Common errors in meta-analysis

Lessons from the Cochrane Review Screening Programme

November 2017

Kerry Dwan

Trusted evidence.
Informed decisions.
Better health.

Objectives

The objectives of this workshop are to highlight common statistical errors made in Cochrane Systematic Reviews, and to provide practical, hands on learning and guidance to help authors and editors address these errors.

- Slides with examples
- Practical Exercises
- General Discussion

Poll

What are your roles in Cochrane?

- Editor
- Author
- Statistician
- Other
- No role in Cochrane yet

Cochrane

Common Errors

- Funny Looking Results
- Analyses
- Errors we may not see

FLR (Funny Looking Results)

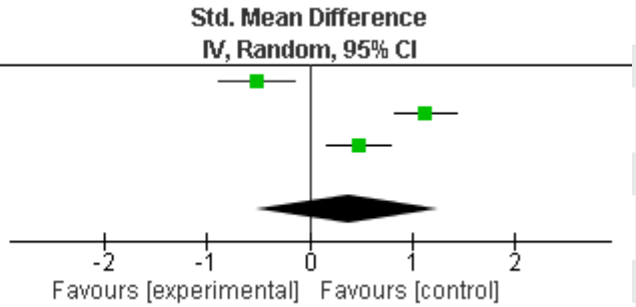
1. Data entry errors/ transposition errors
2. Study weight at odd with sample size
3. Outliers
4. Study ID appearing more than once in a forest plot
5. Reporting at odds with forest plot

FLR #1 -Data Entry Error

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Study 4	0.8	1.7	62	2	2.9	51	32.9%	-0.51 [-0.89, -0.14]
Study 5	2.5	2	92	0.5	1.5	95	33.6%	1.13 [0.82, 1.44]
Study 6	1.8	1.5	85	1	1.8	78	33.5%	0.48 [0.17, 0.79]
Total (95% CI)			**239**			**224**	**100.0%**	**0.37 [-0.52, 1.26]**

Heterogeneity: $Tau^2 = 0.59$; $Chi^2 = 43.69$, df = 2 ($P < 0.00001$); $I^2 = 95\%$

Test for overall effect: $Z = 0.82$ ($P = 0.41$)



Poll

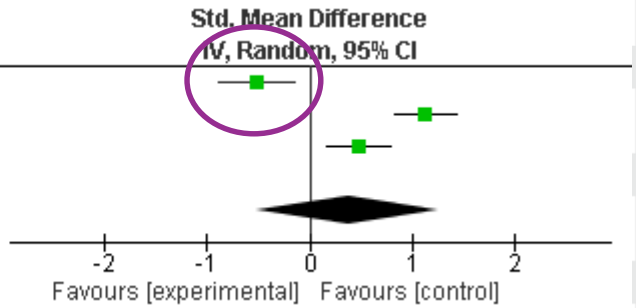
Which study do you think is probably erroneous?

- **Study 4**
- **Study 5**
- **Study 6**

FLR #1 -Data Entry Error

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Study 4	0.8	1.7	62	2	2.9	51	32.9%	-0.51 [-0.89, -0.14]
Study 5	2.5	2	92	0.5	1.5	95	33.6%	1.13 [0.82, 1.44]
Study 6	1.8	1.5	85	1	1.8	78	33.5%	0.48 [0.17, 0.79]
Total (95% CI)			**239**			**224**	**100.0%**	**0.37 [-0.52, 1.26]**

Heterogeneity: Tau² = 0.59; Chi² = 43.69, df = 2 (P < 0.00001); I² = 95%
Test for overall effect: Z = 0.82 (P = 0.41)



Study 4 Data

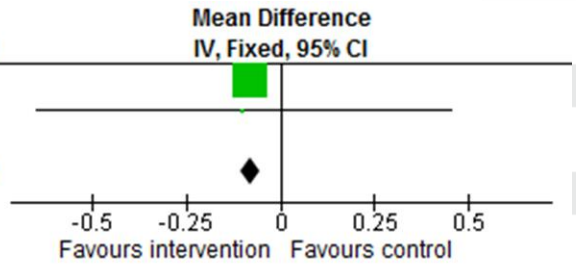
TABLE 2. Comparison of placebo, clomipramine, and haloperidol with baseline for CARS, ESRS, and DOTES[a]

Measure	Baseline Mean (SD)	Placebo Mean (SD)	Clomipramine Mean (SD)	Haloperidol Mean (SD)	*p*
CARS	41.8 (7.1)	39.4 (7.0)	37.8 (8.7)	36.7 (6.1)	0.05[b]
ESRS	6.6 (6.7)	7.9 (7.1)	10.3 (7.3)	7.8 (5.8)	0.35[c]
DOTES	0.6 (2.2)	0.8 (1.7)	2.0 (2.9)	2.3 (3.3)	0.07[d]

FLR #2 – Study weight at odds with sample size

Study or Subgroup	Intervention Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Study 1	0.6	0.001	33	0.68	0.07	34	99.8%	-0.08 [-0.10, -0.06]
Study 2	0.8	0.73	31	0.9	1.58	40	0.2%	-0.10 [-0.65, 0.45]
Total (95% CI)			**64**			**74**	**100.0%**	**-0.08 [-0.10, -0.06]**

Heterogeneity: Chi² = 0.01, df = 1 (P = 0.94); I² = 0%
Test for overall effect: Z = 6.67 (P < 0.00001)



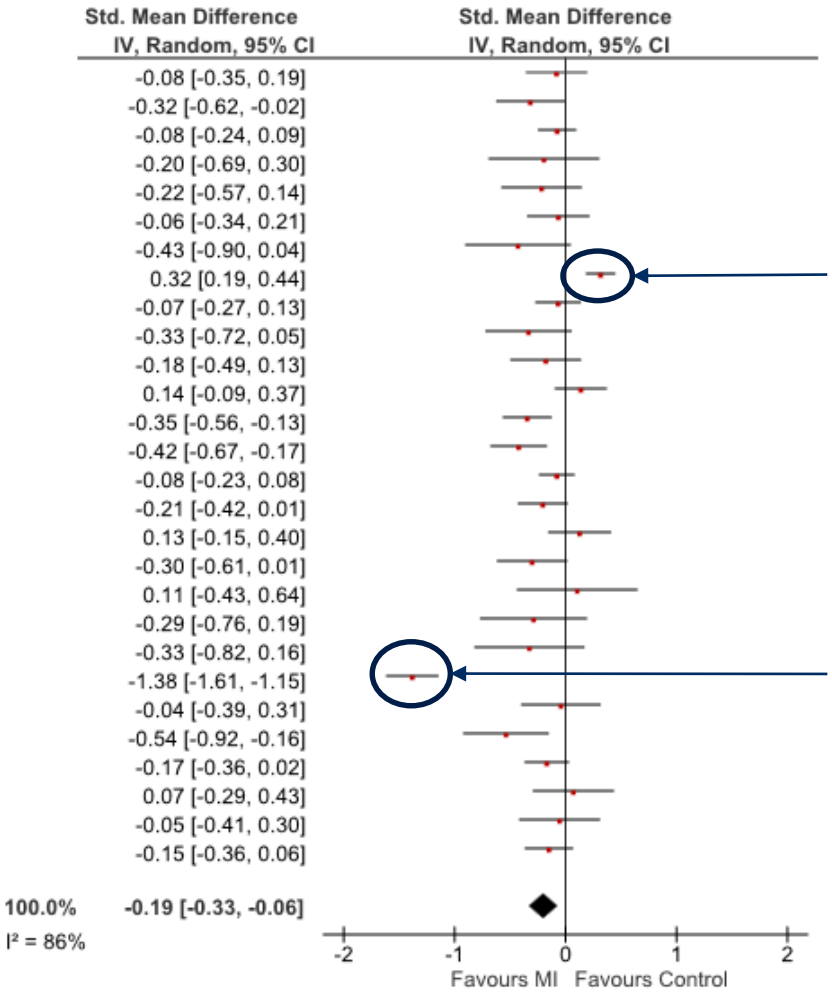
Poll

Which study do you think is probably erroneous?

- **Study 1**
- **Study 2**

Question: why? (type the answer in your question box)

FLR #3 – Outliers

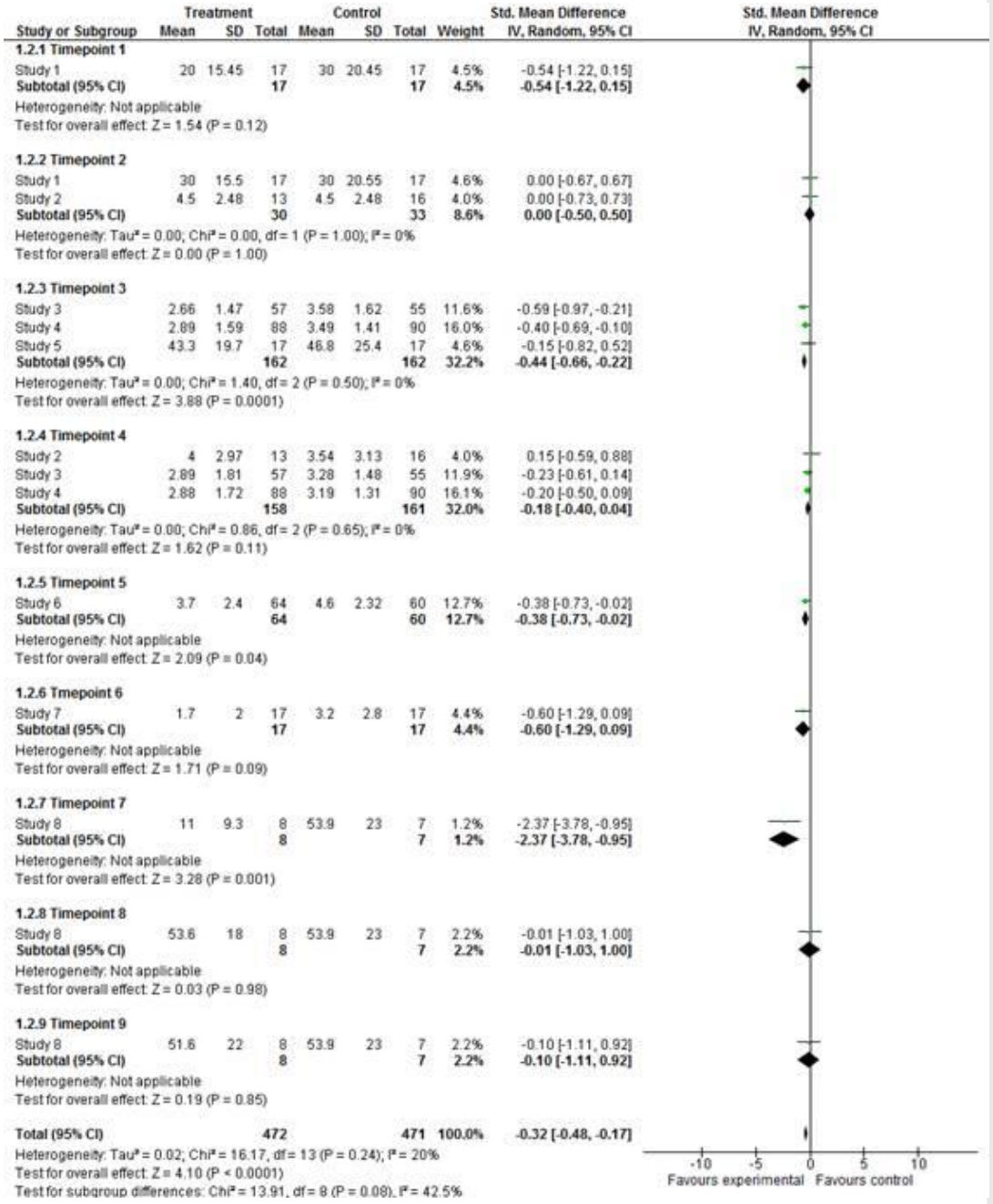


Minus sign left off mean

SEMs used instead of SDs

FLR #4 – Study ID appearing >1 in a forest plot

Question: what is the problem with this? (type the answer in your question box)

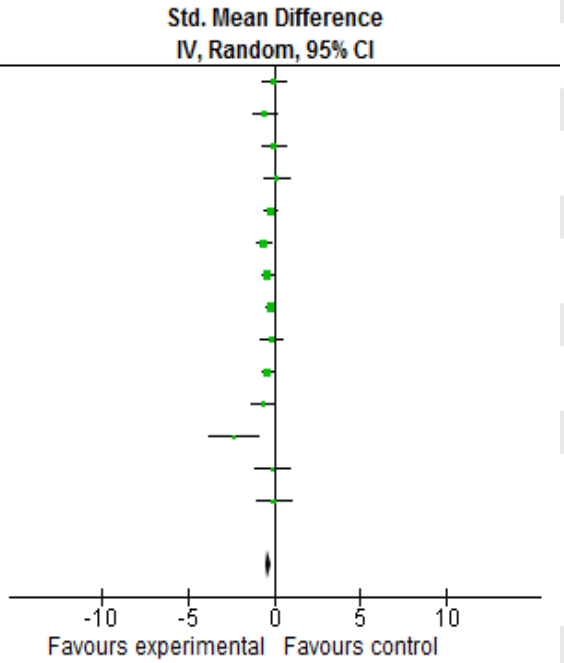


Cochrane

Study or Subgroup	Treatment Mean	Treatment SD	Treatment Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Study 1	30	15.5	17	30	20.55	17	4.6%	0.00 [-0.67, 0.67]
Study 1	20	15.45	17	30	20.45	17	4.5%	-0.54 [-1.22, 0.15]
Study 2	4.5	2.48	13	4.5	2.48	16	4.0%	0.00 [-0.73, 0.73]
Study 2	4	2.97	13	3.54	3.13	16	4.0%	0.15 [-0.59, 0.88]
Study 3	2.89	1.81	57	3.28	1.48	55	11.9%	-0.23 [-0.61, 0.14]
Study 3	2.66	1.47	57	3.58	1.62	55	11.6%	-0.59 [-0.97, -0.21]
Study 4	2.89	1.59	88	3.49	1.41	90	16.0%	-0.40 [-0.69, -0.10]
Study 4	2.88	1.72	88	3.19	1.31	90	16.1%	-0.20 [-0.50, 0.09]
Study 5	43.3	19.7	17	46.8	25.4	17	4.6%	-0.15 [-0.82, 0.52]
Study 6	3.7	2.4	64	4.6	2.32	60	12.7%	-0.38 [-0.73, -0.02]
Study 7	1.7	2	17	3.2	2.8	17	4.4%	-0.60 [-1.29, 0.09]
Study 8	11	9.3	8	53.9	23	7	1.2%	-2.37 [-3.78, -0.95]
Study 8	51.6	22	8	53.9	23	7	2.2%	-0.10 [-1.11, 0.92]
Study 8	53.6	18	8	53.9	23	7	2.2%	-0.01 [-1.03, 1.00]
Total (95% CI)			**472**			**471**	**100.0%**	**-0.32 [-0.48, -0.17]**

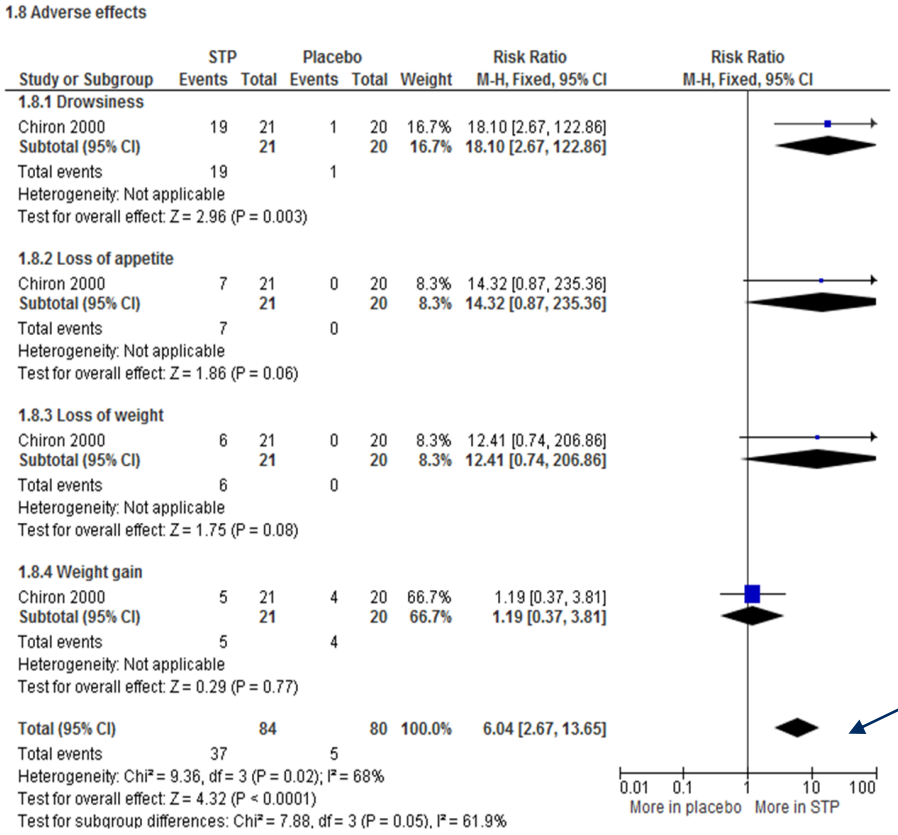
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 16.17$, df = 13 (P = 0.24); $I^2 = 20\%$

Test for overall effect: Z = 4.10 (P < 0.0001)



Studies included multiple times

FLR #5 – Reporting at odds with forest plot



'Higher proportions of participants were reported to experience side effects in the treatment group compared with placebo (100% vs 25%; RR 6.04, 95% CI 2.67 to 13.65)'.

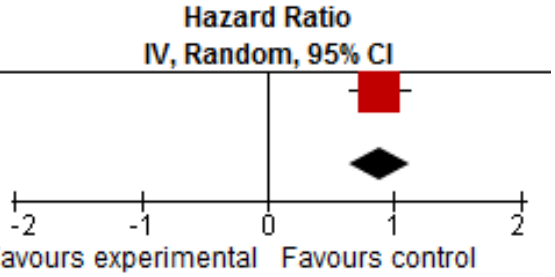
Question: what is the issue here? (type the answer in your question box)

FLR #5 – Reporting at odds with forest plot

'The confidence intervals for the estimated HR include large benefit and moderate harm of intervention (0.88; 95% CI 0.64 to 1.12), P = 0.43'

Study or Subgroup	Hazard Ratio	SE	Intervention Total	Control Total	Weight	Hazard Ratio IV, Random, 95% CI
Study 1	0.88	0.12	5472	3940	100.0%	0.88 [0.64, 1.12]
Total (95% CI)			**5472**	**3940**	**100.0%**	**0.88 [0.64, 1.12]**

Heterogeneity: Not applicable
Test for overall effect: Z = 7.33 (P < 0.00001)



Question: what is the issue here? (type the answer in your question box)

Analysis

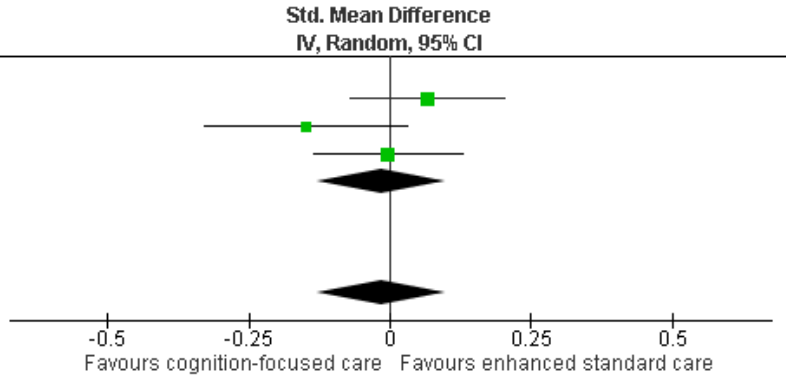
1. Unit of analysis
   - Crossover trials (Nolan et al. PLoS ONE 2016)
   - Cluster trials (Richardson et al. PLoS ONE 2016)
2. Subgroups
   - Post hoc, wrong analysis, incorrect interpretation
   - Adequate number of studies, 10?
   - Specify small number of characteristics in advance with rationale (Donegan et al. PLoS ONE 2016)
3. SMDs and MDs
   - Used incorrectly, not often back transformed
4. Random effects versus fixed effects
   - Inconsistently used

1. Unit of analysis

⊟ *Unit of analysis issues*

We planned to take into account the level at which randomisation occurred, such as cross-over trials, cluster-randomised trials and multiple observations for the same outcome. In case of cross-over trials or cluster-randomised trials, we planned to extract estimates of effect that took into account the correlation of the measurements.

	Cognition-focused care			Enhanced standard care				Std. Mean Difference
Study or Subgroup	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Weight**	**IV, Random, 95% CI**
2.1.1 Community-based studies								
Davies 2008 (1)	14.1	24.07276	437	12.5	24.07276	387	36.2%	0.07 [-0.07, 0.20]
Fisher 2011	1.78	0.96	256	1.93	1.05	227	26.3%	-0.15 [-0.33, 0.03]
Glasgow 2005	27.4	32.91	469	27.5	32.91	417	37.6%	-0.00 [-0.13, 0.13]
Subtotal (95% CI)			**1162**			**1031**	**100.0%**	**-0.02 [-0.13, 0.10]**
Heterogeneity: Tau² = 0.00; Chi² = 3.53, df = 2 (P = 0.17); I² = 43%								
Test for overall effect: Z = 0.28 (P = 0.78)								
Total (95% CI)			**1162**			**1031**	**100.0%**	**-0.02 [-0.13, 0.10]**
Heterogeneity: Tau² = 0.00; Chi² = 3.53, df = 2 (P = 0.17); I² = 43%								
Test for overall effect: Z = 0.28 (P = 0.78)								
Test for subgroup differences: Not applicable								



Footnotes

(1) Median values were reported, SDs values were calculated based on reported change in mean and P value

- Unadjusted data from study reports used in analysis

Practical Exercise 1

Practical Exercise 1 - Feedback

Practical Exercise 1 – Solutions

Figure 1: outcome 1

Study 24	33	38	40	40	2.6%	0.87 [0.76, 0.99]
Study 47	142	228	257	355	2.7%	0.86 [0.76, 0.97]
Study 50	188	265	162	242	2.7%	1.06 [0.94, 1.19]
Study 10	137	188	149	194	2.8%	0.95 [0.84, 1.07]
Study 35	[illegible]				2.8%	0.96 [0.85, 1.07]
Study 61	280	759	600	906	2.8%	0.56 [0.50, 0.62]
Study 6	[illegible]				2.9%	0.96 [0.87, 1.05]
Study 70	325	576	343	578	2.9%	0.95 [0.86, 1.05]
Study 41	[illegible]				3.0%	1.03 [0.95, 1.11]
Study 61	552	887	608	918	3.1%	0.94 [0.88, 1.01]
Study 66	[illegible]				3.1%	0.98 [0.92, 1.05]
Study 64	216	237	217	223	3.2%	0.94 [0.90, 0.98]
Total (95% CI)		**10850**		**10858**	**100.0%**	**0.89 [0.85, 0.93]**
Total events	5482		6217			

Heterogeneity: Tau² = 0.02; Chi² = 217.65, df = 52 (P < 0.00001); I² = 76%
Test for overall effect: Z = 4.89 (P < 0.00001)

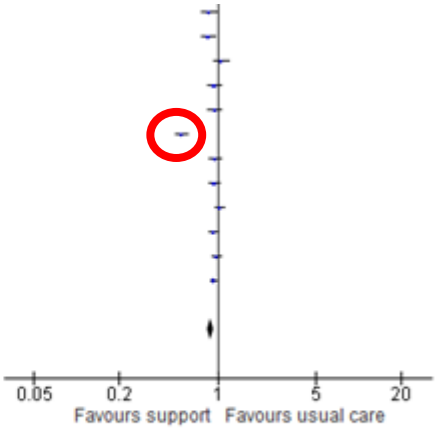
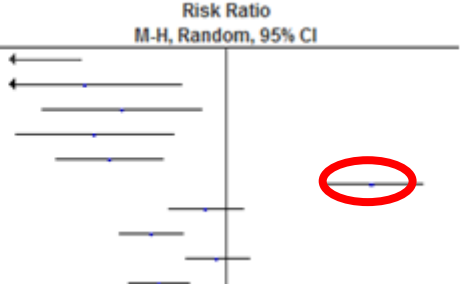


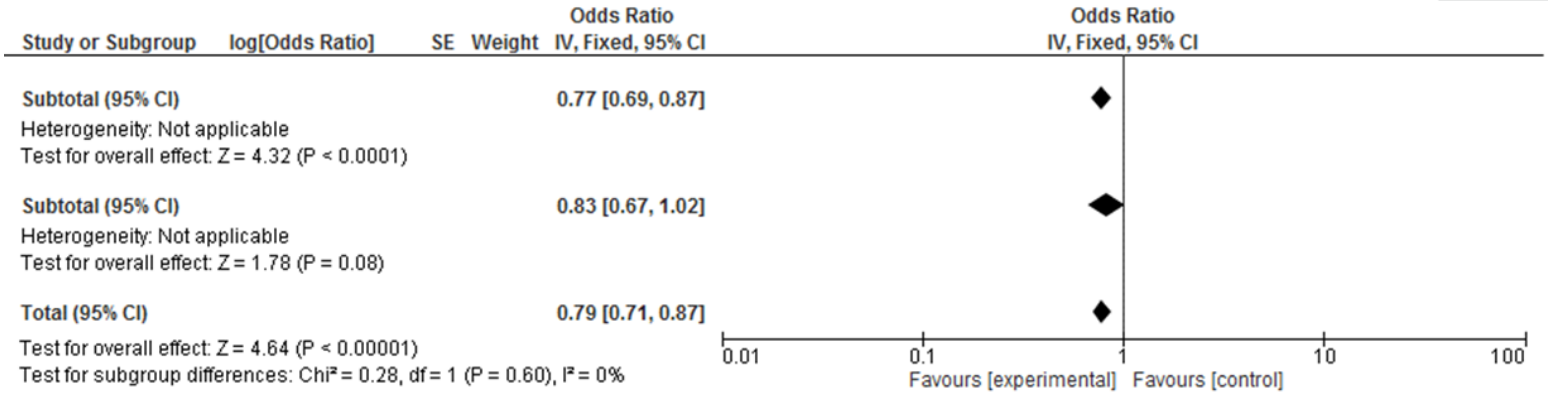
Figure 2: outcome 2

	Support		Usual care			Risk Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	M-H, Random, 95% CI
Study 90	3	50	33	50	0.1%	0.09 [0.03, 0.28]
Study 72	5	27	16	25	0.2%	0.29 [0.12, 0.67]
Study 45	6	19	15	19	0.3%	0.40 [0.20, 0.81]
Study 89a	6	22	20	23	0.3%	0.31 [0.16, 0.63]
Study 79	[illegible]				0.5%	0.36 [0.22, 0.57]
Study 3	20	33	26	157	0.6%	3.66 [2.34, 5.72]
Study 74	[illegible]				0.9%	0.84 [0.61, 1.17]
Study 67	36	80	26	30	1.1%	0.52 [0.39, 0.69]
Study 63	38	69	48	81	1.1%	0.93 [0.70, 1.23]



2. Comparing Subgroups

- use a formal statistical test to compare subgroups



***Abstract**: Our Review suggests that (INTERVENTION) may have more beneficial effects in (SUBGROUP)*

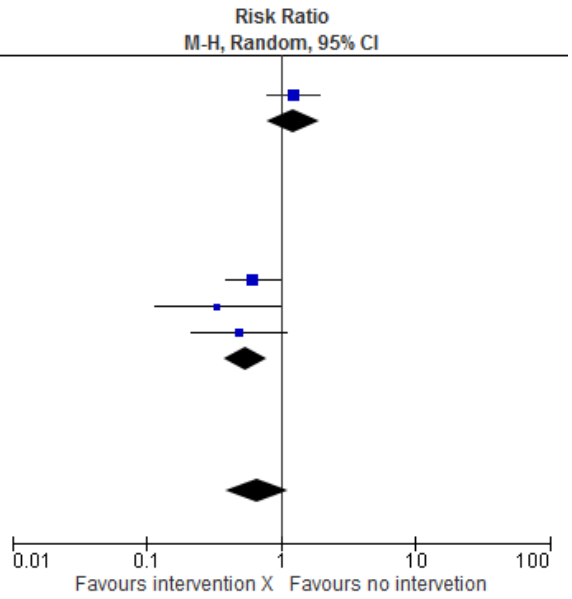
***PLS**: In the further analyses, there is evidence indicated that the effects of (INTERVENTION) in reducing (OUTCOME) rate may be different between (SUBGROUP 1) and (SUBGROUP 2), with more benefits observed in (SUBGROUP 1)*

Practical Exercise 2

Practical Exercise 2 - Feedback

Practical Exercise 2 – Solutions

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Risk Ratio M-H, Random, 95% CI
1.1.1 Intervention X (a)						
Study 3	31	115	26	119	31.7%	1.23 [0.78, 1.94]
Subtotal (95% CI)		**115**		**119**	**31.7%**	**1.23 [0.78, 1.94]**
Total events	31		26			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.91 (P = 0.36)						
1.1.2 Intervention X (b)						
Study 1	20	77	34	81	31.6%	0.62 [0.39, 0.98]
Study 2	4	60	12	60	15.5%	0.33 [0.11, 0.98]
Study 4	7	47	14	46	21.2%	0.49 [0.22, 1.10]
Subtotal (95% CI)		**184**		**187**	**68.3%**	**0.55 [0.38, 0.79]**
Total events	31		60			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.20$, df = 2 (P = 0.55); $I^2 = 0\%$						
Test for overall effect: Z = 3.18 (P = 0.001)						
Total (95% CI)		**299**		**306**	**100.0%**	**0.67 [0.39, 1.14]**
Total events	62		86			
Heterogeneity: $Tau^2 = 0.18$; $Chi^2 = 8.58$, df = 3 (P = 0.04); $I^2 = 65\%$						
Test for overall effect: Z = 1.49 (P = 0.14)						
Test for subgroup differences: $Chi^2 = 7.38$, df = 1 (P = 0.007), $I^2 = 86.5\%$						



Main results

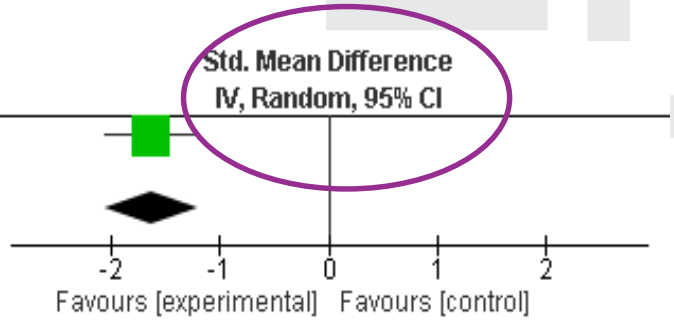
The effect of intervention X on reducing outcome A was uncertain due to the low quality of the evidence (RR 0.67, 95% CI 0.39 to 1.14; 605 participants; 4 studies). Subgroup analysis by type of intervention X provided limited evidence that X (b) may lower the risk of outcome A.

3. MDs and SMDs

"We will convert continuous outcome data into standardised mean differences (SMDs) and present with 95% CIs, as it is assumed that study authors will use different measurement scales. If continuous outcome data is recorded using the same measurement scale, data will be converted into mean differences (MDs) and presented with 95% Cis".

	Experimental			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Study 4	23	12.9	62	46	15.2	51	100.0%	-1.63 [-2.06, -1.20]
Total (95% CI)			**62**			**51**	**100.0%**	**-1.63 [-2.06, -1.20]**

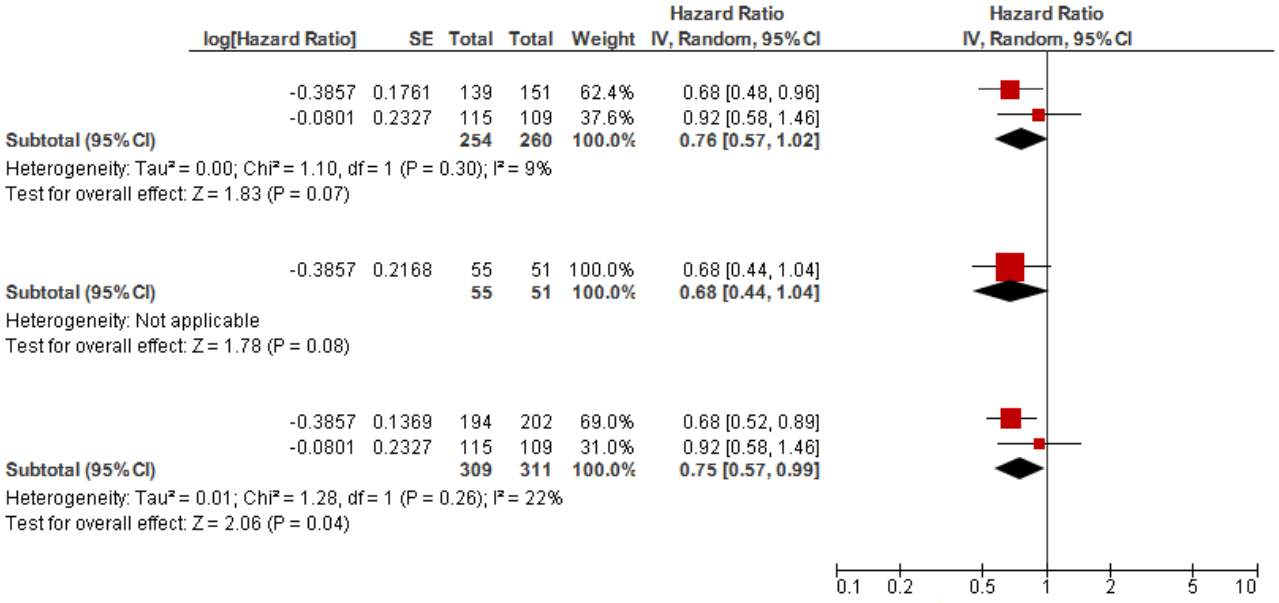
Heterogeneity: Not applicable
Test for overall effect: Z = 7.46 (P < 0.00001)



Question: what is the problem here? (type the answer in your question box)

4. Fixed Effect versus Random Effects

"We considered statistical heterogeneity between trials to be substantial if, following meta-analysis, I^2 was greater than 30% and either T^2 is greater than zero, or there was a low P-value (< 0.10) in the Chi^2 test for heterogeneity. If substantial heterogeneity was identified used the random-effects (RE) model instead of the fixed-effects (FE) model to pool data".



Question: what is the problem here? (type the answer in your question box)

Errors we may not see

- Have any papers been missed?
- Have the right results been copied from the papers?
- Have the standard deviations been confused with standard errors?

Question: Are there any other errors we may not see? (type suggestions in your question box)


Test drive
training.cochrane.org/common-errors

Cochrane

Final Tips

Tips for spotting errors

- Numbers that stand out (perfect homogeneity, single outlying results, sample size does not match with precision relative to other studies)
- For non-standard RCT designs - evidence of how SEs were adjusted (check methods against plots).
- For primary outcomes select the biggest study or the one that has most weight and check the analysis results against the paper.
- For other outcomes pick a study entirely at random and check numbers used against what is available in published trial report or elsewhere. If authors have stated that they got unpublished data then move on to next study.

Discussion

References and resources

Nolan SJ, Hambleton I, Dwan K (2016) The Use and Reporting of the Cross-Over Study Design in Clinical Trials and Systematic Reviews: A Systematic Assessment. PLoS ONE 11(7): e0159014. https://doi.org/10.1371/journal.pone.0159014

Richardson M, Garner P, Donegan S (2016) Cluster Randomised Trials in Cochrane Reviews: Evaluation of Methodological and Reporting Practice. PLoS ONE 11(3): e0151818. https://doi.org/10.1371/journal.pone.0151818

Donegan S, Williams L, Dias S, Tudur-Smith C, Welton N (2015) Exploring Treatment by Covariate Interactions Using Subgroup Analysis and Meta-Regression in Cochrane Reviews: A Review of Recent Practice. PLoS ONE 10(6): e0128804. https://doi.org/10.1371/journal.pone.0128804

MECIR http://methods.cochrane.org/mecir

training.cochrane.org/common-errors

Practical Exercises 3 and 4 - Feedback

Practical Exercise 3 – Solutions

TABLE 5 Total diary score over the first 10 days: mean (SD)

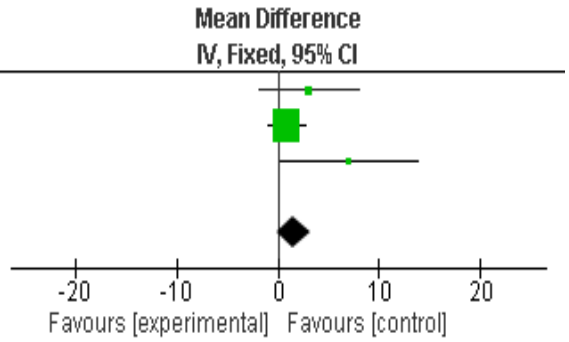
	Group 1 Bed rest and exercise and education (n = 50)	Group 2 Exercise and education (n = 41)	Group 3 Bed rest (n = 47)	Group 4 Control (n = 48)
Improvement	22.27 (5.14)	23.30 (6.92)	21.66 (6.54)	21.54 (6.31)
Activities	24.35 (8.75)	21.34 (9.22)	24.34 (10.04)	20.90 (8.46)
Pain	23.77 (5.22)	25.94 (7.47)	24.15 (7.12)	22.88 (5.88)

Note: Lower total scores indicate a better clinical result

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Brown 2003	19.1	21	145	16	22	146	12.2%	3.10 [-1.84, 8.04]
Gilbert 1995	23.77	5.88	65	22.88	5.22	65	81.3%	0.89 [-1.02, 2.80]
Smith 2015	31	17.21	42	24	17.21	62	6.5%	7.00 [0.26, 13.74]
Total (95% CI)			252			273	100.0%	1.56 [-0.17, 3.28]

Heterogeneity: Chi² = 3.35, df = 2 (P = 0.19); I² = 40%

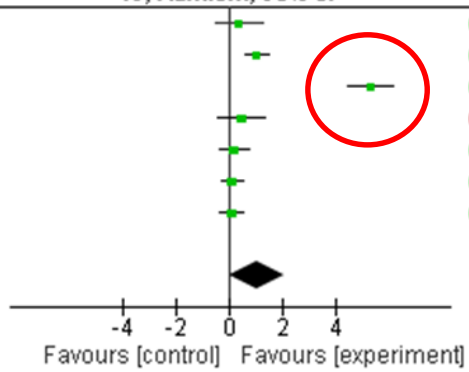
Test for overall effect: Z = 1.77 (P = 0.08)

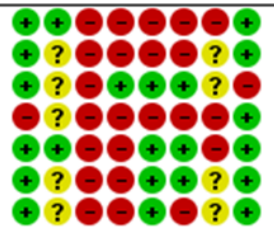


Practical Exercise 4 – Solutions

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Study 1	0.07	0.3	69	-0.04	0.3	5	13.6%	0.36 [-0.55, 1.27]
Study 2	-1.02	4.77	50	-5.92	4.57	52	14.8%	1.04 [0.63, 1.46]
Study 3	3.02	0.08	62	2.59	0.08	40	13.8%	5.33 [4.49, 6.18]
Study 4	-0.45	1.29	11	-1.11	1.45	9	13.6%	0.46 [-0.43, 1.36]
Study 5	-7.07	5.05	28	-7.85	3.59	26	14.6%	0.17 [-0.36, 0.71]
Study 6	-0.5	4.3	37	-0.9	4.2	69	14.9%	0.09 [-0.31, 0.49]
Study 7	0.31	3.08	31	0.06	2.85	41	14.7%	0.08 [-0.38, 0.55]
Total (95% CI)			**288**			**242**	**100.0%**	**1.05 [0.04, 2.07]**

Heterogeneity: Tau² = 1.77; Chi² = 137.94, df = 6 (P < 0.00001); I² = 96%
Test for overall effect: Z = 2.03 (P = 0.04)





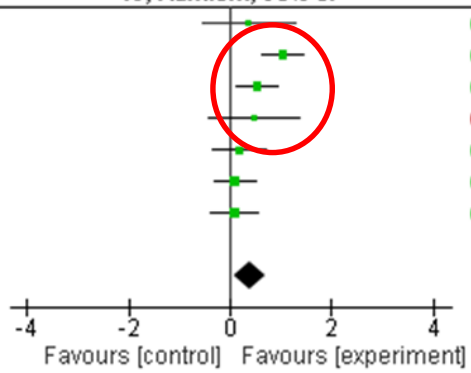
Risk of bias legend

(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Sample Size
(H) Other bias

Practical Exercise 4 – Solutions

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Study 1	0.07	0.3	69	-0.04	0.3	5	7.8%	0.36 [-0.55, 1.27]
Study 2	-1.02	4.77	50	-5.93	4.57	52	17.6%	1.04 [0.63, 1.46]
Study 3	3.02	0.8	62	2.59	0.8	40	17.9%	0.53 [0.13, 0.94]
Study 4	-0.45	1.29	11	-1.11	1.45	9	8.0%	0.46 [-0.43, 1.36]
Study 5	-7.07	5.05	28	-7.85	3.59	26	14.5%	0.17 [-0.36, 0.71]
Study 6	-0.5	4.3	37	-0.9	4.2	69	18.0%	0.09 [-0.31, 0.49]
Study 7	0.31	3.08	31	0.06	2.85	41	16.2%	0.08 [-0.38, 0.55]
Total (95% CI)			**288**			**242**	**100.0%**	**0.40 [0.10, 0.70]**

Heterogeneity: Tau² = 0.09; Chi² = 14.37, df = 6 (P = 0.03); I² = 58%
Test for overall effect: Z = 2.58 (P = 0.010)



Risk of bias legend

(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Sample Size
(H) Other bias