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)

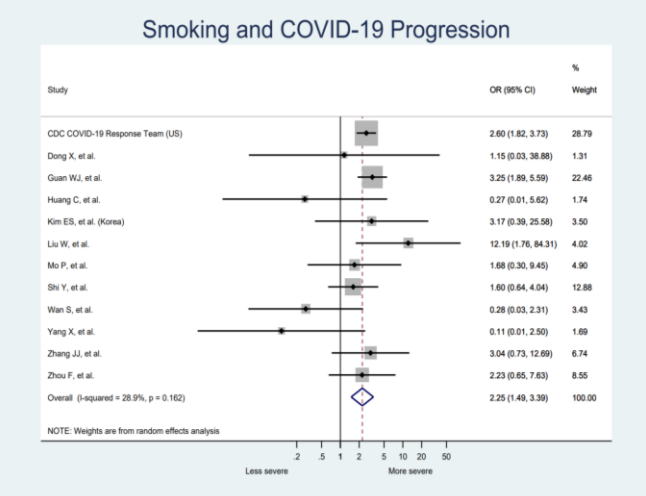
Systematic Review & Meta-analyses

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

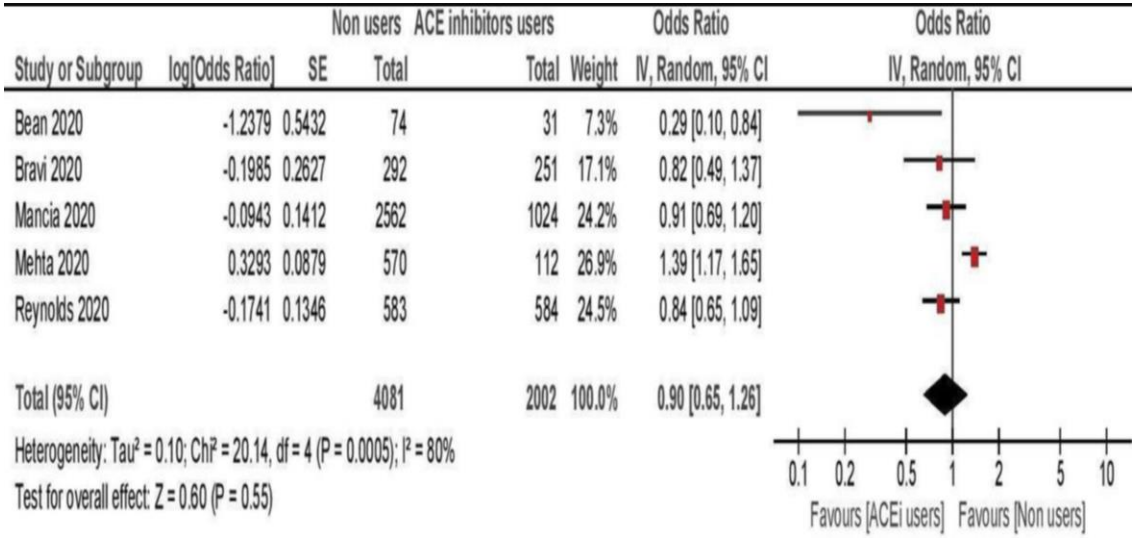
- If there is considerable variation in results, it may be misleading to quote an average value

What is heterogeneity?

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



Patanavanich R, Glantz SA, 2020



Risk of severe/lethal COVID-19 among ACE inhibitors users versus non-users Flacco ME et al. 2020

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:

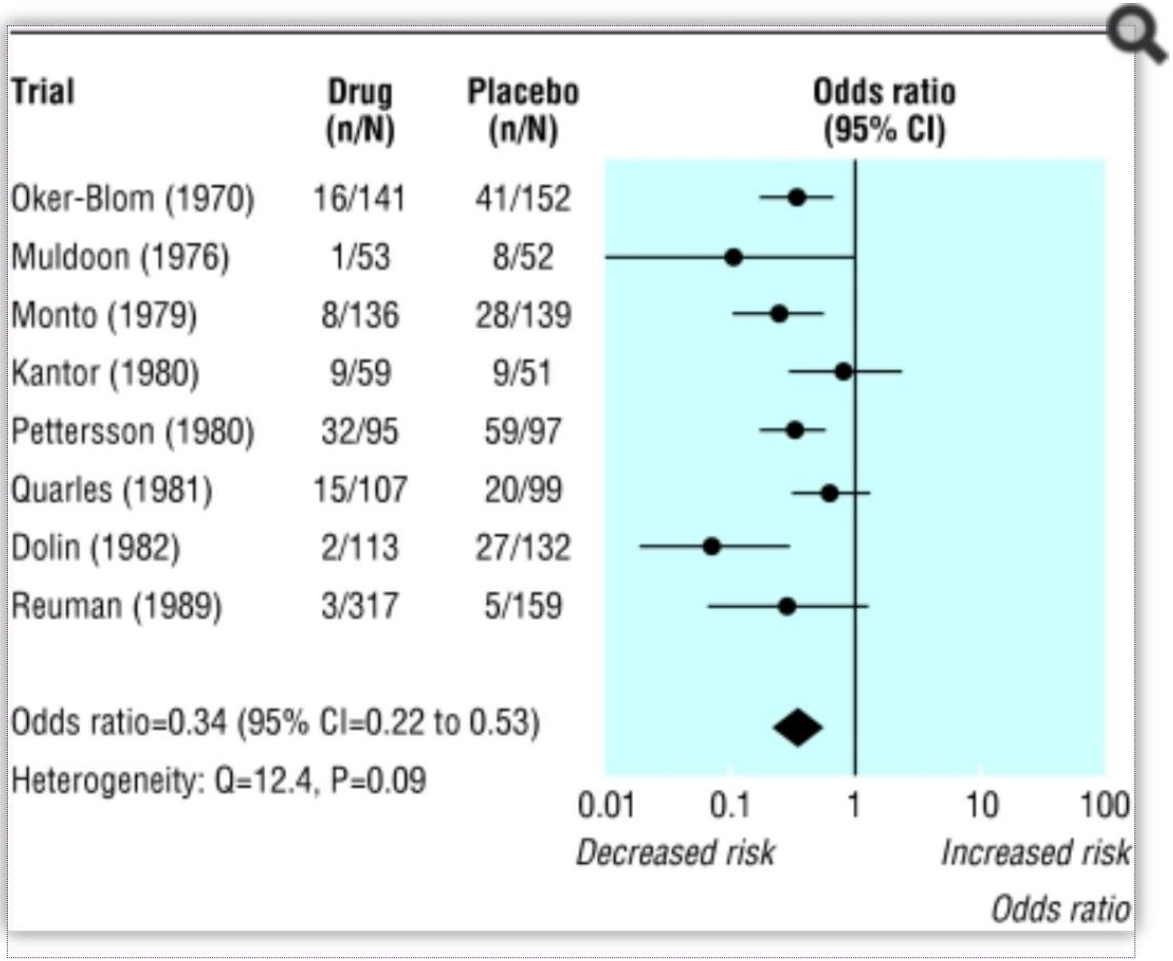
$$Q = \sum[weight_i \times (lnOR_{mh} - lnOR_i)^2]$$

  - If $p < 0.05$, then there is significant heterogeneity.
  - Degrees of freedom: total number of studies-1

Statistical Tests of Homogeneity (heterogeneity)

- Power of such statistical tests is low

  (a non-significant test does not rule out clinically important heterogeneity)
- We might increase the level of significance to 10%



Eight trials of amantadine for prevention of influenza.[11] Outcome is cases of influenza. Summary odds ratios calculated with random effects method

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

$$T^2 = \frac{Q - df}{C}$$

$$C = \sum W_i - (\sum W_i^2 / \sum W_i)$$

T is the standard deviation of true effect size

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

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

The percentage of observed variability in estimated effects which is due to heterogeneity

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’

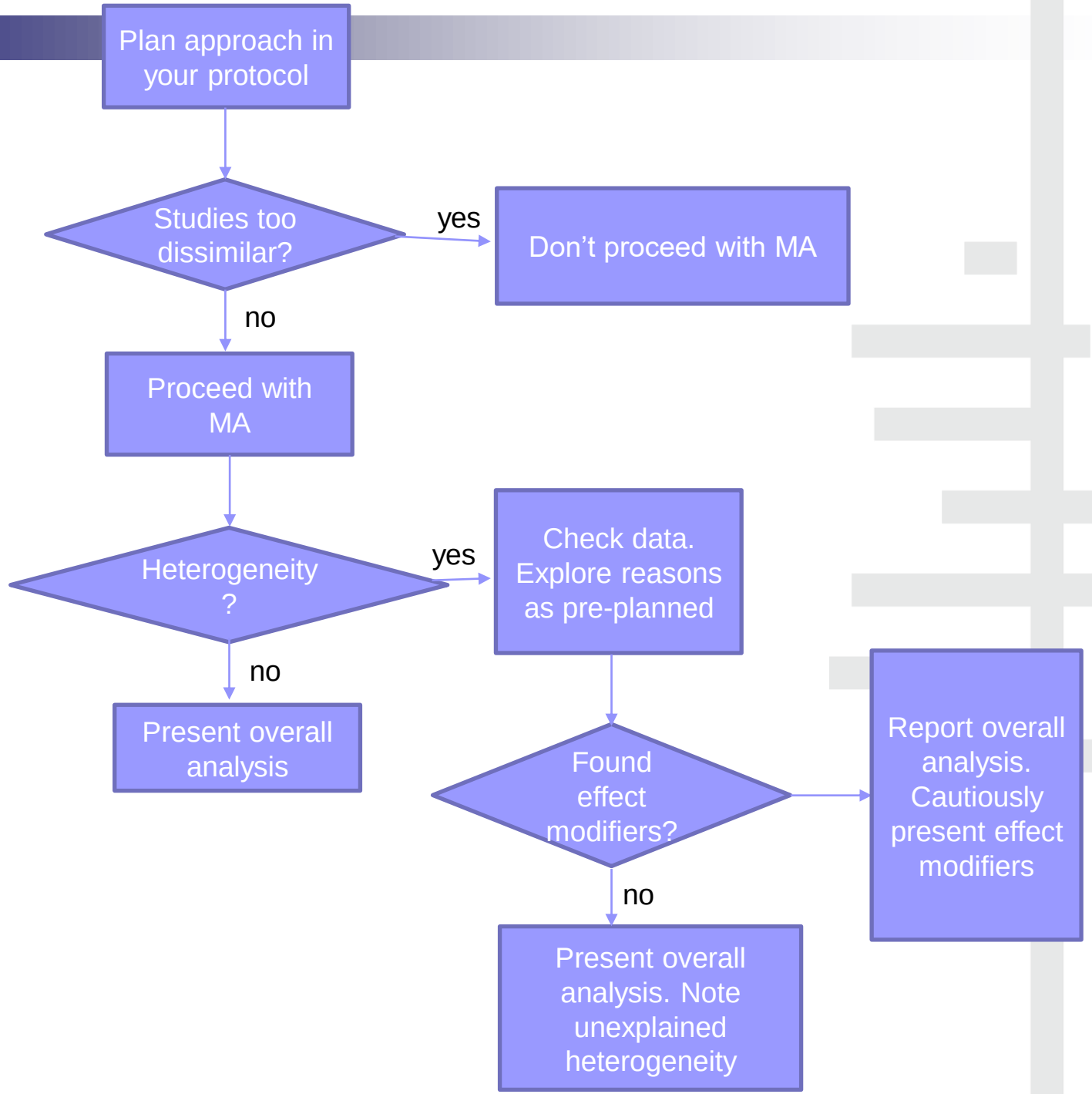
3. Explore heterogeneity through:

   Subgroup analysis

   Meta-regression

4. If Heterogeneity still persist:

   Use Random Effects Model

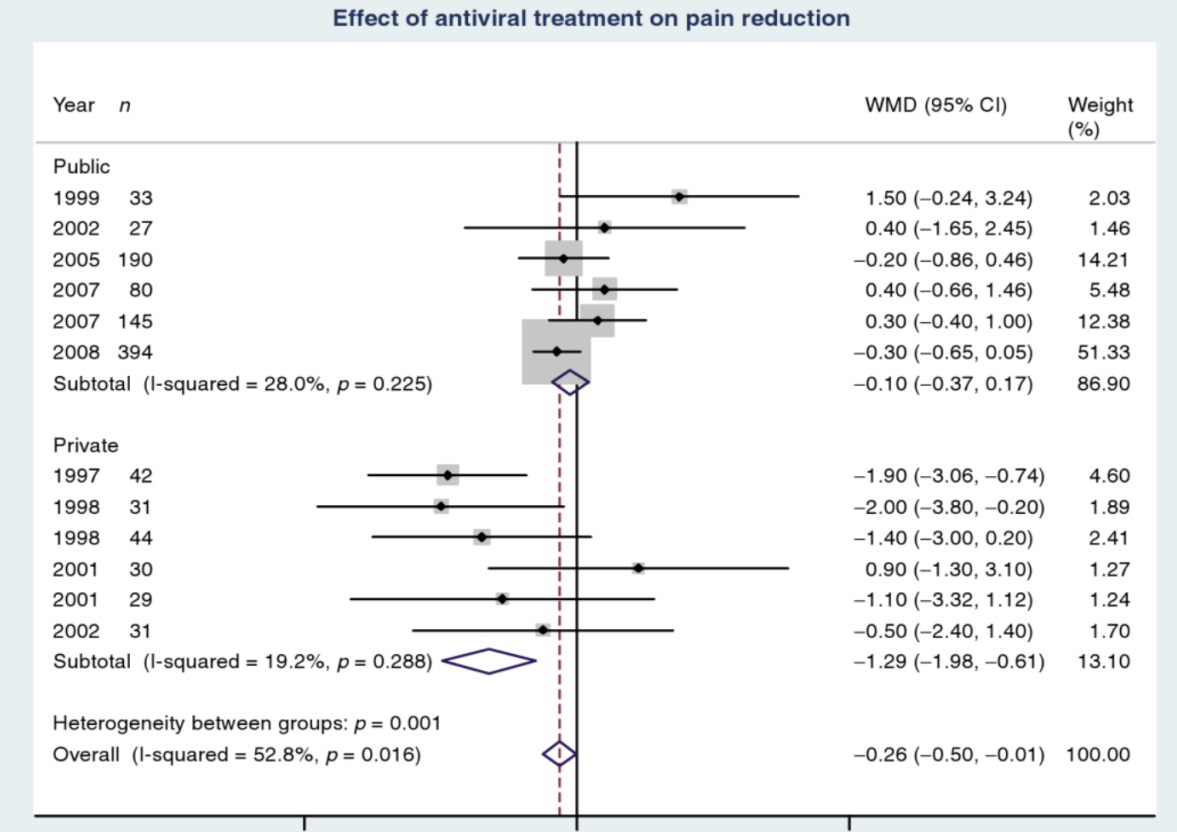

Plan approach in your protocol
Studies too dissimilar?
yes
Don’t proceed with MA
no
Proceed with MA
Heterogeneity ?
yes
Check data. Explore reasons as pre-planned
no
Present overall analysis
Found effect modifiers?
Report overall analysis. Cautiously present effect modifiers
no
Present overall analysis.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

Exploring Heterogeneity

Comparison: Subgroup: Quality of Blinding
Outcome: Lumbar BMD

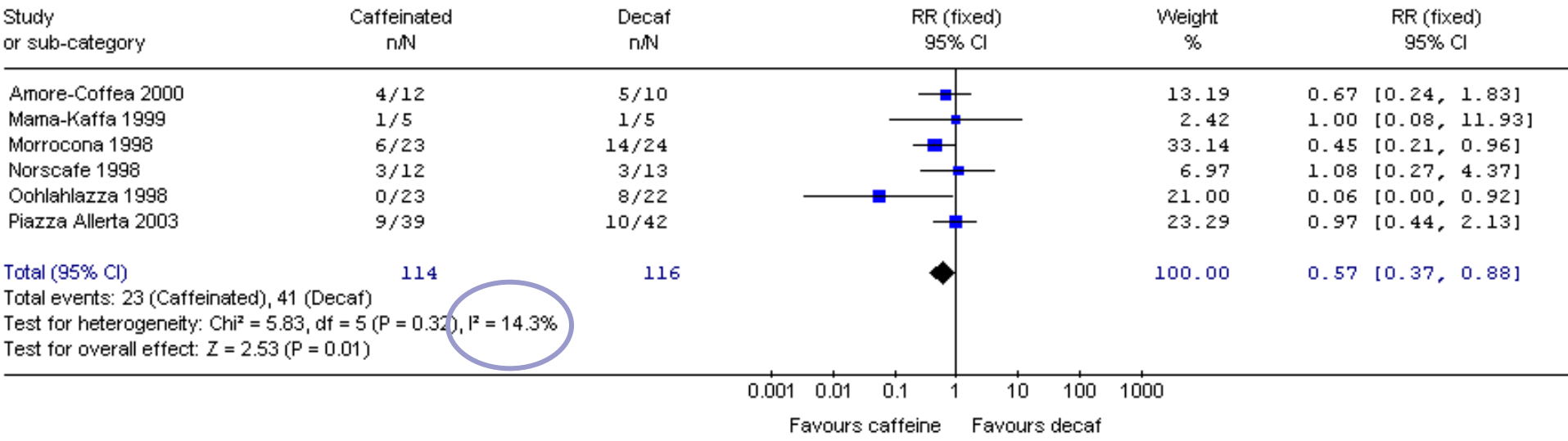
Study	Expt n	Expt mean(sd)	Ctrl n	Ctrl mean(sd)	WMD (95%CI Fixed)	Weight %	WMD (95%CI Fixed)
Blinding = 0							
Evans 1993	15	2.40 (9.10)	11	-4.70 (4.40)		1.7	7.100 [1.811,12.389]
Gurlek 1997	10	4.54 (17.96)	10	0.14 (3.42)		0.4	4.400 [-6.932,15.732]
Montessori 1997	40	6.28 (5.02)	34	-0.03 (9.20)		3.9	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							

Exploring Heterogeneity

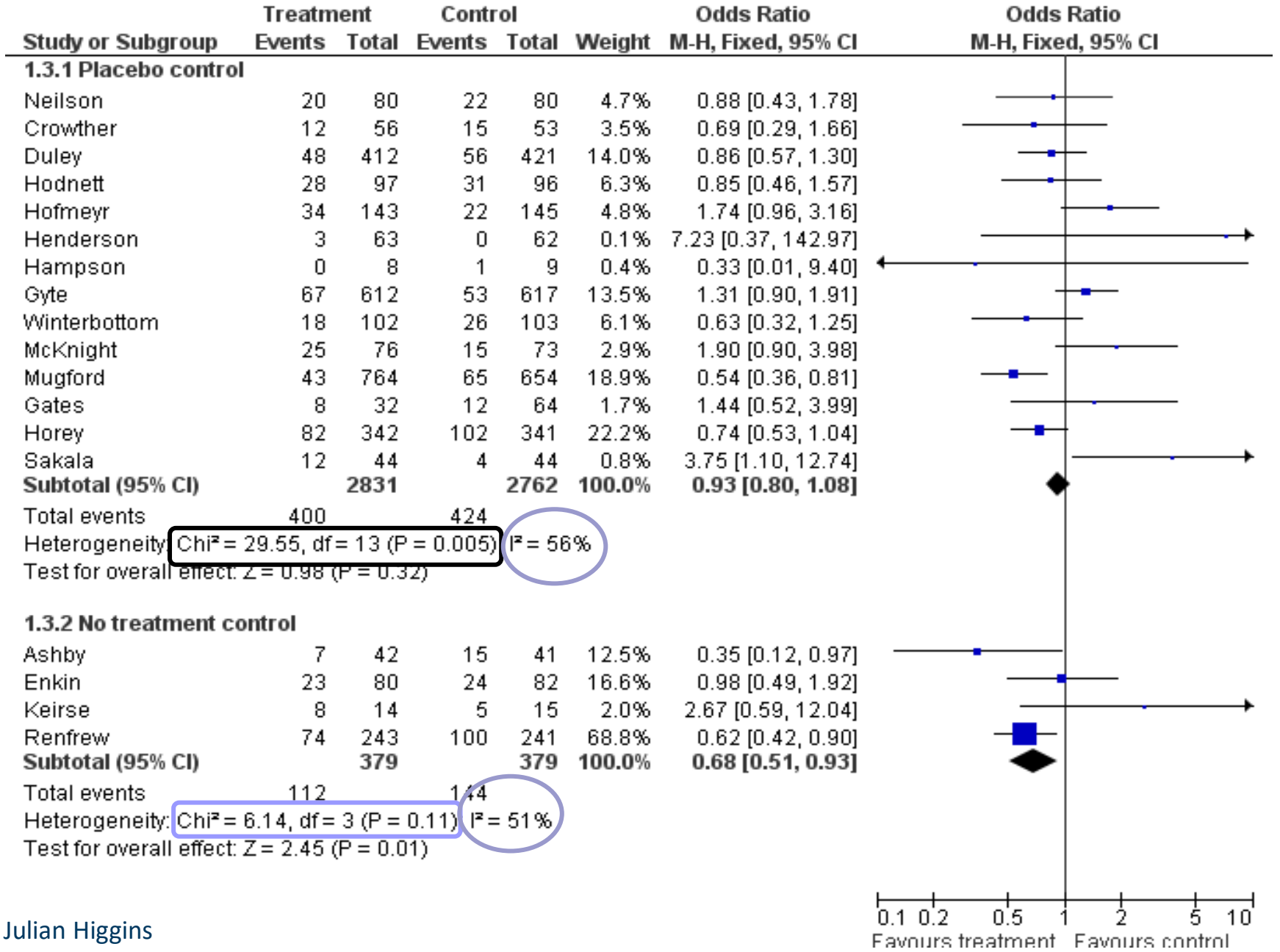


The I^2 statistic

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



Effect on Vit K on bleeding

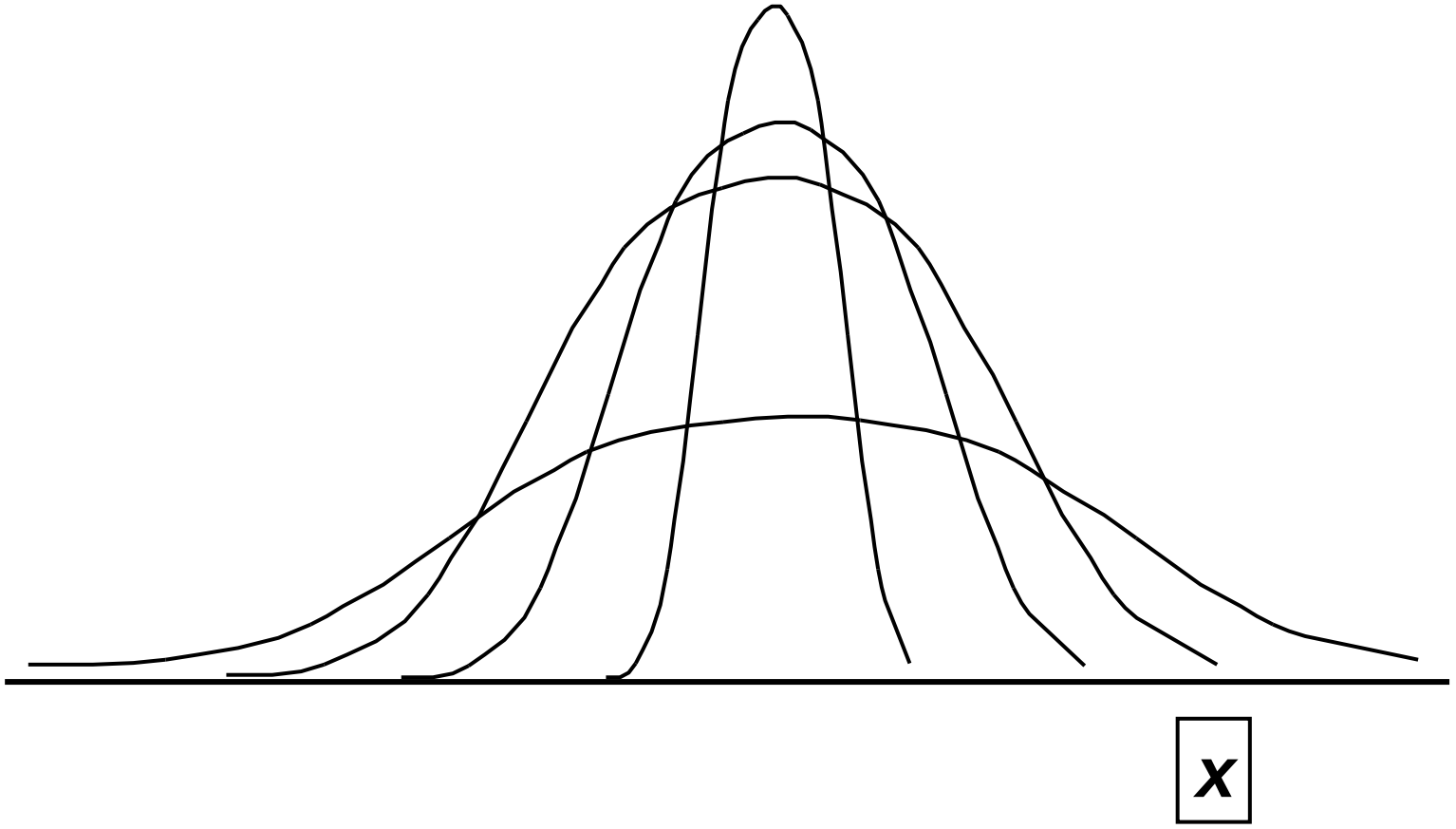


Source: Julian Higgins

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



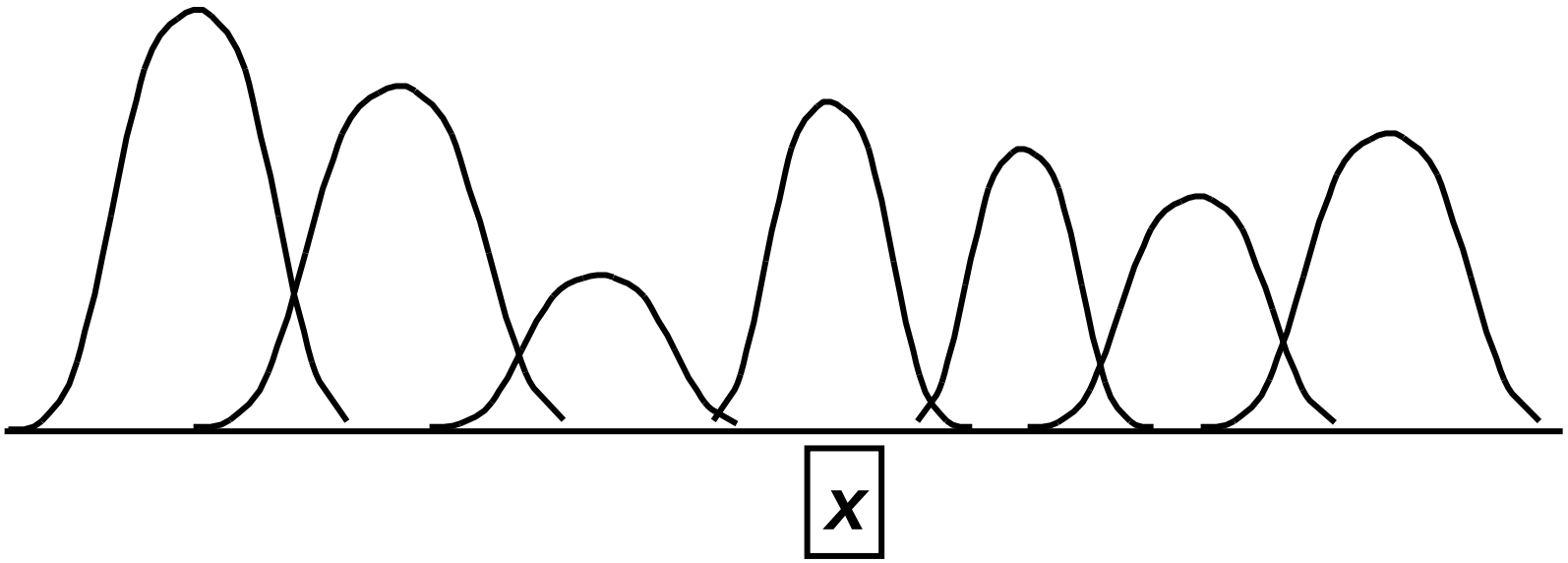
Fixed Effects Model

- 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.

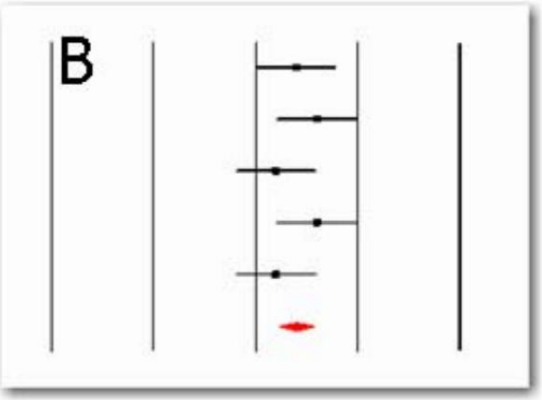
Random-Effects Model



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 = $\frac{1}{\text{Variance + heterogeneity}}$

Between-studies variance is low
because total variance is low



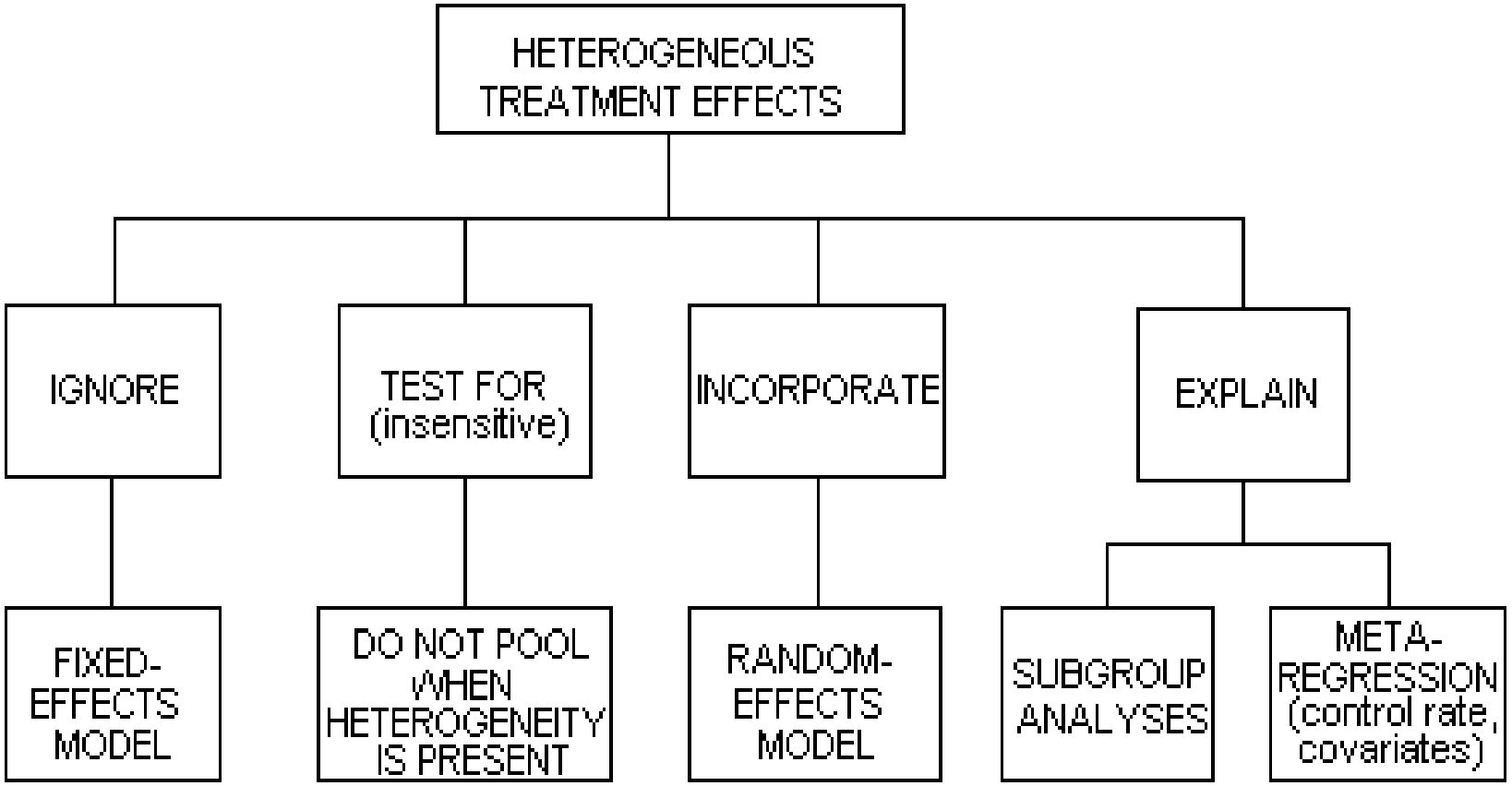
Between-studies variance is low
because within-studies variance is high

Between-studies variance is high
because total variance is high
And within-studies variance is low

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

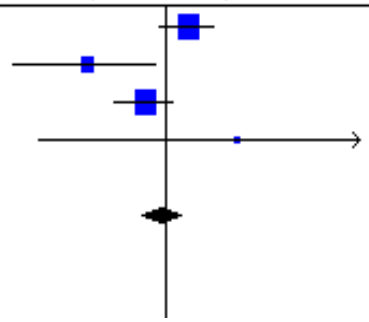


Fixed vs. Random Effects: Discrete Data

Fixed Effects

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

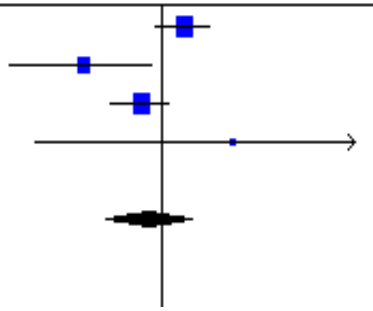
Study	Expt n/N	Ctrl n/N	Relative Risk (95%CI Fixed)	Weight %	RR (95%CI Fixed)
Meunier	69 / 208	37 / 146		42.6	1.31 [0.93,1.84]
Pak	6 / 54	16 / 56		15.4	0.39 [0.16,0.92]
Riggs 1990	33 / 101	42 / 101		41.1	0.79 [0.55,1.13]
Sebert	2 / 35	1 / 41		0.9	2.34 [0.22,24.76]
Total (95%CI)	110 / 398	96 / 344		100.0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



Random Effects

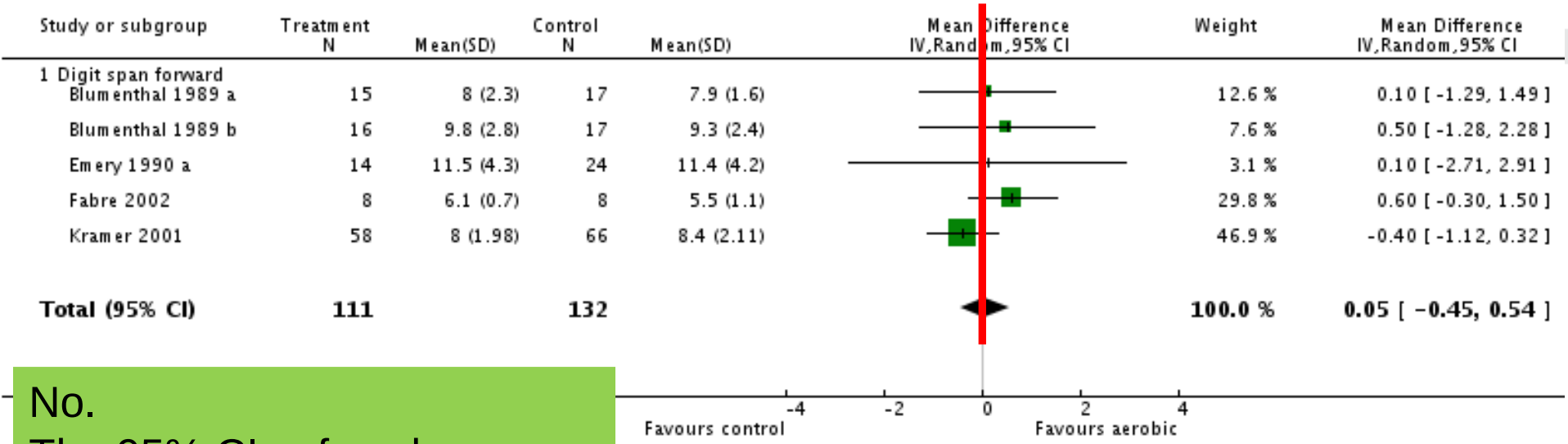
Comparison: **Fluoride vs Placebo - Overall**
Outcome: **No. People with new vertebral fractures - 2 years**

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?

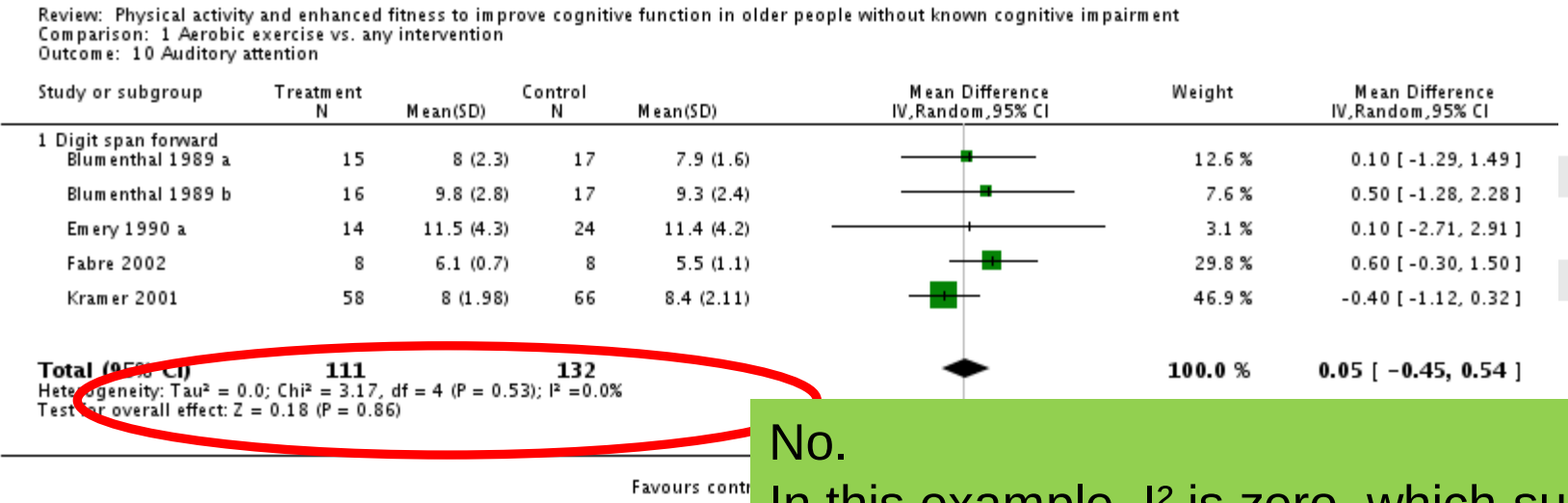
Review: Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment
Comparison: 1 Aerobic exercise vs. any intervention
Outcome: 10 Auditory attention



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?



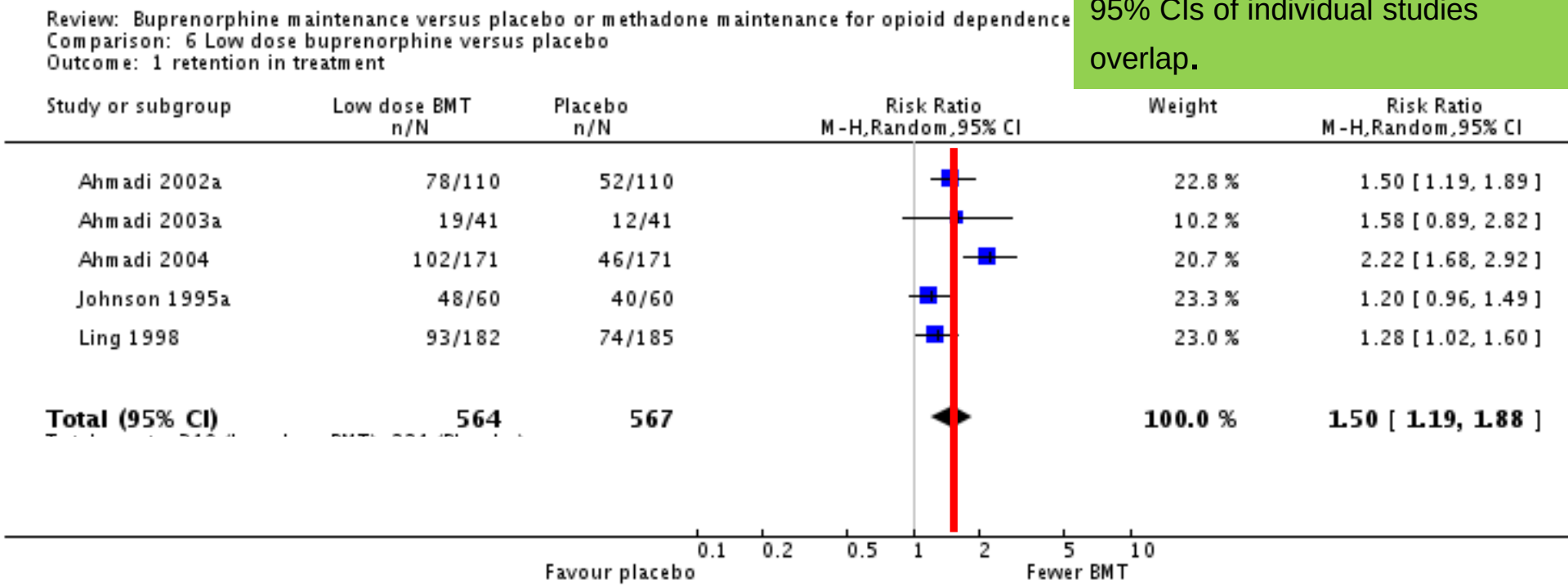
No.

In this example, I^2 is zero, which suggests that the variation between the studies is no more tha 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?

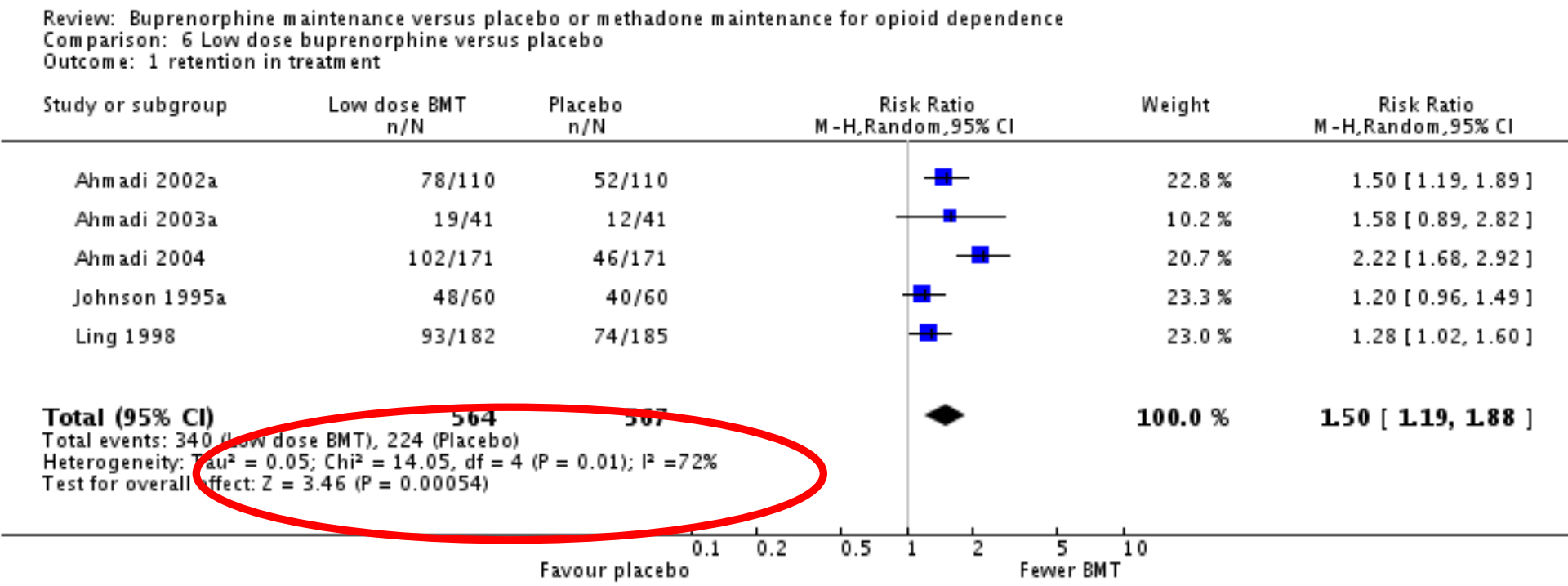
Yes.
In this forest plot, although the effect estimates are all on the right side of the plot, not all of the 95% CIs of individual studies overlap.



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?

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

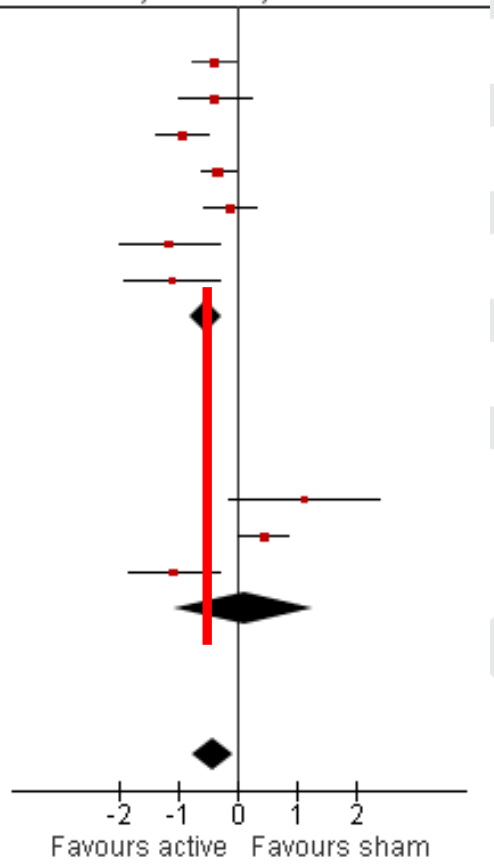


Source: Mattick RP, Kimber J, Breen C,Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database of Systematic Reviews* 2008,

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.

Study or Subgroup	Std. Mean Difference	SE	Weight	Std. Mean Difference IV, Random, 95% CI
1.5.1 Single-dose studies				
André-Obadia 2008 (1)	-0.41092	0.190803	12.5%	-0.41 [-0.78, -0.04]
Hirayama 2006 (2)	-0.38726	0.317881	9.8%	-0.39 [-1.01, 0.24]
Lefaucheur 2001a (3)	-0.9332	0.219521	11.9%	-0.93 [-1.36, -0.50]
Lefaucheur 2008 (4)	-0.334132	0.143793	13.4%	-0.33 [-0.62, -0.05]
Pleger 2004 (5)	-0.138771	0.217836	11.9%	-0.14 [-0.57, 0.29]
Saitoh 2007 (6)	-1.158204	0.42585	7.7%	-1.16 [-1.99, -0.32]
Saitoh 2007 (7)	-1.110603	0.418912	7.8%	-1.11 [-1.93, -0.29]
Subtotal (95% CI)			**75.0%**	**-0.54 [-0.81, -0.28]**
Heterogeneity: Tau² = 0.06; Chi² = 12.76, df = 6 (P = 0.05); I² = 53%				
Test for overall effect: Z = 4.02 (P < 0.0001)				
1.5.2 Multiple-dose studies				
Defrin 2007 (8)	1.12	0.642857	4.8%	1.12 [-0.14, 2.38]
Kang 2009 (9)	0.43402	0.216221	12.0%	0.43 [0.01, 0.86]
Passard 2007 (10)	-1.08	0.392857	8.3%	-1.08 [-1.85, -0.31]
Subtotal (95% CI)			**25.0%**	**0.10 [-1.06, 1.26]**
Heterogeneity: Tau² = 0.87; Chi² = 13.80, df = 2 (P = 0.001); I² = 86%				
Test for overall effect: Z = 0.17 (P = 0.86)				
Total (95% CI)			**100.0%**	**-0.42 [-0.76, -0.09]**
Heterogeneity: Tau² = 0.20; Chi² = 36.83, df = 9 (P < 0.0001); I² = 76%				
Test for overall effect: Z = 2.47 (P = 0.01)				
Test for subgroup differences: Chi² = 1.13, df = 1 (P = 0.29), I² = 11.2%				



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