Publication Bias in Systematic Reviews

Payam Kabiri, MD, PhD
Clinical Epidemiologist
Farid Najafi, MD. PhD.
Epidemiologist

Publication bias

- The likelihood of finding studies is related to the results of those studies

Publication Bias

- "Publication bias refers to the greater likelihood that studies with positive results will be published"

- *JAMA* 2002;287:2825-2828

Publication Bias

- Positive trials are more likely to be submitted for publication
- Positive trials are more likely to be published
- Positive trials are more likely to be published quickly

- Stern and Simes *BMJ 1997;*315:640-645

Publication Bias

- Sterling study: 97% of papers published in 4 psychology journals showed statistically significant results at alpha level 5% !
- Dickerson study: compared published RCTs with unpublished ones .results:55%pub,14% unpub, favoring new therapy!
- Mahoney study:75 reviewers asked to review different versions of a fictitious manuscript. ”introduction” & ”methods” : identical, ”results” & “discussion” : different (+/ambiguous /-). results of reviewers evaluation : manuscripts with “positive” results received higher average scores!

Controlled Clinical Trials
Volume 8, Issue 4, December 1987, Pages 343-353

Publication bias and clinical trials

K. Dickersin *, S. Chan **, T.C. Chalmersx, H.S. Sacks, H. Smith Jr.

Show more

https://doi.org/10.1016/0197-2456(87)90155-3

Abstract

A study was performed to evaluate the extent to which the medical literature may be misleading as a result of selective publication of randomized clinical trials (RCTs) with results showing a statistically significant treatment effect. Three hundred eighteen authors of published trials were asked whether they had participated in any unpublished RCTs. The 156 respondents reported 271 unpublished and 1041 published trials. Of the 178 completed unpublished RCTs with a trend specified, 26 (14%) favored the new therapy compared to 423 of 767 (55%) published reports ($p < 0.001$). For trials that were completed but not published, the major reasons for nonpublication were “negative” results and lack of interest. From the data provided, it appears that non-publication was primarily a result of failure to write up and submit the trial results rather than rejection of submitted manuscripts. The results of this study imply the existence of a publication bias of importance both to meta-analysis and the interpretation of statistically significant positive trials.

Cognitive Therapy and Research
June 1977, Volume 1, Issue 2, pp 161–175 | Cite as

Publication prejudices: An experimental study of confirmatory bias in the peer review system

Authors | Authors and affiliations

Michael J. Mahoney

Abstract

Confirmatory bias is the tendency to emphasize and believe experiences which support one's views and to ignore or discredit those which do not. The effects of this tendency have been repeatedly documented in clinical research. However, its ramifications for the behavior of scientists have yet to be adequately explored. For example, although publication is a critical element in determining the contribution and impact of scientific findings, little research attention has been devoted to the variables operative in journal review policies. In the present study, 75 journal reviewers were asked to referee manuscripts which described identical experimental procedures but which reported positive, negative, mixed, or no results. In addition to showing poor interrater agreement, reviewers were strongly biased against manuscripts which reported results contrary to their theoretical perspective. The implications of these findings for epistemology and the peer review system are briefly addressed.

Publication Bias

- 1)…if they had reached sig.
- 2) positive result
- 3) interesting results for both reviewers & authors!
- 4) language bias (ENG) in being included in a meta-analysis.

How to Bypass Publication Bias

- Searching Libraries for Thesis & Research Reports
- Searching Registries
- Searching Grey Literature
- Searching especial Journals like:

"*Journal of Negative results in Biomedicine*"

Funnel plots

- A funnel plot is a scatter plot of treatment effect (Effect Size) against a measure of study size.
- Useful when the number of studies is not less than 10

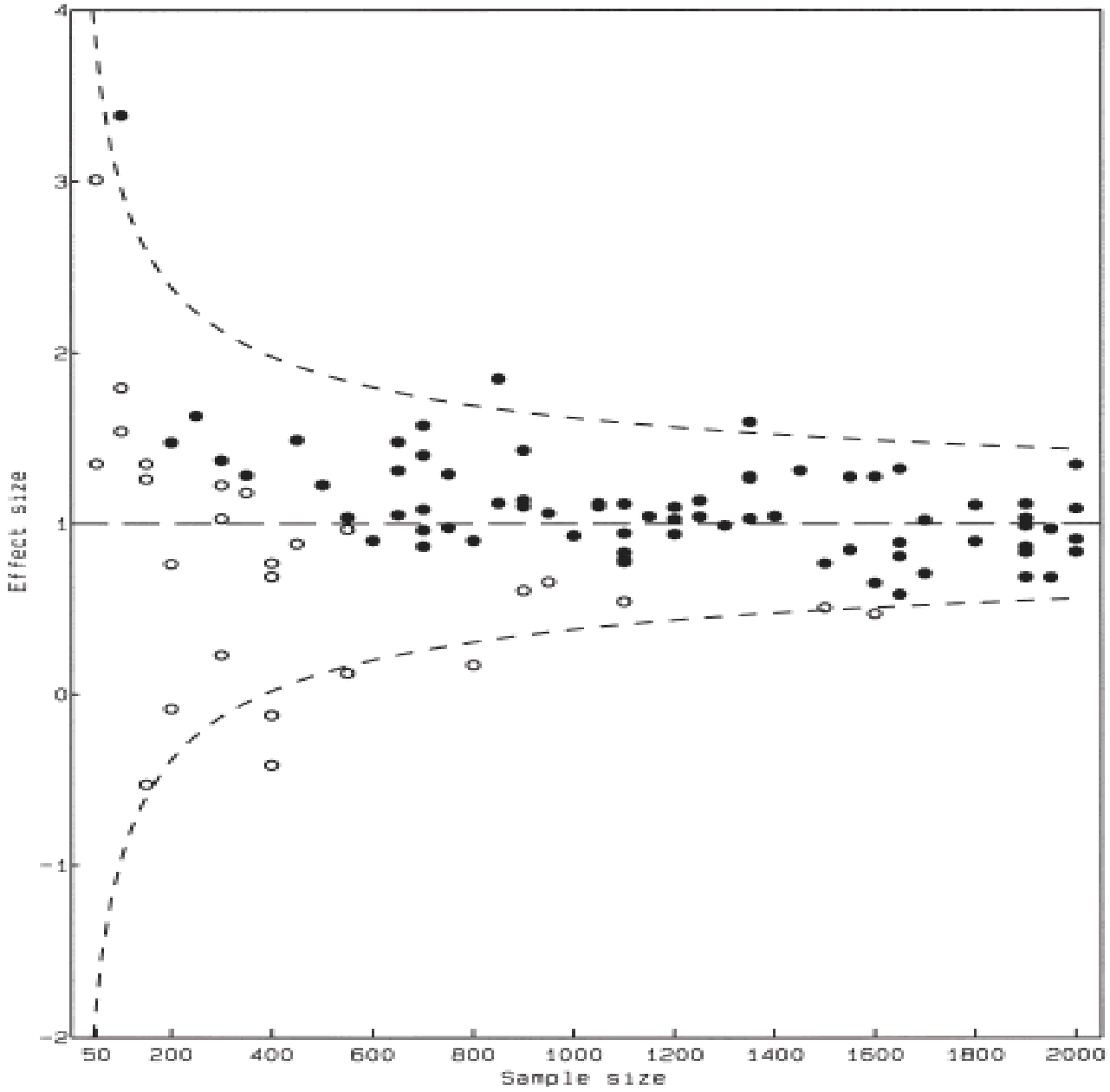
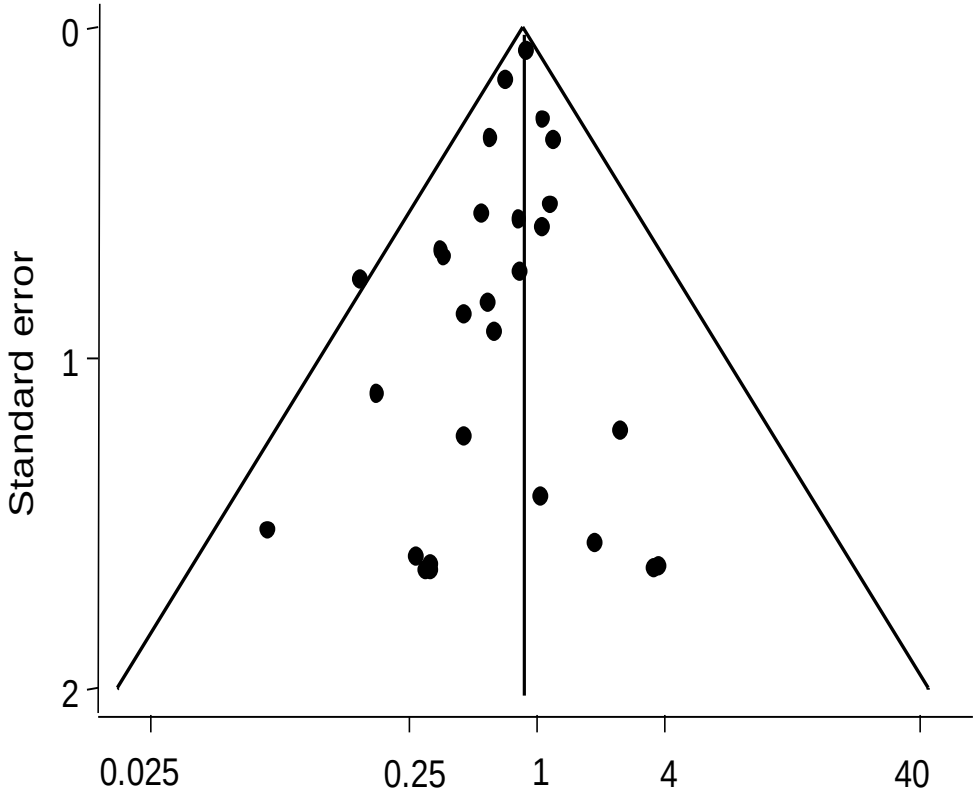


Fig. 1. Simulated funnel plot. (●) Effect size significantly increased ($P < 0.05$). (○) Effect size not significant. (— —) Expected value of effect size. (– –) Expected 95% confidence region for samples.

Funnel plots

- A funnel plot is a scatter plot of treatment effect (Effect Size) against a measure of study size.

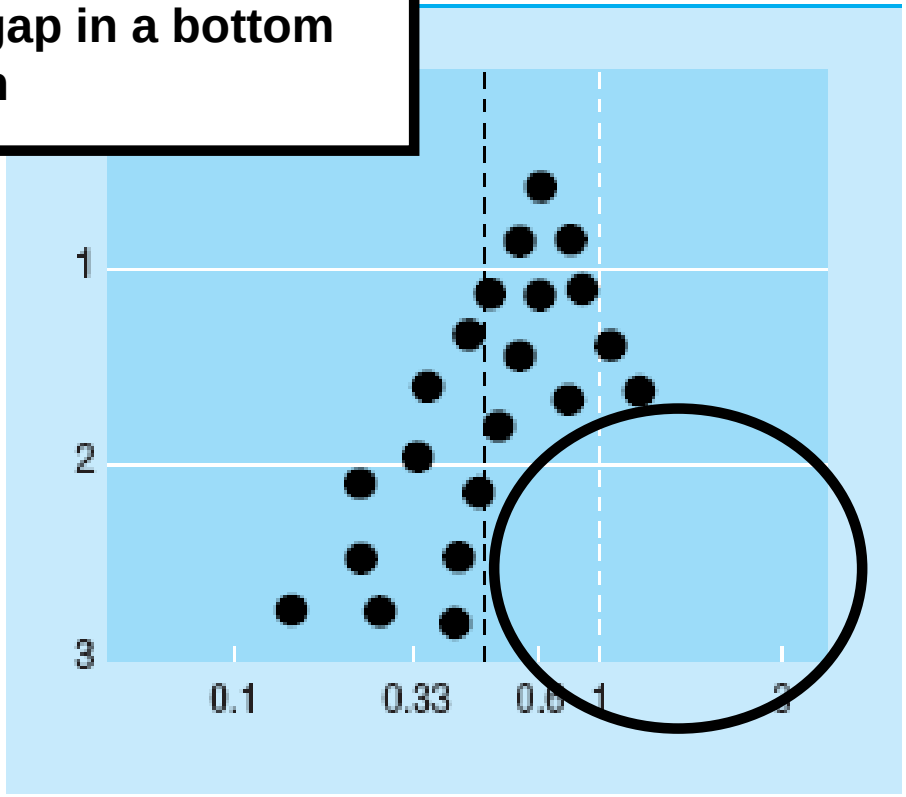


Why Funnel?

- Precision in the estimation of the true treatment effect increases as the sample size increases.
- Small studies scatter more widely at the bottom of the graph
- In the absence of bias the plot should resemble a symmetrical inverted funnel

Publication Bias

Asymmetrical appearance of the funnel plot with a gap in a bottom corner of the graph



Publication Bias

- In this situation the effect calculated in a meta-analysis will overestimate the treatment effect
- The more pronounced the asymmetry, the more likely it is that the amount of bias will be substantial.

Possible sources of asymmetry in funnel plots

1. Selection biases
   - Publication bias
   - Location biases
2. Poor methodological quality of smaller studies
   - Poor methodological design
   - Inadequate analysis
   - Fraud
3. True heterogeneity
   - Size of effect differs according to study size (for example, due to differences in the intensity of interventions or differences in underlying risk between studies of different sizes)
4. Chance

Examples for true heterogeneity

- Higher benefit in patients at high risk for outcome which is affected by the intervention and these high risk patients are usually more likely to be included in early, small studies
- Standard treatment improve over time and smaller trial begin before larger trial

Publication bias Approaches

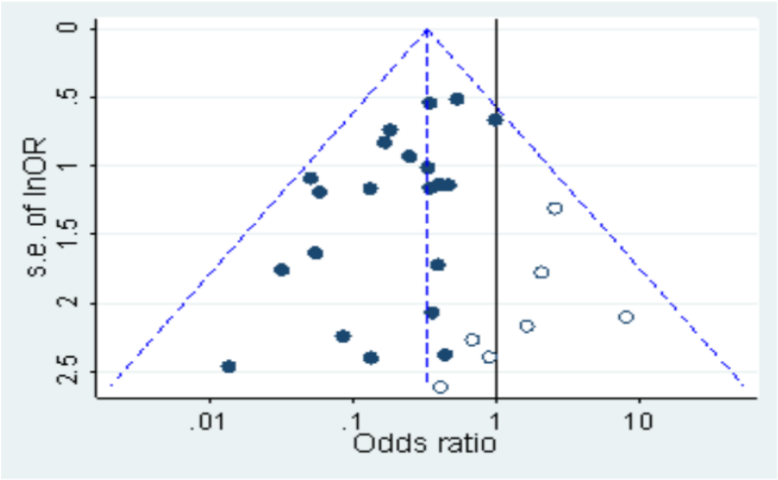
- Attempt to Retrieve all Studies
- Worst Case Adjustment
  - Number of unpublished negative studies to negate a "positive" meta-analysis:
  - $X = [N \times (ES) / 1.645]^2 - N$
    - where: N = number of studies in meta-analysis,
    - ES = effect size
- Example:
  - If N = 25, and ES = 0.6 then X = 58.2
  - Almost 60 unpublished negative studies would be required to negate the meta-analysis of 25 studies.

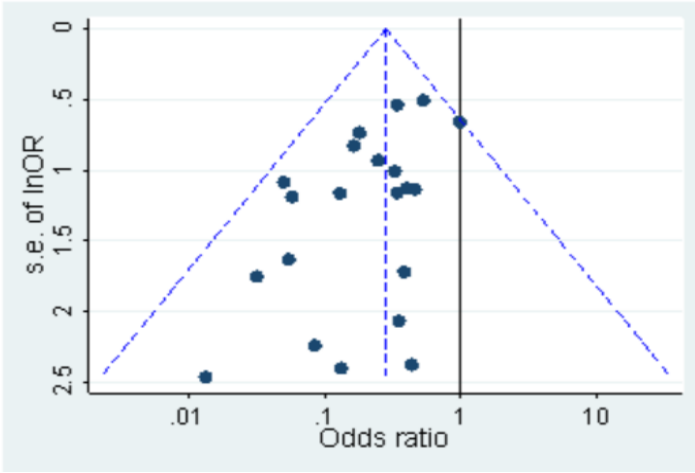
Poor methodological quality

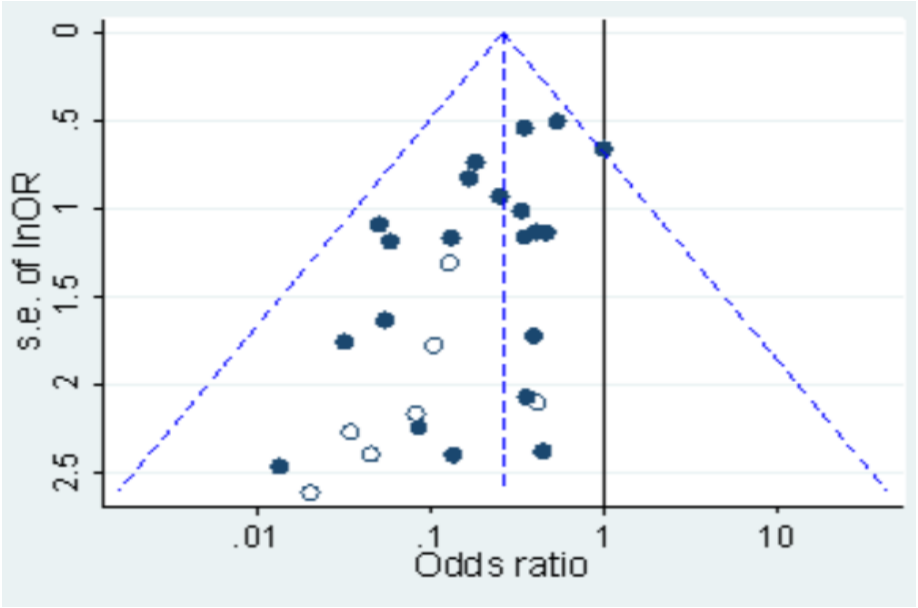
- Smaller studies are, on average, conducted and analyzed with less methodological rigor than larger studies.
- Trials of lower quality also tend to show larger treatment effects
- Trials which, if conducted and analyzed properly, would have been ‘negative’ may thus become ‘positive’

Figure 10.4.a: Hypothetical funnel plots

Panel A: symmetrical plot in the absence of bias. Panel B: asymmetrical plot in the presence of reporting bias. Panel C: asymmetrical plot in the presence of bias because some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated intervention effect estimates.







Meta-Analysis

Payam Kabiri, MD. PhD.

Clinical Epidemiologist

Farid Najafi, MD. PhD.
Epidemiologist

Meta-Analysis

- Meta-analysis is a statistical analysis of a collection of studies
- Meta-analysis methods focus on contrasting and comparing results from different studies in anticipation of identifying consistent patterns and sources of disagreements among these results
- Primary objective:
  - Synthetic goal (estimation of summary effect)

    vs.
  - Analytic goal (estimation of differences)

Systematic Review & Meta-analyses

- A systematic review need not contain any meta-analyses.

- If there is considerable variation in results, it may be misleading to quote an average value

What is heterogeneity?

Variability in effect size estimates which exceeds that expected from sampling error alone.

Heterogeneity

Sources of variety of varieties are:

- Study diversity (difference in participant, intervention and outcome)
- Methodological diversity (study design and risk of bias)
- Statistical heterogeneity (result from two above mentioned sources)

Sources of Variation over Studies

- Inter-study variation may exist
- Sampling error may vary among studies (sample size)
- Characteristics may differ among studies (population, intervention)

Heterogeneity

How to Identify it:

- Common sense

  are the populations, interventions and outcomes in each of the included studies sufficiently similar

- Statistical tests

Statistical Tests of Homogeneity (heterogeneity)

- Homogeneity calculations
  - H_o = studies are homogeneous
  - Based on testing the sum of weighted differences between the summary effect and individual effects
  - Calculate Mantel Haenszel Q, where:

$$Q = \sum[weight_i \times (lnOR_{mh} - lnOR_i)^2]$$

  - If $p < 0.05$, then there is significant heterogeneity.

Statistical Tests of Homogeneity (heterogeneity)

- Power of such statistical tests is low

  (a non-significant test does not rule out clinically important heterogeneity)

Tau^2 (t^2)

- Total variance= between studies variance + within studies variance
- Tau^2 is a sign of between studies
- Higher Tau^2 shows higher heterogeniety

I^2

- I^2 reports the quantitative value for heterogeneity (by Higgins)
- The values are between 0.00% to 100%
- 0.00% means there is no heterogeneity
- 0.00%-25% low heterogeneity
- 26%-50% moderate heterogeneity
- >50% high heterogeneity

$$I^2 = \left(\frac{Q - df}{Q}\right) * 100$$

Statistical Models

For Calculating overall effects, there are two Statistical Models:

- Fixed effects model (FEM)
- Random effects model (REM)

How to deal with Heterogeneity

- If homogenous, use fixed effects model
  - random will give same results
  - fixed is computationally simpler
- If heterogeneous…then first ask why?!
  - In the face of heterogeneity, focus of analysis should be to describe possible sources of variability
  - attempt to identify sources of important subgroup differences

How to Deal with Heterogeneity

1. No Heterogeneity:

   Use Fixed Effects Model

2. If Heterogeneity is there:

   Do not ‘pool at all’

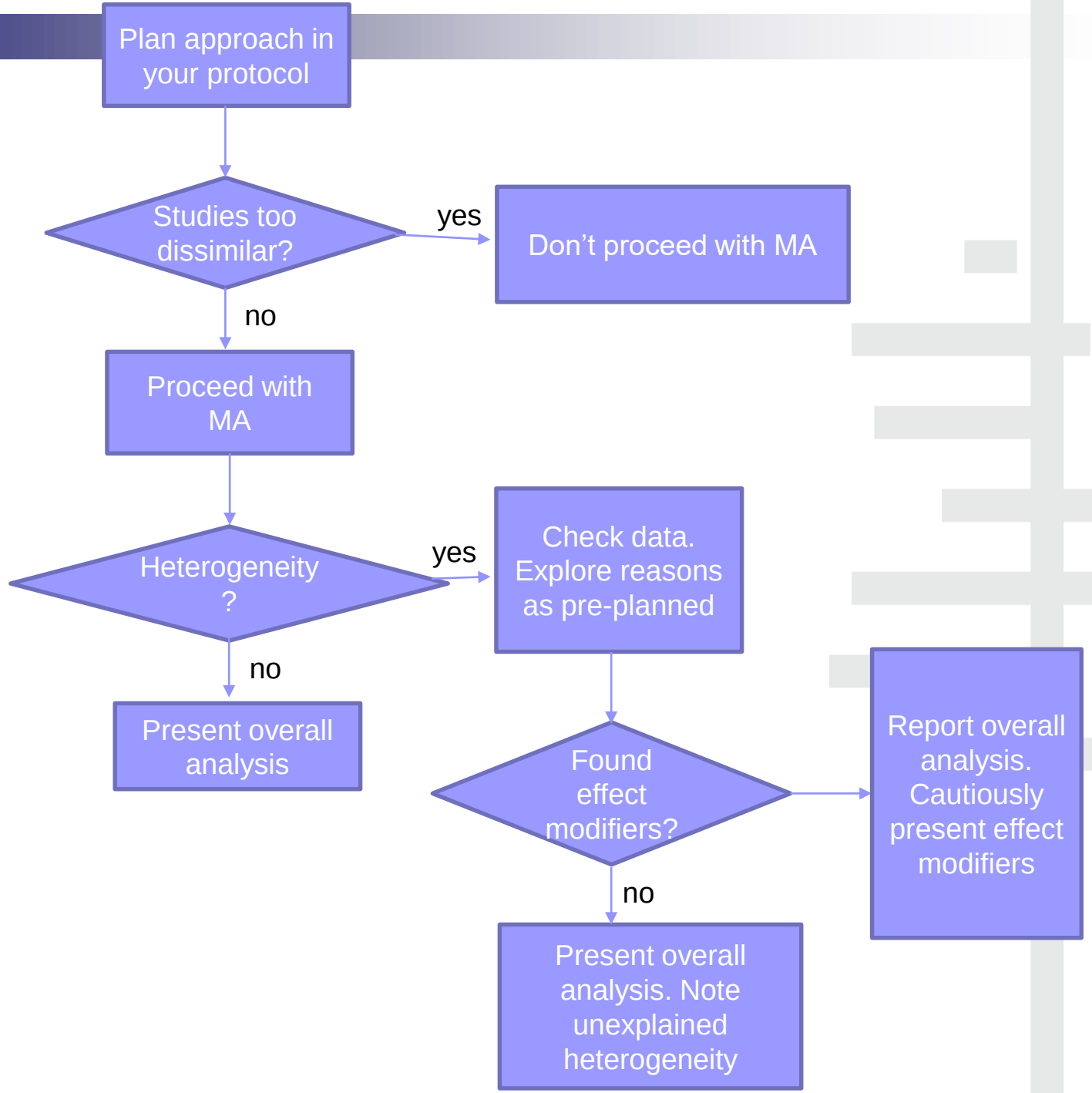
3. Explore heterogeneity through:

   Subgroup analysis

   Meta-regression

4. If Heterogeneity still persist:

   Use Random Effects Model


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

Exploring Heterogeneity

Comparison: Subgroup: Quality of Blinding
Outcome: Lumbar BMD

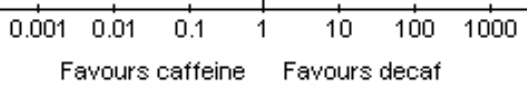
Study	Expt n	Expt mean(sd)	Ctrl n	Ctrl mean(sd)	WMD (95%CI Fixed)	Weight %	WMD (95%CI Fixed)
Blinding = 0							
Evans 1993	15	2.40 (9.10)	11	-4.70 (4.40)		1.7	7.100 [1.811,12.389]
Gurlek 1997	10	4.54 (17.96)	10	0.14 (3.42)		0.4	4.400 [-6.932,15.732]
Montessori 1997	40	6.28 (5.02)	34	-0.03 (9.20)		3.9	6.310 [2.848,9.772]
Wimalawansa 95	14	4.22 (3.93)	14	-2.25 (3.55)		6.0	6.470 [3.696,9.244]
Wimalawansa 98	16	4.30 (2.80)	16	-0.90 (2.40)		14.1	5.200 [3.393,7.007]
Subtotal (95%CI)	95		85			26.0	5.767 [4.435,7.100]
Chi-square 1.02 (df=4) Z=8.48							
Blinding = 1							
Herd 1997	64	2.14 (3.76)	71	-1.72 (3.45)		30.9	3.860 [2.638,5.082]
Meunier 1997	25	0.58 (4.15)	24	-2.34 (4.02)		8.8	2.920 [0.632,5.208]
Pouilles 1997	43	0.06 (5.90)	43	-2.46 (4.44)		9.5	2.520 [0.313,4.727]
Storm 1990	22	4.80 (7.79)	21	-4.50 (7.97)		2.1	9.300 [4.587,14.013]
Watts 1990	92	4.20 (7.67)	90	1.38 (7.98)		8.9	2.820 [0.545,5.095]
Watts B 1990	93	5.20 (6.75)	88	1.47 (5.83)		13.7	3.730 [1.895,5.565]
Subtotal (95%CI)	339		337			74.0	3.579 [2.789,4.370]
Chi-square 7.52 (df=5) Z=8.88							
Total (95%CI)	434		422			100.0	4.148 [3.469,4.828]
Chi-square 16.20 (df=10) Z=11.96							

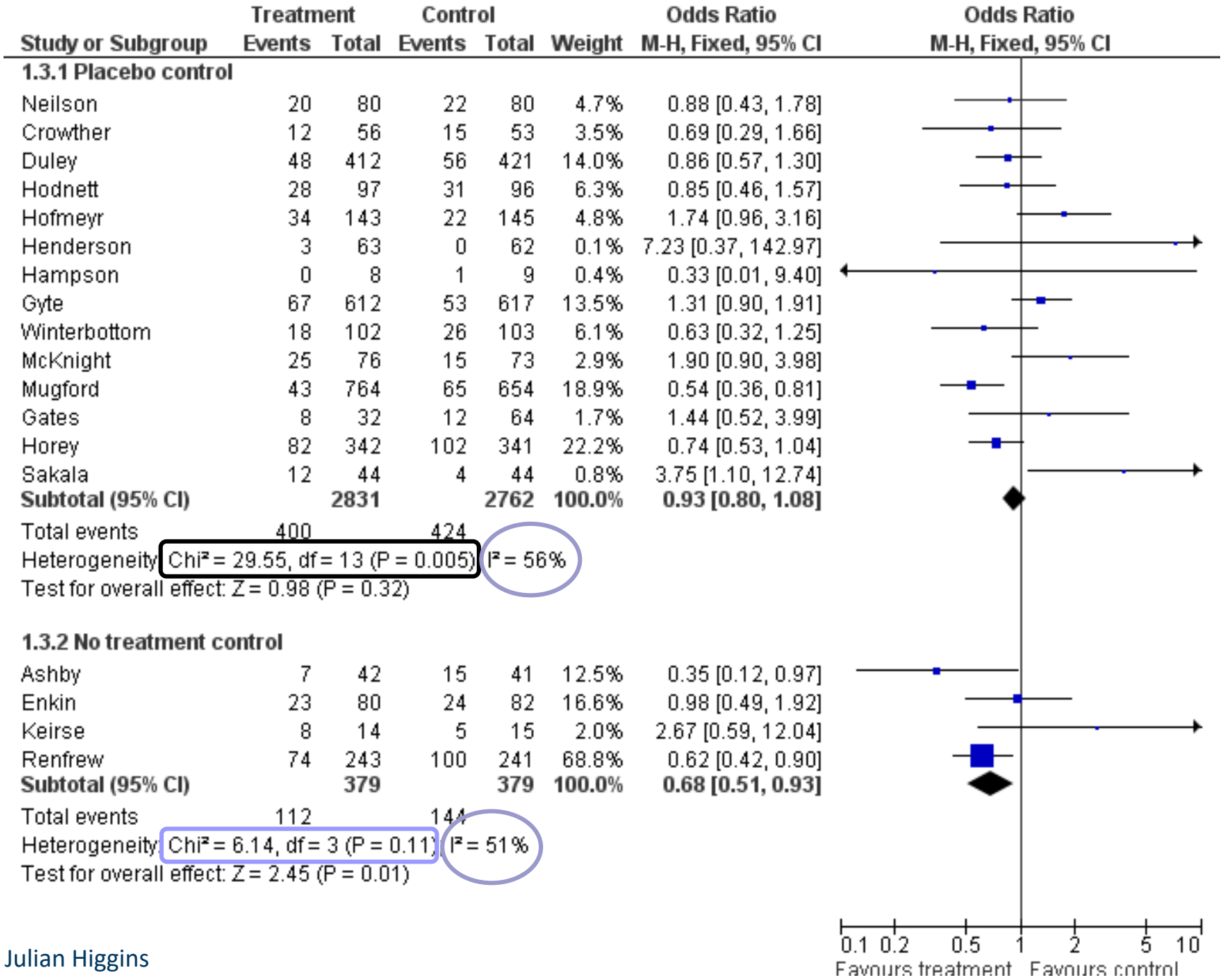
The I^2 statistic

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

Study or sub-category	Caffeinated n/N	Decaf n/N	RR (fixed) 95% CI	Weight %	RR (fixed) 95% CI
Amore-Coffea 2000	4/12	5/10		13.19	0.67 [0.24, 1.83]
Mama-Kaffa 1999	1/5	1/5		2.42	1.00 [0.08, 11.93]
Morrocona 1998	6/23	14/24		33.14	0.45 [0.21, 0.96]
Norscafe 1998	3/12	3/13		6.97	1.08 [0.27, 4.37]
Oohlahlazza 1998	0/23	8/22		21.00	0.06 [0.00, 0.92]
Piazza Allerta 2003	9/39	10/42		23.29	0.97 [0.44, 2.13]
Total (95% CI)	114	116		100.00	0.57 [0.37, 0.88]

Total events: 23 (Caffeinated), 41 (Decaf)
Test for heterogeneity: Chi² = 5.83, df = 5 (P = 0.32), I² = 14.3%
Test for overall effect: Z = 2.53 (P = 0.01)



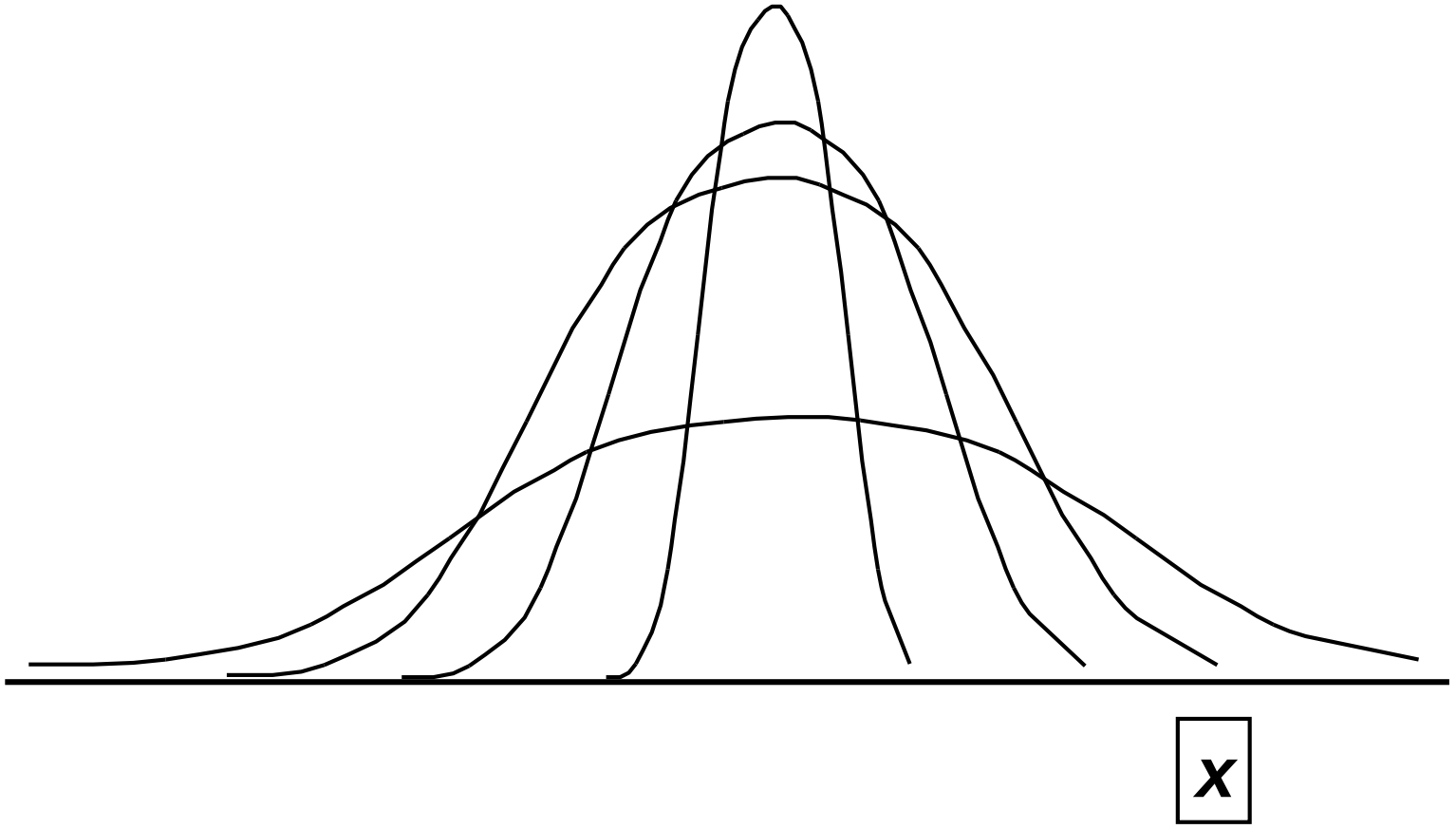


Source: Julian Higgins

Fixed effects model

- All trials are measuring a single, true effect
- The reason for any difference between the effect in an individual trial and this true effect is chance

Fixed-Effects Model



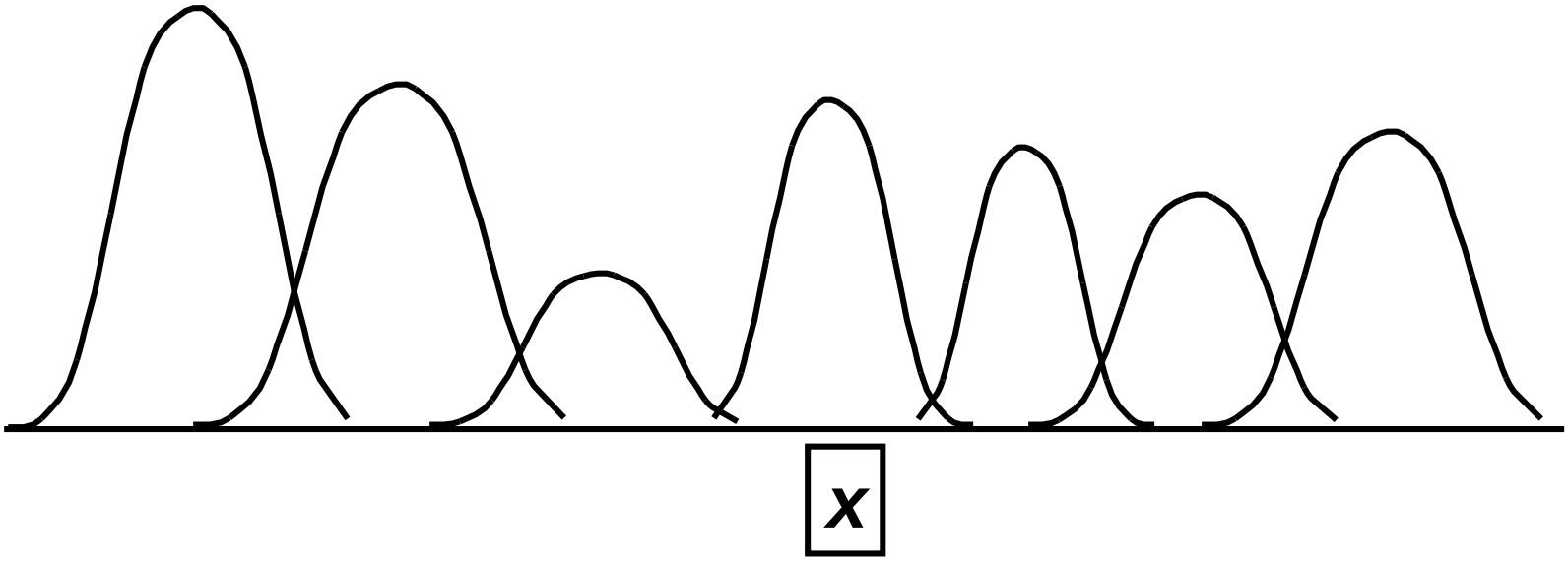
Fixed Effects Model

- Require from each study
  - effect estimate; and
  - standard error of effect estimate
- Combine these using a weighted average:
  - pooled estimate = $\frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}}$
  - where weight = 1 / variance of estimate
- Assumes a common underlying effect behind every trial

Random Effects models

- consider both *between-study* and *within-study* variability.
- Each trial is measuring a different, true effect
- The true effects for each trial are normally distributed
- There is a true average effect
- The reason for any difference between the effect in an individual trial and this average effect is both the difference between the true effect for the trial and this average, and chance.

Random-Effects Model



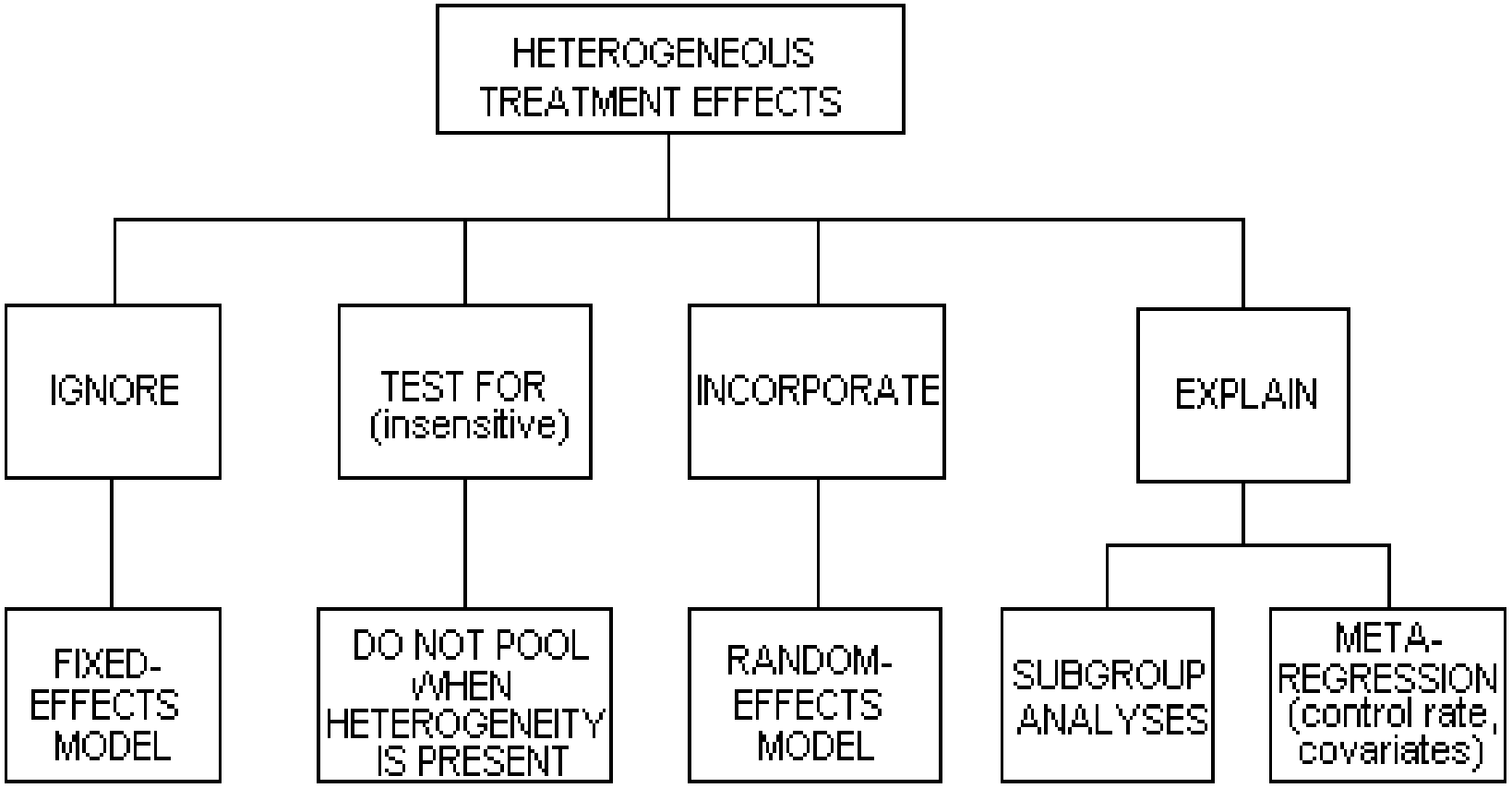
Random-Effects Model

- Assume true effect estimates really vary across studies
- Two sources of variation:
  - within studies (between patients)
  - between studies (heterogeneity)
- What the software does is Revise weights to take into account both components of variation:
- Weight = $\frac{1}{\text{Variance + heterogeneity}}$

Random-Effects Model

- When heterogeneity exists we get:
  - a different pooled estimate (but not necessarily) with a different interpretation
  - a wider confidence interval
  - a larger p-value

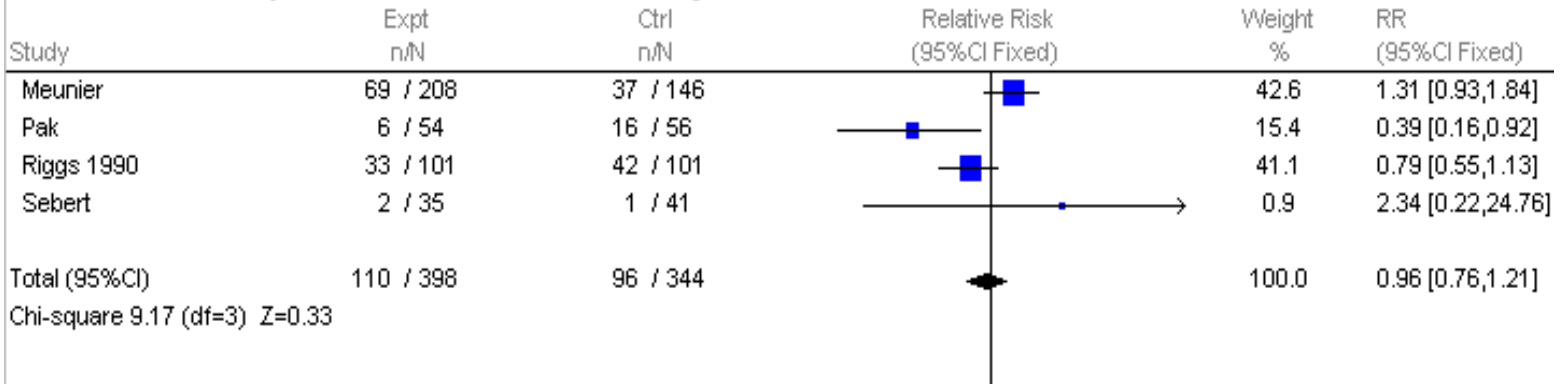
Generic Inferential Framework



Fixed vs. Random Effects: Discrete Data

Fixed Effects

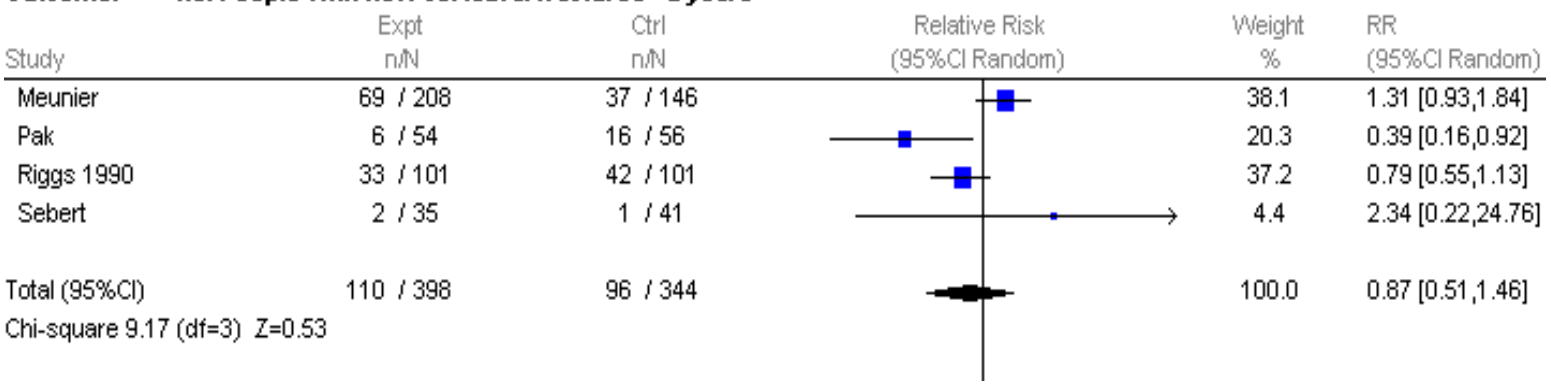
Comparison: **Fluoride vs Placebo - Overall**
Outcome: **No. People with new vertebral fractures - 2 years**



Study	Expt n/N	Ctrl n/N	Relative Risk (95%CI Fixed)	Weight %	RR (95%CI Fixed)
Meunier	69 / 208	37 / 146		42.6	1.31 [0.93,1.84]
Pak	6 / 54	16 / 56		15.4	0.39 [0.16,0.92]
Riggs 1990	33 / 101	42 / 101		41.1	0.79 [0.55,1.13]
Sebert	2 / 35	1 / 41		0.9	2.34 [0.22,24.76]
Total (95%CI)	110 / 398	96 / 344		100.0	0.96 [0.76,1.21]
Chi-square 9.17 (df=3) Z=0.33					

Random Effects

Comparison: **Fluoride vs Placebo - Overall**
Outcome: **No. People with new vertebral fractures - 2 years**

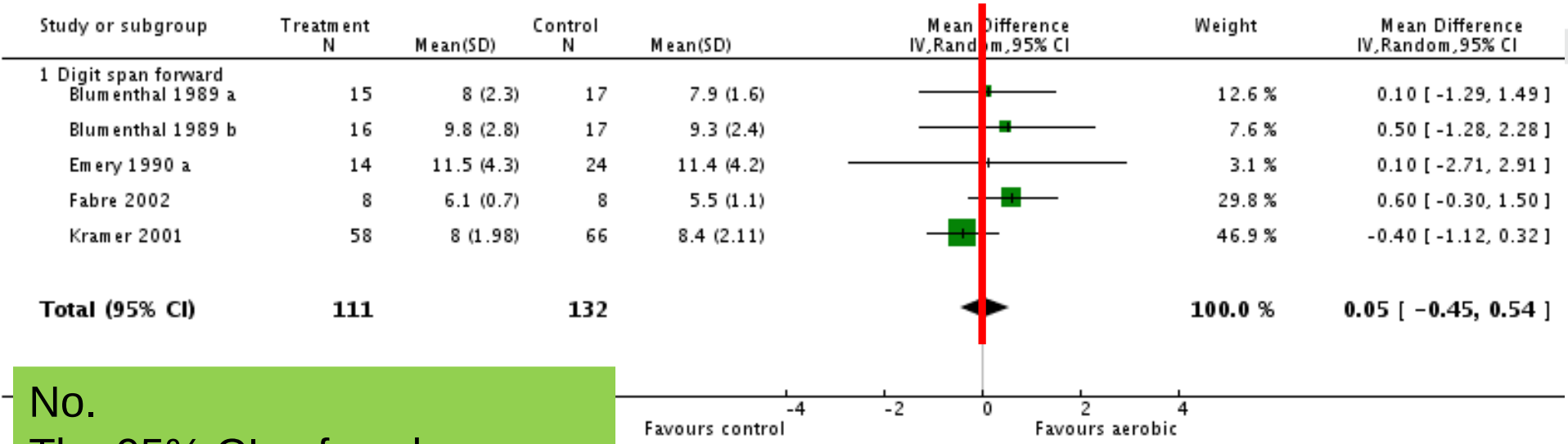


Study	Expt n/N	Ctrl n/N	Relative Risk (95%CI Random)	Weight %	RR (95%CI Random)
Meunier	69 / 208	37 / 146		38.1	1.31 [0.93,1.84]
Pak	6 / 54	16 / 56		20.3	0.39 [0.16,0.92]
Riggs 1990	33 / 101	42 / 101		37.2	0.79 [0.55,1.13]
Sebert	2 / 35	1 / 41		4.4	2.34 [0.22,24.76]
Total (95%CI)	110 / 398	96 / 344		100.0	0.87 [0.51,1.46]
Chi-square 9.17 (df=3) Z=0.53					

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

Study or subgroup	Treatment N	Mean(SD)	Control N	Mean(SD)	Weight	Mean Difference IV,Random,95% CI
1 Digit span forward						
Blumenthal 1989 a	15	8 (2.3)	17	7.9 (1.6)	12.6 %	0.10 [-1.29, 1.49]
Blumenthal 1989 b	16	9.8 (2.8)	17	9.3 (2.4)	7.6 %	0.50 [-1.28, 2.28]
Emery 1990 a	14	11.5 (4.3)	24	11.4 (4.2)	3.1 %	0.10 [-2.71, 2.91]
Fabre 2002	8	6.1 (0.7)	8	5.5 (1.1)	29.8 %	0.60 [-0.30, 1.50]
Kramer 2001	58	8 (1.98)	66	8.4 (2.11)	46.9 %	-0.40 [-1.12, 0.32]
Total (95% CI)	**111**		**132**		**100.0 %**	**0.05 [-0.45, 0.54]**

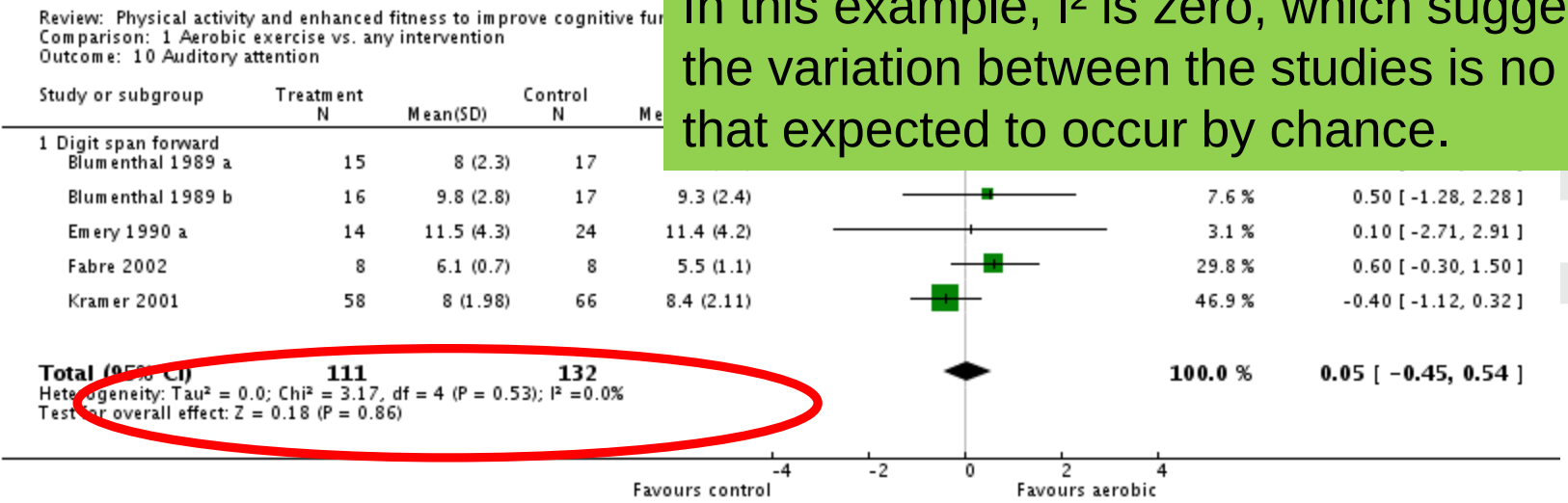


No.
The 95% CIs of each individual study overlap

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.

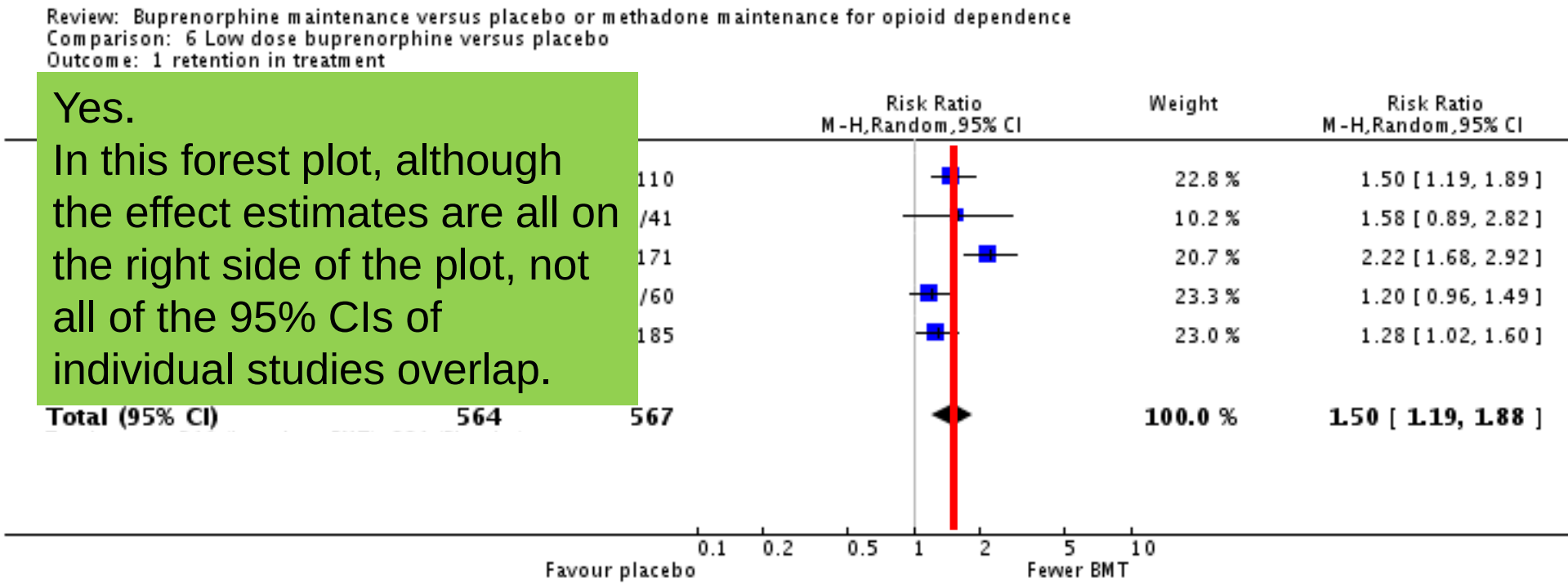
Do the statistics show heterogeneity?

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?



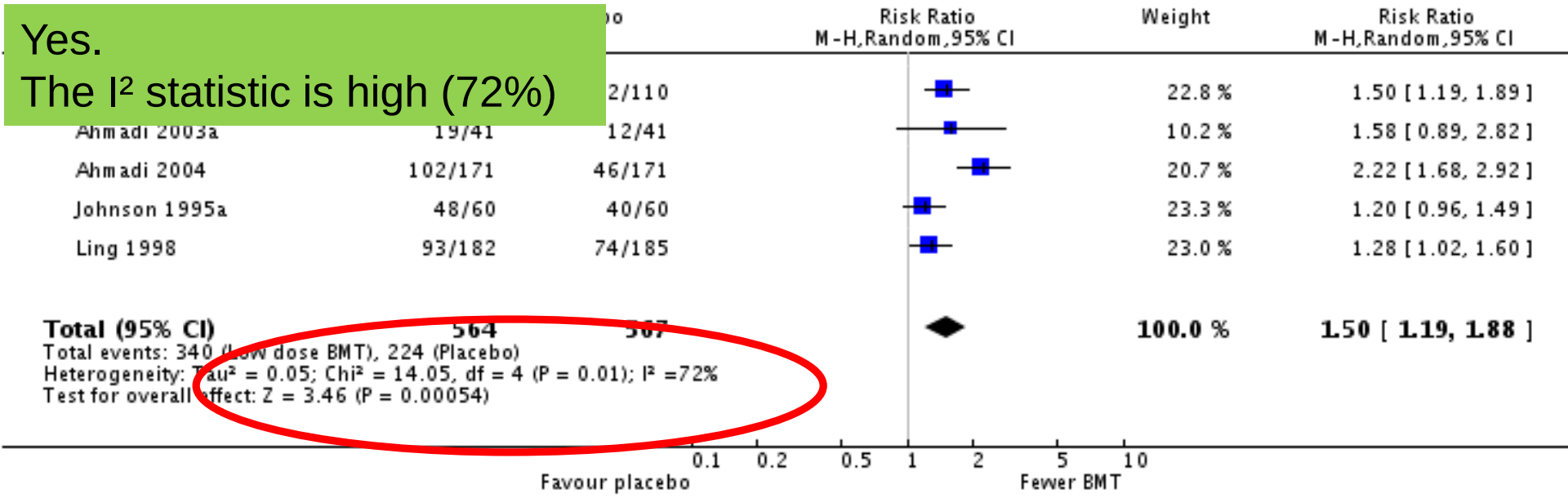
Yes.
In this forest plot, although the effect estimates are all on the right side of the plot, not all of the 95% CIs of individual studies overlap.

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,

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

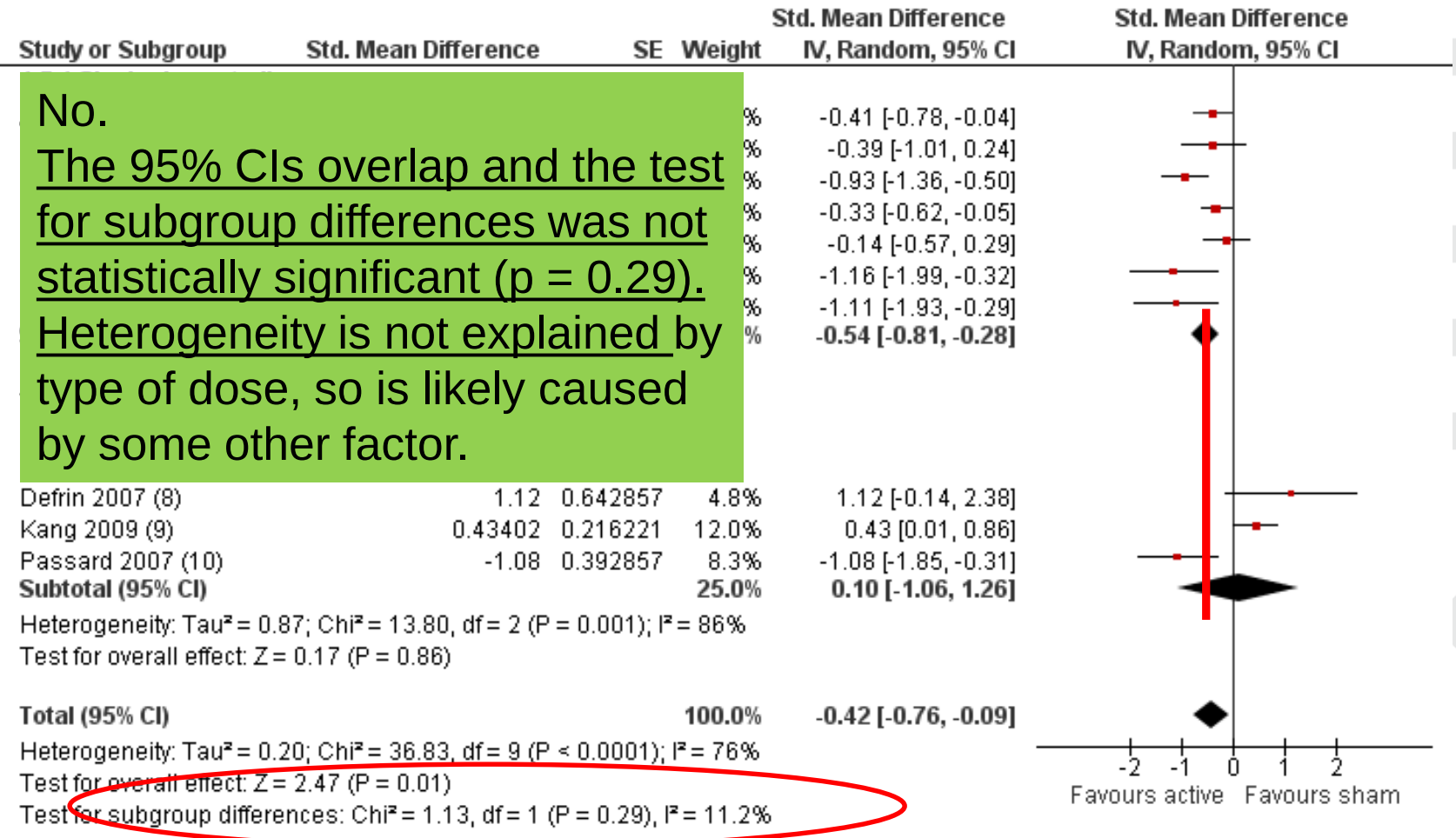
Yes.
The I^2 statistic is high (72%)



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,

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.

اگر میل داشتید Email بزنید !

fnajafi@kums.ac.ir
Farid_n32@yahoo.com