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)

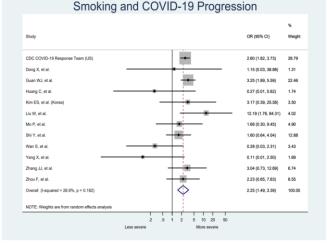
Systematic Review & Meta-analyses

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.



			Non users	ACE inhibitors users		Odds Ratio	Odds Ratio
Study or Subgroup	log[Odds Ratio]	SE	Total	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI
Bean 2020	-1.2379	0.5432	74	31	7.3%	0.29 [0.10, 0.84]	
Bravi 2020	-0.1985	0.2627	292	251	17.1%	0.82 [0.49, 1.37]	
Mancia 2020	-0.0943	0.1412	2562	1024	24.2%	0.91 [0.69, 1.20]	-
Mehta 2020	0.3293	0.0879	570	112	26.9%	1.39 [1.17, 1.65]	+
Reynolds 2020	-0.1741	0.1346	583	584	24.5%	0.84 [0.65, 1.09]	*
Total (95% CI)			4081	2002	100.0%	0.90 [0.65, 1.26]	•
Heterogeneity: Tau ² =	0.10; Ch ² = 20.14,	df = 4 (P	= 0.0005); 2	= 80%			
Test for overall effect:	Z = 0.60 (P = 0.55)						0.1 0.2 0.5 1 2 5 10 Favours [ACEi users] Favours [Non users]

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)

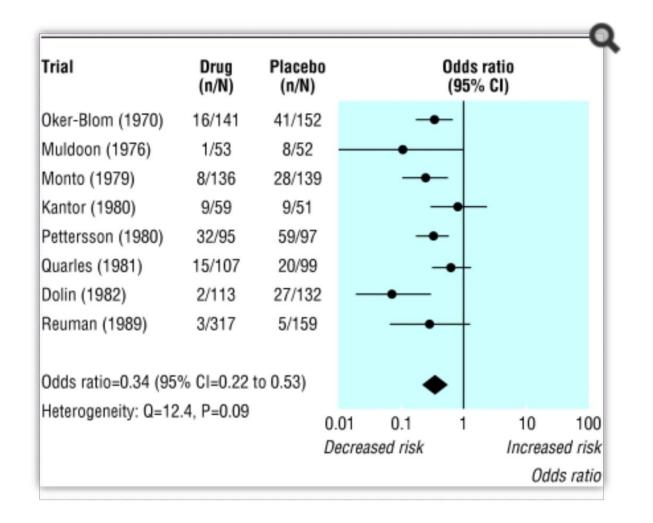
- Homogeneity calculations
 - \Box H_o = studies are homogeneous
 - Based on testing the sum of weighted differences between the summary effect and individual effects
 - Calculate Mantel Haenszel Q, where:

 $Q = \sum [weight_i x (InOR_{mh} - InOR_i)^2]$

- \Box If p< 0.05, then there is significant heterogeneity.
- ⁸ 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

$Tau^2\left(t^2\right)$

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

$$T^2 = \frac{Q - df}{C} \qquad \qquad C = \sum W_i - \left(\sum W_i^2 / \sum W_i\right)$$

T is the standard deviation of true effect size

I^2

- I² 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:

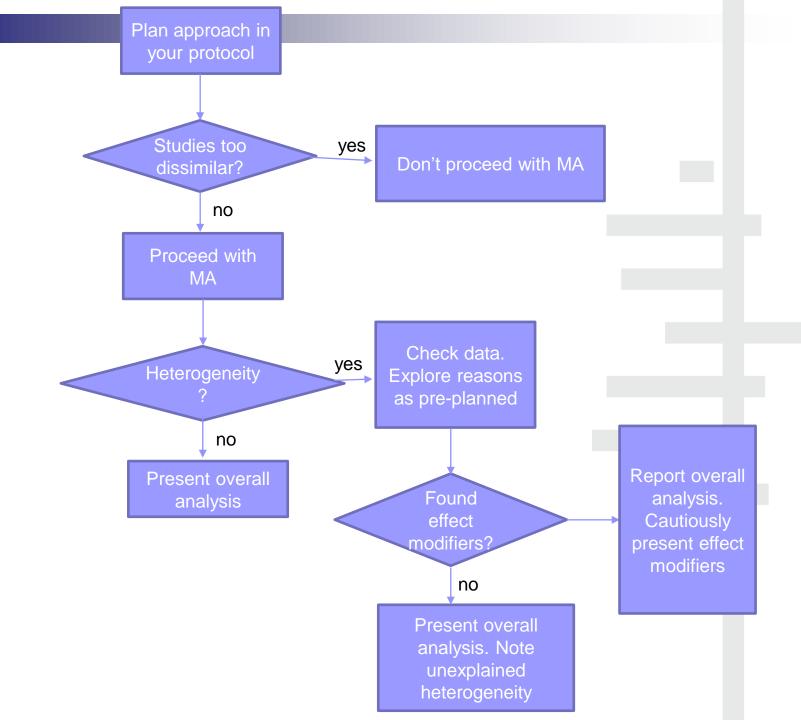
Do not 'pool at all'

3. Explore heterogeneity through:

Subgroup analysis

Meta-regression

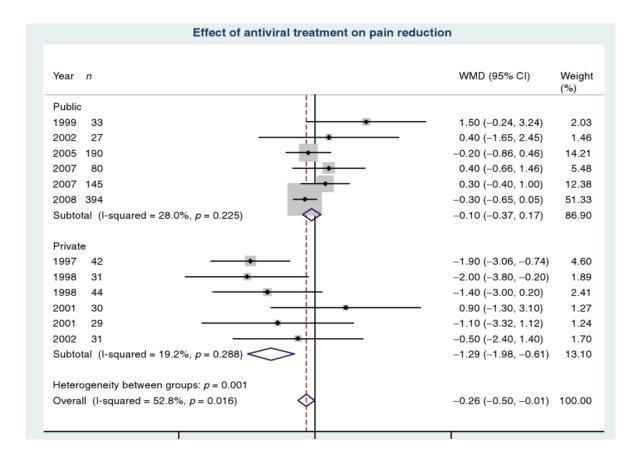
4. If Heterogeneity still persist: Use Random Effects Model



Exploring Heterogeneity

-		ality of Blinding					
Outcome: L	umbar BMD	— .	-				
	Expt	Expt	Ctrl	Ctrl	WMD (050) OLEWISH	Weight	WMD (050(CLEined))
Study	Π	mean(sd)	Π	mean(sd)	(95%Cl Fixed)	%	(95%Cl Fixed)
Blinding = 0							
Evans 1993	15	2.40 (9.10)	11	-4.70 (4.40)	$ \longrightarrow$	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	5 14	4.22 (3.93)	14	-2.25 (3.55)	- _	6.0	6.470 [3.696,9.244]
Wimalawansa 98	3 16	4.30 (2.80)	16	-0.90 (2.40)		14.1	5.200 [3.393,7.007]
Subtotal (95%Cl)	95		85			26.0	5.767 [4.435,7.100]
Chi-square 1.02 (d	lf=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%Cl)	339		337			74.0	3.579 [2.789,4.370]
Chi-square 7.52 (d	lf=5) Z=8.88						
Total (95%Cl)	434		422		•	100.0	4.148 [3.469,4.828]
Chi-square 16.20 ((df=10) Z=11.9	96					

Exploring Heterogeneity



The I² statistic

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

RR (fixed) Caffeinated Study Decaf Weight RR (fixed) or sub-category n/N n/N 95% CI % 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 0.45 [0.21, 0.96] 6/23 14/24 33.14 Norscafe 1998 3/12 3/13 6.97 1.08 [0.27, 4.37] Oohlahlazza 1998 0.06 [0.00, 0.92] 0/23 8/22 21.00 Piazza Allerta 2003 23.29 0.97 [0.44, 2.13] 9/39 10/42 Total (95% CI) 0.57 [0.37, 0.88] 114 116 100.00 Total events: 23 (Caffeinated), 41 (Decaf) Test for heterogeneity: $Chi^2 = 5.83$, df = 5 (P = 0.32), $l^2 = 14.3\%$ Test for overall effect: Z = 2.53 (P = 0.01) 0.001 0.01 0.1 10 100 1000 1

Favours caffeine Favours decaf

Effect on Vit K on bleeding

	Treatm	nent	Contr	ol		Odds Ratio	Odds Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	M-H, Fixed, 95% Cl	M-H, Fixed, 95% Cl
1.3.1 Placebo control							
Neilson	20	80	22	80	4.7%	0.88 [0.43, 1.78]	
Crowther	12	56	15	53	3.5%	0.69 [0.29, 1.66]	
Duley	48	412	56	421	14.0%	0.86 [0.57, 1.30]	
Hodnett	28	97	31	96	6.3%	0.85 [0.46, 1.57]	
Hofmeyr	34	143	22	145	4.8%	1.74 [0.96, 3.16]	
Henderson	3	63	0	62	0.1%	7.23 [0.37, 142.97]	
Hampson	0	8	1	9	0.4%	0.33 [0.01, 9.40]	•
Gyte	67	612	53	617	13.5%	1.31 [0.90, 1.91]	+
Winterbottom	18	102	26	103	6.1%	0.63 [0.32, 1.25]	
McKnight	25	76	15	73	2.9%	1.90 [0.90, 3.98]	
Mugford	43	764	65	654	18.9%	0.54 [0.36, 0.81]	_ _
Gates	8	32	12	64	1.7%	1.44 [0.52, 3.99]	
Horey	82	342	102	341	22.2%	0.74 [0.53, 1.04]	
Sakala	12	44	4	44	0.8%	3.75 [1.10, 12.74]	
Subtotal (95% CI)		2831		2762	100.0%	0.93 [0.80, 1.08]	◆
Total events	400		424				
Heterogeneity: Chi ² = 3	29.55, df	= 13 (P	' = 0.005)	I² = 58	6%		
Test for overall effect: .	Z = 0.98 ((P = 0.3	(2)				
1.3.2 No treatment co	ntrol						
Ashby	7	42	15	41	12.5%	0.35 [0.12, 0.97]	_
Enkin	23	80	24	82	16.6%	0.98 [0.49, 1.92]	
Keirse	8	14	5	15	2.0%	2.67 [0.59, 12.04]	
Renfrew	74	243	100	241	68.8%	0.62 [0.42, 0.90]	
Subtotal (95% CI)		379		379	100.0%	0.68 [0.51, 0.93]	•
Total events	112		144				_
Heterogeneity: Chi ² = I	6.14, df=	3 (P =	0.11) I [≥] =	= 51%)			
Test for overall effect: 2		-					
		-	-				
Julian Higgins							0.1 0.2 0.5 1 2 5 10

Favours treatment Favours control

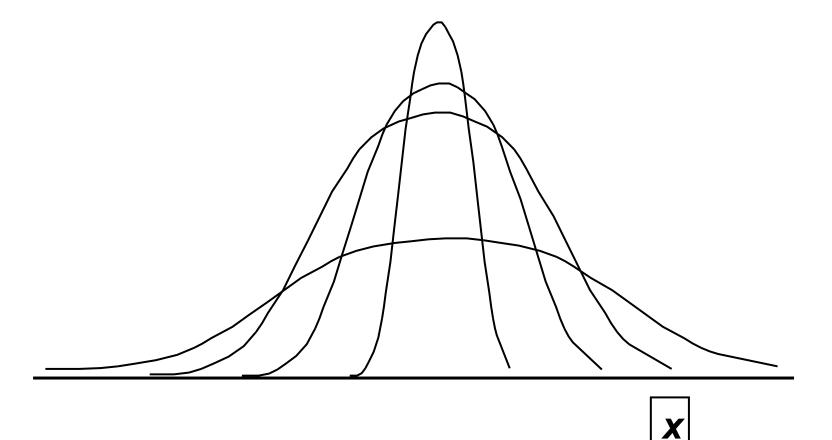
Source: Julian Higgins

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



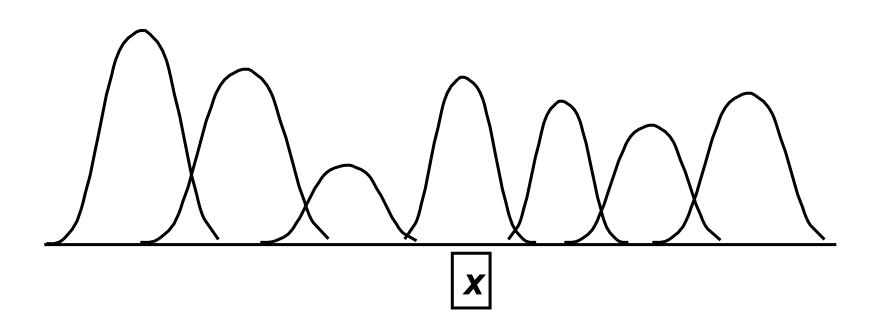
Fixed Effects Model

- Require from each study effect estimate; and standard error of effect estimate Combine these using a weighted average: sum of (estimate × weight) \Box 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.

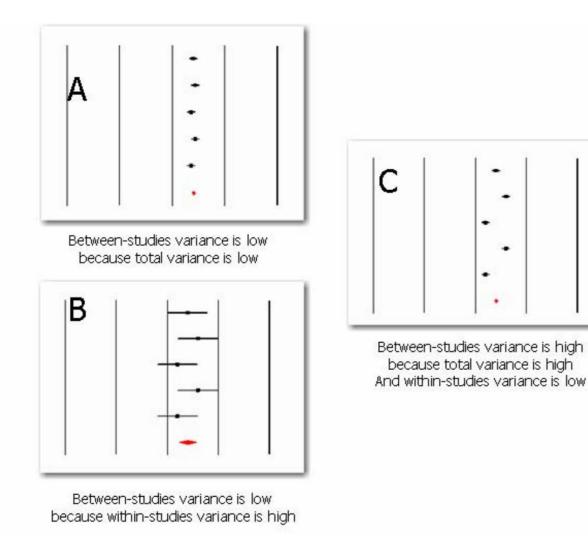
Random-Effects Model



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 =

1 Variance + heterogeneity



www.Meta-Analysis.com © 2007 Borenstein, Hedges, Rothstein | 13

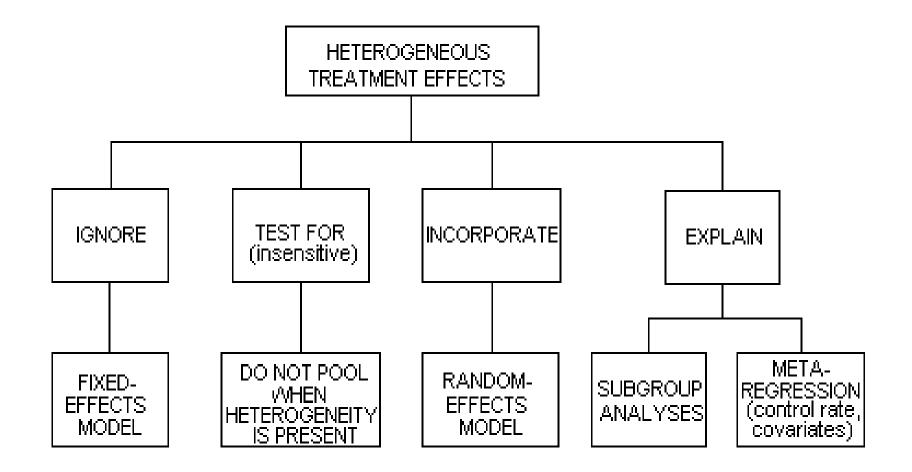
٠

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

Comparison: Fluoride vs Placebo - Overall

Outcome:	No. People with new vertebra	al fractures - 2 years			
	Expt	Ctrl	Relative Risk	Weight	RR
Study	n/N	n/N	(95%Cl Fixed)	%	(95%Cl 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%Cl) Chi-square 9.1	110 / 398 7 (df=3) Z=0.33	96 / 344	+	100.0	0.96 [0.76,1.21]

Random Effects

Comparison: Fluoride vs Placebo - Overall

Outcome:	No. People with new vertebral fractures - 2 years
----------	---

	Expt	Ctrl	Relative Risk	Weight	RR
Study	n/N	n/N	(95%Cl Random)	%	(95%Cl 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%Cl) Chi-square 9.17 (df=3) 2	110 / 398 (=0.53	96 / 344	-	100.0	0.87 [0.51,1.46]

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)	Mean Difference IV,Rand m,95% Cl	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)	<mark>-</mark>	- 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% Cl individual st			-	-4 Favours control	-2 0 2 Favours	4 aerobic	

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?

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)	Mean Difference IV,Random,95% Cl	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) Hete ogeneity: Tau ² = 0.0 Test or overall effect: Z =			132 53); l ² =0.0%		No.	100.0 %	0.05 [-0.45, 0.54]	erecto that
					In this example the variation be that expected to	etween t	he studies is	~ ~

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?

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

Review: Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence 95% CIS Comparison: 6 Low dose buprenorphine versus placebo Outcome: 1 retention in treatment OVErlap.

Study or subgroup	Low dose BMT n/N	Placebo n/N	Risk Ratio M-H,Random,95% Cl	Weight	Risk Ratio M - H, Random , 95% CI
Ahmadi 2002a	78/110	52/110		22.8 %	1.50 [1.19, 1.89]
Ahmadi 2003a	19/41	12/41		10.2 %	1.58 [0.89, 2.82]
Ahmadi 2004	102/171	46/171		20.7 %	2.22 [1.68, 2.92]
Johnson 1995a	48/60	40/60		23.3 %	1.20 [0.96, 1.49]
Ling 1998	93/182	74/185		23.0 %	1.28 [1.02, 1.60]
Total (95% CI)	564	567		100.0 %	1.50 [1.19, 1.88]
		0. Favour placebo		5 10 ver BMT	

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 I² statistic is high (72%)

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

Study or subgroup	Low dose BMT n/N	Placebo n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% Cl
Ahmadi 2002a	78/110	52/110		22.8 %	1.50 [1.19, 1.89]
Ahmadi 2003a	19/41	12/41		10.2%	1.58 [0.89, 2.82]
Ahmadi 2004	102/171	46/171		20.7 %	2.22 [1.68, 2.92]
Johnson 1995a	48/60	40/60		23.3 %	1.20 [0.96, 1.49]
Ling 1998	93/182	74/185	-	23.0 %	1.28 [1.02, 1.60]
	564 dose BMT), 224 (Placebo) 0.05; Chi ² = 14.05, df = 4 = 3.46 (P = 0.00054)			100.0 %	1.50 [1.19, 1.88]
		0.1 Favour placebo	1 0.2 0.5 1 2 5 Fewer	10 BMT	

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

				Std. Mean Difference	Std. Me other factor.	
Study or Subgroup	Std. Mean Difference	SE	Weight	IV, Random, 95% Cl	IV, Random, 95% Cl	
1.5.1 Single-dose studies	6					
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		-1.11 [-1.93, -0.29]		
Subtotal (95% CI)			75.0%	-0.54 [-0.81, -0.28]	•	
Heterogeneity: Tau ² = 0.0	6; Chi² = 12.76, df = 6 (P	? = 0.05); l ² =	= 53%			
Test for overall effect: Z =	4.02 (P < 0.0001)					
1.5.2 Multiple-dose studi						
Defrin 2007 (8)		0.642857	4.8%	1.12 [-0.14, 2.38]	↓ → →	
Kang 2009 (9)		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 = 0.8		? = 0.001); l ^a	'= 86%		•	
Test for overall effect: Z =	0.17 (P = 0.86)					
T-4-1/05// OD			400.00			
Total (95% CI)			100.0%	-0.42 [-0.76, -0.09]	🖛	
Heterogeneity: Tau ² = 0.2		? < 0.0001);	I² = 76%		-2 -1 0 1 2	
Test for everall effect: Z =	· ·				Favours active Favours sham	
Test for subgroup differer	nces: Chi² = 1.13, df = 1	(P = 0.29), l	²=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.



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