Do we need another trial on exercise in patients with knee osteoarthritis?

No new trials on exercise in knee OA

Arianne P Verhagen (1,2), Manuela Ferreira (3), Elja A.E. Reijneveld-van de Vendel (1), Carolien H. Teirlinck (1), Jos Runhaar (1), Marienke van Middelkoop (1), Lotte Hermsen (4), Ingrid B de Groot (4), Sita MA Bierma-Zeinstra (1)

1) *Dept General Practice, Erasmus Medical Centre University, Rotterdam;*
2) *Discipline of Physiotherapy, Graduate School of Health, University of Technology Sydney, Australia*
3) *Institute of Bone and Joint Research, The Kolling Institute, Faculty of Medicine and Health, The University of Sydney, Australia*
4) *National Health Care Institute, The Netherlands (Zorginstituut Nederland, ZIN)*

Conflict of interest: All authors declare there is no conflict of interest.

Word count: abstract: 223; text: 1994; nr references: 19; nr figures: 2

Correspondence: Arianne P Verhagen; Discipline of Physiotherapy, Graduate School of Health, University of Technology Sydney, Australia**; email: Arianne.verhagen@uts.edu.au**

ABSTRACT

Objective. We aim to investigate if we need additional trials on exercise in knee osteoarthritis (OA) to accept a certain effect size to be a 'true' effect size, and new studies are not needed anymore.

Design. We performed a secondary analyses of a meta-analysis of studies on patients with knee osteoarthritis, on pain immediately post treatment. We performed five different analysis: a) we evaluated publication bias, b) we performed subgroup analysis, c) a sensitivity analysis based on the overall risk of bias (RoB) score, d) a cumulative meta-analysis and e) we developed an extended funnel plot to explore the potential impact of a new study on the summary effect estimate.

Results. We included 42 studies with in total 6863 patients. The analyses showed that a) there is no clear publication bias, b) subgrouping did not affect the overall effect estimate, c) the effect estimate of exercise is more consistent (no heterogeneity) in the studies of low RoB, d) the benefit of exercise was clear since 2010 and e) the extended funnel plot suggests that an additional study has a none or very limited impact to change the current effect estimate.

Conclusion. Exercise is effective and clinically worthwhile in reducing pain immediately post treatment compared to no or minimal interventions in patients with knee OA and adding new data will unlikely change this conclusion.

BACKGROUND

In The Netherlands, the basic healthcare insurance is mandatory. Within this basic health insurance package elementary healthcare and emergency healthcare as well as proven (cost)-effective treatments are reimbursed by the healthcare insurers. Additional healthcare insurance for other treatments (e.g. exercise treatments, manual therapy, acupuncture, homeopathy) is voluntary. When this system was introduced about 15 years ago, most physiotherapy treatments, including exercise therapy for osteoarthritis patients were not included in the basic health insurance package. Therefore, the patients need to pay for these treatments themselves or need to have an additional healthcare insurance that covers these treatments. Based on questions from patients with osteoarthritis (OA) and the Royal Dutch Physiotherapy Association (KNGF), the Minister of Health, Welfare and Sports requested in 2016 informed advice from the National Health Care Institute to evaluate whether exercise therapy for patients with knee (or hip) OA should be added to the basic health insurance package.

At that time, two Cochrane systematic reviews on exercise therapy versus no or minimal treatment in patients with knee OA summarised the evidence on these treatments [1,2]. These reviews reported a small, but statistically significant post treatment effect in favour of exercise therapy concerning pain reduction and increase in function (standardised mean differences (SMD) between 0.4-0.6). Based on this evidence, multiple national and international guidelines recommend exercise therapy as an effective treatment for patients with knee OA [3,4,5].

Nevertheless, for a decision in the Dutch situation the Dutch Minister wanted to know whether there was a clinically meaningful effect of exercise therapy compared to no or minimal interventions for knee and hip OA patients in the Dutch situation as provided by Dutch physiotherapists. Therefore, we updated both Cochrane reviews, using Cochrane methodology, but with stricter selection criteria concerning the interventions and control interventions compared to the existing reviews [6]. The effect estimate appeared to be comparable to the ones presented in both Cochrane reviews, and of low quality due to the large numbers of studies with high risk of bias (design) and heterogeneity (inconsistency).

Therefore, the aim of the present study, given the multiple studies and meta-analyses that reached a comparable outcome concerning exercise effectiveness in patients with knee OA, was to investigate whether new data is needed before we accept this effect size to be a ‘true’ effect size. Although the current study question was initiated in a Dutch context, the answer concerns a broader audience and is relevant for all clinicians and researchers in interested in the care for patients with knee OA.

METHOD

Design Secondary analysis of the updated systematic review and meta-analysis of studies on patients with knee OA only [1,2,6]. We selected randomised clinical trials evaluating exercise compared to no or minimal treatments in patients with knee OA. The intervention should be exercise treatment as provided by Dutch physiotherapists (no Tai Chi or home exercises) and the control intervention should be 'usual care' (e.g. advice), no treatment (e.g. waiting list), a minimal intervention (e.g. medication), or non-supervised exercise therapy (e.g. home-based exercise therapy). Studies with passive interventions in the control groups, such as hot packs or ultrasound are excluded [6]. The date of last search of the update was 31 August 2016. We only use the data on pain immediately post treatment, as for function the results are comparable and just a few studies (5-9 studies at any follow-up moment, see Appendix 1) reported follow-up data.

Analysis For the initial analyses, we performed meta-analyses using a random effects model. From the original publications we extracted data on means and standard deviations. Standardised mean differences (SMDs) and their 95% confidence intervals (CI) were used to calculate treatment effect sizes of each included study and these were presented in forest plots. Furthermore, we used the GRADE approach to grade the quality of the evidence as recommended by the Cochrane Collaboration [6,7]. For the secondary analysis, we first evaluated whether the analysis suffered from publication bias by plotting the effect estimates in a funnel plot [8,9]. Next, we performed subgroup analyses based on a priory defined subgroups to evaluate whether the effect estimate differed between subgroups [10,11]. We defined subgroups concerning a) severity of the complaint (clinical OA, clinical and radiological OA, patients on a waiting list for total knee replacement), b) duration of intervention (short: ≤ 12 weeks versus long: > 12 weeks), c) land- or water-based exercises, d) individual supervision or group exercises and e) fully supervised versus not fully supervised exercises [6]. Third, we performed a sensitivity analysis based on the overall risk of bias (RoB) score [11,12]. RoB was assessed using the tool developed by the Cochrane Back and Neck Group and defined comparable with one of the Cochrane reviews [2,12]. We a priori defined studies with low RoB when they had a clear and concealed randomisation procedure and an intention to treat analysis. Fourth, we performed a cumulative meta-analysis to see from which point in time the effect estimate reached statistical significance and to see where the effect estimate becomes stable over time and extra studies are presumed redundant [13]. For this we added studies per year in which they were published to the pooled result. All these analyses were done in RevMan (https://community.cochrane.org/help/tools-and-software/revman-5). Lastly, we developed an extended funnel plot to explore the potential impact of a new study on the summary effect estimate of a meta-analysis [14,15]. With this method regions of the funnel plot are calculated that indicate in

which region a new study would have to be located to change the effect estimate of the meta-analysis markedly; e.g. from statistically significant or clinically meaningful to non-significant/meaningful or vice versa. These analyses have been done in Stata.

RESULTS

Description of the updated systematic review

For the updated systematic review and meta-analysis, we included in total 52 studies on patient with knee OA; we excluded 26 of the studies included in both Cochrane reviews based on the stricter selection criteria and we found an additional 16 new studies after the date of last search in both Cochrane reviews (2014) [6].

In total, 42 out of the 52 studies reported the outcome measure 'pain immediately post treatment' with an effect estimate (SMD) of 0.5 (95% confidence interval (CI): 0.37 to 0.63) [6, Appendix 1]. This effect estimate is comparable with the ones presented in both Cochrane reviews.

The 42 studies included in total 6863 patients (mean 132 patients per study). The number of patients per intervention group varied between 6 and 467; 17 studies included less than 25 patients in one or more study groups, and are considered small. On average the new studies (10 out of 42) were smaller compared to the older ones (mean 119.4 patients compared to mean 137.6 patients in the older ones) included in the original Cochrane reviews [1,2].

Publication bias

In the scatterplot there is no apparent funnel in the plot as all studies have comparable sample sizes [6, Appendix 2]. The regression line is vertical, so we cautiously conclude there is no clear publication bias present.

Subgroup analysis

Only individually supervised exercises showed to be somewhat more effective (SMD=0.61 (95%CI: 0.43 to 0.80); 23 studies) compared to group exercises (SMD=0.37 (95%CI: 0.20 to 0.54); 19 studies) [6]. All other subgroups showed no statistical significant or clinical relevant differences in effect estimate. Therefore, we conclude that subgrouping did not affect the overall effect estimate, but that the effect of exercise is higher in individual exercise therapy compared to group exercise therapy.

Sensitivity analysis

Out of the 42 studies, 13 were rated as low RoB. Pooling the results of the studies with low RoB only resulted in a slightly higher effect estimate compared to the overall effect: SMD=0.54 (95%CI: 0.43 to 0.66) [6, Appendix 3]. Also, the heterogeneity decreased from 69% (moderate heterogeneity) to 5% (no heterogeneity) in the analysis of low RoB studies only. We conclude that the studies with low RoB provide more consistent estimates than studies with high-risk of bias.

Cumulative meta-analysis

Figure 1 showed that the effect estimate was statistically significant in favor of exercise from the first study onwards. Furthermore, the effect estimate levelled towards an SMD=0.5 since 2010 with a rather stable confidence interval. Therefore, we conclude that the benefit of exercise was clear since 1998 when several studies (n=5) showed consistent results. In addition, we conclude that since 2010 extra studies seemed redundant, as extra studies did not have any effect on the effect estimate nor the confidence interval.

Please add figure 1 here

Extended funnel plots

In figure 2 we show the regions where a new study needs to be located in the funnel plot to change the summary estimate from 'clearly clinically worthwhile' (set at SMD=0.37 [16]; red area) to 'clearly not worthwhile' (blue), or 'unclear if worthwhile' (green). No study, no matter how large, is able to change the current conclusion from 'clearly clinically worthwhile' (red area) to 'clearly not worthwhile' (should be the blue area) as there is no blue area in the plot. The extended funnel plot suggests that an additional study has a very limited impact to change the current effect estimate to 'unclear if worthwhile'.

Please add figure 2 here

DISCUSSION

Based on the findings of the extra analyses, we determined that the conclusion that exercise is effective and clinically worthwhile concerning pain post-treatment compared to no or minimal interventions in patients with knee OA is a robust finding and one (or more) new trials will unlikely change this conclusion. The effect estimates are larger in studies where physiotherapists treat patients individually and more consistent in the studies of low risk of bias.

We found an abundance of relatively small studies evaluating exercise compared to no or minimal treatment in patients with knee OA. New studies were even smaller compared to the older ones. One of the advantages of systematic reviews and meta-analysis is that by combining small and underpowered studies one can come to a clearer and better powered summary effect estimate. Furthermore, our cumulative meta-analysis revealed that after 2010 it can be regarded unethical randomizing patients to no or minimal intervention control groups as exercise clearly is effective on decreasing pain and can therefore be regarded as the standard treatment. Our results support the conclusion of a network meta-analysis published in 2013, although our conclusion is based on different analyses [17].

Nevertheless, our results only concern treatment effects immediately post treatment. Long-term effects were only assessed in a minority of studies. Also, in this study we are not able to make any statements on the efficacy of exercise. There does not exist any robust evidence on the efficacy of the exercise element, compared to placebo. This is mainly due to the fact that it is difficult, if not impossible, to develop a credible exercise placebo intervention. Currently other study questions regarding exercises in patients with knee OA seem to be more important, such as: 'which type of exercises for which type of patient is most effective', 'what is the additional effect of exercise in a combination of treatments', 'does exercise decrease or postpone total knee surgery', or 'what is the best strategy to implement exercise treatment into the osteoarthritis care'. Just a few studies addressing these issues are available at the moment [18,19]. Nevertheless, a search into the trial registers showed that in 2016 still nine trials comparing exercise to no or minimal treatment with pain as an outcome in patients with knee OA are being executed, all primarily planning only to assess post-treatment outcomes. Clinicians and grant organizations should be strongly discouraged in designing and granting any new trial on this study question, as it is unlikely that a new study, even a large one, would change the pooled effect estimate post-treatment.

CONCLUSION

The current conclusion that exercise, supervised by physiotherapists, is a clinically relevant and statistically significant effective intervention in reducing pain compared to no or minimal treatment for patients with knee OA is rather robust and can be accepted as 'true'. No further studies on this issue are needed as additional data will not likely be able to change this conclusion.

CONTRIBUTORS AND SOURCES

This paper is based on an update of Cochrane reviews which was requested by Zorginstituut Nederland (ZIN). ZIN coordinated the study and controlled the quality of the process (LH,IG). The study was executed by the Erasmus Medical Centre University. Several authors involved either have ample experience in performing (Cochrane) systematic reviews (APV, SMAB-Z, MM, JR), several authors are recognized experts in the field of osteoarthritis (SMAB-Z, JR, MM), two authors helped with the data-extraction (CHT, EAER-V) and one author helped with the secondary analysis (MF). The first author (APV) is the guarantor of this manuscript.

ACKNOWLEDGEMENTS:

There are no other contributors to this work to be acknowledged, nor did we receive funding for this study.

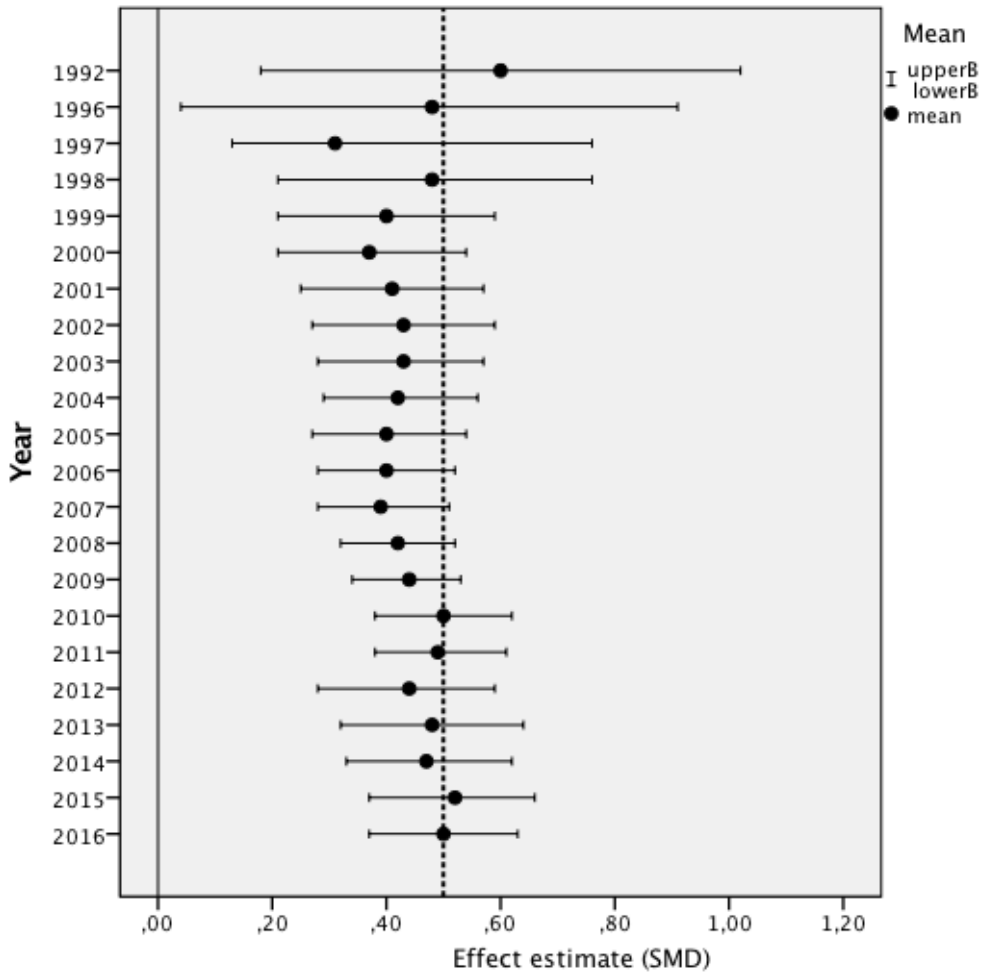
CONFLICT OF INTEREST: all authors declare they have no conflict of interest

References

1. Bartels EM, Juhl CB, Christensen R, Hagen KB, Danneskiold-Samsøe B, Dagfinrud H, Lund H. Aquatic exercise for the treatment of knee and hip osteoarthritis. Cochrane Database Syst Rev. 201623;3:CD005523.
2. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev. 2015;1:CD004376.
3. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum. 2014;43(6):701-12.
4. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, Hawker GA, Henrotin Y, Hunter DJ, Kawaguchi H, Kwoh K, Lohmander S, Rannou F, Roos EM, Underwood M. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014;22(3):363-88.
5. NICE, 2014. Osteoarthritis: Care and Management in adults (www.nice.org.uk/Guidance/CG177)
6. ZIN: https://www.zorginstituutnederland.nl/actueel/nieuws/2017/03/13/advies-zorginstituut-oefentherapie-bij-artrose-in-basispakket
7. Guyatt GH, Oxman AD, Vist GE, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6
8. Egger M, Smith GD, Scheider M, Minder C. Bias in meta-analysis detected by a simple graphical test. BMJ 1997;315:629–34.
9. Mavridis D, Salanti G. Exploring and accounting for publication bias in mental health: a brief overview of methods. Evid Based Ment Health. 2014;17(1):11-5.
10. Song F, Bachmann MO. Cumulative subgroup analysis to reduce waste in clinical research for individualised medicine. BMC Med. 2016;14(1):197.
11. Cochrane Handbook: https://training.cochrane.org/handbook
12. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, Bronfort G, van Tulder MW; Editorial Board of the Cochrane Back, Neck Group. 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. Spine 2015;40:1660-73.
13. Muellerleile P, Mullen B. Sufficiency and stability of evidence for public health interventions using cumulative meta-analysis. Am J Public Health. 2006;96(3):515-22.
14. Langan D, Higgins JP, Gregory W, Sutton AJ. Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis. Journal of Clinical Epidemiology 2012;65:511-19.
15. Ferreira ML, Herbert RD, Crowther MJ, Verhagen A, Sutton AJ. When is a further clinical trial justified? BMJ. 2012;345:e5913
16. Wandel S, Jüni P, Tendal B, Nüesch E, Villiger PM, Welton NJ, Reichenbach S, Trelle S. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. BMJ. 2010;341:c4675.

17. Uthman OA, van der Windt DA, Jordan JL, Dziedzic KS, Healey EL, Peat GM, Foster NE. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. BMJ. 2013 Sep 20;347:f5555. doi: 10.1136/bmj.f5555.
18. Skou ST, Roos EM, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O, Rasmussen S. A Randomized, Controlled Trial of Total Knee Replacement. N Engl J Med. 2015;373(17):1597-606.
19. Svege I, Nordsletten L, Fernandes L, Risberg MA. Exercise therapy may postpone total hip replacement surgery in patients with hip osteoarthritis: a long-term follow-up of a randomised trial. Ann Rheum Dis. 2015;74(1):164-9.

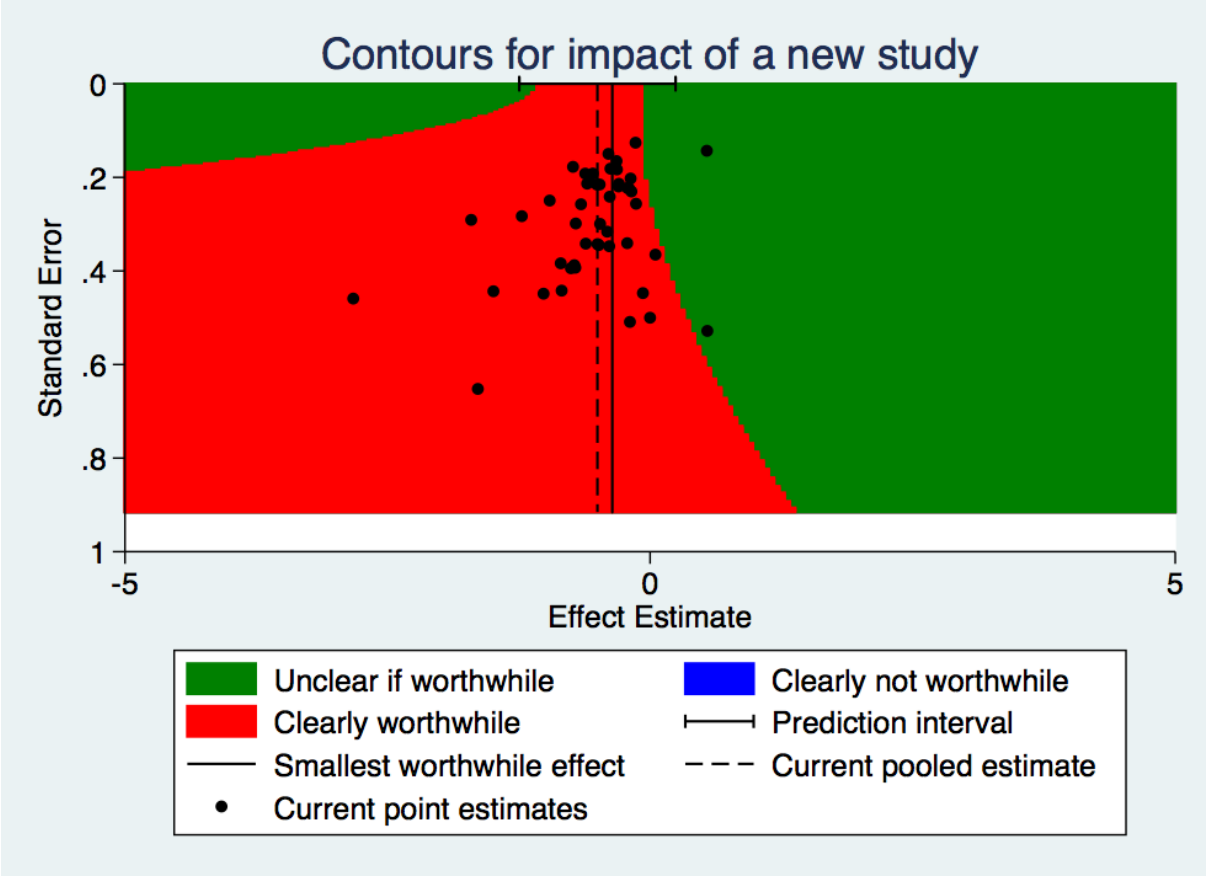
Figure 1: Cumulative meta-analysis



Straight line: no effect

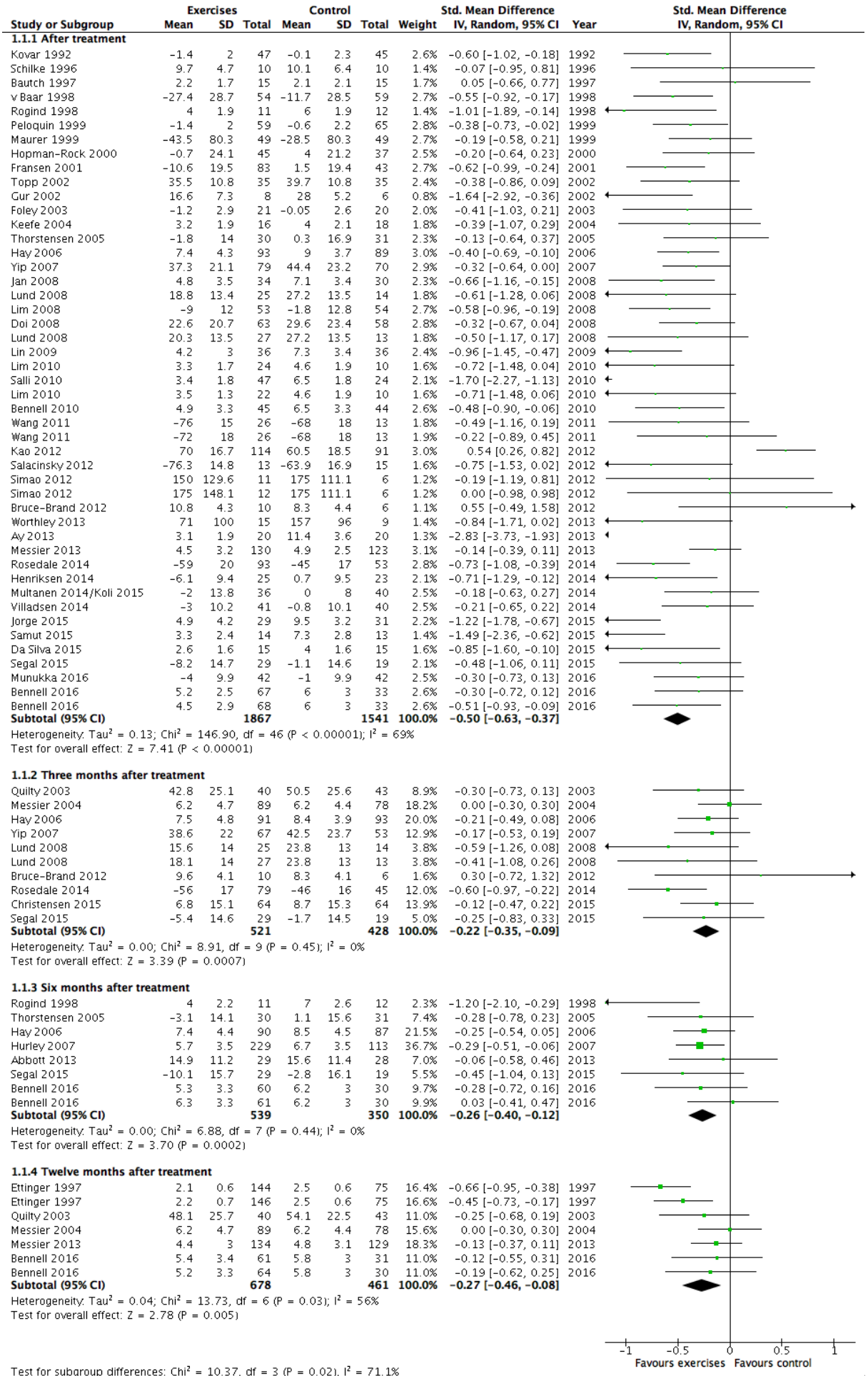
Dotted line: overall effect estimate

Figure 2: Extended funnel plot



290 Appendix 1: Effectiveness of exercise on pain, post treatment

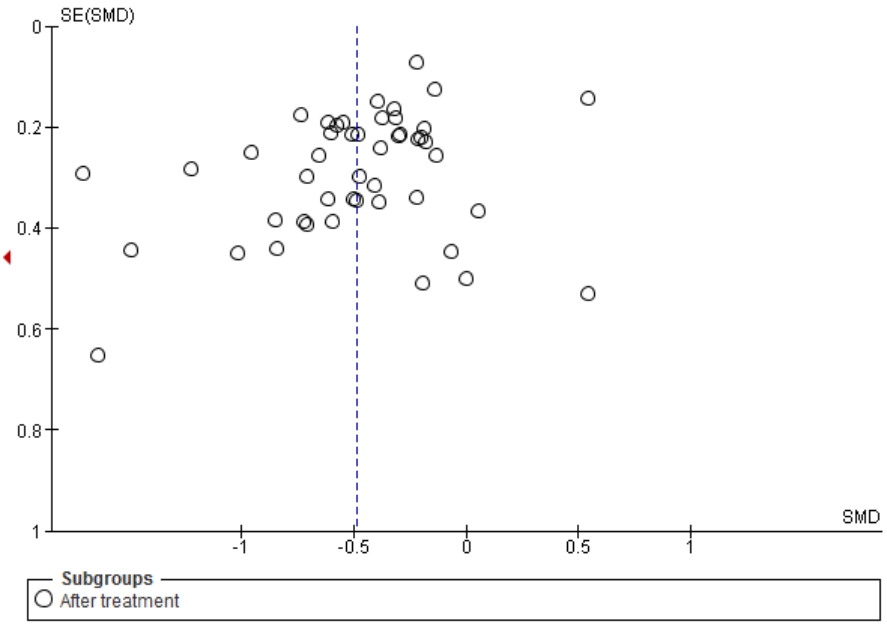
291


Exercises
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI Year
IV, Random, 95% CI
1.1.1 After treatment
Kovar 1992 -1.4 2 47 -0.1 2.3 45 2.6% -0.60 [-1.02, -0.18] 1992
Schilke 1996 9.7 4.7 10 10.1 6.4 10 1.4% -0.07 [-0.95, 0.81] 1996
Bautch 1997 2.2 1.7 15 2.1 2.1 15 1.7% 0.05 [-0.66, 0.77] 1997
v Baar 1998 -27.4 28.7 54 -11.7 28.5 59 2.7% -0.55 [-0.92, -0.17] 1998
Rogind 1998 4 1.9 11 6 1.9 12 1.4% -1.01 [-1.89, -0.14] 1998
Peloquin 1999 -1.4 2 59 -0.6 2.2 65 2.8% -0.38 [-0.73, -0.02] 1999
Maurer 1999 -43.5 80.3 49 -28.5 80.3 49 2.7% -0.19 [-0.58, 0.21] 1999
Hopman-Rock 2000 -0.7 24.1 45 4 21.2 37 2.5% -0.20 [-0.64, 0.23] 2000
Fransen 2001 -10.6 19.5 83 1.5 19.4 43 2.7% -0.62 [-0.99, -0.24] 2001
Topp 2002 35.5 10.8 35 39.7 10.8 35 2.4% -0.38 [-0.86, 0.09] 2002
Gur 2002 16.6 7.3 8 28 5.2 6 0.8% -1.64 [-2.92, -0.36] 2002
Foley 2003 -1.2 2.9 21 -0.05 2.6 20 2.0% -0.41 [-1.03, 0.21] 2003
Keefe 2004 3.2 1.9 16 4 2.1 18 1.8% -0.39 [-1.07, 0.29] 2004
Thorstensen 2005 -1.8 14 30 0.3 16.9 31 2.3% -0.13 [-0.64, 0.37] 2005
Hay 2006 7.4 4.3 93 9 3.7 89 3.0% -0.40 [-0.69, -0.10] 2006
Yip 2007 37.3 21.1 79 44.4 23.2 70 2.9% -0.32 [-0.64, 0.00] 2007
Jan 2008 4.8 3.5 34 7.1 3.4 30 2.3% -0.66 [-1.16, -0.15] 2008
Lund 2008 18.8 13.4 25 27.2 13.5 14 1.8% -0.61 [-1.28, 0.06] 2008
Lim 2008 -9 12 53 -1.8 12.8 54 2.7% -0.58 [-0.96, -0.19] 2008
Doi 2008 22.6 20.7 63 29.6 23.4 58 2.8% -0.32 [-0.67, 0.04] 2008
Lund 2008 20.3 13.5 27 27.2 13.5 13 1.8% -0.50 [-1.17, 0.17] 2008
Lin 2009 4.2 3 36 7.3 3.4 36 2.4% -0.96 [-1.45, -0.47] 2009
Lim 2010 3.3 1.7 24 4.6 1.9 10 1.6% -0.72 [-1.48, 0.04] 2010
Salli 2010 3.4 1.8 47 6.5 1.8 24 2.1% -1.70 [-2.27, -1.13] 2010
Lim 2010 3.5 1.3 22 4.6 1.9 10 1.6% -0.71 [-1.48, 0.06] 2010
Bennell 2010 4.9 3.3 45 6.5 3.3 44 2.6% -0.48 [-0.90, -0.06] 2010
Wang 2011 -76 15 26 -68 18 13 1.8% -0.49 [-1.16, 0.19] 2011
Wang 2011 -72 18 26 -68 18 13 1.9% -0.22 [-0.89, 0.45] 2011
Kao 2012 70 16.7 114 60.5 18.5 91 3.0% 0.54 [0.26, 0.82] 2012
Salacinsky 2012 -76.3 14.8 13 -63.9 16.9 15 1.6% -0.75 [-1.53, 0.02] 2012
Simao 2012 150 129.6 11 175 111.1 6 1.2% -0.19 [-1.19, 0.81] 2012
Simao 2012 175 148.1 12 175 111.1 6 1.2% 0.00 [-0.98, 0.98] 2012
Bruce-Brand 2012 10.8 4.3 10 8.3 4.4 6 1.1% 0.55 [-0.49, 1.58] 2012
Worthley 2013 71 100 15 157 96 9 1.4% -0.84 [-1.71, 0.02] 2013
Ay 2013 3.1 1.9 20 11.4 3.6 20 1.3% -2.83 [-3.73, -1.93] 2013
Messier 2013 4.5 3.2 130 4.9 2.5 123 3.1% -0.14 [-0.39, 0.11] 2013
Rosedale 2014 -59 20 93 -45 17 53 2.8% -0.73 [-1.08, -0.39] 2014
Henriksen 2014 -6.1 9.4 25 0.7 9.5 23 2.1% -0.71 [-1.29, -0.12] 2014
Multanen 2014/Koli 2015 -2 13.8 36 0 8 40 2.5% -0.18 [-0.63, 0.27] 2014
Villadsen 2014 -3 10.2 41 -0.8 10.1 40 2.5% -0.21 [-0.65, 0.22] 2014
Jorge 2015 4.9 4.2 29 9.5 3.2 31 2.2% -1.22 [-1.78, -0.67] 2015
Samut 2015 3.3 2.4 14 7.3 2.8 13 1.4% -1.49 [-2.36, -0.62] 2015
Da Silva 2015 2.6 1.6 15 4 1.6 15 1.6% -0.85 [-1.60, -0.10] 2015
Segal 2015 -8.2 14.7 29 -1.1 14.6 19 2.1% -0.48 [-1.06, 0.11] 2015
Munukka 2016 -4 9.9 42 -1 9.9 42 2.5% -0.30 [-0.73, 0.13] 2016
Bennell 2016 5.2 2.5 67 6 3 33 2.6% -0.30 [-0.72, 0.12] 2016
Bennell 2016 4.5 2.9 68 6 3 33 2.6% -0.51 [-0.93, -0.09] 2016
Subtotal (95% CI) 1867 1541 100.0% -0.50 [-0.63, -0.37]
Heterogeneity: Tau² = 0.13; Chi² = 146.90, df = 46 (P < 0.00001); I² = 69%
Test for overall effect: Z = 7.41 (P < 0.00001)
1.1.2 Three months after treatment
Quilty 2003 42.8 25.1 40 50.5 25.6 43 8.9% -0.30 [-0.73, 0.13] 2003
Messier 2004 6.2 4.7 89 6.2 4.4 78 18.2% 0.00 [-0.30, 0.30] 2004
Hay 2006 7.5 4.8 91 8.4 3.9 93 20.0% -0.21 [-0.49, 0.08] 2006
Yip 2007 38.6 22 67 42.5 23.7 53 12.9% -0.17 [-0.53, 0.19] 2007
Lund 2008 15.6 14 25 23.8 13 14 3.8% -0.59 [-1.26, 0.08] 2008
Lund 2008 18.1 14 27 23.8 13 13 3.8% -0.41 [-1.08, 0.26] 2008
Bruce-Brand 2012 9.6 4.1 10 8.3 4.1 6 1.6% 0.30 [-0.72, 1.32] 2012
Rosedale 2014 -56 17 79 -46 16 45 12.0% -0.60 [-0.97, -0.22] 2014
Christensen 2015 6.8 15.1 64 8.7 15.3 64 13.9% -0.12 [-0.47, 0.22] 2015
Segal 2015 -5.4 14.6 29 -1.7 14.5 19 5.0% -0.25 [-0.83, 0.33] 2015
Subtotal (95% CI) 521 428 100.0% -0.22 [-0.35, -0.09]
Heterogeneity: Tau² = 0.00; Chi² = 8.91, df = 9 (P = 0.45); I² = 0%
Test for overall effect: Z = 3.39 (P = 0.0007)
1.1.3 Six months after treatment
Rogind 1998 4 2.2 11 7 2.6 12 2.3% -1.20 [-2.10, -0.29] 1998
Thorstensen 2005 -3.1 14.1 30 1.1 15.6 31 7.4% -0.28 [-0.78, 0.23] 2005
Hay 2006 7.4 4.4 90 8.5 4.5 87 21.5% -0.25 [-0.54, 0.05] 2006
Hurley 2007 5.7 3.5 229 6.7 3.5 113 36.7% -0.29 [-0.51, -0.06] 2007
Abbott 2013 14.9 11.2 29 15.6 11.4 28 7.0% -0.06 [-0.58, 0.46] 2013
Segal 2015 -10.1 15.7 29 -2.8 16.1 19 5.5% -0.45 [-1.04, 0.13] 2015
Bennell 2016 5.3 3.3 60 6.2 3 30 9.7% -0.28 [-0.72, 0.16] 2016
Bennell 2016 6.3 3.3 61 6.2 3 30 9.9% 0.03 [-0.41, 0.47] 2016
Subtotal (95% CI) 539 350 100.0% -0.26 [-0.40, -0.12]
Heterogeneity: Tau² = 0.00; Chi² = 6.88, df = 7 (P = 0.44); I² = 0%
Test for overall effect: Z = 3.70 (P = 0.0002)
1.1.4 Twelve months after treatment
Ettinger 1997 2.1 0.6 144 2.5 0.6 75 16.4% -0.66 [-0.95, -0.38] 1997
Ettinger 1997 2.2 0.7 146 2.5 0.6 75 16.6% -0.45 [-0.73, -0.17] 1997
Quilty 2003 48.1 25.7 40 54.1 22.5 43 11.0% -0.25 [-0.68, 0.19] 2003
Messier 2004 6.2 4.7 89 6.2 4.4 78 15.6% 0.00 [-0.30, 0.30] 2004
Messier 2013 4.4 3 134 4.8 3.1 129 18.3% -0.13 [-0.37, 0.11] 2013
Bennell 2016 5.4 3.4 61 5.8 3 31 11.0% -0.12 [-0.55, 0.31] 2016
Bennell 2016 5.2 3.3 64 5.8 3 30 11.0% -0.19 [-0.62, 0.25] 2016
Subtotal (95% CI) 678 461 100.0% -0.27 [-0.46, -0.08]
Heterogeneity: Tau² = 0.04; Chi² = 13.73, df = 6 (P = 0.03); I² = 56%
Test for overall effect: Z = 2.78 (P = 0.005)
Test for subgroup differences: Chi² = 10.37, df = 3 (P = 0.02), I² = 71.1%
-1 -0.5 0 0.5 1
Favours exercises Favours control

Risk of bias legend:

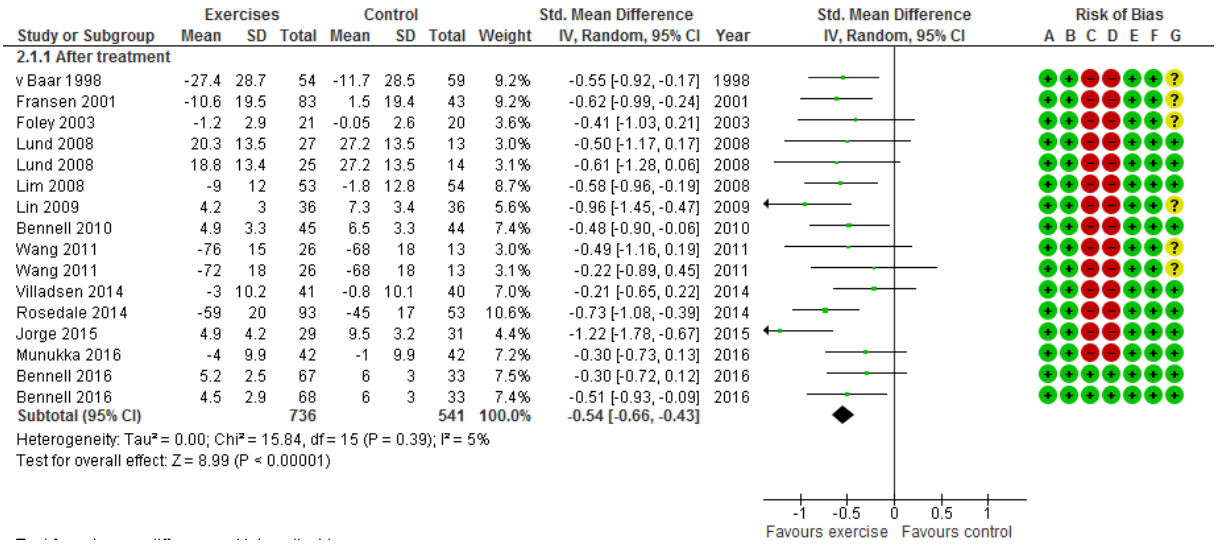
A: random sequence generation (selection bias); B: allocation concealment (selection bias); C: blinding of participants and personnel (performance bias); D: blinding of outcome assessment (detection bias); E: incomplete outcome data (attrition bias); F: selective reporting (reporting bias); G other bias

Appendix 2: Funnel plot for publication bias, pain post treatment



Dotted line = regression line

Appendix 3: Effectiveness of exercise in studies with low risk of bias, pain post treatment



Study or Subgroup	Exercises Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Year
2.1.1 After treatment									
v Baar 1998	-27.4	28.7	54	-11.7	28.5	59	9.2%	-0.55 [-0.92, -0.17]	1998
Fransen 2001	-10.6	19.5	83	1.5	19.4	43	9.2%	-0.62 [-0.99, -0.24]	2001
Foley 2003	-1.2	2.9	21	-0.05	2.6	20	3.6%	-0.41 [-1.03, 0.21]	2003
Lund 2008	20.3	13.5	27	27.2	13.5	13	3.0%	-0.50 [-1.17, 0.17]	2008
Lund 2008	18.8	13.4	25	27.2	13.5	14	3.1%	-0.61 [-1.28, 0.06]	2008
Lim 2008	-9	12	53	-1.8	12.8	54	8.7%	-0.58 [-0.96, -0.19]	2008
Lin 2009	4.2	3	36	7.3	3.4	36	5.6%	-0.96 [-1.45, -0.47]	2009
Bennell 2010	4.9	3.3	45	6.5	3.3	44	7.4%	-0.48 [-0.90, -0.06]	2010
Wang 2011	-76	15	26	-68	18	13	3.0%	-0.49 [-1.16, 0.19]	2011
Wang 2011	-72	18	26	-68	18	13	3.1%	-0.22 [-0.89, 0.45]	2011
Villadsen 2014	-3	10.2	41	-0.8	10.1	40	7.0%	-0.21 [-0.65, 0.22]	2014
Rosedale 2014	-59	20	93	-45	17	53	10.6%	-0.73 [-1.08, -0.39]	2014
Jorge 2015	4.9	4.2	29	9.5	3.2	31	4.4%	-1.22 [-1.78, -0.67]	2015
Munukka 2016	-4	9.9	42	-1	9.9	42	7.2%	-0.30 [-0.73, 0.13]	2016
Bennell 2016	5.2	2.5	67	6	3	33	7.5%	-0.30 [-0.72, 0.12]	2016
Bennell 2016	4.5	2.9	68	6	3	33	7.4%	-0.51 [-0.93, -0.09]	2016
Subtotal (95% CI)			**736**			**541**	**100.0%**	**-0.54 [-0.66, -0.43]**	

Heterogeneity: Tau² = 0.00; Chi² = 15.84, df = 15 (P = 0.39); I² = 5%
Test for overall effect: Z = 8.99 (P < 0.00001)

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 other bias