyses**. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.1 (updated September 2020). Cochrane, 2020. Available from www.training.cochrane.org/handbook.

Acknowledgements

- **Compiled by Sharon Kramer**
- **Based on Cochrane online training modules developed by the University of Portsmouth and the Cochrane UK.**