Practical Exercise 1 - Unit of analysis issues and Data Extraction

Consider the following excerpt from a recent review, and the following Forest plots – Which studies would concern you? What would you do?

Cluster-randomised trials

There are 14 cluster-randomised trials in the analyses along with individually randomised trials. Their sample sizes have been adjusted using the methods described in the *Handbook* and by Donner 2000 using an estimate of the intracluster correlation coefficient (ICC) derived from the trial (if possible), or from another source. If ICCs from other sources were used, we have noted this and carried out sensitivity analyses to investigate the effect of variation in ICC. We have synthesised the findings from individually- and cluster-randomised trials provided that there was little heterogeneity between the study designs and the interaction between the effect of intervention and the choice of randomisation unit was considered to be unlikely.

Trials with multiple groups

In order to avoid 'double counting', the data provided by studies that involved one comparison group but two interventions groups had the number of events and number of participants halved.

Figure 1: outcome 1

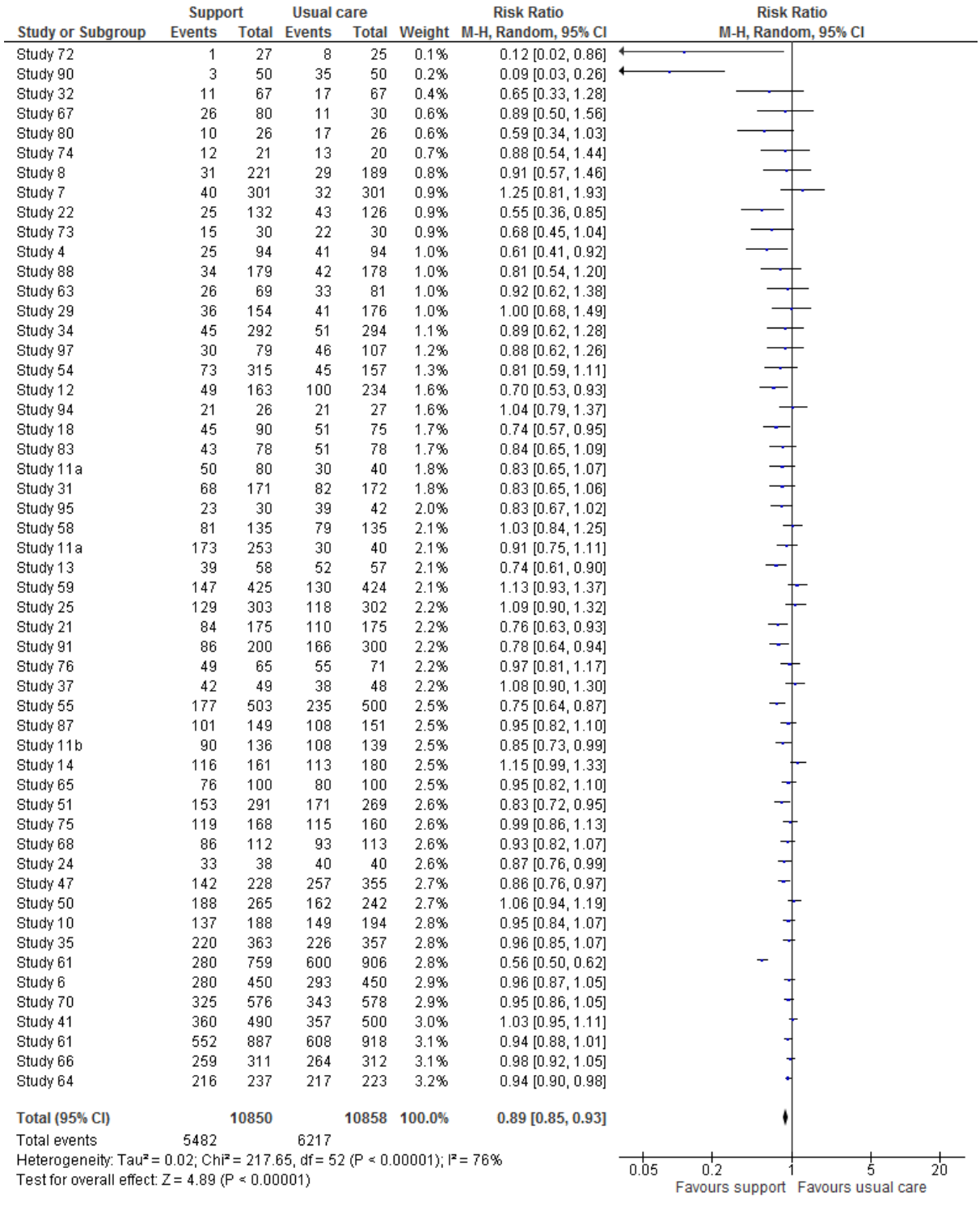
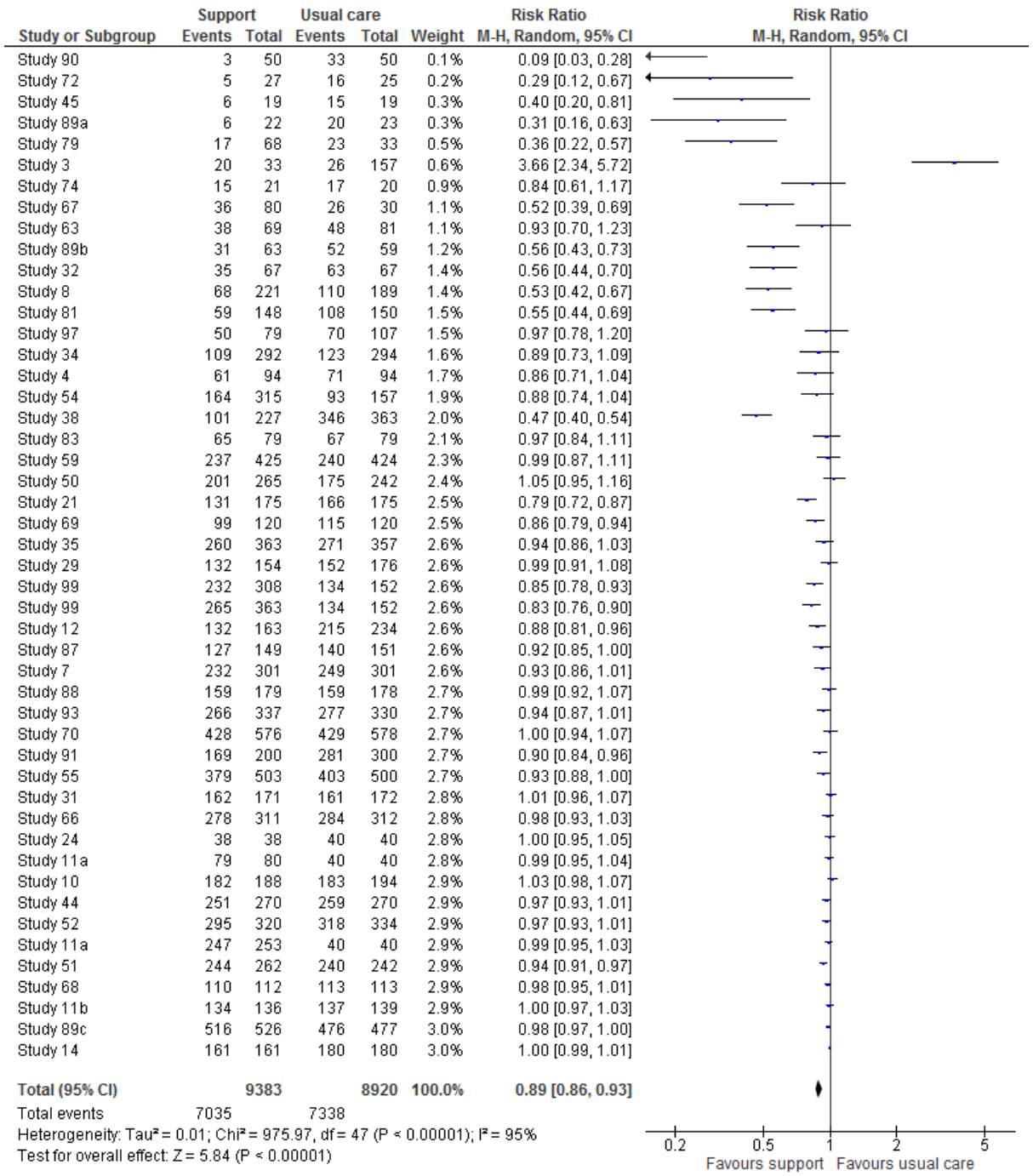


Figure 2: outcome 2



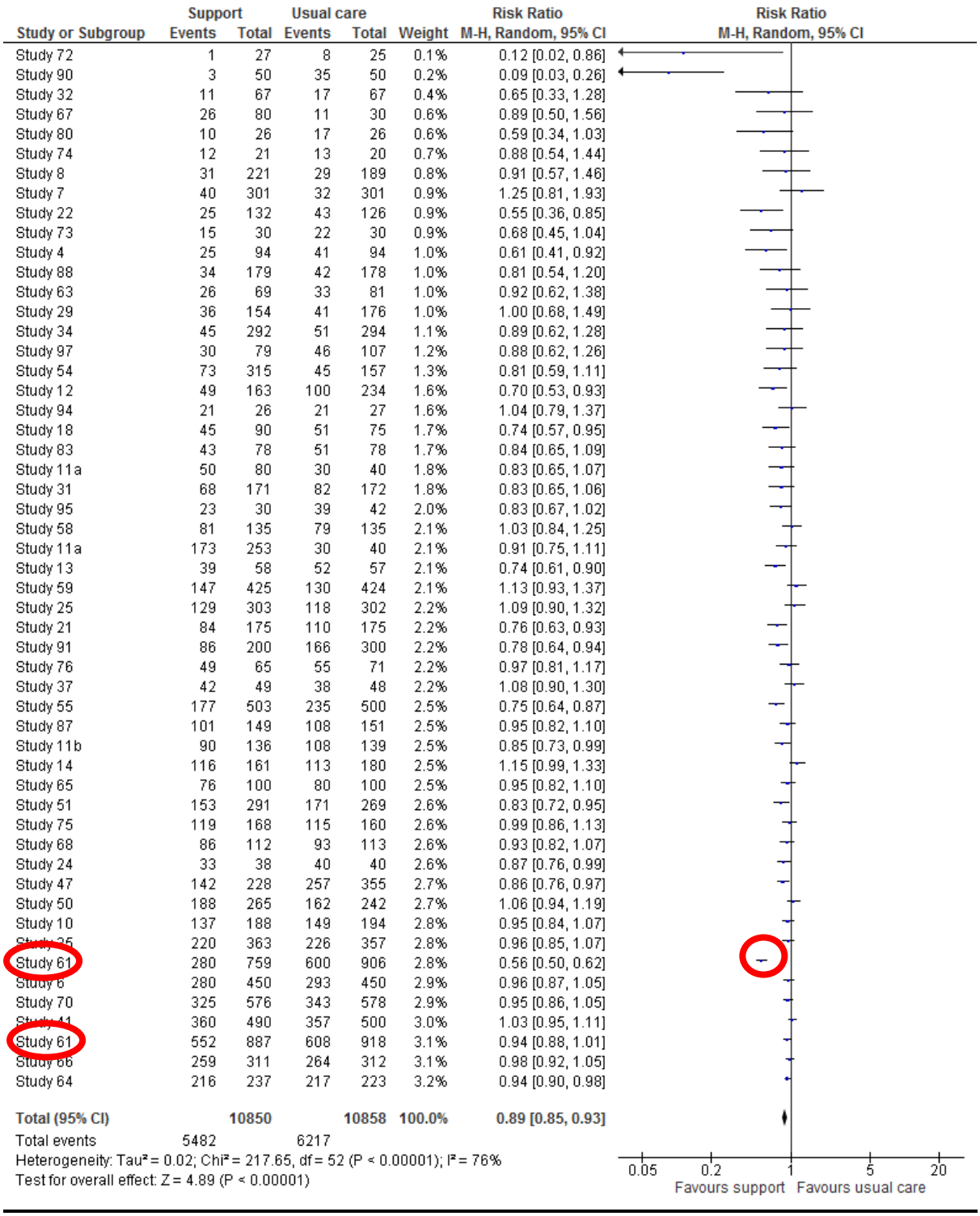
Study or Subgroup	Support Events	Support Total	Usual care Events	Usual care Total	Weight	Risk Ratio M-H, Random, 95% CI
Study 90	3	50	33	50	0.1%	0.09 [0.03, 0.28]
Study 72	5	27	16	25	0.2%	0.29 [0.12, 0.67]
Study 45	6	19	15	19	0.3%	0.40 [0.20, 0.81]
Study 89a	6	22	20	23	0.3%	0.31 [0.16, 0.63]
Study 79	17	68	23	33	0.5%	0.36 [0.22, 0.57]
Study 3	20	33	26	157	0.6%	3.66 [2.34, 5.72]
Study 74	15	21	17	20	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]
Study 89b	31	63	52	59	1.2%	0.56 [0.43, 0.73]
Study 32	35	67	63	67	1.4%	0.56 [0.44, 0.70]
Study 8	68	221	110	189	1.4%	0.53 [0.42, 0.67]
Study 81	59	148	108	150	1.5%	0.55 [0.44, 0.69]
Study 97	50	79	70	107	1.5%	0.97 [0.78, 1.20]
Study 34	109	292	123	294	1.6%	0.89 [0.73, 1.09]
Study 4	61	94	71	94	1.7%	0.86 [0.71, 1.04]
Study 54	164	315	93	157	1.9%	0.88 [0.74, 1.04]
Study 38	101	227	346	363	2.0%	0.47 [0.40, 0.54]
Study 83	65	79	67	79	2.1%	0.97 [0.84, 1.11]
Study 59	237	425	240	424	2.3%	0.99 [0.87, 1.11]
Study 50	201	265	175	242	2.4%	1.05 [0.95, 1.16]
Study 21	131	175	166	175	2.5%	0.79 [0.72, 0.87]
Study 69	99	120	115	120	2.5%	0.86 [0.79, 0.94]
Study 35	260	363	271	357	2.6%	0.94 [0.86, 1.03]
Study 29	132	154	152	176	2.6%	0.99 [0.91, 1.08]
Study 99	232	308	134	152	2.6%	0.85 [0.78, 0.93]
Study 99	265	363	134	152	2.6%	0.83 [0.76, 0.90]
Study 12	132	163	215	234	2.6%	0.88 [0.81, 0.96]
Study 87	127	149	140	151	2.6%	0.92 [0.85, 1.00]
Study 7	232	301	249	301	2.6%	0.93 [0.86, 1.01]
Study 88	159	179	159	178	2.7%	0.99 [0.92, 1.07]
Study 93	266	337	277	330	2.7%	0.94 [0.87, 1.01]
Study 70	428	576	429	578	2.7%	1.00 [0.94, 1.07]
Study 91	169	200	281	300	2.7%	0.90 [0.84, 0.96]
Study 55	379	503	403	500	2.7%	0.93 [0.88, 1.00]
Study 31	162	171	161	172	2.8%	1.01 [0.96, 1.07]
Study 66	278	311	284	312	2.8%	0.98 [0.93, 1.03]
Study 24	38	38	40	40	2.8%	1.00 [0.95, 1.05]
Study 11a	79	80	40	40	2.8%	0.99 [0.95, 1.04]
Study 10	182	188	183	194	2.9%	1.03 [0.98, 1.07]
Study 44	251	270	259	270	2.9%	0.97 [0.93, 1.01]
Study 52	295	320	318	334	2.9%	0.97 [0.93, 1.01]
Study 11a	247	253	40	40	2.9%	0.99 [0.95, 1.03]
Study 51	244	262	240	242	2.9%	0.94 [0.91, 0.97]
Study 68	110	112	113	113	2.9%	0.98 [0.95, 1.01]
Study 11b	134	136	137	139	2.9%	1.00 [0.97, 1.03]
Study 89c	516	526	476	477	3.0%	0.98 [0.97, 1.00]
Study 14	161	161	180	180	3.0%	1.00 [0.99, 1.01]
Total (95% CI)		**9383**		**8920**	**100.0%**	**0.89 [0.86, 0.93]**
Total events	7035		7338			

Heterogeneity: Tau² = 0.01; Chi² = 975.97, df = 47 (P < 0.00001); I² = 95%
Test for overall effect: Z = 5.84 (P < 0.00001)

ANSWERS

In figure 1, study 61 is a multiarm trial and also a cluster trial and although it has been adjusted for clustering, the usual care arm has not been split in half and so participants are double counted. Also the number of events in the support arm in the first entry is incorrect but this difficult to see without access to the trial report.

Figure 3: outcome 1



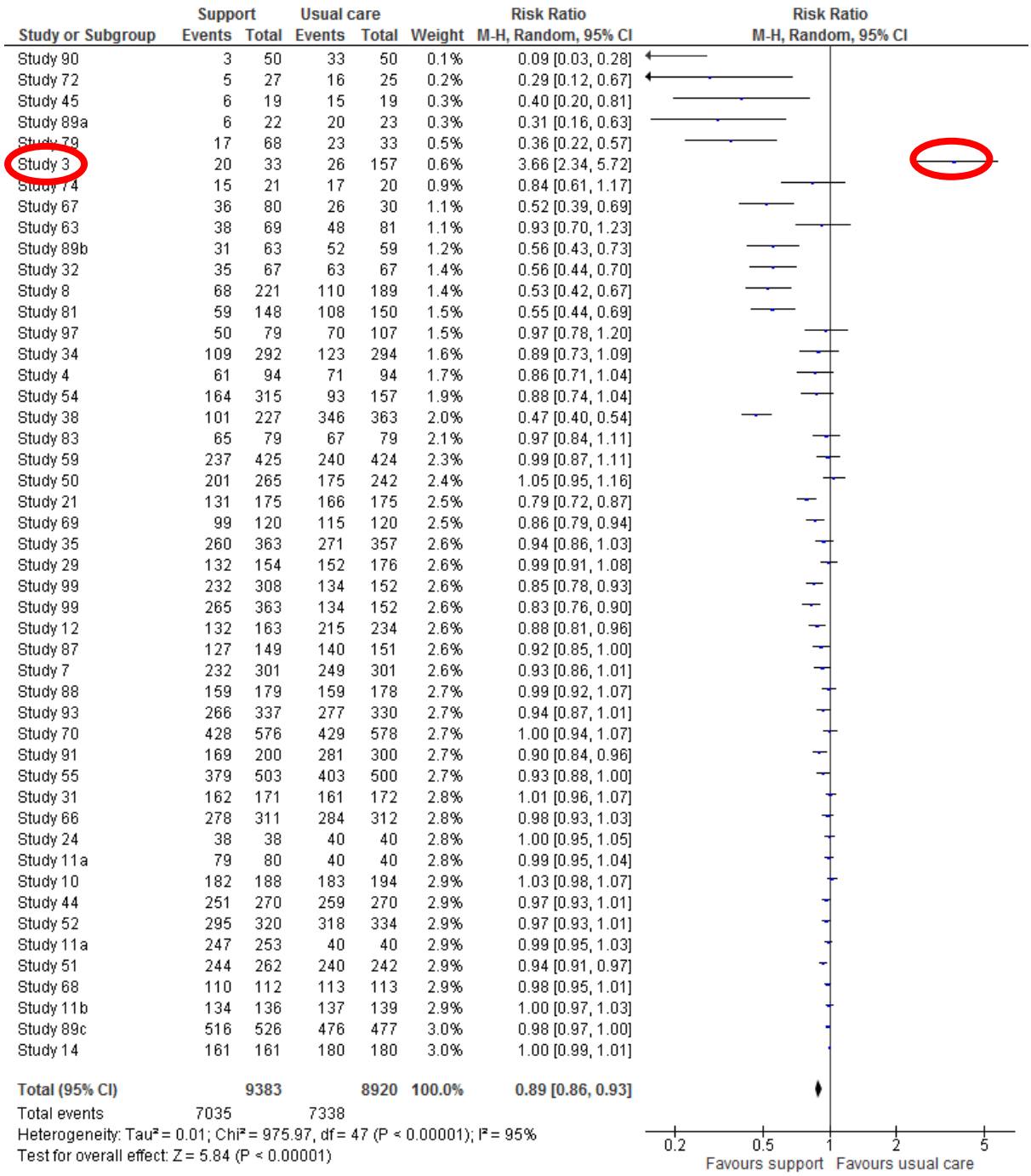
Study or Subgroup	Support Events	Support Total	Usual care Events	Usual care Total	Weight	Risk Ratio M-H, Random, 95% CI
Study 72	1	27	8	25	0.1%	0.12 [0.02, 0.86]
Study 90	3	50	35	50	0.2%	0.09 [0.03, 0.26]
Study 32	11	67	17	67	0.4%	0.65 [0.33, 1.28]
Study 67	26	80	11	30	0.6%	0.89 [0.50, 1.56]
Study 80	10	26	17	26	0.6%	0.59 [0.34, 1.03]
Study 74	12	21	13	20	0.7%	0.88 [0.54, 1.44]
Study 8	31	221	29	189	0.8%	0.91 [0.57, 1.46]
Study 7	40	301	32	301	0.9%	1.25 [0.81, 1.93]
Study 22	25	132	43	126	0.9%	0.55 [0.36, 0.85]
Study 73	15	30	22	30	0.9%	0.68 [0.45, 1.04]
Study 4	25	94	41	94	1.0%	0.61 [0.41, 0.92]
Study 88	34	179	42	178	1.0%	0.81 [0.54, 1.20]
Study 63	26	69	33	81	1.0%	0.92 [0.62, 1.38]
Study 29	36	154	41	176	1.0%	1.00 [0.68, 1.49]
Study 34	45	292	51	294	1.1%	0.89 [0.62, 1.28]
Study 97	30	79	46	107	1.2%	0.88 [0.62, 1.26]
Study 54	73	315	45	157	1.3%	0.81 [0.59, 1.11]
Study 12	49	163	100	234	1.6%	0.70 [0.53, 0.93]
Study 94	21	26	21	27	1.6%	1.04 [0.79, 1.37]
Study 18	45	90	51	75	1.7%	0.74 [0.57, 0.95]
Study 83	43	78	51	78	1.7%	0.84 [0.65, 1.09]
Study 11a	50	80	30	40	1.8%	0.83 [0.65, 1.07]
Study 31	68	171	82	172	1.8%	0.83 [0.65, 1.06]
Study 95	23	30	39	42	2.0%	0.83 [0.67, 1.02]
Study 58	81	135	79	135	2.1%	1.03 [0.84, 1.25]
Study 11a	173	253	30	40	2.1%	0.91 [0.75, 1.11]
Study 13	39	58	52	57	2.1%	0.74 [0.61, 0.90]
Study 59	147	425	130	424	2.1%	1.13 [0.93, 1.37]
Study 25	129	303	118	302	2.2%	1.09 [0.90, 1.32]
Study 21	84	175	110	175	2.2%	0.76 [0.63, 0.93]
Study 91	86	200	166	300	2.2%	0.78 [0.64, 0.94]
Study 76	49	65	55	71	2.2%	0.97 [0.81, 1.17]
Study 37	42	49	38	48	2.2%	1.08 [0.90, 1.30]
Study 55	177	503	235	500	2.5%	0.75 [0.64, 0.87]
Study 87	101	149	108	151	2.5%	0.95 [0.82, 1.10]
Study 11b	90	136	108	139	2.5%	0.85 [0.73, 0.99]
Study 14	116	161	113	180	2.5%	1.15 [0.99, 1.33]
Study 65	76	100	80	100	2.5%	0.95 [0.82, 1.10]
Study 51	153	291	171	269	2.6%	0.83 [0.72, 0.95]
Study 75	119	168	115	160	2.6%	0.99 [0.86, 1.13]
Study 68	86	112	93	113	2.6%	0.93 [0.82, 1.07]
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]
[illegible]	220	363	226	357	2.8%	0.96 [0.85, 1.07]
Study 61	280	759	600	906	2.8%	0.56 [0.50, 0.62]
[illegible]	280	450	293	450	2.9%	0.96 [0.87, 1.05]
Study 70	325	576	343	578	2.9%	0.95 [0.86, 1.05]
[illegible]	360	490	357	500	3.0%	1.03 [0.95, 1.11]
Study 61	552	887	608	918	3.1%	0.94 [0.88, 1.01]
[illegible]	259	311	264	312	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)

In figure 2, study 3 is an outlier. The trial report was accessed and the data checked. The total in the usual care arm was incorrect, it should have been 33 not 157

Figure 4: outcome 2



Practical Exercise 2 – Subgroup Analyses

Intervention X is a device which is available in two variations – X (a) and X (b).

- What concerns might you have about the interpretation of the following analyses of intervention X?
- What changes might you suggest to the analyses and the interpretation?

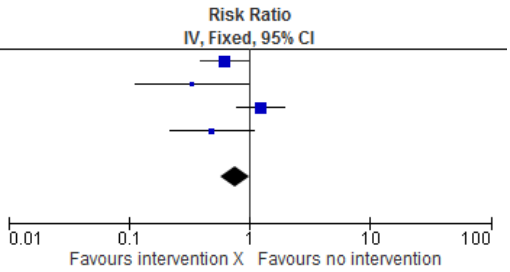
1 – Intervention X vs no intervention

1.1 Outcome A

	Intervention X		No intervention			Risk Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	IV, Fixed, 95% CI
Study 1	20	77	34	81	39.9%	0.62 [0.39, 0.98]
Study 2	4	60	12	60	7.2%	0.33 [0.11, 0.98]
Study 3	31	115	26	119	40.3%	1.23 [0.78, 1.94]
Study 4	7	47	14	46	12.6%	0.49 [0.22, 1.10]
Total (95% CI)		**299**		**306**	**100.0%**	**0.76 [0.57, 1.01]**
Total events	62		86			

Heterogeneity: Chi² = 8.55, df = 3 (P = 0.04); I² = 65%

Test for overall effect: Z = 1.88 (P = 0.06)



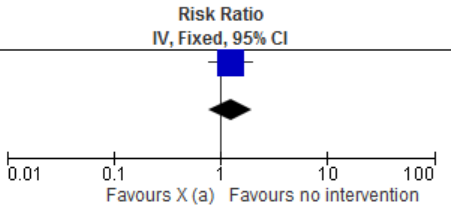
2 – Intervention X (a) vs no intervention

2.1 Outcome A

	Intervention X (a)		No intervention			Risk Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	IV, Fixed, 95% CI
Study 3	31	115	26	119	100.0%	1.23 [0.78, 1.94]
Total (95% CI)		**115**		**119**	**100.0%**	**1.23 [0.78, 1.94]**
Total events	31		26			

Heterogeneity: Not applicable

Test for overall effect: Z = 0.91 (P = 0.36)



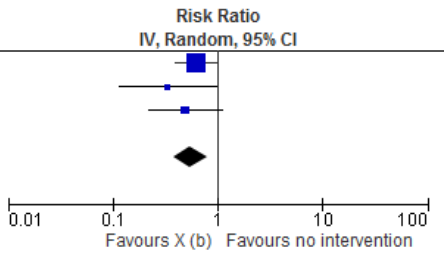
3 – Intervention X (b) vs no intervention

3.1 Outcome A

	Intervention X (b)		No intervention			Risk Ratio
Study or Subgroup	Events	Total	Events	Total	Weight	IV, Random, 95% CI
Study 1	20	77	34	81	66.8%	0.62 [0.39, 0.98]
Study 2	4	60	12	60	12.0%	0.33 [0.11, 0.98]
Study 4	7	47	14	46	21.1%	0.49 [0.22, 1.10]
Total (95% CI)		**184**		**187**	**100.0%**	**0.55 [0.38, 0.79]**
Total events	31		60			

Heterogeneity: Tau² = 0.00; Chi² = 1.17, df = 2 (P = 0.56); I² = 0%

Test for overall effect: Z = 3.18 (P = 0.001)



Main results

Overall, intervention X was not associated with a statistically significant change in outcome A. In a subgroup analysis it was found that intervention X (b) may be associated with a significant reduction in outcome A.

Authors' conclusions

This systematic review suggests that the use of intervention X (b) may be beneficial. However only a limited number of RCTs with rather small sample sizes were available. Further RCTs on intervention X are needed.

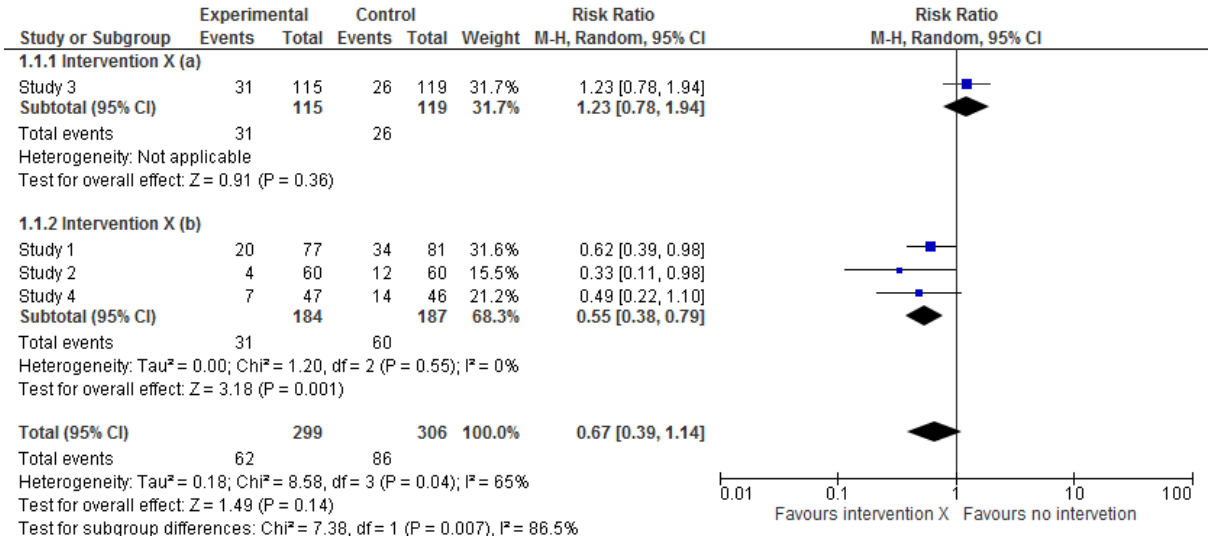
ANSWERS

The data analyses are inappropriately structured. Fixed and random effects models have been inconsistently applied. The analyses should be restructured so that X (a) vs no intervention and intervention X (b) vs no intervention are presented as subgroups of the intervention X vs no intervention comparison, rather than as individual comparisons. This will allow for tests for subgroup differences to be applied and appropriate conclusions to be drawn about subgroup differences. The appropriate model should be used, following the methods specified for data synthesis in the protocol.

Appropriately structured forest plot and improved interpretation:

1 – Intervention X vs no intervention

1.1 Outcome A



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.

Authors' conclusions

This systematic review has identified limited evidence on the effect of intervention X. We have not been able to identify convincing direct evidence of superiority of X (b) over X (a). We found a limited number of RCTs with small sample sizes. Further RCTs on intervention X are needed.

Practical Exercise 3 – Data Entry Errors

A review author wishes to assess the effect of an 'Exercise' intervention on the outcome 'Pain'.

Consider the following Forest Plot and the table from the published paper Gilbert 1995 from which data was extracted. Can you spot any data extraction errors?

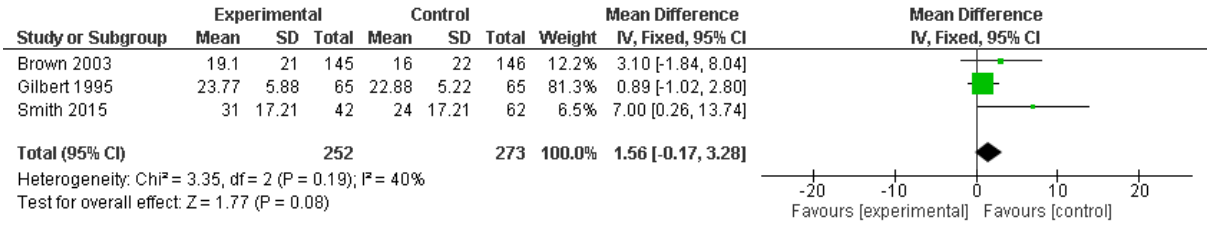
Original Study Data – Gilbert 1995

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

Forest Plot



ANSWERS

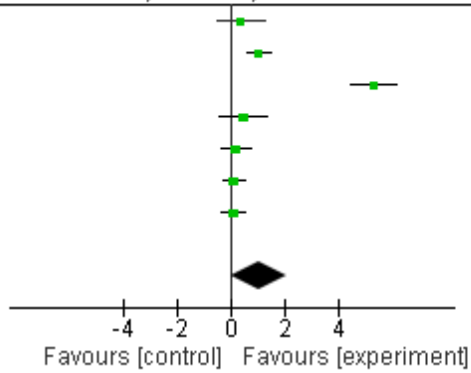
- Sample sizes in the forest plot are incorrect – they should read 50 and 48 (not 65 and 65)
- The standard deviations are mixed up (experimental SD is 5.22, control is 5.88)
- The wrong intervention group data has been extracted from the paper. Authors should have extracted Group 2 data (Exercise + Education) and not Group 1 data (Exercise + Education and Bed Rest). If we compare Group 1 data with control, we cannot know if the effect is due to exercise + education, or to bed rest.

Practical Exercise 4 – Outliers

Analysis - Anxiety

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	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	-5.93	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

Results > Effects of Intervention

Overall, a large effect was observed for intervention participants (n = 288) compared to usual care (n = 242) in the pooled analysis of Anxiety for all seven studies (SMD, 1.05, 95% CI, 0.04 to 2.07) (Figure 7). Study 3 contributed the largest effect difference and this may be due to potential bias in the timing of the assessment prior to initiation of the intervention. Anxiety is usually high in anticipation of treatment and this change in effect may be due to the reduction in this anticipatory anxiety rather than as a result of the intervention. However, our confidence in this effect is low as high heterogeneity based on the I^2 test that was statistically significant (I^2 = 96%, Chi2 = 137.94, df = 6, p <0.001). We downgraded the quality of evidence for anxiety by two levels (very serious (-2)) for risk of bias in study design given that 4 of the 7 studies had selective outcome reporting and 5 of the 7 studies had unclear allocation concealment, and we downgraded one more level (-1) due to important inconsistency (Summary of findings table 1).

Practical Exercise Questions

1. Which study would you consider an 'Outlier' in this Forest Plot?
2. Have the authors addressed this outlier in their description/interpretation of the result?
3. Is the authors interpretation of this Forest Plot sufficient and appropriate? If so, why? If not, why not?

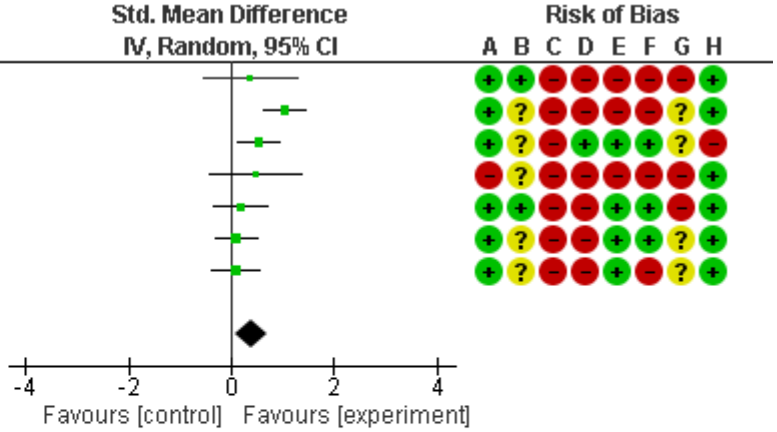
Practice Exercise Answers

1. Study 3
2. Yes – They attribute it to the timing of the assessment prior to initiation of the intervention, leading to a 'reduction in this anticipatory anxiety'. They also downgrade the quality/certainty of the result for inconsistency.
3. No – even with this explanation, it is a very notable outlier, and 96% is a very difficult level of heterogeneity to accept. Authors first instinct was to assume that the data was correct and that the outlier can be explained. Editors should still question this high level of heterogeneity and conduct our investigation?

   In this case, we did question it. As it turned out, authors have entered the Standard Deviations incorrectly (They should read 0.8, not 0.08). When the correct SDs are entered, the graph would appear as follows;

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	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

Note that the I^2 has dropped from 96% to 58%.

Learning Point: Don't automatically accept outliers/high heterogeneity just because the authors think they can explain it