Cochrane Library

Cochrane Database of Systematic Reviews

Exercise for osteoarthritis of the knee (Review)

Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL

Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL.
Exercise for osteoarthritis of the knee.
Cochrane Database of Systematic Reviews 2015, Issue 1. Art. No.: CD004376.
DOI: 10.1002/14651858.CD004376.pub3.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Exercise for osteoarthritis of the knee

Marlene Fransen[1], Sara McConnell[2], Alison R Harmer[3], Martin Van der Esch[4], Milena Simic[5], Kim L Bennell[6]

[1]Faculty of Health Sciences, University of Sydney, Sydney, Australia. [2]Department of Medicine, St Joseph's Health Care Centre, Toronto, Canada. [3]Faculty of Health Sciences, Clinical and Rehabilitation Sciences Research Group, University of Sydney, Sydney, Australia. [4]Department of Rehabilitation, Reade, Centre for Rehabilitation and Rheumatology, Amsterdam, Netherlands. [5]Discipline of Physiotherapy, Faculty of Health Sciences, The University of Sydney, Lidcombe, Australia. [6]Department of Physiotherapy, Melbourne School of Health Sciences, The University of Melbourne, Melbourne, Australia

Contact: Marlene Fransen, Faculty of Health Sciences, University of Sydney, Room 0212, Cumberland Campus C42, Sydney, New South Wales, 1825, Australia. marlene.fransen@sydney.edu.au.

Editorial group: Cochrane Musculoskeletal Group.
Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 1, 2015.

Citation: Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews* 2015, Issue 1. Art. No.: CD004376. DOI: 10.1002/14651858.CD004376.pub3.

ABSTRACT

Background

Knee osteoarthritis (OA) is a major public health issue because it causes chronic pain, reduces physical function and diminishes quality of life. Ageing of the population and increased global prevalence of obesity are anticipated to dramatically increase the prevalence of knee OA and its associated impairments. No cure for knee OA is known, but exercise therapy is among the dominant non-pharmacological interventions recommended by international guidelines.

Objectives

To determine whether land-based therapeutic exercise is beneficial for people with knee OA in terms of reduced joint pain or improved physical function and quality of life.

Search methods

Five electronic databases were searched, up until May 2013.

Selection criteria

All randomised controlled trials (RCTs) randomly assigning individuals and comparing groups treated with some form of land-based therapeutic exercise (as opposed to exercise conducted in the water) with a non-exercise group or a non-treatment control group.

Data collection and analysis

Three teams of two review authors independently extracted data, assessed risk of bias for each study and assessed the quality of the body of evidence for each outcome using the GRADE (Grades of Recommendation, Assessment, Development and Evaluation) approach. We conducted analyses on continuous outcomes (pain, physical function and quality of life) immediately after treatment and on dichotomous outcomes (proportion of study withdrawals) at the end of the study; we also conducted analyses on the sustained effects of exercise on pain and function (two to six months, and longer than six months).

Main results

In total, we extracted data from 54 studies. Overall, 19 (20%) studies reported adequate random sequence generation and allocation concealment and adequately accounted for incomplete outcome data; we considered these studies to have an overall low risk of bias. Studies were largely free from selection bias, but research results may be vulnerable to performance and detection bias, as only four of

the RCTs reported blinding of participants to treatment allocation, and, although most RCTs reported blinded outcome assessment, pain, physical function and quality of life were participant self-reported.

High-quality evidence from 44 trials (3537 participants) indicates that exercise reduced pain (standardised mean difference (SMD) -0.49, 95% confidence interval (CI) -0.39 to -0.59) immediately after treatment. Pain was estimated at 44 points on a 0 to 100-point scale (0 indicated no pain) in the control group; exercise reduced pain by an equivalent of 12 points (95% CI 10 to 15 points). Moderate-quality evidence from 44 trials (3913 participants) showed that exercise improved physical function (SMD -0.52, 95% CI -0.39 to -0.64) immediately after treatment. Physical function was estimated at 38 points on a 0 to 100-point scale (0 indicated no loss of physical function) in the control group; exercise improved physical function by an equivalent of 10 points (95% CI 8 to 13 points). High-quality evidence from 13 studies (1073 participants) revealed that exercise improved quality of life (SMD 0.28, 95% CI 0.15 to 0.40) immediately after treatment. Quality of life was estimated at 43 points on a 0 to 100-point scale (100 indicated best quality of life) in the control group; exercise improved quality of life by an equivalent of 4 points (95% CI 2 to 5 points).

High-quality evidence from 45 studies (4607 participants) showed a comparable likelihood of withdrawal from exercise allocation (event rate 14%) compared with the control group (event rate 15%), and this difference was not significant: odds ratio (OR) 0.93 (95% CI 0.75 to 1.15). Eight studies reported adverse events, all of which were related to increased knee or low back pain attributed to the exercise intervention provided. No study reported a serious adverse event.

In addition, 12 included studies provided two to six-month post-treatment sustainability data on 1468 participants for knee pain and on 1279 (10 studies) participants for physical function. These studies indicated sustainability of treatment effect for pain (SMD -0.24, 95% CI -0.35 to -0.14), with an equivalent reduction of 6 (3 to 9) points on 0 to 100-point scale, and of physical function (SMD -0.15 95% CI -0.26 to -0.04), with an equivalent improvement of 3 (1 to 5) points on 0 to 100-point scale.

Marked variability was noted across included studies among participants recruited, symptom duration, exercise interventions assessed and important aspects of study methodology. Individually delivered programmes tended to result in greater reductions in pain and improvements in physical function, compared to class-based exercise programmes or home-based programmes; however between-study heterogeneity was marked within the individually provided treatment delivery subgroup.

Authors' conclusions

High-quality evidence indicates that land-based therapeutic exercise provides short-term benefit that is sustained for at least two to six months after cessation of formal treatment in terms of reduced knee pain, and moderate-quality evidence shows improvement in physical function among people with knee OA. The magnitude of the treatment effect would be considered moderate (immediate) to small (two to six months) but comparable with estimates reported for non-steroidal anti-inflammatory drugs. Confidence intervals around demonstrated pooled results for pain reduction and improvement in physical function do not exclude a minimal clinically important treatment effect. Since the participants in most trials were aware of their treatment, this may have contributed to their improvement. Despite the lack of blinding we did not downgrade the quality of evidence for risk of performance or detection bias. This reflects our belief that further research in this area is unlikely to change the findings of our review.

PLAIN LANGUAGE SUMMARY

Exercise for osteoarthritis of the knee

Background: What is OA of the knee, and what is exercise?

Osteoarthritis (OA) is a disease of joints, such as the hip. When the joint loses cartilage, the bone grows to try to repair the damage. However, instead of making things better, the bone grows abnormally and makes things worse. For example, the bone can become misshapen and make the joint painful and unstable. Doctors used to think that OA simply resulted in thinning of the cartilage. However, it is now known that OA is a disease of the whole joint.

Exercise can be any activity that enhances or maintains muscle strength, physical fitness and overall health. People exercise for many reasons; they may exercise to lose weight, to strengthen muscles or to relieve the symptoms of OA.

Study characteristics

This summary of an update of a Cochrane review presents what we know from research about the effects of exercise for people with OA of the knee. After searching for all relevant studies up to May 2013, we added 23 new studies since the last version of the review, now including 54 studies (3913 participants), most on mild to moderate symptomatic knee OA. Except for five studies in which participants enrolled in a Tai Chi-based programme, most participants underwent land-based exercise programmes consisting of traditional muscle strengthening, functional training and aerobic fitness programmes, which were individually supervised or were provided during a class; these individuals were compared with people who did not exercise. Evidence from 44 studies (3537 participants) shows the effects of exercise immediately after treatment; 12 studies provided data on two to six-month post-treatment sustainability. Here we report only results for the immediate treatment period.

Key results

Pain on a scale of 0 to 100 points (lower scores mean reduced pain).

• People who completed an exercise programme rated their pain at 12 (10 to 15) points lower at end of treatment (12% absolute improvement) compared with people who did not exercise.

• People who completed an exercise programme rated their pain at 32 points.

• People who did not exercise rated their pain at 44 points.

Physical function on a scale of 0 to 100 points (lower score means better physical function).

• People who completed an exercise programme rated their physical function at 10 points (8 to 13 points) lower at end of treatment (10% absolute improvement) compared with people who did not exercise.

• People who completed an exercise programme rated their physical function at 28 points.

• People who did not exercise rated their physical function at 38 points.

Quality of life on a scale of 0 to 100 points (higher score means better quality of life).

• Overall, people who completed an exercise programme rated their quality of life at 4 points (2 to 5 points) higher at the end of treatment (4% absolute improvement).

• People who completed an exercise programme rated their quality of life at 47 points.

• People who did not exercise rated their quality of life at 43 points.

Withdrawals.

• One fewer persons out of 100 dropped out of the exercise programme (1% absolute decrease).

• Out of 100 people in exercise programmes, 14 dropped out.

• Out of 100 people who did not exercise, 15 dropped out.

Quality of the evidence

High-quality evidence shows that among people with knee OA, exercise moderately reduced pain immediately after cessation of treatment and improved quality of life only slightly, without an increase in dropouts. Further research is unlikely to change the estimate of these results.

Moderate-quality evidence indicates that exercise moderately improved physical function immediately after cessation of treatment. Further research may change the estimate of these results.

Most clinical studies have provided no precise information on side effects such as injuries or falls sustained during exercise, but we would expect these to be rare. Eight studies reported increased knee or low back pain attributed to the exercise programme, and all identified studies reported no injuries.

SUMMARY OF FINDINGS

Summary of findings for the main comparison. Immediate post-treatment effects of exercise for osteoarthritis of the knee

Immediate post-treatment effects of exercise for osteoarthritis of the knee

Patient or population: patients with knee OA
Settings: clinic or community
Intervention: land-based exercise
Comparison: no exercise

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	No exercise	Land-based exercise				
Pain Self-report questionnaires. Scale from 0-100 (0 represents no pain)	Mean pain in the control groups was **44 points**	Mean pain in intervention groups was **0.49 standard deviations lower** (0.39-0.59 lower) This translates to an absolute mean reduction of 12 (10-15) points compared with control group on a 0-100 scale[a]		3537 (44 studies)	⊕⊕⊕⊕ **High**	SMD -0.49 (-0.39 to -0.59) Absolute reduction in pain 12% (10%-15%); relative change 27% (21%-32%)[a] NNTB 4 (3-5)[b]
Physical function Self-report questionnaire. Scale from 0-100 (0 represents no physical disability)	Mean physical function in control groups was **38 points**	Mean physical function in intervention groups was **0.52 standard deviations lower** (0.39-0.64 lower) This translates to an absolute mean improvement of 10 (8-13) points on a 0-100 scale[c]		3913 (44 studies)	⊕⊕⊕⊖ **Moderate** [d]	SMD -0.52 (-0.39 to -0.64) Absolute improvement 10% (8%-13%); relative improvement 26% (20%-32%)[c] NNTB 4 (3-5)[b]
Quality of life Self-report questionnaire. Scale from 0-100 (100 is maximum quality of life)	Mean quality of life in control groups was **43 points**	Mean quality of life in intervention groups was **0.28 standard deviations higher** (0.15-0.4 higher) This translates to an absolute improvement of 4 (2-5) points on a 0-100 scale[e]		1073 (13 studies)	⊕⊕⊕⊕ **High**	SMD 0.28 (0.15-0.40) Absolute improvement 4% (2%-5%); relative improvement 9% (5%-13%)[e] NNTB 8 (5-14)[b]

Study withdrawals or dropouts	153 per 1000	137 per 1000		4607 (44 studies)	⊕⊕⊕⊕ **High**	OR 0.93 (0.75-1.15) Absolute risk reduction: 1% fewer events with exercise (2% fewer-2% more); relative risk reduction 6% fewer events with exercise (21% fewer-12% more) NNTH n/a[b]

*The basis for the **assumed risk** (e.g. median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **GRADE:** Grades of Recommendation, Assessment, Development and Evaluation; **KOOS:** Knee Osteoarthritis Outcome Scale; **NNTB:** Number needed to treat for an additional beneficial outcome; **NNTH:** Number needed to treat for an additional harmful outcome; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence.
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[a]Calculations based on the control group baseline mean (SD) pain: 44.3 (24.4) points on 0-100 scale (from Yip 2007).
[b]Number needed to treat for an additional beneficial outcome (NNTB) or harmful outcome (NNTH) not applicable (n/a) when result was not statistically significant. Number needed to treat (NNT) for continuous outcomes calculated using the Wells calculator (from the CMSG Editorial office; http://musculoskeletal.cochrane.org/), and for dichotomous outcomes using the Cates NNT calculator (www.nntonline.net/visualrx/).
[c]Calculations based on the control group baseline mean (SD) function: 40.0 (20.0) points on 0-100 scale (from Hurley 2007).
[d]Physical function downgraded for inconsistency (heterogeneity, I^2 = 68%).
[e]Calculated on the basis of the control group baseline mean (SD): 39.2 (13.1) points on 0-100 KOOS subscale (from Lund 2008).

BACKGROUND

Description of the condition

Osteoarthritis (OA), the most common rheumatic disease, primarily affects the articular cartilage and the subchondral bone of a synovial joint, eventually resulting in joint failure. The most typical radiographic features include formation of osteophytes at the joint margins, joint space narrowing, subchondral sclerosis, subchondral cyst formation and chondrocalcinosis (Scott 1993). It has been estimated that about 40% to 80% of people with radiographic changes will have symptomatic disease. Symptomatic knee OA is highly prevalent among older people worldwide (10% to 30%), especially in rural regions, where occupational physical demands are high (Busija 2010).

People with symptomatic OA of the knee describe deep, aching pain. In early disease, pain is intermittent and most often is associated with joint use. For many people, symptomatic disease progresses, and the pain becomes more chronic and may occur at rest and during the night. The joint feels 'stiff,' resulting in typical pain and difficulty when movement is initiated after a period of rest. Individuals with advanced disease may experience crepitus or deep 'creaking' sounds on movement and often limited range of joint motion. People with progressive symptomatic knee OA experience increasing difficulty with daily functional activities. In fact, knee OA is more responsible than any other disease for disability in walking, stair climbing and housekeeping among non-institutionalised people 50 years of age and older (Davis 1991; Guccione 1994; van Dijk 2006). Ultimately, chronic OA involving lower limb joints leads to reduced physical fitness with resultant increased risk of cardiometabolic co-morbidity (Minor 1988; Philbin 1995; Nielen 2012) and early mortality (Hochberg 2008).

Description of the intervention

Therapeutic exercise covers a range of targeted physical activities that directly aim to improve muscle strength, joint range of motion and aerobic fitness.

How the intervention might work

Currently, no cure for OA is known. However, disease-related factors, such as impaired muscle function and reduced fitness, are potentially amenable to exercise (Buchner 1992; Fiatarone 1993).

Exercise takes a multitude of forms and results in numerous systemic and local effects, some of which have been investigated among people with knee OA.

Among people with knee OA, improving muscle strength is one of the main aims of exercise, given that weakness is common. Strength training of sufficient dosage can address muscle weakness by improving muscle mass and/or recruitment. However, among patient groups, pain must be considered and may be a barrier leading to underdosage of the strength stimulus. Enhanced strength of the lower limb may lessen knee forces, reduce pain and improve physical function (Bennell 2008; Dekker 2013). Increased muscle strength may modify biomechanics, resulting in a decreased joint loading rate or localised stress in the articular cartilage, thereby playing an important role in both initiation and progression of knee OA (Cooper 1995; Felson 1995; Kujala 1995; McAlindon 1999; Rangger 1995; Slemenda 1997; Zhang 1996).

Poor physical fitness is another impairment reported among people with knee OA. Physiological reserve for aerobic capacity is enhanced primarily by increasing muscle oxidative capacity. Aerobic exercise (e.g. walking, cycling) of sufficient intensity increases muscle oxidative enzymes and muscle capillarisation, hence increasing peak oxygen uptake. Higher oxygen uptake is inversely related to morbidity and mortality and renders every submaximal daily task easier (in terms of effort). Thus, improved fitness may enhance quality of life by allowing a greater range of available daily tasks, thereby improving physical function.

Why it is important to do this review

International guidelines advocate various non-pharmacological treatments, including exercise, for first-line treatment of people with OA (Zhang 2010 Nelson 2013). This is an update of a previous Cochrane review (Fransen 2008).

OBJECTIVES

To determine whether land-based therapeutic exercise is beneficial for people with knee OA in terms of reduced joint pain or improved physical function and quality of life.

METHODS

Criteria for considering studies for this review

Types of studies

Randomised or quasi-randomised controlled trials, published in the English language, comparing groups given some form of land-based therapeutic exercise versus a non-exercise group.

Types of participants

Male and female adults given an established diagnosis of knee OA according to accepted criteria (Altman 1991), or who self-reported knee OA on the basis of chronic joint pain (with or without radiographic confirmation).

Types of interventions

Any land-based non-perioperative therapeutic exercise regimens aimed at relieving the symptoms of OA, regardless of content, duration, frequency or intensity. The comparator (control) group could be an active (given any non-exercise intervention) or no treatment (including waiting list) group.

Types of outcome measures

In accordance with international consensus regarding the core set of outcome measures for phase III clinical trials in OA (Bellamy 1997), each randomised clinical trial had to include assessment of at least one of the following.

1. Knee pain.
2. Self-reported physical function.
3. Quality of life.

These outcomes were assessed at three time points: immediately at the end of treatment (post-treatment), two to six months after cessation of monitored study treatment and longer than six months after cessation of monitored study treatment. Each included study was required to report measurement of outcomes in at least one of these time periods.

We also noted the number of participants withdrawing from the study before post-treatment assessment and the number of participants experiencing adverse events, if provided.

Search methods for identification of studies

Electronic searches

Five electronic databases were searched from inception to May 2013: MEDLINE (Appendix 1), EMBASE (Appendix 2), the Cochrane Central Register of Controlled Trials (CENTRAL) (Appendix 3), the Cumulative Index to Nursing and Allied Health Literature (CINAHL) (Appendix 4) and the Physiotherapy Evidence Database (PEDro) (Appendix 5).

We also included a search of ClinicalTrials.gov (www.ClinicalTrials.gov) and the World Health Organization (WHO) trials portal (www.who.int/ictrp/en/).

Searching other resources

We searched the reference lists of identified included studies as well.

Data collection and analysis

Selection of studies

Three teams of two review authors (MF, SM, AH, MVdE, MS, KB) independently screened retrieved clinical studies for inclusion. If agreement was not achieved at any stage, a third review author from one of the other two teams adjudicated.

Data extraction and management

Three teams of two review authors (MF, SM, AH, MVdE, MS, KB) extracted data from all included studies and conducted the risk of bias assessment. If agreement was not achieved at any stage, a third review author from one of the other two teams adjudicated.

If a trial provided data from more than one pain scale, we extracted data from the pain scale that is highest on the list below according to a previously described hierarchy of pain-related outcomes (Juni 2006; Reichenbach 2007).

1. Global pain.
2. Pain on walking.
3. Western Ontario and McMaster Osteoarthritis Index (WOMAC) osteoarthritis pain subscore.
4. Composite pain scores other than WOMAC.
5. Pain on activities other than walking.
6. Pain at rest or pain during the night.
7. WOMAC global algofunctional score.
8. Lequesne Osteoarthritis Index global score.
9. Other algofunctional scale.

Data on more than one physical function scale, when reported in a trial, were extracted according to the hierarchy presented below.

1. Global disability score.
2. Walking disability.
3. WOMAC disability subscore.
4. Composite disability scores other than WOMAC.
5. Disability other than walking.
6. WOMAC global scale.
7. Lequesne Osteoarthritis Index global score.
8. Other algofunctional scale.

If data on more than one quality of life scale were reported in a trial, data were extracted according to the hierarchy presented below.

1. Short Form (SF)-36, Mental Component Summary (MCS).
2. SF-12 MCS.
3. EuroQol.
4. Sickness Impact Profile (SIP).
5. Nottingham Health Profile (NHP).
6. Other quality of life scales.

Assessment of risk of bias in included studies

We assessed risk of bias in included studies in accordance with methods recommended by The Cochrane Collaboration (Risk of bias in included studies).

We assessed risk of bias according to the following domains.

1. Random sequence generation.
2. Allocation concealment.
3. Blinding of participants and personnel.
4. Blinding of outcome assessment, subjective self-reported outcomes (pain, physical function, quality of life)
5. Blinding of outcome assessment, other outcomes
6. Incomplete outcome data.
7. Selective outcome reporting.

We graded each potential source of bias as high, low or unclear and provide justification for our judgement in the 'Risk of bias' table.

We summarised the risk of bias judgements across different studies for each of the seven domains listed.

We presented the figures generated by the 'Risk of bias' tool to provide summary assessments of risk of bias (Figure 1).

Figure 1. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

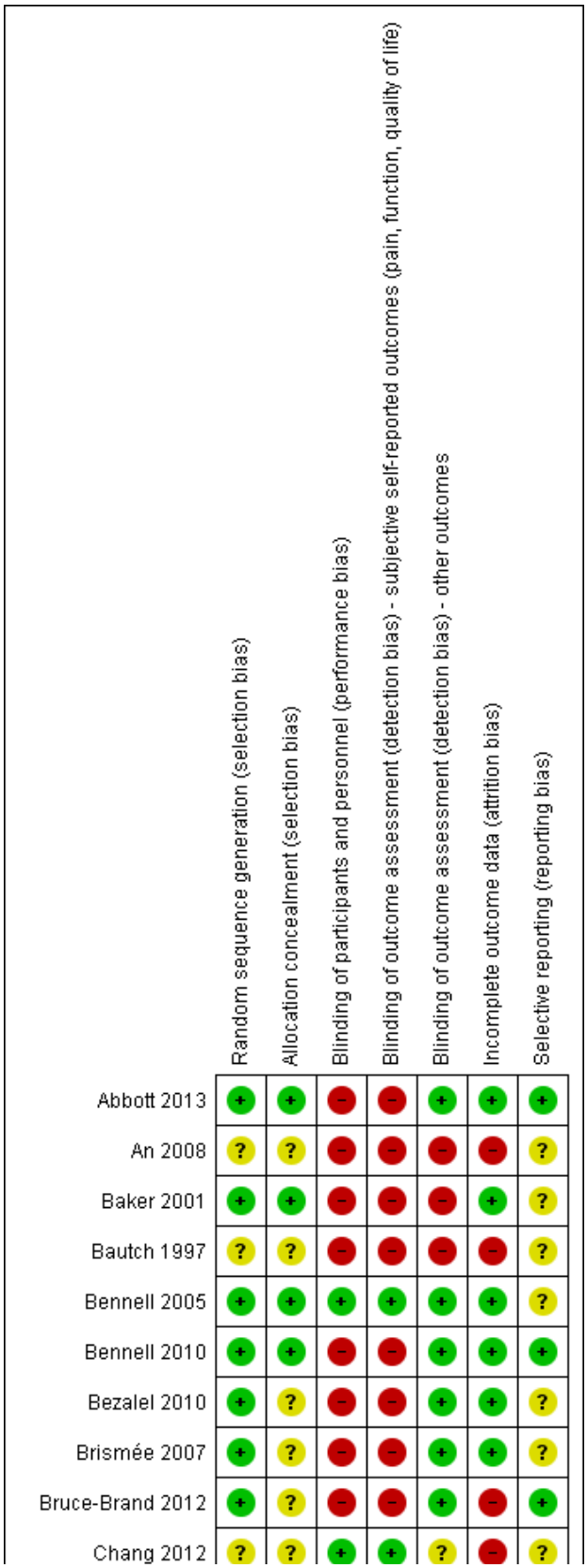


Figure 1. (Continued)

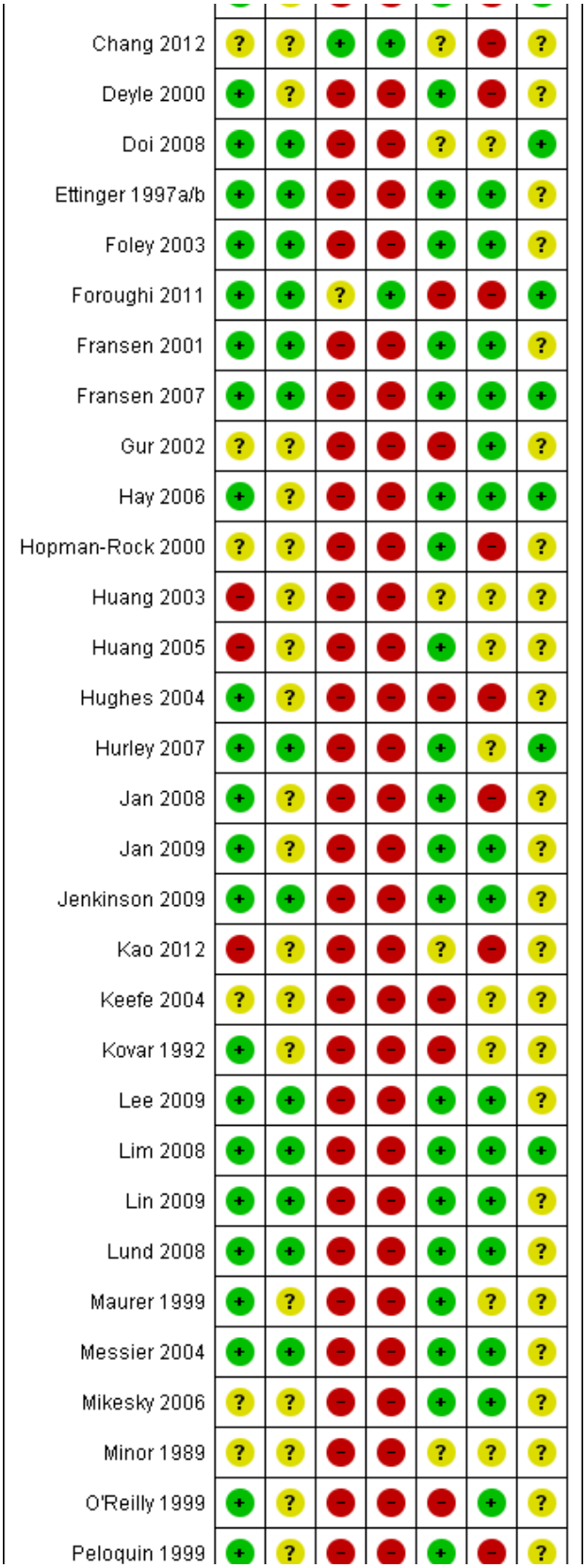
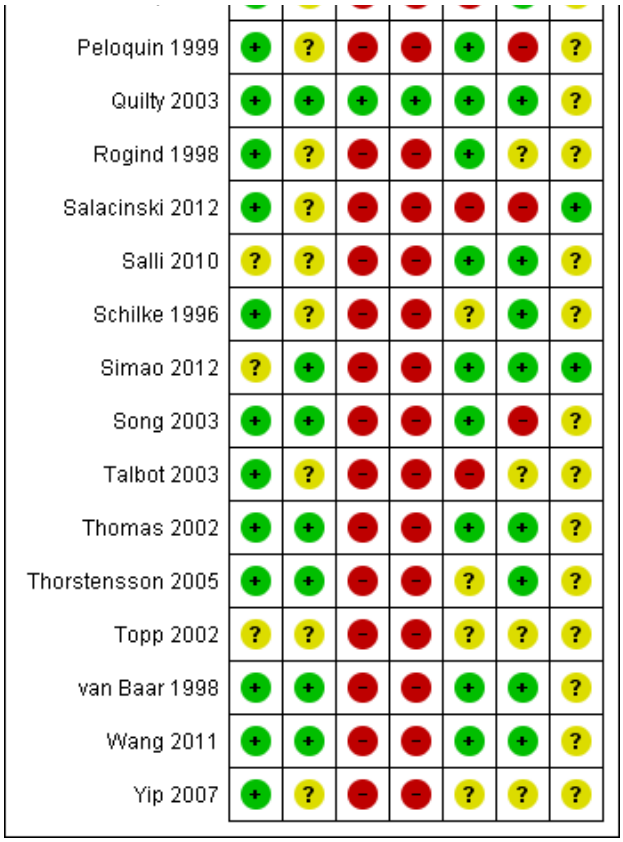


Figure 1. (Continued)



If the three domains of random sequence generation, allocation concealment and incomplete outcome data (selection bias and attrition bias) were adequately met in a study, we judged the overall risk of bias as low for that study.

Measures of treatment effect

As studies used a variety of continuous scales to evaluate pain, physical function and quality of life outcomes, a unitless measure of treatment effect size was needed to allow the results of various randomised controlled trials (RCTs) to be combined. We used standardised mean differences (SMDs) to calculate treatment effect sizes from the end of treatment, or change scores and related standard deviation (SD) scores, when possible. Treatment effect size therefore is a unitless measure providing an indication of size in terms of its variability. Outcomes pooled using SMDs were reexpressed as equivalent mean differences by multiplying by a representative control group (high weighting in pooled analyses) baseline SD. We pooled the Mantel-Haenszel odds ratio (OR) to calculate the effects of treatment allocation on study withdrawal before the first outcome assessment.

Unit of analysis issues

The unit of analysis was the participant; thus no unit of analysis issues are described.

Dealing with missing data

No data were missing. We contacted study authors when data could not be extrapolated in the desired form from the published manuscript.

Assessment of heterogeneity

In a random-effects model, overall effects are adjusted to include an estimate of the degree of variation between studies, or heterogeneity, in intervention effect (Tau^2) (Deeks 2011). The Chi^2 test assesses whether differences in results are beyond those that can be attributed to sampling error (chance). The impact of heterogeneity on meta-analysis results is quantified by the I^2 statistic. This statistic describes the percentage of variability in effect estimates that is due to heterogeneity rather than to chance (Deeks 2011): 30% to 60% probably represents moderate heterogeneity, and > 50% is usually considered as representing substantial heterogeneity.

Assessment of reporting biases

For studies published after 1 July 2005, we screened the Clinical Trials Register at the International Clinical Trials Registry Platform of the World Health Organization (http://apps.who.int/trialssearch) to obtain the a priori trial protocol. We evaluated whether selective reporting of outcomes occurred (outcome reporting bias).

To assess for potential small-study effects in meta-analyses (i.e. intervention effect is more beneficial in smaller studies), we compared effect estimates derived from a random-effects model with those obtained from a fixed-effect model of meta-analysis. In the presence of small-study effects, the random-effects model will provide a more beneficial estimate of the intervention than the fixed-effect model (Sterne 2011).

Data synthesis

We used the random-effects model to combine outcomes.

Summary of findings table

We created a 'Summary of findings' table by using the following outcomes: immediate post-treatment pain, physical function, quality of life, withdrawals due to adverse events and total adverse events. We used GRADEpro software and the five GRADE (Grades of Recommendation, Assessment, Development and Evaluation) considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of a body of evidence for stated outcomes (Schünemann 2011a; Schünemann 2011b).

Outcomes pooled using SMDs were reexpressed as absolute mean differences (or changes) by multiplying by a representative control group baseline SD from a trial using a familiar instrument and dividing by points of the measurement scale expressed as a percentage.

In the Comments column of the 'Summary of findings' table, we have presented the absolute percent difference, the relative percent change from baseline and the number needed to treat for an additional beneficial outcome (NNTB) (the NNTB is provided only for outcomes with statistically significant differences between intervention and control groups).

For continuous outcomes, absolute risk difference was calculated as mean difference between intervention and control groups given in original measurement units (divided by the scale), expressed as a percentage; the relative difference was calculated as the absolute change (or mean difference) divided by the baseline mean of the control group from a representative trial. The NNTB for continuous measures was calculated using the Wells calculator (available at the CMSG Editorial office; http://musculoskeletal.cochrane.org/).

We assumed a minimal clinically important difference (MCID) of 15 points on a 0 to 100-point pain scale, and of 10 points on a 0 to 100-point function scale.

Subgroup analysis and investigation of heterogeneity

The influence of using end of treatment or change scores was evaluated for the investigation of heterogeneity.

Subgroup analyses were conducted to explore possible differences in pooled SMDs for immediate post-treatment pain and physical function according to:

1. treatment content (quadriceps exercises only, lower limb strengthening, strengthening and aerobics, walking programme, other programmes),

2. treatment delivery mode (individual, class-based, home programme) and

3. number of face-to-face contact occasions (< 12, ≥ 12).

These sub-groups were chosen to reflect differences in dosage and content of the exercise programs using crude metrics that were usually available in all the study reports.

Sensitivity analysis

1.We assessed the effect of potential selection and attrition bias on immediate post-treatment pain and physical function outcomes.

2. We assessed the effect of potential detection bias on immediate post-treatment pain and physical function outcomes.

RESULTS

Description of studies

Results of the search

Of 212 retrieved RCTs identified by the literature search, 54 met the inclusion criteria (Abbott 2013; An 2008; Baker 2001; Bautch 1997; Bennell 2005; Bennell 2010; Bezalel 2010; Brismée 2007; Bruce-Brand 2012; Chang 2012; Deyle 2000; Doi 2008; Ettinger 1997a/b; Foley 2003; Foroughi 2011; Fransen 2001; Fransen 2007; Gur 2002; Hay 2006; Hopman-Rock 2000; Huang 2003; Huang 2005; Hughes 2004; Hurley 2007; Jan 2008; Jan 2009; Jenkinson 2009; Kao 2012; Keefe 2004; Kovar 1992; Lee 2009; Lim 2008; Lin 2009; Lund 2008; Maurer 1999; Messier 2004; Mikesky 2006; Minor 1989; O'Reilly 1999; Peloquin 1999; Quilty 2003; Rogind 1998; Salacinski 2012; Salli 2010; Schilke 1996; Simao 2012; Song 2003; Talbot 2003; Thomas 2002; Thorstensson 2005; Topp 2002; van Baar 1998; Wang 2011; Yip 2007). Details for each of the included studies are outlined in Characteristics of included studies.

One of the 54 studies included two clearly different exercise intervention groups and was treated as two trials, with sample size of the control group equally divided between the two exercise intervention groups: aerobic walking and resistance training (Ettinger 1997a/b). Five of the included studies recruited people with a diagnosis of hip or knee OA (Foley 2003; Fransen 2007; Hopman-Rock 2000; van Baar 1998; Abbott 2013). These five studies provided data specific for participants with knee OA. Five studies allocated participants to two (Gur 2002; Jan 2008; Jan 2009; Salli 2010) or three (Huang 2003) different forms of muscle strengthening. As control groups in both studies were relatively small, the mean effects of exercise allocations were combined and were compared with those of the control group. One study (Huang 2005) described two allocations combining exercise with ultrasound or hyaluronan. Only the exercise alone allocation was considered in the current review. Two studies described four treatment allocations (Messier 2004; Jenkinson 2009), two of which included a weight reduction programme. Only the exercise alone allocation versus the control group was considered in the current review. One study (Mikesky 2006) included participants without knee pain. Data were provided by the study author on 37 participants with knee pain and confirmed knee OA. One study (Keefe 2004) described four allocations, two involving a spouse-assisted coping strategy intervention. Only the exercise alone groups and the control groups were evaluated in the current review. Two studies included (in addition to a more traditional exercise programme) a proprioceptive training allocation (Lin 2009) and an allocation to squatting on a vibratory platform (Simao 2012). One study stratified results according to varus or normal knee alignment (Lim 2008). These results were averaged for the two stratifications.

Included studies

Marked variability among the 54 included studies was noted with regard to study participants recruited, timing of outcomes

assessments, exercise interventions assessed and important aspects of study methodology. Most studies recruited between 50 and 150 participants. However, 19 (35%) studies recruited fewer than 25 participants in one or both allocation groups (An 2008; Baker 2001; Bautch 1997; Brismée 2007; Bruce-Brand 2012; Chang 2012; Foley 2003; Gur 2002; Keefe 2004; Lee 2009; Mikesky 2006; Minor 1989; Rogind 1998; Salacinski 2012; Salli 2010; Schilke 1996; Simao 2012; Song 2003; Talbot 2003), whereas five studies recruited more than 200 participants (Abbott 2013; Hurley 2007; Jenkinson 2009; Kao 2012; Thomas 2002), one of which recruited 750 participants (Thomas 2002).

Sample recruitment varied widely, with studies recruiting exclusively community volunteers (An 2008; Bennell 2005; Bennell 2010; Brismée 2007; Ettinger 1997a/b; Foroughi 2011; Fransen 2007; Hughes 2004; Kao 2012; Lim 2008; O'Reilly 1999; Peloquin 1999; Quilty 2003; Salacinski 2012; Wang 2011), patients drawn from specialist rheumatology or orthopaedic clinics (Bezalel 2010; Bruce-Brand 2012; Doi 2008; Foley 2003; Jan 2008; Jan 2009; Lin 2009; Schilke 1996; Song 2003; Thorstensson 2005; Yip 2007), a mix of community volunteers and patients from specialist clinics or referred by general practitioners (Abbott 2013; Bautch 1997; Jenkinson 2009; Keefe 2004; Lund 2008; Minor 1989), patients referred by general practitioners (Hay 2006; Hurley 2007; Thomas 2002; van Baar 1998) or patients from physiotherapy waiting lists (Deyle 2000; Fransen 2001).

In two studies, approximately 50% of the sample reported a symptom duration of less than a year (Chang 2012; van Baar 1998), whilst a few other studies reported a mean symptom duration longer than 10 years (Foroughi 2011; Maurer 1999; Minor 1989). Many studies did not report symptom duration. Most studies stated that the American College of Rheumatology diagnostic criteria were used for study inclusion. However, 'knee pain in the past week' (O'Reilly 1999), 'knee pain in the last month' (Jenkinson 2009) or patellofemoral knee pain (Quilty 2003) was sufficient in three studies. In one study, patients with OA diagnosed via arthroscopy or who were on the waiting list for total knee replacement were included (Bruce-Brand 2012). Five studies required radiographic disease of at least Kellgren and Lawrence Grade III for study participation (Bruce-Brand 2012; Doi 2008; Lim 2008; Rogind 1998; Thorstensson 2005), whereas other studies included only participants with radiographic disease of Kellgren and Lawrence Grade III or less (Chang 2012; Jan 2008; Jan 2009; Lin 2009; Salacinski 2012). Many study cohorts comprised participants who were overweight (body mass index (BMI) 25 to 29.9 kg/m^2) or obese (BMI ≥ 30 kg/m^2). Consequently, mean BMI (reported or calculated from mean weight and height data) was in the normal range in only a few studies (Doi 2008; Jan 2008; Jan 2009; Lin 2009; Salacinski 2012). Two studies targeted only overweight or obese participants (BMI ≥ 28 kg/m^2), resulting in cohorts with a mean BMI of 34 kg/m^2 (Messier 2004) and a median BMI of 33.6 kg/m^2 (Jenkinson 2009). This range of recruitment strategies and inclusion criteria resulted in wide variability in baseline radiographic and symptomatic disease severity between studies, when reported.

Many studies did not report medication use. One study excluded people taking non-steroidal anti-inflammatory drugs (NSAIDs) (Bautch 1997), whereas another included only people currently taking NSAIDs at least twice a week (Kovar 1992). Cessation of NSAID use was required for the duration of one study (Jan 2008). Another study offered paracetamol as required (up to 2 g per day) to all participants (Salli 2010). Sticky patch analgesia was available as required for all participants in a study in which the control group was taking NSAIDs (Doi 2008). One study stratified allocation groups according to glucosamine or chondroitin use (Foroughi 2011).

A wide range of therapeutic exercise programmes were assessed. Delivery mode varied between one-on-one (individual) programmes (Analysis 6.1; Analysis 7.1) and exercise programmes undertaken most often by the participant at home (Analysis 6.3; Analysis 7.3). However, many 'home' programmes incorporated home visits by a trained nurse or a community physiotherapist. Also, most individual treatments and class-based programmes provided a home exercise programme. Only one study included allocation to individual treatment or to a class-based programme (Fransen 2001). Results for each of these allocations were presented in the original manuscript for all participants (including those originally allocated to a waiting list control) and were presented as such for this comparison.

Complexity of content and mode of exercise varied considerably between studies. Simple quadriceps muscle strengthening (i.e. supine or seated knee extension using leg weight only) was used by one study (Doi 2008), whereas another study initially used very simple exercises (e.g. straightening knee over rolled towel) and progressed to functional exercises after several months (Jenkinson 2009). One study (Simao 2012) used squat exercises alone to strengthen multiple lower limb muscles, and another used multiple sitting and standing exercises with body weight only (Wang 2011). Other studies, although often using a combination of exercise equipment, used mainly elastic resistance bands (Bennell 2010; Bruce-Brand 2012; Chang 2012; Topp 2002), free weights (Ettinger 1997a/b; Lim 2008) or resistance machines (Foley 2003; Foroughi 2011; Fransen 2001; Gur 2002; Huang 2003; Huang 2005; Jan 2008; Jan 2009; Maurer 1999; Mikesky 2006; Salli 2010; Schilke 1996). A number of studies employed complex, multi-modal programmes including manual therapy, upper limb and/or truncal muscle strengthening and balance co-ordination (Abbott 2013; Bennell 2005; Deyle 2000; Peloquin 1999; Rogind 1998; van Baar 1998), in addition to lower limb muscle strengthening. Aerobic walking (Ettinger 1997a/b; Kovar 1992; Messier 2004; Minor 1989; Talbot 2003) or cycling programmes (Salacinski 2012) were the focus of some studies.Five studies evaluated Tai Chi classes (Brismée 2007; Fransen 2007; Lee 2009; Song 2003; Yip 2007), and one study used Baduanjin exercises (An 2008). Exercises were not clearly described in one study (Kao 2012), and in another the website that provided exercise descriptions was not available (Hurley 2007). Overall, the exercise content of studies evidenced much variability, and many studies did not provide a clear rationale for choice of exercise.

Along with delivery mode and content, treatment 'dosage' (duration, frequency, intensity) varied widely between studies. Monitored treatment sessions, presented in individual or class-based format, ranged from 20 to 60 minutes. Exercise frequency for monitored classes or for individual clinic sessions in most studies was two to three times per week; however, frequency varied between once per week (Bezalel 2010; Hopman-Rock 2000; Kao 2012; Topp 2002; Yip 2007) and five times per week (An 2008). Concurrent monitored clinic classes and home programmes were provided in a few studies (Abbott 2013; Bennell 2010; Bruce-Brand 2012; Topp 2002), thus potentially increasing the overall frequency of weekly exercise. The total number of monitored exercise sessions provided ranged from none (Talbot 2003) to 72 (Foroughi 2011).

Four studies prescribed daily home exercise (Doi 2008; Jenkinson 2009; O'Reilly 1999; Thomas 2002), and one study monitored daily pedometer step counts (Talbot 2003). Total treatment duration for monitored classes or individual clinic sessions ranged from one month (Bezalel 2010; Deyle 2000) to six months (Foroughi 2011). Two studies prescribed home programmes for up to two years (Jenkinson 2009; Thomas 2002).

Prescribed exercise generally was of moderate to moderately high intensity, although some studies failed to report whether exercise intensity was maintained or progressed during the course of exercise training. Intensity achieved during strength training using free or limb weights or Theraband was commonly a 10-repetition maximum (10RM) with varying numbers of sets (Bennell 2010; Chang 2012; Ettinger 1997a/b; Lim 2008) or was at least moderate (Bruce-Brand 2012; Topp 2002; Wang 2011). One study ensured that strength exercise was conducted at least at 60% maximum heart rate (HRmax); this was progressed to the highest tolerable intensity (Thorstensson 2005). Muscle strength training conducted using a variety of resistance machines was generally very well quantified and ranged from 50% 1RM (Lin 2009), through 60% to 80% 1RM (Foley 2003; Foroughi 2011; Jan 2008; Jan 2009; Mikesky 2006), to maximum effort at various isokinetic speeds (Gur 2002; Huang 2003; Huang 2005; Maurer 1999; Salli 2010; Schilke 1996). For some studies, although strength exercises were described, exercise intensity was not quantified (Bezalel 2010; Doi 2008; Kao 2012; O'Reilly 1999; Thomas 2002). Aerobic exercise intensity, achieved via walking programmes, ranged from low (Bautch 1997; Talbot 2003) to moderate (50% to 70% heart rate reserve (HRR) or 60% to 80% HRmax) (Ettinger 1997a/b; Minor 1989). One study used moderate-intensity (70% HRmax) stationary cycling (Salacinski 2012). Another few studies used moderate-intensity walking (40% to 60% HRmax or 50% to 85% HRR) or cycling (50% to 60% HRmax) and resistance training in the same session (Fransen 2001; Hughes 2004; Keefe 2004; Messier 2004; Peloquin 1999). Tai Chi exercises were used in five studies (Brismée 2007; Fransen 2007; Lee 2009; Song 2003; Yip 2007), and Baduanjin (Qigong) exercises in one study (An 2008), but intensity was not measured (via heart rate or rating of perceived exertion). Other studies employed complex programmes of physiotherapy, exercise and other strategies, rendering overall assessment of exercise intensity difficult.

Thirty-six of the 54 included studies (67%) used the Western Ontario and McMaster Universities Arthritis Index (WOMAC) to evaluate knee pain or self-reported physical function. A variety of scales were used by the other studies. Thirteen studies used visual analogue scales (VASs) to measure pain (Abbott 2013; Bautch 1997; Bennell 2005; Brismée 2007; Gur 2002; Hopman-Rock 2000; Huang 2003; Huang 2005; Lund 2008; Quilty 2003; Rogind 1998; Salacinski 2012; Salli 2010). Only three studies included a separate participant global assessment of treatment effectiveness (Kao 2012; van Baar 1998; Yip 2007).

Excluded studies

A total of 151 studies were excluded for reasons given in the Characteristics of excluded studies table (Ageberg 2010; Aglamis 2008; Aglamis 2009; Akyol 2010; Alfredo 2012; Anwer 2011; Aoki 2009; Atamaz 2006; Atamaz 2012; Boocock 2009; Borjesson 1996; Brosseau 2012; Bulthuis 2007; Bulthuis 2008; Callaghan 1995; Cetin 2008; Chaipinyo 2009; Chamberlain 1982; Cheing 2002; Cheing 2004; Ciolac 2011; Coupe 2007; Crotty 2009; Deyle 2005; Dias 2003; Diracoglu 2005; Duman 2012; Durmus 2007; Durmus 2012; Ebnezar 2012; Ebnezar 2012a; Evcik 2002; Evgeniadis 2008; Eyigor 2004; Farr 2010; Feinglass 2012; Fitzgerald 2011; Forestier 2010; Foroughi 2011a; Foster 2007; Gaal 2008; Gaudreault 2011; Gill 2009; Green 1993; Gremion 2009; Haslam 2001; Helmark 2010; Helmark 2012; Hinman 2007; Hiyama 2012; Hoeksma 2004; Huang 2005b; Hughes 2010; Hurley 1998; Hurley 2007a; Hurley 2012; Jan 1991; Jan 2008a; Jessep 2009; Karagulle 2007; Kawasaki 2008; Kawasaki 2009; King 2008; Konishi 2009; Kreindler 1989; Kuptniratsaikul 2002; Lankhorst 1982; Lim 2002; Lim 2010; Lin 2004; Lin 2007; Liu 2008; Mangione 1999; Marra 2012; Mascarin 2012; McCarthy 2004; McKnight 2010; McQuade 2011; Messier 1997; Messier 2000a; Messier 2000b; Messier 2007; Messier 2008; Miller 2012; Moss 2007; Murphy 2008; Neves 2011; Ng 2010; Nicklas 2004; Ozdincler 2005; Penninx 2001; Penninx 2002; Pereira, 2011; Petersen 2010; Petersen 2011; Peterson 1993; Petrella 2000; Pietrosimone 2010; Pietrosimone 2012; Pisters 2010; Pisters 2010a; Piva 2011; Piyakhachornrot 2011; Quirk 1985; Rattanachaiyanont 2008; Ravaud 2004; Reid 2010; Reid 2011; Rejeski 1998; Sayers 2012; Schlenk 2011; Scopaz 2009; Selfe 2008; Sen 2004; Sevick 2009; Shakoor 2007; Shakoor 2010; Shen 2008; Silva 2008; Sled 2010; Song 2010; Soni 2012; Stitik 2007; Stitik 2007a; Sullivan 1998; Swank 2011; Sylvester 1989; Teixeira 2011; Thiengwittayaporn 2009; Toda 2001; Tok 2011; Topp 2009; Tsauo 2008; Tunay 2010; Tuzun 2004; van Baar 2001; Van Gool 2005; Veenhof 2007; Walls 2010; Wang 2006; Wang 2007; Wang 2007a; Wang 2009; Weng 2009; Whitehurst 2011; Williamson 2007; Williamson 2007a; Wyatt 2001; Yilmaz 2010; Yip 2007a; Yip 2008).

Risk of bias in included studies

According to the above criteria (methodological quality assessment), a total of 19 (20%) studies could be considered as achieving 'low risk of bias' from the published report (Abbott 2013; Baker 2001; Bennell 2005; Bennell 2010; Ettinger 1997a/b; Foley 2003; Fransen 2001; Fransen 2007; Jenkinson 2009; Lee 2009; Lim 2008; Lin 2009; Lund 2008; Messier 2004; Quilty 2003; Thomas 2002; Thorstensson 2005; van Baar 1998; Wang 2011). Five of these studies provided sustainability (two to six months or longer than six months) data only (Abbott 2013; Jenkinson 2009; Messier 2004; Quilty 2003; Thomas 2002) (Figure 1).

Allocation

Although most studies reported the methods used to generate randomisation, allocation concealment procedures were less frequently described (Figure 1).

Blinding

Only four of the 54 included studies claimed blinding of study participants (Bennell 2005; Chang 2012; Foroughi 2011; Quilty 2003). Bennell 2005 used sham ultrasound (US) with non-active gel as the placebo treatment; Chang 2012 had both allocations randomised to general physiotherapy with the addition of Theraband exercises for the experimental group. Foroughi 2011 provided low-resistance, non-progressive 'sham exercise'. The fourth study uniquely used a Zelen randomisation, leading the control group to be unaware of participation in a randomised trial (Quilty 2003).

Just over half (57%) of the 54 studies clearly stated that the outcomes assessor was blinded to group allocation. However, as outcomes evaluated in this review were participant self-report (pain, physical function, quality of life), and given that participants

were mostly not blinded to allocation status, vulnerability to biased reporting may still be present.

Incomplete outcome data

Just over half of the studies (29/54) reported minimal loss to follow-up or utilised imputation methods (usually last observation carried forward) to perform 'intention-to-treat' analyses.

Selective reporting

The presence of reporting bias was simply based on study registration. As this criterion would cause earlier studies to be at a disadvantage (before study registration requirements), the risk of bias was judged as 'uncertain' for unregistered studies. Therefore this criterion also was not considered in the overall estimate of study bias.

Effects of interventions

See: **Summary of findings for the main comparison** Immediate post-treatment effects of exercise for osteoarthritis of the knee

At the time of the original review, several attempts were made to contact seven study authors to obtain additional data. Four study authors responded, and two were able to provide requested results for the location of OA in the knee (Hopman-Rock 2000; van Baar 1998), one was able to provide WOMAC scores disaggregated for pain and physical function (Deyle 2000) and one was able to provide change scores for each allocation group (Thomas 2002). No contact could be established with the other three study authors. Therefore, for one study a misprint assumption was made on one 'impossible' standard error of the mean score (Bautch 1997). For another study, two baseline standard deviations had to be extrapolated from a study of similar size using the same self-report questionnaires (Maurer 1999). For the third study, post-treatment results for the control group were used as the baseline for the active treatment groups (two-group analysis) (Ettinger 1997a/b). For updated reviews, three studies that recruited participants with OA of the hip and/or OA of the knee (Abbott 2013; Foley 2003; Fransen 2007) provided data disaggregated according to the most symptomatic joint (hip or knee).

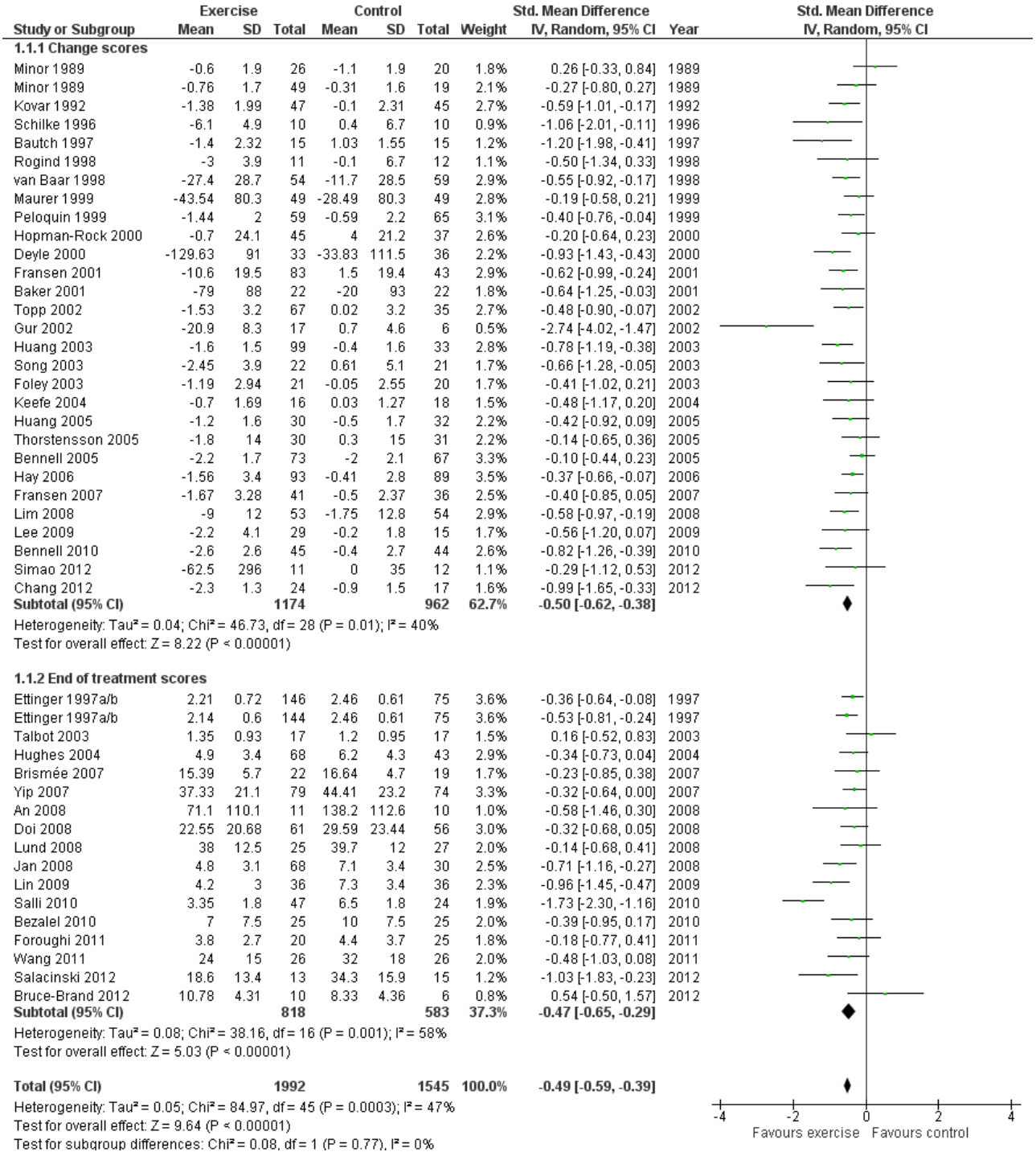
Comparison 1

Immediate post-treatment effects

Pain

Forty-four studies provided data on 3537 participants (Figure 2) (Analysis 1.1). Pooled results of these 44 studies demonstrated statistically significant benefit, with an SMD of 0.49 (95% CI 0.39 to 0.59). This effect size would be considered moderate (Cohen 1977) and was equivalent to a reduction of 12 points (95% CI 10 to 15 points) on a 0 to 100-point VAS pain scale (0 means no pain). Between-study heterogeneity was moderate (I^2 = 47%). No significant difference was noted between the SMD extrapolated from change scores and from end of treatment scores (P value 0.77) (I^2 = 0%).

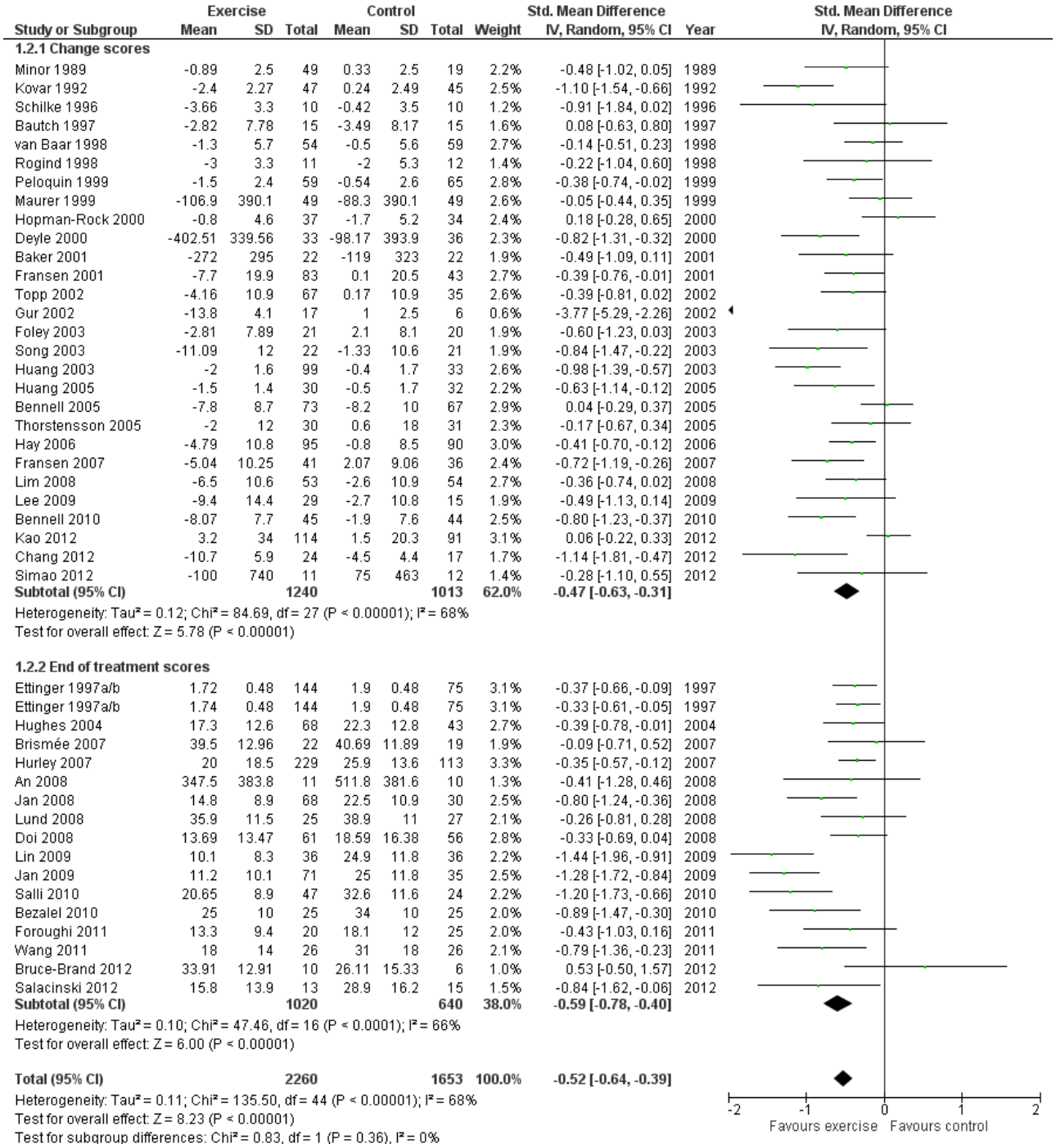
Figure 2. Forest plot of comparison: 1 Post treatment, outcome: 1.1 Pain.



Physical function

Forty-four studies provided data on 3913 participants (Figure 3) (Analysis 1.2). Pooled results of these 44 studies demonstrated statistically significant benefit, with an SMD of 0.52 (95% CI 0.39 to 0.64). This effect size would be considered moderate (Cohen 1977) and was equivalent to an improvement of 10 points (95% CI 8 to 13 points) on a 0 to 100-point scale. Between-study heterogeneity was substantial (I^2 = 68%). No significant difference was noted between change and end of treatment scores (P value 0.36) (I^2 = 0%).

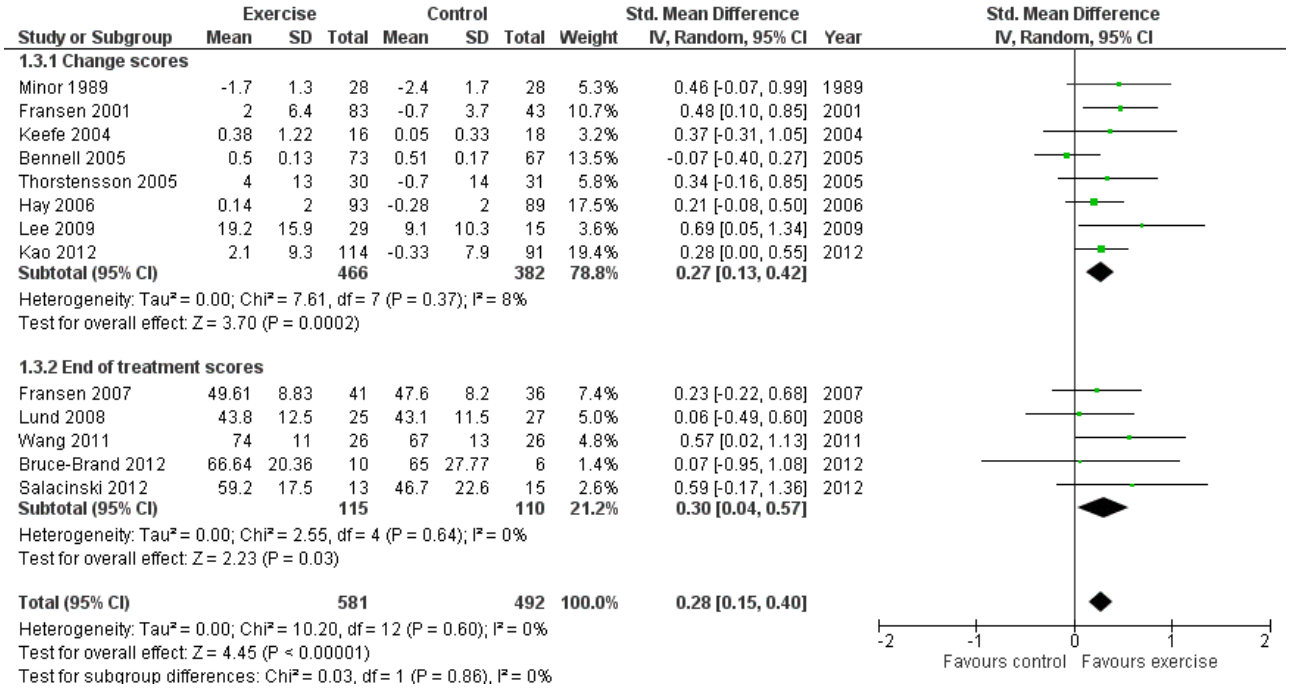
Figure 3. Forest plot of comparison: 1 Post treatment, outcome: 1.2 Physical function.



Quality of life

Thirteen studies provided data on 1073 participants (Figure 4) (Analysis 1.3). Pooled results of these 13 studies demonstrated statistically significant benefit, with an SMD of 0.28 (95% CI 0.15 to 0.40). This effect size would be considered small (Cohen 1977) and was equivalent to an improvement of 4 points (95% CI 2 to 5 points) on a 0 to 100-point scale. Between-study heterogeneity was negligible (I^2 = 0%). No significant difference was noted between change scores and end of treatment scores (P value 0.86) (I^2 = 0%).

Figure 4. Forest plot of comparison: 1 Post treatment, outcome: 1.3 Quality of life.



Study withdrawals

Forty-five studies provided data on study withdrawals at the time of the first post-treatment assessment (Analysis 1.4). Of these 45 studies, only whole sample estimates (knee and hip OA) were available for two studies (Foley 2003; van Baar 2001). No significantly increased risk of study withdrawal was noted in the exercise allocation group (14%) compared with the control group (15%) (OR 0.93, 95% CI 0.75 to 1.15).

Comparison 2

Treatment sustainability (two to six months)

Pain

Twelve studies provided data on 1468 participants (Analysis 2.1). Pooled results demonstrated statistically significant benefit (SMD 0.24, 95% CI 0.14 to 0.35). This effect size would be considered small - equivalent to a reduction of 6 (95% CI 3 to 9) points on a 0 to 100-point scale. Between-study heterogeneity was absent (I^2 = 0%). No significant difference was noted between change scores and end of treatment scores (P value 0.40) (I^2 = 0%).

Physical function

Ten studies provided data on 1279 participants (Analysis 2.2). Pooled results demonstrated statistically significant benefit (SMD 0.15, 95% CI 0.04 to 0.26). This effect size would be considered small —equivalent to an improvement of 3 (95% CI 1 to 5) points on a 0 to 100-point scale. Between-study heterogeneity was absent (I^2 = 0%). No significant difference was noted between change scores and end of treatment scores (P value 0.95) (I^2 = 0%).

Comparison 3

Treatment sustainability (longer than six months)

Pain

After exclusion of two studies with extremely outlying results (Huang 2003; Huang 2005), six studies provided data on 1104 participants (Analysis 3.1). Pooled results demonstrated a non-significant effect (SMD 0.08, 95% CI -0.15 to 0.30). Between-study heterogeneity was moderate (I^2 = 43%). No significant difference was noted between change scores and end of treatment scores (P value 0.73) (I^2 = 0%).

Physical function

After exclusion of two studies with extremely outlying results (Huang 2003; Huang 2005), six studies provided data on 1098 participants (Analysis 3.2). Pooled results demonstrated statistically significant benefit (SMD 0.20, 95% CI 0.08 to 0.32). This effect size would be considered small - equivalent to an improvement of 4 (95% CI 2 to 6) points on a 0 to 100-point scale. Between-study heterogeneity was absent (I^2 = 0%). No significant difference was noted between change scores and end of treatment scores (P value 0.96) (I^2 = 0%).

Subgroup Analyses

Comparison 4

Treatment content

Pain

Studies providing immediate post-treatment assessments for pain were classified into five categories according to their exercise programme content Analysis 4.1: quadriceps strengthening only (nine studies, 620 participants); lower limb strengthening (12 studies, 863 participants , combination strengthening and aerobic

exercise (10 studies, 920 participants); walking programmes (four studies, 351 participants) and 'other programmes' (e.g. Tai Chi) (10 studies, 733 participants). Each of the treatment content subgroups reported significantly reduced pain. No significant differences were noted between the various exercise programmes in mean pooled SMD ranged from 0.35 for 'other programmes' to 0.50 to 0.64 for various strengthening/aerobic programmes - equivalent to improvements of 9 points ('other programmes') to 12 to 16 points (various strengthening/aerobic programs) on a 0 10 100-point scale. Within-group between-study heterogeneity was substantial for the quadriceps strengthening (70%) and lower limb strengthening (61%) programmes.

Exclusion of two extreme outliers (simple quadriceps strengthening (Salli 2010) and lower limb strengthening (Gur 2002)) reduced the SMD and within-group heterogeneity to 0.49 and 0.47 (26% and 37%), respectively

Physical function

Studies providing immediate post-treatment assessments for physical function were similarly classified Analysis 4.2: quadriceps strengthening only (10 studies, 726 participants), lower limb strengthening (13 studies, 1066 participants), combination strengthening and aerobic exercise (10 studies, 1231 participants), walking programmes (three studies, 317 participants) and 'other programmes' (most Tai Chi or complex non-specific programmes) (10 studies, 915 participants). Each of the treatment content subgroups reported significantly improved physical function. No significant differences were noted between the various exercise programmes in mean pooled SMD which ranged from 0.27 for 'other programmes' to 0.74 for quadriceps strengthening only - equivalent to improvements of 5 points ('other programmes') to 15 points (quadriceps strengthening only) on a 0 to 100-point scale. Within-group between-study heterogeneity was considerable for many of the subgroups (quadriceps only I^2 = 73%; lower limb strengthening I^2 = 76%) and could not be reduced (by more than 25%) by exclusion of any one study.

Comparison 5

Treatment delivery mode

Pain

Studies providing immediate post-treatment assessments for pain were categorised according to three treatment delivery modes Analysis 5.1: individual treatments (14 studies, 1133 participants), class-based programmes (24 studies, 1905 participants) and 'home' programmes (seven studies, 550 participants). Pooled analysis demonstrated that each of the treatment delivery modes provided significant reductions in pain: individual treatments: SMD 0.76, 95% CI 0.52 to 1.01; exercise classes: SMD 0.42, 95% CI 0.33 to 0.51; and home programmes: SMD 0.38, 95% CI 0.21 to 0.55. These effect sizes ranged from large (individual treatments) to small (home programmes) - equivalent to improvements of 19 (95% CI 13 to 25) points for individual treatments, 10 (95% CI 8 to 12) points for exercise classes and 9 (95% CI 5 to 13) points for home programmes on a 0 to 100-point scale. Between-study heterogeneity for the category of individual treatments was substantial (I^2 = 72%) and was negligible for class-based programmes and home programmes (I^2 = 0%). A statistically significant difference was detected between the three modes of delivery (P value 0.03) (Analysis 5.1).

After exclusion of two extreme outliers in the individual treatments category (Gur 2002; Salli 2010), SMD 0.61, 95% CI 0.42 to 0.79) and between study heterogeneity, I^2 = 49%, were considerably reduced, and no statistically significant difference among the three modes of treatment delivery could be detected (P value 0.14).

Physical function

Studies providing immediate post-treatment assessments of physical function were similarly categorised Analysis 5.2: individual treatments (16 studies, 1493 participants), class-based programmes (24 studies, 2152 participants) and 'home' programmes (seven studies, 699 participants). Pooled analysis demonstrated that each of the treatment delivery modes provided significant reductions in pain: individual treatments: SMD 0.76, 95% CI 0.50 to 1.03; exercise classes: SMD 0.38, 95% CI 0.26 to 0.49; and home programmes: SMD 0.37, 95% CI 0.21 to 0.53. These effect sizes would be considered large (individual treatments) to small (exercise classes, home programmes) - equivalent to improvements of 16 (95% CI 10 to 21) points for individual treatments, 8 (95% CI 5 to 10) points for exercise classes and 7 (95% CI 4 to 11) points for home programmes on a 0 to 100-point scale. Between-study heterogeneity for the category of individual treatments was substantial (I^2 = 84%) but was moderate for class-based programmes (I^2 = 33%) and minimal for home programmes (I^2 = 8%). A statistically significant difference was detected among the three modes of delivery in terms of physical function (P value 0.03) Analysis 5.2. Even after exclusion of one extreme outlier in individual treatments (Gur 2002), heterogeneity remained substantial (I^2 = 78%), but differences among the three modes of delivery failed to achieve statistical significance (P value 0.06).

Comparison 6

Number of contact occasions

Pain

Studies providing immediate post-treatment pain assessments were dichotomised according to the number of face-to-face contact occasions (in clinics or as home visits) with the healthcare professional supervising or monitoring the exercise programme Analysis 6.1: fewer than 12 contact occasions (10 studies, 1019 participants) versus 12 or more contact occasions (34 studies, 2468 participants).

Both categories achieved significant benefit: fewer than 12 occasions: SMD 0.40, 95% CI 0.24 to 0.56; 12 or more contact occasions: SMD 0.55, 95% CI 0.45 to 0.66. Although '12 or more occasions' did result in a larger SMD, the effect size of both categories would be considered moderate - equivalent to improvements of 10 (95% CI 6 to 14) points for fewer than 12 occasions and 13 (95% CI 11 to 16) points for 12 or more contact occasions on a 0 to 100-point scale. Between-study heterogeneity was moderate (I^2= 35% and 43%). No significant difference could be detected between the two categories of contact occasions in terms of pain (P value 0.15) (Analysis 6.1).

Physical function

Studies providing immediate post-treatment assessments of physical function Analysis 6.2: fewer than 12 occasions (nine studies, 1033 participants) versus 12 or more contact occasions

(33 studies, 2432 participants). Both categories achieved significant benefit: fewer than 12 contact occasions: SMD 0.33, 95% CI 0.09 to 0.57; 12 or more contact occasions: SMD 0.55, 95% CI 0.41 to 0.60. The category of '12 or more occasions' did result in a larger SMD (moderate effect size) compared with the 'fewer than 12 occasions' category (small effect size) - equivalent to improvements of 7 (95% CI 2 to 11) points for fewer than 12 occasions and 11 (95% CI 8 to 12) points for 12 or more contact occasions, on a 0 to 100-point scale. However, between-study heterogeneity was considerable for each category (I^2 = 72% and 60%), with no influential outliers (reducing heterogeneity > 25%). Differences between the two categories of contact occasions failed to achieve statistical significance (P value 0.09).

Sensitivity Analyses

Comparison 7

Selection and attrition bias

Pain

If random sequence generation, allocation concealment and incomplete outcome data domains were adequately met by a study, we judged the overall risk of bias as low for that study (14 studies, 1458 participants) (Analysis 7.1). All other included studies were categorised as 'uncertain or high risk of bias' (30 studies, 2029 participants). The pooled effects restricted to the 'low-risk ' studies still indicated a significant reduction in pain (SMD 0.47, 95% CI 0.36 to 0.59) - equivalent to improvements of 12 (95% CI 9 to 15) points on a 0 to 100-point scale, and very similar to the pooled effects with all studies included (SMD 0.49; 95% CI 0.39 to 0.59) Analysis 1.1 Between-study heterogeneity was negligible for studies with low risk of bias (I^2 = 14%) but substantial for studies categorised as having uncertain or high risk (I^2 = 52%).

Physical function

On the basis of the same criteria, 14 studies (456 participants) were categorised as 'low risk' while 30 studies (2457 participants were categosed as having 'uncertain or high risk' of bias Analysis 7.2. Pooled SMDs for 'low-risk ' studies indicated a significant treatment effect: SMD 0.45 (95% CI 0.28 to 0.63) - equivalent to improvements of 9 (95% CI 6 to 13) points on a 0 to 100-point scale, and very similar to the pooled effect with all studies included (SMD 0.52; 95% CI 0.39 to 0.64) Analysis 1.2. Between-study heterogeneity was substantial for both categories (I^2 = 57% and 72%).

Detection bias

Pain

If participants were stated to be blinded to treatment allocation, we considered the study as low risk for detection bias (3 studies, 226 participants) Analysis 7.3 . All other included studies were categorised as 'uncertain or high risk of bias (41 studies, 3261 participants). The mean effect for 'low risk' studies (SMD 0.37) was lower than the mean pooled effect with all studies included (SMD 0.49), but equivalent to a mean reduction in pain of 9 points on a 0 to 100-point scale. However the 95% CI around the mean SMD for the 'low risk' studies included the possibility of 'no effect' (95% CI -0.13 to 0.87). The small number of 'low risk' studies on basis of participant blinding resulted in extremely wide 95% CIs around the SMD and substantial between-study heterogeneity (I^2 = 64%).

Physical function

On basis of the same criteria, 3 studies (226 participants) were categorised as 'low risk' while 41 studies (3687 participants) were categorised as having 'uncertain or high risk' of bias Analysis 7.4. The mean effect for the 'low risk' studies (SMD 0.46) was very similar to the mean pooled effect with all studies included (SMD 0.52) and equivalent to a mean improvement in physical function of 9 points on a 0 to 100-point scale. However the 95% CI around the mean SMD for the 'low risk studies included the possibility of 'no effect' (95% CI -0.22 to 1.14). Again the small number of 'low risk' studies on basis of participant blinding resulted in extremely wide 95% CIs and substantial between-study heterogeneity (I^2 = 80%).

Comparisons 1 through 7

Both mean effect sizes and 95% CIs tended to be slightly smaller with a fixed-effect model than with the random-effects model used in this meta-analysis. However, this difference was never clinically meaningful or statistically significant. The only exceptions were Analysis 7.3; Analysis 7.4, where the fixed effect model resulted in markedly smaller SMDs for the 'low risk' categories.

Adverse events

Only eleven RCTs specifically reported on adverse events (Abbott 2013; Bennell 2010; Chang 2012; Foley 2003; Foroughi 2011; Fransen 2007; Hurley 2007; Jan 2009; Lim 2008; Lund 2008; van Baar 1998).

Abbott 2013 "detected no trial related adverse events," and van Baar 1998 stated that one participant receiving exercise reported adverse effects. Foley 2003 reported four withdrawals in the exercise group due to increased pain (two people), increased blood pressure (one person) and doctor's advice (one person) compared with one withdrawal due to illness in the control group. Fransen 2007 reported one withdrawal in the Tai Chi allocation group that was due to increased low back pain. The largest numbers of adverse events were reported by Bennell 2010 (five), Hurley 2007 (five), Jan 2009 (five), Lim 2008 (10) and Lund 2008 (11). All reported events were related to increased back, hip or knee pain among participants allocated to exercise. No serious adverse events were reported in any of the included studies.

DISCUSSION

Summary of main results

This systematic review is an update of a previous Cochrane review, published in 2008, which included 32 RCTs. An additional 22 randomised controlled trials have been included in this update for a total of 54 trials, providing data on 5362 participants for outcomes on pain and on 5222 participants for outcomes on physical function. Overall, meta-analysis demonstrated that evaluated land-based therapeutic exercise programmes resulted in an immediate mean treatment benefit for knee pain (SMD 0.49, 95% CI 0.39 to 0.59), physical function (SMD 0.52, 95% CI 0.39 to 0.64) and quality of life (SMD 0.28, 95% CI 0.15 to 0.40). These mean immediate treatment benefits, extrapolated from 44 randomised controlled clinical trials involving 3537 participants for pain and 3913 participants for physical function, would be considered moderate - equivalent to 12 (95% CI 10 to 15) points and 10 (95% CI 8 to 13) points for pain and physical function, respectively, on a 0 to 100-point scale. Treatment benefit for quality of life, extrapolated from 13 trials involving 1073 participants, would be considered small -

equivalent to 4 points (95% CI 2 to 5 points). The benefit for pain is comparable with reported estimates for current simple analgesics and non-steroidal anti-inflammatory drugs taken for knee pain (Zhang 2010). Confidence intervals around demonstrated pooled results for pain reduction and improvement in physical function do not exclude a minimal clinically important treatment effect (15 points for pain and 10 points for physical function on a 0 to 100-point scale). If the meta-analysis result for immediate post-treatment pain is restricted to those 14 studies, with a total of 1458 participants, evaluated as having low risk of selection and attrition bias, exercise still demonstrated significant benefit (SMD 0.47, 95% CI 0.36 to 0.59) of moderate size - equivalent to 11 (95% CI 9 to 15) points on a 0 to 100-point scale. Similar results were found for physical function when restricted to the 14 studies, with a total of 1456 participants, evaluated as having low risk of bias (SMD 0.45, 95% CI 0.28 to 0.63) - equivalent to 9 (95% CI 6 to 13) points on a 0 to 100-point scale.

A new analysis added to this Cochrane review is an evaluation of the effects of exercise on quality of life. A relatively small number of studies (13; 24%) evaluated immediate post-treatment quality of life by using a variety of measures. Five studies reported the Mental Component Summary (MCS) of the Short Form-36 (SF-36) health survey, three studies reported the Knee Osteoarthritis Outcome Scale quality of life subscale, two studies evaluated the depression component of the Arthritis Impact Measurement Scales and one study each reported the Hospital Anxiety Depression Scale, SF-12 MCS and Assessment of Quality of Life. These measures have been validated for use in people with knee OA and have demonstrated generally good responsiveness (Brazier 1999; Liang 1990; Monticone 2013). A small beneficial effect of exercise on quality of life was identified immediately post treatment for people with knee OA. Because of the limited number of studies reporting follow-up quality of life outcomes, meta-analysis of treatment sustainability for quality of life could not be performed in this review.

The pain-relieving benefit of exercise declined at two to six months post exercise but was still significant, as evidenced in 12 studies involving 1468 participants (SMD 0.24, 95% CI 0.14 to 0.35). However, pain benefits were lost longer than six months post exercise, as was found in six studies involving 1104 participants (SMD 0.08, 95% CI -0.15 to 0.30). A small but significant treatment benefit for physical function remained two to six months following exercise, as extrapolated from 10 studies involving 1279 participants (SMD 0.15, 95% CI 0.04 to 0.26), as well as at time points longer than six months, as evidenced in six studies involving 1098 participants (SMD 0.20, 95% CI 0.08 to 0.32). These results suggest that although the pain-relieving benefit of exercise is not maintained six or more months after treatment, improvements in physical function are better sustained.

Overall completeness and applicability of evidence

Because of marked heterogeneity within evaluated exercise programmes, sub-group analyses were conducted according to the stated main focus of the evaluated exercise programme, the mode of treatment delivery and the number of directly supervised treatment occasions. Although these subgroup analyses should be viewed as exploratory, as they are non-randomised comparisons, some interesting findings were derived. A range of exercise types can be utilised in clinical practice, with lower limb muscle strengthening and general aerobic exercise recommended by most international guidelines (Hochberg 2012; McAlindon 2014). Few studies have attempted to directly compare different types of exercise. One study compared aerobic walking and muscle strengthening, but lack of study power for this particular research question led to inconclusive results (Ettinger 1997a/b). Two other studies compared different strengthening regimens: weight bearing quadriceps exercises versus non-weight bearing quadriceps exercises in one study (Jan 2009), and concentric-eccentric strengthening exercises versus isometric strengthening exercises in the other (Salli 2010). Neither study found significant differences between types of strengthening exercises. It is interesting to note that meta-analyses also could not demonstrate significant differences in the magnitude of treatment effects for pain and physical function between the various exercise programmes Analysis 4.1; Analysis 4.2. However, for both pain and physical function, exercise programmes classified as “other” (which included Tai Chi or complex non-specific exercise programmes involving coordination, stretching or balancing exercises) yielded small benefits (pain: SMD 0.35, 95% CI 0.20 to 0.49; physical function: SMD 0.27, 95% CI 0.07 to 0.47) and seemed to be less effective than strengthening and aerobic exercise. This may reflect the limited focus of these other exercise programmes on specific muscle groups, or it may reflect lower exercise intensity (which was not measured or was not quantifiable for most of these programmes). For physical function in particular, exercise involving quadriceps strengthening alone (10 studies) was the most beneficial, yielding an effect size considered large (SMD 0.74, 95% CI 0.41 to 1.07). Medium effects on physical function were identified for exercise programmes that employed general lower limb strengthening (SMD 0.54, 95% CI 0.26 to 0.83) and strengthening combined with aerobic exercise (SMD 0.52, 95% CI 0.36 to 0.67). Small benefits were detected for walking exercise programmes (SMD 0.35, 95% CI 0.11 to 0.58), although this result was obtained with pooled data from only three studies. Although a program focusing on quadriceps strengthening yielded the greatest effect on physical function, no statistically significant differences between programmes were noted.

We examined the influence of the exercise programme delivery mode Analysis 5.1; Analysis 5.2. Although studies assessing home programmes (SMD 0.38) and class-based programmes (SMD 0.42) demonstrated effect sizes for pain that were consistently smaller than those for more closely supervised individual treatments (SMD 0.76), differences between the various forms of treatment delivery were not statistically significant after two extreme outliers were removed from the individual treatments category. For physical function, individual treatments also yielded a large effect size, and exercise classes and home programmes yielded small effect sizes but failed to achieve statistical significance between the three delivery modes (P value 0.06) after an extreme outlier had been excluded from the individual treatments category. It should be noted that substantial heterogeneity was demonstrated with individual treatment delivery, and this may reflect the varying numbers of individual contact sessions or the different exercise programmes.

The magnitude of the treatment effect for both pain and physical function was influenced by the number of face-to-face contact occasions with the healthcare professional supervising or monitoring the exercise programme Analysis 6.1; Analysis 6.2. However, unlike in the previous Cochrane review, the difference between fewer than 12 occasions and 12 or more occasions failed

to reach statistical significance; this is likely due to considerable between-study heterogeneity. Taken together, results suggest that most people with knee OA need some form of ongoing monitoring or supervision to optimise clinical benefits of exercise treatment. We chose to classify exposure to exercise interventions on the basis of the number of contact occasions, not according to duration of treatment (e.g. number of weeks). Although no ideal method of classifying exercise therapy exposure is known, the number of contact occasions was chosen, as it provided a quantitative outcome for the number of potential progressions through the exercise programme. A threshold of 12 sessions was chosen because a large number of studies reported two-weekly sessions over six weeks or three-weekly sessions over four weeks, suggesting 12 as a relevant number for dichotomising data.

Exercise 'dosage,' which is a factor of frequency, intensity and programme duration, varied considerably between the studies included in this review. Uncertainties in actual dosage arise as a result of the dependence of exercise intensity not only upon exercise prescription but also upon individual exertion. The influence of programme duration upon dosage is difficult to quantify, with simple addition not providing a sufficient physiologically plausible model. Only one of the included studies attempted to evaluate the influence of exercise dosage on outcomes by comparing high- and low-intensity resistance training of the knee flexor and extensor muscles while controlling for total exercise workload (Jan 2008). Investigators found no significant differences in pain or physical function between groups, although the study was considered to have a moderate to high risk of bias. Furthermore, studies with comparable exercise programme content were insufficient to provide a meaningful subgroup analysis of the influence of exercise dosage on treatment effectiveness. Therefore, specific recommendations cannot be made regarding optimal dosage (frequency, intensity, duration).

Quality of the evidence

Overall quality of the body of evidence was assessed as high when the GRADE approach was applied for pain and quality of life. Although a potential study limitation may exist for evidence on pain and quality of life (a potential for performance and detection bias that may overestimate effect sizes), we did not consider it substantial enough to downgrade the evidence. Evidence underpinning physical function was moderate and was downgraded because of imprecision (marked heterogeneity between study findings).

For immediate post-treatment pain and physical function, 14 of 42 studies (33%) were categorised as having low risk of selection and attrition bias (random sequence generation, allocation concealment and incomplete outcome data domains adequately met). Apart from adequate randomisation procedures and allocation concealment and limited loss to follow-up, blinding of participants when outcome measures are self-report would provide the best chance that trial results will be free of selection, performance, attrition and detection bias. Blinding of study participants is difficult to achieve in studies evaluating exercise programmes. Using 'sham' exercise as the control intervention can introduce ethical concerns (substantial wasted time for control participants attending an ineffective programme) and is likely to be fairly transparent to most people with OA.

Regarding other methodological criteria, findings included the following: Most studies (40; 74%) reported using random sequence generation; 33 studies (61%) reported using blinded outcomes assessment (for other outcomes); only 24 studies (44%) reported adequate allocation concealment and 29 studies (54%) provided complete outcome data. When pooling the results according to risk of selection and attrition bias, the mean treatment effect size for immediate post-treatment pain and physical function was similar for ‘low-risk’ studies Analysis 7.1; Analysis 7.2 compared with the pooled treatment effects with all studies included Analysis 1.1; Analysis 1.2. The overall estimate of low risk of selection and attrition bias was comparable for the 22 studies identified in the update (eight 'low risk of bias' studies; 36%) and the 32 studies identified in the previous Cochrane review (11 'low risk of bias' studies; 34%). While the pooled results for the 'low risk' (detection bias) group indicated a lower mean effect for pain and physical function, the confidence intervals indicate a finding of uncertainty (not of 'no effect') as the confidence intervals do not exclude a clinically important effect.

Potential biases in the review process

Some important caveats to this review must be stated. First, given that the comparator in many studies was a no treatment control group, and that blinding of participants was not performed in almost all trials, the well-documented strong placebo effects for self-reported outcomes in knee OA (Zhang 2010) have not been controlled for in the exercise studies. Thus it is not possible to determine the exact magnitude of beneficial effects. The second issue concerns the responsiveness of self-reported pain and physical function measures. Many of the studies included in this systematic review recruited a majority of participants with early or mild symptomatic disease. Although people with early disease frequently demonstrate reduced muscle strength and aerobic capacity compared with their age- and gender-matched peers without symptomatic OA, these physiological impairments often are not yet large enough to translate into reportable difficulties on simple questionnaires. This lack of reportable difficulties would considerably reduce the potential range of improvement that was possible (ceiling effect) on self-report questionnaires in people with early or mild disease. One of the potential benefits of exercise in people with early disease, such as increased physiological reserve capacity, will not be captured by these questionnaires. Objective measures of physical performance not only strengthen the methodological quality of a study when masking to allocation is unattainable for the participant, they also potentially provide data that can be used to better discriminate between people with early disease in whom disease-related impairments have not yet developed into self-reported functional limitations or disability. Thus, reporting of both objective physiological measures and self-reported assessments in an individual study is desirable.

Several limitations of this review have been identified. We conducted an extensive literature search. Because resources were limited, we extracted data only from studies published in the English language, potentially excluding other evidence. Four studies were published in a language other than English (Carlos 2012; Ghroubi 2008; Oida 2008; Rosa 2012), and we were unable to source full text for two studies (Eungpinichpong 1997; Keogan 2007). These studies await classification. However, the possibility of publication bias could not be ruled out, as we did not attempt to retrieve unpublished studies.

The effectiveness of exercise was investigated only for measures of self-reported pain, physical function and quality of life. However, regular exercise has been demonstrated to offer many other overall physical and mental health benefits, apart from those related to OA-induced disease impairments. Therefore this review likely underestimates the overall beneficial effects of exercise amongst people with knee OA. Mediating effects of exercise dosage and disease severity on the effectiveness of exercise could not be ascertained because of large variability in reported data.

Agreements and disagreements with other studies or reviews

Updated results of this meta-analysis concur with previously identified benefits of exercise for pain and physical function among people with knee OA. However, effect sizes are greater than those reported in the previous Cochrane review (SMD 0.40, 95% CI 0.30 to 0.50 for pain; SMD 0.37, 95% CI 0.25 to 0.49 for physical function). A moderate effect size for pain was noted, whereas the previous small effect size for physical function has increased and now would be classified as moderate. The larger effects identified in this review are likely due to separation of findings into those noted immediately post treatment and those reported at a follow-up time point, which could not be done in the previous review, given the smaller study numbers. Hence the larger effects are a reflection of superior results immediately following treatment.

AUTHORS' CONCLUSIONS

Implications for practice

High-quality evidence suggests that land-based therapeutic exercise provides benefit in terms of reduced knee pain and quality of life and moderate-quality evidence of improved physical function among people with knee OA.Since the participants in most trials were aware of their treatment, this may have contributed to their improvement. Despite the lack of blinding we did not downgrade the quality of evidence for risk of performance or detection bias. This reflects our belief that further research in this area is unlikely to change the findings of our review.

Healthcare professionals and people with OA can be reassured that any type of exercise programme that is done regularly and is closely monitored by healthcare professionals can improve pain and physical function related to knee OA in the short term. This allows a great deal of choice, ranging from individual physiotherapy-led sessions and exercise classes to home-based programmes. Exercise programmes that were individually provided appeared to be associated with greater improvements in knee pain and physical function.

Results of this meta-analysis are restricted to evaluation of symptomatic benefits. Regular exercise has the potential to modify structural disease progression among people with knee OA, but this was not evaluated in this review and remains an unanswered question in the literature.

Implications for research

Treatment effect size for many of the studies was modest. Multi-faceted interventions that incorporate exercise strategies into patient care may provide greater benefit and should be tested.

1. Identify possible predictors of patient responsiveness to therapeutic exercise, such as radiographic disease severity, symptom duration, outcomes expectancy, psychological well being, obesity, knee stability, etc.
2. Develop multi-armed placebo-controlled randomised clinical trials to help provide evidence of optimal exercise content and dosage.
3. Initiate research to assess the long-term effectiveness of exercise for people with knee OA in terms of structural disease progression.

ACKNOWLEDGEMENTS

Louise Falzon, Mt Sinai Medical Centre, New York, for designing the literature search strategy.

Dr Renea Johnston, Managing Editor, Australian Editorial Base, Cochrane Musculoskeletal Review Group, for overall guidance and expert advice.

Tamara Rader, Cochrane Musculoskeletal Review Group, for designing the updated literature search strategy.

REFERENCES

References to studies included in this review

Abbott 2013 *{published data only}*

Abbott JH, Robertson MC, Chapple C, Pinto D, Wright AA, Leon de la Barras L, et al. Manual therapy, exercise therapy, or both, in addition to usual care for osteoarthritis of the hip or knee: a randomized controlled trial. I: Clinical effectiveness. *Osteoarthritis and Cartilage* 2013;**21**:525-34.

An 2008 *{published data only}*

An B, Dai K, Zhu Z, Wang Y, Hao Y, Tang T, et al. Baduanjin alleviates the symptoms of knee osteoarthritis. *Journal of Alternative & Complementary Medicine* 2008;**14**(2):167-74.

Baker 2001 *{published data only}*

Baker KR, Nelson ME, Felson DT, Layne JE, Sarno R, Roubenoff R. The efficacy of home based progressive strength training in older adults with knee osteoarthritis: a randomized controlled trial. *Journal of Rheumatology* 2001;**28**:1655-65.

Bautch 1997 *{published data only}*

Bautch JC, Malone DG, Vailas AC. Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. *Arthritis Care & Research* 1997;**10**:48-55.

Bennell 2005 *{published data only}*

Bennell KL, Himan RS, Metcalf BR, Buchbinder R, McConnell J, McColl G, et al. Efficacy of physiotherapy management of knee joint osteoarthritis: a randomised, double blind, placebo controlled trial. *Annals of the Rheumatic Diseases* 2005;**64**:906-12.

Bennell 2010 *{published data only}*

Bennell KL, Hunt MA, Wrigley TV, Hunter DJ, McManus FJ, Hodges PW, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. *Osteoarthritis and Cartilage* 2010;**18**(5):621-8.

Bezalel 2010 *{published data only}*

Bezalel T, Carmeli E, Katz-Leurer M. The effect of a group education programme on pain and function through knowledge acquisition and home-based exercise among patients with knee osteoarthritis: a parallel randomised single-blind clinical trial. *Physiotherapy* 2010;**96**(2):137-43.

Brismée 2007 *{published data only}*

Brismée J, Paige RL, Chyu M, Boatright JD, Hagar JM, McCaleb JA, et al. Group and home-based Tai Chi in elderly subjects with knee osteoarthritis: a randomized controlled trial. *Clinical Rehabilitation* 2007;**21**(2):99-111.

Bruce-Brand 2012 *{published data only}*

Bruce-Brand RA, Walls RJ, Ong JC, Emerson BS, O'Byrne JM, Moyna NM. Effects of home-based resistance training and neuromuscular electrical stimulation in knee osteoarthritis: a randomized controlled trial. *BMC Musculoskeletal Disorders* 2012;**13**:118.

Chang 2012 *{published data only}*

Chang TF, Liou TH, Chen CH, Huang YC, Chang KH. Effects of elastic-band exercise on lower-extremity function among female patients with osteoarthritis of the knee. *Disability and Rehabilitation* 2012;**34**(20):1727-35.

Deyle 2000 *{published data only}*

Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB, Allison SC. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. *Annals of Internal Medicine* 2000;**132**(3):173-81.

Doi 2008 *{published data only}*

Doi T, Akai M, Fujino K, Iwaya T, Kurosawa H, Hayashi K, et al. Effect of home exercise of quadriceps on knee osteoarthritis compared with nonsteroidal antiinflammatory drugs: a randomized controlled trial. *American Journal of Physical Medicine & Rehabilitation* 2008;**87**(4):258-69.

Ettinger 1997a/b *{published data only}*

Ettinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T, et al. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). *JAMA* 1997;**277**:25-31.

Foley 2003 *{published data only}*

Foley A, Halbert J, Hewitt T, Crotty M, Halbert J, Hewitt T, et al. Does hydrotherapy improve strength and physical function in patients with osteoarthritis—a randomised controlled trial comparing a gym based and a hydrotherapy based strengthening program. *Annals of the Rheumatic Diseases* 2003;**62**:1162-7.

Foroughi 2011 *{published data only}*

Foroughi N, Smith RM, Lange AK, Baker MK, Fiatarone Singh MA, Vanwanseele B. Lower limb muscle strengthening does not change frontal plane moments in women with knee osteoarthritis: a randomized controlled trial. *Clinical Biomechanics* 2011;**26**(2):167-74.

Fransen 2001 *{published data only}*

Fransen M, Crosbie J, Edmonds J. Physical therapy is effective for patients with osteoarthritis of the knee: a randomized controlled trial. *Journal of Rheumatology* 2001;**28**:156-64.

Fransen 2007 *{published data only}*

Fransen M, Nairn L, Winstanley J, Lam P, Edmonds J. The Physical Activity for Osteoarthritis Management (PAFORM) study. A randomised controlled clinical trial evaluating hydrotherapy and Tai Chi classes. *Arthritis Care & Research* 2007;**57**:407-14.

Gur 2002 *{published data only}*

Gur H, Cakin N, Akova B, Okay E, Kucukoglu S. Concentric versus combined concentric-eccentric isokinetic training: effects on functional capacity and symptoms in patients with osteoarthrosis of the knee. *Archives of Physical and Medical Rehabilitation* 2003;**83**:308-16.

Hay 2006 *{published data only}*

Hay EM, Foster NE, Thomas E, Peat G, Phelan M, Yates HE, et al. Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial. *BMJ* 2006;**333**:995.

Hopman-Rock 2000 *{published data only}*

Hopman-Rock M, Westhoff M. The effects of a health educational and exercise program for older adults with osteoarthritis of the hip or knee. *Journal of Rheumatology* 2000;**27**:1947-54.

Huang 2003 *{published data only}*

Huang M-H, Lin Y-S, Yang R-C, Lee C-L. A comparison of various therapeutic exercises on the functional status of patients with knee osteoarthritis. *Seminars in Arthritis and Rheumatism* 2003;**32**:398-406.

Huang 2005 *{published data only}*

Huang M-H, Yang R-C, Lee C-L, Chen T-W, Wang M-C. Preliminary results of integrated therapy for patients with knee osteoarthritis. *Arthritis Care & Research* 2005;**53**:812-20.

Hughes 2004 *{published data only}*

Hughes SL, Seymour RB, Campbell R, Pollak N, Huber G, Sharma L. Impact of the Fit and Strong intervention on older adults with osteoarthritis. *The Gerontologist* 2004;**44**:217-28.

Hurley 2007 *{published data only}*

Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Patel A, Williamson E, et al. Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial. *Arthritis & Rheumatism* 2007;**57**(7):1211-9.

Jan 2008 *{published data only}*

Jan M, Lin J, Liau J, Lin Y, Lin D. Investigation of clinical effects of high- and low-resistance training for patients with knee osteoarthritis: a randomized controlled trial. *Physical Therapy* 2008;**88**(4):427-36.

Jan 2009 *{published data only}*

Jan M, Lin C, Lin Y, Lin J, Lin D. Effects of weight-bearing versus nonweight-bearing exercise on function, walking speed, and position sense in participants with knee osteoarthritis: a randomized controlled trial. *Archives of Physical Medicine & Rehabilitation* 2009;**90**(6):897-904.

Jenkinson 2009 *{published data only}*

Jenkinson CM, Doherty M, Avery AJ, Read A, Taylor MA, Sach TH, et al. Effects of dietary intervention and quadriceps strengthening exercises on pain and function in overweight people with knee pain: randomised controlled trial. *BMJ* 2009;**339**:b3170.

Kao 2012 *{published data only}*

Kao MJ, Wu MP, Tsai MW, Chang WW, Wu SF. The effectiveness of a self-management program on quality of life for knee osteoarthritis (OA) patients. *Archives of Gerontology and Geriatrics* 2012;**54**(2):317-24.

Keefe 2004 *{published data only}*

Keefe FJ, Blumenthal J, Baucom D, Affleck G, Waugh R, Caldwell DS, et al. Effects of spouse-assisted coping skills training and exercise training in patients with osteoarthritic knee pain: a randomized controlled study. *Pain* 2004;**110**:539-49.

Kovar 1992 *{published data only}*

Kovar PA, Allegrante JP, MacKenzie CR, Peterson MGE, Gutin B, Charlson ME. Supervised fitness walking in patients with osteoarthritis of the knee. *Annals of Internal Medicine* 1992;**116**:529-34.

Lee 2009 *{published data only}*

Lee H, Park H, Chae Y, Kim S, Kim J, Yin C. Tai Chi Qigong for the quality of life of patients with knee osteoarthritis: a pilot, randomized, waiting list controlled trial. *Clinical Rehabilitation* 2009;**23**(6):504-11.

Lim 2008 *{published data only}*

Lim BW, Hinman RS, Wrigley TV, Sharma L, Bennell KL. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. *Arthritis Care & Research* 2008;**59**(7):943-51.

Lin 2009 *{published data only}*

Lin D, Lin CJ, Lin Y, Jan M. Efficacy of 2 non-weight-bearing interventions, proprioception training versus strength training, for patients with knee osteoarthritis: a randomized clinical trial. *Journal of Orthopaedic & Sports Physical Therapy* 2009;**39**(6):450-7.

Lund 2008 *{published data only}*

Lund H, Weile U, Christensen R, Rostock B, Downey A, Bartels EM, et al. A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. *Journal of Rehabilitation Medicine* 2008;**40**(2):137-44.

Maurer 1999 *{published data only}*

Maurer BT, Stern AG, Kinossian B, Cook KD, Schumacher HR. Osteoarthritis of the knee: isokinetic quadriceps exercise versus an educational intervention. *Archives of Physical Medicine and Rehabilitation* 1999;**80**:1293-9.

Messier 2004 *{published data only}*

Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, et al. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis. *Arthritis & Rheumatism* 2004;**50**:1501-10.

Mikesky 2006 *{published data only}*

Mikesky AE, Mazzuca SA, Brandt KD, Perkins SM, Damush T, Lane KA. Effects of strength training on the incidence and progression of knee osteoarthritis. *Arthritis Care & Research* 2006;**55**:690-9.

Minor 1989 *{published data only}*

Minor MA, Hewett JE, Webel RR, Anderson SK, Kay DR. Efficacy of physical conditioning exercise in patients with

rheumatoid arthritis and osteoarthritis. *Arthritis & Rheumatism* 1989;**32**:1396-405.

O'Reilly 1999 *{published data only}*

O'Reilly SC, Muir KR, Doherty M. Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomised controlled trial. *Annals of the Rheumatic Diseases* 1999;**58**:15-9.

Peloquin 1999 *{published data only}*

Peloquin L, Bravo G, Gauthier P, Lacombe G, Billiard J-S. Effects of a cross-training exercise program in persons with osteoarthritis of the knee. A randomized controlled trial. *Journal of Clinical Rheumatology* 1999;**5**:126-36.

Quilty 2003 *{published data only}*

Quilty B, Tucker M, Campbell R, Dieppe P. Physiotherapy, including quadriceps exercises and patellar taping, for knee osteoarthritis with predominant patello-femoral joint involvement: randomized controlled trial. *Journal of Rheumatology* 2003;**30**:1311-7.

Rogind 1998 *{published data only}*

Rogind H, Bibow-Nielsen B, Jensen B, Moller HC, Frimodt-Moller H, Bliddal H. The effects of a physical training program on patients with osteoarthritis of the knees. *Archives of Physical Medicine and Rehabilitation* 1998;**79**:1421-7.

Salacinski 2012 *{published data only}*

Salacinski AJ, Krohn K, Lewis SF, Holland ML, Ireland K, Marchetti G. The effects of group cycling on gait and pain-related disability in individuals with mild-to-moderate knee osteoarthritis: a randomized controlled trial. *Journal of Orthopaedic & Sports Physical Therapy* 2012;**42**(12):985-95.

Salli 2010 *{published data only}*

Salli A, Sahin N, Baskent A, Ugurlu H. The effect of two exercise programs on various functional outcome measures in patients with osteoarthritis of the knee: a randomized controlled clinical trial. *Isokinetics & Exercise Science* 2010;**18**(4):201-9.

Schilke 1996 *{published data only}*

Schilke JM, Johnson GO, Housh TJ, O'Dell JR. Effects of muscle-strength training on the functional status of patients with osteoarthritis of the knee joint. *Nursing Research* 1996;**45**:68-72.

Simao 2012 *{published data only}*

Simao AP, Avelar NC, Tossige-Gomes R, Neves CD, Mendonca VA, Miranda AS, et al. Functional performance and inflammatory cytokines after squat exercises and whole-body vibration in elderly individuals with knee osteoarthritis. *Archives of Physical Medicine & Rehabilitation* 2012;**93**(10):1692-700.

Song 2003 *{published data only}*

Song R, Lee E-O, Lam P, Bae S-C. Effects of Tai Chi exercise on pain, balance, muscle strength, and perceived difficulties in physical functioning in older women with osteoarthritis: a randomized clinical trial. *Journal of Rheumatology* 2003;**30**:2039-44.

Talbot 2003 *{published data only}*

* Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. *Journal of the American Geriatric Society* 2003;**51**:387-92.

Thomas 2002 *{published data only}*

Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. *BMJ* 2002;**325**:752-7.

Thorstensson 2005 *{published data only}*

Thorstensson CA, Roos EM, Petersson IF, Ekdahl C. Six-week high-intensity exercise program for middle-aged patients with knee osteoarthritis: a randomized controlled trial. *BMC Musculoskeletal Disorders* 2005;**6**:27.

Topp 2002 *{published data only}*

Topp R, Woolley S, Horuyak J, Khuder S, Kahaleh B. The effect of dynamic versus isometric resistance training on pain and functioning among adults with osteoarthritis of the knee. *Archives of Physical and Medical Rehabilitation* 2002;**83**:1187-95.

van Baar 1998 *{published data only}*

van Baar ME, Dekker J, Oostendorp RAB, Bijl D, Voorn TB, Lemmens JAM, et al. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. *Journal of Rheumatology* 1998;**25**:2432-9.

Wang 2011 *{published data only}*

Wang TJ, Lee SC, Liang SY, Tung HH, Wu SF, Lin YP. Comparing the efficacy of aquatic exercises and land-based exercises for patients with knee osteoarthritis. *Journal of Clinical Nursing* 2011;**20**(17-18):2609-22.

Yip 2007 *{published data only}*

Yip YB, Wit JW, Fung KKY, Wong DYS, Chong SYC, Chung LH, et al. Impact of an arthritis self-management programme with an added exercise component for osteoarthritic knee sufferers on improving pain, functional outcomes, and use of health care services: an experimental study. *Patient Education Counseling* 2007;**65**:113-21.

References to studies excluded from this review

Ageberg 2010 *{published data only}*

Ageberg E, Link A, Roos EM. Feasibility of neuromuscular training in patients with severe hip or knee OA: the individualized goal-based NEMEX-TJR training program. *BMC Musculoskeletal Disorders* 2010;**11**:126.

Aglamis 2008 *{published data only}*

Aglamis B, Toraman NF, Yaman H. The effect of a 12-week supervised multicomponent exercise program on knee OA in Turkish women. *Journal of Back & Musculoskeletal Rehabilitation* 2008;**21**(2):121-8.

Aglamis 2009 *{published data only}*

Aglamis B, Toraman NF, Yaman H. Change of quality of life due to exercise training in knee osteoarthritis: SF-36 and WOMAC. *Journal of Back & Musculoskeletal Rehabilitation* 2009;**22**(1):43-8.

Akyol 2010 *{published data only}*

Akyol Y, Durmus D, Alayli G, Tander B, Bek Y, Canturk F, et al. Does short-wave diathermy increase the effectiveness of isokinetic exercise on pain, function, knee muscle strength, quality of life, and depression in the patients with knee osteoarthritis? A randomized controlled clinical study. *European Journal of Physical & Rehabilitation Medicine* 2010;**46**(3):325-36.

Alfredo 2012 *{published data only}*

Alfredo PP, Bjordal JM, Dreyer SH, Ferreira, Zaguetti G, Ovanessian V, et al. Efficacy of low level laser therapy associated with exercises in knee osteoarthritis: a randomized double-blind study. *Clinical Rehabilitation* 2012;**26**(6):523-33.

Anwer 2011 *{published data only}*

Anwer S, Quddus N, Miraj M, Equebal A. Effectiveness of electromyographic biofeedback training on quadriceps muscle strength in osteoarthritis of knee. *Hong Kong Physiotherapy Journal* 2011;**29**(2):86-93.

Aoki 2009 *{published data only}*

Aoki O, Tsumura N, Kimura A, Okuyama S, Takikawa S, Hirata S. Home stretching exercise is effective for improving knee range of motion and gait in patients with knee osteoarthritis. *Journal of Physical Therapy Science* 2009;**21**(2):113-9.

Atamaz 2006 *{published data only}*

Atamaz F, Kirazli Y, Akkoc Y, Atamaz F, Kirazli Y, Akkoc Y. A comparison of two different intra-articular hyaluronan drugs and physical therapy in the management of knee osteoarthritis. *Rheumatology International* 2006;**26**:873-8.

Atamaz 2012 *{published data only}*

Atamaz FC, Durmaz B, Baydar M, Demircioglu OY, Iyiyapici A, Kuran B, et al. Comparison of the efficacy of transcutaneous electrical nerve stimulation, interferential currents, and shortwave diathermy in knee osteoarthritis: a double-blind, randomized, controlled, multicenter study. *Archives of Physical Medicine & Rehabilitation* 2012;**93**(5):748-56.

Boocock 2009 *{published data only}*

Boocock M, McNair P, Cicuttini F, Stuart A, Sinclair T. The short-term effects of running on the deformation of knee articular cartilage and its relationship to biomechanical loads at the knee. *Osteoarthritis and Cartilage* 2009;**17**(7):883-90.

Borjesson 1996 *{published data only}*

Borjesson M, Robertson E, Weidenhielm L, Mattsson E, Olsson E. Physiotherapy in knee osteoarthrosis effect on pain and walking. *Physiotherapy Research International* 1996;**1**:89-97.

Brosseau 2012 *{published data only}*

Brosseau L, Wells GA, Kenny GP, Reid R, Maetzel A, Tugwell P, et al. The implementation of a community-based aerobic walking program for mild to moderate knee osteoarthritis (OA): a knowledge translation (KT) randomized controlled trial (RCT): Part I: The Uptake of the Ottawa Panel clinical practice guidelines (CPGs). *BMC Public Health* 2012;**12**:871.

Bulthuis 2007 *{published data only}*

Bulthuis Y, Drossaers-Bakker KW, Taal E, Rasker J, Oostveen J, van't Pad Bosch P, et al. Arthritis patients show long-term benefits from 3 weeks intensive exercise training directly following hospital discharge. *Rheumatology* 2007;**46**(11):1712-7.

Bulthuis 2008 *{published data only}*

Bulthuis Y, Mohammad S, Braakman-Jansen LMA, Drossaers-Bakker KW, van de Laar MAFJ. Cost-effectiveness of intensive exercise therapy directly following hospital discharge in patients with arthritis: results of a randomized controlled clinical trial. *Arthritis & Rheumatism* 2008;**59**(2):247-54.

Callaghan 1995 *{published data only}*

Callaghan MJ, Oldham JA, Hunt J. An evaluation of exercise regimes for patients with osteoarthritis of the knee: a single-blind randomized controlled trial. *Clinical Rehabilitation* 1995;**9**:213-8.

Cetin 2008 *{published data only}*

Cetin N, Aytar A, Atalay A, Akman MN. Comparing hot pack, short-wave diathermy, ultrasound, and TENS on isokinetic strength, pain, and functional status of women with osteoarthritic knees: a single-blind, randomized, controlled trial. *American Journal of Physical Medicine & Rehabilitation* 2008;**87**(6):443-51.

Chaipinyo 2009 *{published data only}*

Chaipinyo K, Karoonsupcharoen O. No difference between home-based strength training and home-based balance training on pain in patients with knee osteoarthritis: a randomised trial. *Australian Journal of Physiotherapy* 2009;**55**(1):25-30.

Chamberlain 1982 *{published data only}*

Chamberlain MA, Care G, Harfield B. Physiotherapy in osteoarthrosis of the knees. A controlled trial of hospital versus home exercises. *International Rehabilitation Medicine* 1982;**4**:101-6.

Cheing 2002 *{published data only}*

Cheing GLY, Hui-Chan CWY, Chan KM. Does four weeks of TENS and/or isometric exercise produce cumulative reduction of osteoarthritic knee pain?. *Clinical Rehabilitation* 2002;**16**:749-60.

Cheing 2004 *{published data only}*

Cheing GLY, Hui-Chan CWY. Would the addition of TENS to exercise training produce better physical performance outcomes in people with knee osteoarthritis than either intervention alone?. *Clinical Rehabilitation* 2004;**18**:487.

Ciolac 2011 *{published data only}*

Ciolac EG, Greve JMDA. Muscle strength and exercise intensity adaptation to resistance training in older women with knee osteoarthritis and total knee arthroplasty. *Clinics (Sao Paulo, Brazil)* 2011;**66**(12):2079-84.

Coupe 2007 *{published data only}*

Coupe VM, Veenhof C, van Tulder MW, Dekker J, Bijlsma JW, Van den Ende CH. The cost effectiveness of behavioural graded activity in patients with osteoarthritis of hip and/or knee. *Annals of the Rheumatic Diseases* 2007;**66**(2):215-21.

Crotty 2009 *{published data only}*

Crotty M, Prendergast J, Battersby MW, Rowett D, Graves SE, Leach G, et al. Self-management and peer support among people with arthritis on a hospital joint replacement waiting list: a randomised controlled trial. *Osteoarthritis and Cartilage* 2009;**17**(11):1428-33.

Deyle 2005 *{published data only}*

Deyle GD, Allison SC, Matekel RL, Ryder MG, Stang JM, Gohdes DD, et al. Physical therapy treatment effectiveness for osteoarthritis of the knee: a randomized comparison of supervised clinical exercise and manual therapy procedures versus a home exercise program. *Physical Therapy* 2005;**12**:1301-17.

Dias 2003 *{published data only}*

Dias RC, Ramos LR, Dias JM. Impact of an exercise and walking protocol on the quality of life of the elderly with osteoarthritis of the knee. *Physiotherapy Research International* 2003;**8**:121-30.

Diracoglu 2005 *{published data only}*

Diracoglu D, Aydin R, Baskent A, Celik A, Diracoglu D, Aydin R, et al. Effects of kinesthesia and balance exercises in knee osteoarthritis. *Journal of Clinical Rheumatology* 2005;**11**:303-10.

Duman 2012 *{published data only}*

Duman I, Taskaynatan MA, Mohur H, Tan AK. Assessment of the impact of proprioceptive exercises on balance and proprioception in patients with advanced knee osteoarthritis. *Rheumatology international* 2012;**32**(12):3793-8.

Durmus 2007 *{published data only}*

Durmus D, Alayli G, Canturk F. Effects of quadriceps electrical stimulation program on clinical parameters in the patients with knee osteoarthritis. *Clinical Rheumatology* 2007;**26**:674-8.

Durmus 2012 *{published data only}*

Durmus D, Alayli G, Bayrak IK, Canturk F. Assessment of the effect of glucosamine sulfate and exercise on knee cartilage using magnetic resonance imaging in patients with knee osteoarthritis: a randomized controlled clinical trial. *Journal of Back & Musculoskeletal Rehabilitation* 2012;**25**(4):275-84.

Ebnezar 2012 *{published data only}*

Ebnezar J, Nagarathna R, Yogitha B, Nagendra HR. Effects of an integrated approach of hatha yoga therapy on functional disability, pain, and flexibility in osteoarthritis of the knee joint: a randomized controlled study. *Journal of Alternative & Complementary Medicine* 2012;**18**(5):463-72.

Ebnezar 2012a *{published data only}*

Ebnezar J, Nagarathna R, Yogitha B, Nagendra HR. Effect of integrated yoga therapy on pain, morning stiffness and anxiety in osteoarthritis of the knee joint: a randomized control study. *International Journal of Yoga* 2012;**5**(1):28-36.

Evcik 2002 *{published data only}*

Evcik D, Sonel B. Effectiveness of a home-based exercise therapy and walking program on osteoarthritis of the knee. *Rheumatology International* 2002;**22**:103-6.

Evgeniadis 2008 *{published data only}*

Evgeniadis G, Beneka A, Malliou P, Mavromoustakos S, Godolias G. Effects of pre- or postoperative therapeutic exercise on the quality of life, before and after total knee arthroplasty for osteoarthritis. *Journal of Back & Musculoskeletal Rehabilitation* 2008;**21**(3):161-9.

Eyigor 2004 *{published data only}*

Eyigor S, Hepguler S, Capaci K. A comparison of muscle training methods in patients with knee osteoarthritis. *Clinical Rheumatology* 2004;**23**:109-15.

Farr 2010 *{published data only}*

Farr JN, Going SB, McKnight PE, Kasle S, Cussler EC, Cornett M. Progressive resistance training improves overall physical activity levels in patients with early osteoarthritis of the knee: a randomized controlled trial. *Physical Therapy* 2010;**90**(3):356-66.

Feinglass 2012 *{published data only}*

Feinglass J, Song J, Semanik P, Lee J, Manheim L, Dunlop D, et al. Association of functional status with changes in physical activity: insights from a behavioral intervention for participants with arthritis. *Archives of Physical Medicine & Rehabilitation* 2012;**93**(1):172-5.

Fitzgerald 2011 *{published data only}*

Fitzgerald GK, Piva SR, Gil AB, Wisniewski SR, Oddis CV, Irrgang JJ. Agility and perturbation training techniques in exercise therapy for reducing pain and improving function in people with knee osteoarthritis: a randomized clinical trial. *Physical Therapy* 2011;**91**(4):452-69.

Forestier 2010 *{published data only}*

Forestier R, Desfour H, Tessier JM, Francon A, Foote AM, Genty C, et al. Spa therapy in the treatment of knee osteoarthritis: a large randomised multicentre trial. *Annals of the Rheumatic Diseases* 2010;**69**(4):660-5.

Foroughi 2011a *{published data only}*

Foroughi N, Smith RM, Lange AK, Singh MA, Vanwanseele B. Progressive resistance training and dynamic alignment in osteoarthritis: a single-blind randomised controlled trial. *Clinical Biomechanics (Bristol, Avon)* 2011;**26**(1):71-7.

Foster 2007 *{published data only}*

Foster NE, Thomas E, Barlas P, Hill JC, Young J, Mason E, et al. Acupuncture as an adjunct to exercise based physiotherapy for osteoarthritis of the knee: randomised controlled trial. *BMJ* 2007;**335**(7617):436.

Gaal 2008 *{published data only}*

Gaal J, Varga J, Szekanecz Z, Kurko J, Ficzere A, Bodolay E, et al. Balneotherapy in elderly patients: effect on pain from degenerative knee and spine conditions and on quality of life. *Israel Medical Association Journal* 2008;**10**(5):365-9.

Gaudreault 2011 *{published data only}*

Gaudreault N, Mezghani N, Turcot K, Hagemeister N, Boivin K, de Guise JA. Effects of physiotherapy treatment on knee osteoarthritis gait data using principal component analysis. *Clinical Biomechanics* 2011;**26**(3):284-91.

Gill 2009 *{published data only}*

Gill SD, McBurney H, Schulz DL. Land-based versus pool-based exercise for people awaiting joint replacement surgery of the hip or knee: results of a randomized controlled trial. *Archives of Physical Medicine & Rehabilitation* 2009;**90**(3):388-94.

Green 1993 *{published data only}*

Green J, McKenna F, Redfern EJ, Chamberlain MA. Home exercises are as effective as outpatient hydrotherapy for osteoarthritis of the hip. *British Journal of Rheumatology* 1993;**32**:812-5.

Gremion 2009 *{published data only}*

Gremion G, Gaillard D, Leyvraz PF, Jolles BM. Effect of biomagnetic therapy versus physiotherapy for treatment of knee osteoarthritis: a randomized controlled trial. *Journal of Rehabilitation Medicine (Stiftelsen Rehabiliteringsinformation)* 2009;**41**(13):1090-5.

Haslam 2001 *{published data only}*

Haslam R. A comparison of acupuncture with advice and exercises on the symptomatic treatment of osteoarthritis of the hip—a randomised controlled trial. *Acupuncture Medicine* 2001;**19**:19-26.

Helmark 2010 *{published data only}*

Helmark IC, Mikkelsen UR, Borglum J, Rothe A, Petersen MC, Andersen O, et al. Exercise increases interleukin-10 levels both intraarticularly and peri-synovially in patients with knee osteoarthritis: a randomized controlled trial. *Arthritis Research & Therapy* 2010;**12**(4):R126.

Helmark 2012 *{published data only}*

Helmark IC, Petersen MCH, Christensen HE, Kjaer M, Langberg H. Moderate loading of the human osteoarthritic knee joint leads to lowering of intraarticular cartilage oligomeric matrix protein. *Rheumatology International* 2012;**32**(4):1009-14.

Hinman 2007 *{published data only}*

Hinman RS, Heywood SE, Day AR. Aquatic physical therapy for hip and knee osteoarthritis: results of a single-blind randomized controlled trial. *Physical Therapy* 2007;**87**:32-43.

Hiyama 2012 *{published data only}*

Hiyama Y, Yamada M, Kitagawa A, Tei N, Okada S. A four-week walking exercise programme in patients with knee osteoarthritis improves the ability of dual-task performance: a randomized controlled trial. *Clinical Rehabilitation* 2012;**26**(5):403-12.

Hoeksma 2004 *{published data only}*

Hoeksma HL, Dekker J, Ronday HK, Heering A, van der Lubbe N, Vel C, et al. Comparison of manual therapy and exercise therapy in osteoarthritis of the hip: a randomized clinical trial. *Arthritis & Rheumatism* 2004;**51**:722-9.

Huang 2005b *{published data only}*

Huang M-H, Lin Y-H, Lee C-L, Yang R-C. Use of ultrasound to increase effectiveness of isokinetic exercise for knee osteoarthritis. *Archives of Physical and Medical Rehabilitation* 2005;**86**:1545-51.

Hughes 2010 *{published data only}*

Hughes SL, Seymour RB, Campbell RT, Desai P, Huber G, Chang HJ. Fit and Strong!: bolstering maintenance of physical activity among older adults with lower-extremity osteoarthritis. *American Journal of Health Behavior* 2010;**34**(6):750-63.

Hurley 1998 *{published data only}*

Hurley MV, Scott DL. Improvements in quadriceps sensorimotor function and disability of patients with knee osteoarthritis following a clinically practicable exercise regime. *British Journal of Rheumatology* 1998;**37**:1181-7.

Hurley 2007a *{published data only}*

Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Williamson E, Jones RH, et al. Economic evaluation of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain. *Arthritis & Rheumatism* 2007;**57**(7):1220-9.

Hurley 2012 *{published data only}*

Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. *Arthritis Care & Research* 2012;**64**(2):238-47.

Jan 1991 *{published data only}*

Jan M-H, Lai J-S. The effects of physiotherapy on osteoarthritic knees of females. *Journal of the Formosan Medical Association* 1991;**90**:1008-13.

Jan 2008a *{published data only}*

Jan M, Tang P, Lin J, Teng S, Lin Y, Lin D. Efficacy of a target-matching foot-stepping exercise on proprioception and function in patients with knee osteoarthritis. *Journal of Orthopaedic & Sports Physical Therapy* 2008;**38**(1):19-25.

Jessep 2009 *{published data only}*

Jessep SA, Walsh NE, Ratcliffe J, Hurley MV. Long-term clinical benefits and costs of an integrated rehabilitation programme compared with outpatient physiotherapy for chronic knee pain. *Physiotherapy* 2009;**95**(2):94-102.

Karagulle 2007 *{published data only}*

Karagulle M, Karagulle MZ, Karagulle O, Donmez A, Turan M. A 10-day course of SPA therapy is beneficial for people with severe knee osteoarthritis. A 24-week randomised, controlled pilot study. *Clinical Rheumatology* 2007;**26**(12):2063-71.

Kawasaki 2008 *{published data only}*

Kawasaki T, Kurosawa H, Ikeda H, Kim SG, Osawa A, Takazawa Y, et al. Additive effects of glucosamine or risedronate for the treatment of osteoarthritis of the knee combined with home exercise: a prospective randomized 18-month trial. *Journal of Bone and Mineral Metabolism* 2008;**26**(3):279-87.

Kawasaki 2009 *{published data only}*

Kawasaki T, Kurosawa H, Ikeda H, Takazawa Y, Ishijima M, Kubota M, et al. Therapeutic home exercise versus intraarticular hyaluronate injection for osteoarthritis of the knee: 6-month prospective randomized open-labelled trial. *Journal of Orthopaedic Science* 2009;**14**(2):182-91.

King 2008 *{published data only}*

King LK, Birmingham TB, Kean CO, Jones IC, Bryant DM, Giffin JR. Resistance training for medial compartment knee osteoarthritis and malalignment. *Medicine & Science in Sports & Exercise* 2008;**40**(8):1376-84.

Konishi 2009 *{published data only}*

Konishi I, Tanabe N, Seki N, Suzuki H, Okamura T, Shinoda K, et al. Physiotherapy program through home visits for community-dwelling elderly Japanese women with mild knee pain. *Tohoku Journal of Experimental Medicine* 2009;**219**(2):91-9.

Kreindler 1989 *{published data only}*

Kreindler H, Lewis CB, Rush S, Schaefer K. Effects of three exercise protocols on strength of persons with osteoarthritis of the knee. *Topics in Geriatriatric Rehabilitation* 1989;**4**:32-9.

Kuptniratsaikul 2002 *{published data only}*

Kuptniratsaikul V, Orchatara T, Nilganuwong S, Visanu T. The efficacy of a muscle exercise program to improve functional performance of the knee in patients with osteoarthritis. *Journal of the Medical Association of Thailand* 2002;**85**:33-9.

Lankhorst 1982 *{published data only}*

Lankhorst GJ, van de Stadt RJ, van der Korst JK, Hinlopen-Bonrath E, Griffioen FM, de Boer W. Relationship of isometric knee extension torque and functional variables in osteoarthrosis of the knee. *Scandinavian Journal of Rehabilitation Medicine* 1982;**14**:7-10.

Lim 2002 *{published data only}*

Lim BW. A comparative study of open and closed kinetic chain exercise regimes in patients with knee osteoarthritis. *Physiotherapy Singapore* 2002;**5**:34-40.

Lim 2010 *{published data only}*

Lim JY, Tchai E, Jang SN. Effectiveness of aquatic exercise for obese patients with knee osteoarthritis: a randomized controlled trial. *PM&R* 2010;**2**(8):723-31; quiz 93.

Lin 2004 *{published data only}*

Lin SY, Davey RC, Cochrane T. Community rehabilitation for older adults with osteoarthritis of the lower limb: a controlled clinical trial. *Clinical Rehabilitation* 2004;**18**:92-101.

Lin 2007 *{published data only}*

Lin DH, Lin YF, Chai HM, Han YC, Jan MH. Comparison of proprioceptive functions between computerized proprioception facilitation exercise and closed kinetic chain exercise in patients with knee osteoarthritis. *Clinical Rheumatology* 2007;**26**(4):520-8.

Liu 2008 *{published data only}*

Liu Y, Guo L, Ma S. Treatment of 256 cases of osteoarthritis of knee joint with Guo Jianhua's four-step therapy. *Journal of Traditional Chinese Medicine* 2008;**28**(2):114-7.

Mangione 1999 *{published data only}*

Mangione KK, McCully K, Gloviak A, Lefebvre I, Hofmann M, Craik R. The effects of high-intensity and low-intensity cycle ergometry in older adults with knee osteoarthritis. *Journal of Gerontology* 1999;**54A**:M184-90.

Marra 2012 *{published data only}*

Marra CA, Cibere J, Grubisic M, Grindrod KA, Gastonguay L, Thomas JM, et al. Pharmacist-initiated intervention trial in osteoarthritis: a multidisciplinary intervention for knee osteoarthritis. *Arthritis Care & Research* 2012;**64**(12):1837-45.

Mascarin 2012 *{published data only}*

Mascarin NC, Vancini RL, Andrade ML, Magalhaes Ede P, de Lira CA, Coimbra IB. Effects of kinesiotherapy, ultrasound and electrotherapy in management of bilateral knee osteoarthritis: prospective clinical trial. *BMC Musculoskeletal Disorders* 2012;**13**:182.

McCarthy 2004 *{published data only}*

McCarthy CJ, Mills PM, Pullen R, Roberts C, Silman A, Oldham JA. Supplementing a home exercise programme with a class-based exercise programme is more effective than home exercise alone in the treatment of knee osteoarthritis. *Rheumatology* 2004;**43**:880-6.

McKnight 2010 *{published data only}*

McKnight PE, Kasle S, Going S, Villanueva I, Cornett M, Farr J, et al. A comparison of strength training, self-management, and the combination for early osteoarthritis of the knee. *Arthritis Care & Research* 2010;**62**(1):45-53.

McQuade 2011 *{published data only}*

McQuade KJ, De Oliveira AS. Effects of progressive resistance strength training on knee biomechanics during single leg step-up in persons with mild knee osteoarthritis. *Clinical Biomechanics* 2011;**26**(7):741-8.

Messier 1997 *{published data only}*

Messier SP, Thompson CD, Ettinger WH Jr. Effects of long-term aerobic or weight training regimes on gait in an older, osteoarthritic population. *Journal of Applied Biomechanics* 1997;**13**:205-25.

Messier 2000a *{published data only}*

Messier SP, Loeser RF, Mitchell MN, Valle G, Morgan TP, Rejeski WJ, Ettinger WH. Exercise and weight loss in obese older adults with knee osteoarthritis: a preliminary study. *Journal of the American Geriatric Society* 2000;**48**:1062-72.

Messier 2000b *{published data only}*

Messier SP, .Royer TD, Craven TE, O'Toole ML, Burns R, Ettinger WH. Long-term exercise and its effect on balance in older, osteoarthritis adults: results from the Fitness, Arthritis and Seniors Trial (FAST). *Journal of the American Geriatric Society* 2000;**48**:131-8.

Messier 2007 *{published data only}*

Messier SP, Mihalko S, Loeser RF, Legault C, Jolla J, Pfruender J, et al. Glucosamine/chondroitin combined with exercise for the treatment of knee osteoarthritis: a preliminary study. *Osteoarthritis and Cartilage* 2007;**15**(11):1256-66.

Messier 2008 *{published data only}*

Messier SP, Legault C, Lenz ME, Thonar EJ, Loeser RF. Effect of an exercise and dietary intervention on serum biomarkers in overweight and obese adults with osteoarthritis of the knee. *Osteoarthritis and Cartilage* 2008;**16**(9):1047-53.

Miller 2012 *{published data only}*

Miller GD, Nicklas BJ, Davis CC, Legault C, Messier SP. Basal growth hormone concentration increased following a weight loss focused dietary intervention in older overweight and obese women. *Journal of Nutrition, Health & Aging* 2012;**16**(2):169-74.

Moss 2007 *{published data only}*

Moss P, Sluka K, Wright A. The initial effects of knee joint mobilization on osteoarthritic hyperalgesia. *Manual Therapy* 2007;**12**:109-18.

Murphy 2008 *{published data only}*

Murphy SL, Strasburg DM, Lyden AK, Smith DM, Koliba JF, Dadabhoy DP, et al. Effects of activity strategy training on pain and physical activity in older adults with knee or hip osteoarthritis: a pilot study. *Arthritis & Rheumatism* 2008;**59**(10):1480-7.

Neves 2011 *{published data only}*

Neves M, Gualano B, Roschel H, Fuller R, Benatti FB, Pinto ALDS, et al. Beneficial effect of creatine supplementation in knee osteoarthritis. *Medicine & Science in Sports & Exercise* 2011;**43**(8):1538-43.

Ng 2010 *{published data only}*

Ng NT, Heesch KC, Brown WJ. Efficacy of a progressive walking program and glucosamine sulphate supplementation on osteoarthritic symptoms of the hip and knee: a feasibility trial. *Arthritis Research & Therapy* 2010;**12**(1):R25.

Nicklas 2004 *{published data only}*

Nicklas BJ, Ambrosius W, Messier SP, Miller GD, Penninx BW, Loeser RF, et al. Diet-induced weight loss, exercise, and chronic inflammation in older, obese adults: a randomized controlled clinical trial. *American Journal of Clinical Nutrition* 2004;**79**:544-51.

Ozdincler 2005 *{published data only}*

Ozdincler AR, Yeldan I, Kinali P. The effects of closed kinetic chain exercise on pain and functional performance of patients with knee osteoarthritis. *The Pain Clinic* 2005;**1**:107-15.

Penninx 2001 *{published data only}*

Penninx B, Messier SP, Rejeski WJ, Williamson JD, DiBari M, Cavazzini C, et al. Physical exercise and prevention of disability in activities of daily living in older persons with osteoarthritis. *Archives of Internal Medicine* 2001;**19**:2309-16.

Penninx 2002 *{published data only}*

Penninx BW, Rejeski WJ, Pandya J, Miller ME, Di Bari M, Applegate WB, et al. Exercise and depressive symptoms: a comparison of aerobic and resistance exercise effects on emotional and physical function in older persons with high and low depressive symptomatology. *Journal of Gerontology* 2002;**57B**:P127-32.

Pereira, 2011 *{published data only}*

Pereira Simão AP, Tossige-Gomes R, Cunha N, Rocha-Vieira E, Coimbra CC, et al. The effect of adding whole-body vibration to squat training on the functional performance and self-report of disease status in elderly patients with knee osteoarthritis: a randomized, controlled clinical study. *Journal of Alternative & Complementary Medicine* 2011;**17**(12):1149-55.

Petersen 2010 *{published data only}*

Petersen SG, Saxne T, Heinegard D, Hansen M, Holm L, Koskinen S, et al. Glucosamine but not ibuprofen alters cartilage turnover in osteoarthritis patients in response to physical training. *Osteoarthritis and Cartilage* 2010;**18**(1):34-40.

Petersen 2011 *{published data only}*

Petersen SG, Beyer N, Hansen M, Holm L, Aagaard P, Mackey AL, et al. Nonsteroidal anti-inflammatory drug or glucosamine reduced pain and improved muscle strength with resistance training in a randomized controlled trial of knee osteoarthritis patients. *Archives of Physical Medicine & Rehabilitation* 2011;**92**(8):1185-93.

Peterson 1993 *{published data only}*

Peterson MGE, Kovar-Toledano PA, Otis JC, Allegrante JP, Mackenzie CR, Gutin B, Kross MA. Effect of a walking program on gait characteristics in patients with osteoarthritis. *Arthritis Care & Research* 1993;**6**:11-6.

Petrella 2000 *{published data only}*

Petrella RJ, Bartha C. Home based exercise therapy for older patients with knee osteoarthritis: a randomized clinical trial. *Journal of Rheumatology* 2000;**27**:2215-21.

Pietrosimone 2010 *{published data only}*

Pietrosimone BG, Saliba SA, Hart JM, Hertel J, Kerrigan DC, Ingersoll CD. Effects of disinhibitory transcutaneous electrical nerve stimulation and therapeutic exercise on sagittal plane peak knee kinematics and kinetics in people with knee osteoarthritis during gait: a randomized controlled trial. *Clinical Rehabilitation* 2010;**24**(12):1091-101.

Pietrosimone 2012 *{published data only}*

Pietrosimone BG, Saliba SA. Changes in voluntary quadriceps activation predict changes in quadriceps strength after therapeutic exercise in patients with knee osteoarthritis. *Knee* 2012;**19**(6):939-43.

Pisters 2010 *{published data only}*

Pisters MF, Veenhof C, de Bakker DH, Schellevis FG, Dekker J. Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a cluster-randomised trial. *Journal of Physiotherapy* 2010;**56**(1):41-7.

Pisters 2010a *{published data only}*

Pisters MF, Veenhof C, Schellevis FG, De Bakker DH, Dekker J. Long-term effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized controlled trial comparing two different physical therapy interventions. *Osteoarthritis and Cartilage* 2010;**18**(8):1019-26.

Piva 2011 *{published data only}*

Piva SR, Fitzgerald GK. Effects of impairment-based exercise on performance of specific self-reported functional tasks in individuals with knee osteoarthritis. *Physical Therapy* 2011;**91**(12):1752-65.

Piyakhachornrot 2011 *{published data only}*

Piyakhachornrot N, Aree-Ue S, Putwatana P, Kawinwonggowit V. Impact of an integrated health education and exercise program in middle-aged Thai adults with osteoarthritis of the knee. *Orthopaedic Nursing* 2011;**30**(2):134-42.

Quirk 1985 *{published data only}*

Quirk AS, Newman RJ, Newman KJ. An evaluation of interferential therapy, shortwave diathermy and exercise in the treatment of osteoarthrosis of the knee. *Physiotherapy* 1985;**71**:55-7.

Rattanachaiyanont 2008 *{published data only}*

Rattanachaiyanont M, Kuptniratsaikul V. No additional benefit of shortwave diathermy over exercise program for knee osteoarthritis in peri-/post-menopausal women: an equivalence trial. *Osteoarthritis and Cartilage* 2008;**16**(7):823-8.

Ravaud 2004 *{published data only}*

Ravaud P, Giraudeau B, Logeart I, Larguier JS, Rolland D, et al. Management of osteoarthritis with an unsupervised home based exercise programme and/or patient administered assessment tools. A cluster randomised controlled trial with a 2x2 factorial design. *Annals of the Rheumatic Diseases* 2004;**63**:703-8.

Reid 2010 *{published data only}*

Reid DA, McNair PJ. Effects of an acute hamstring stretch in people with and without osteoarthritis of the knee. *Physiotherapy* 2010;**96**(1):14-21.

Reid 2011 *{published data only}*

Reid DA, McNair PJ. Effects of a six week lower limb stretching programme on range of motion, peak passive torque and stiffness in people and without osteoarthritis of the knee. *New Zealand Journal of Physiotherapy* 2011;**39**(1):5-12.

Rejeski 1998 *{published data only}*

Rejeski WJ, Ettinger WH, Martin K, Morgan T. Treating disability in knee osteoarthritis with exercise therapy: a central role for self-efficacy and pain. *Arthritis Care & Research* 1998;**11**:94-101.

Sayers 2012 *{published data only}*

Sayers SP, Gibson K, Cook CR. Effect of high-speed power training on muscle performance, function, and pain in older adults with knee osteoarthritis: a pilot investigation. *Arthritis Care & Research* 2012;**64**(1):46-53.

Schlenk 2011 *{published data only}*

Schlenk EA, Lias JL, Sereika SM, Dunbar-Jacob J, Kwoh CK. Improving physical activity and function in overweight and obese older adults with osteoarthritis of the knee: a feasibility study. *Rehabilitation Nursing Journal* 2011;**36**(1):32-42.

Scopaz 2009 *{published data only}*

Scopaz KA, Piva SR, Gil AB, Woollard JD, Oddis CV, Fitzgerald GK. Effect of baseline quadriceps activation on changes in quadriceps strength after exercise therapy in subjects with knee osteoarthritis. *Arthritis & Rheumatism: Arthritis Care & Research* 2009;**61**(7):951-7.

Selfe 2008 *{published data only}*

Selfe TK, Bourguignon C, Taylor AG. Effects of noninvasive interactive neurostimulation on symptoms of osteoarthritis of the knee: a randomized, sham-controlled pilot study. *Journal of Alternative & Complementary Medicine* 2008;**14**(9):1075-81.

Sen 2004 *{published data only}*

Sen A, Gocen Z, Unver B, Karatosun V, Gunal I. The frequency of visits by the physiotherapist of patients receiving home-based exercise therapy for knee osteoarthritis. *Knee* 2004;**11**:151-3.

Sevick 2009 *{published data only}*

Sevick MA, Miller GD, Loeser RF, Williamson JD, Messier SP. Cost-effectiveness of exercise and diet in overweight and obese adults with knee osteoarthritis. *Medicine & Science in Sports & Exercise* 2009;**41**(6):1167-74.

Shakoor 2007 *{published data only}*

Shakoor MA, Taslim MA, Hossain MS. Effects of activity modification on the patients with osteoarthritis of the knee. *Bangladesh Medical Research Council Bulletin* 2007;**33**(2):55-9.

Shakoor 2010 *{published data only}*

Shakoor MA, Rahman MS, Azad AK, Islam MS. Effects of isometric quadriceps muscle strengthening exercise on chronic osteoarthritis of the knee. *Bangladesh Medical Research Council Bulletin* 2010;**36**(1):20-2.

Shen 2008 *{published data only}*

Shen CL, James CR, Chyu MC, Bixby WR, Brismée JM, Zumwalt MA, et al. Effects of Tai Chi on gait kinematics, physical function, and pain in elderly with knee osteoarthritis—a pilot study. *American Journal of Chinese Medicine* 2008;**36**(2):219-32.

Silva 2008 *{published data only}*

Silva LE, Valim V, Pessanha AP, Oliveira LM, Myamoto S, Jones A, et al. Hydrotherapy versus conventional land-based exercise for the management of patients with osteoarthritis of the knee: a randomized clinical trial. *Physical Therapy* 2008;**88**(1):12-21.

Sled 2010 *{published data only}*

Sled EA, Khoja L, Deluzio KJ, Olney SJ, Culham EG. Effect of a home program of hip abductor exercises on knee joint loading, strength, function, and pain in people with knee osteoarthritis: a clinical trial. *Physical Therapy* 2010;**90**(6):895-904.

Song 2010 *{published data only}*

Song R, Roberts BL, Lee E, Lam P, Bae S. A randomized study of the effects of t'ai chi on muscle strength, bone mineral density, and fear of falling in women with osteoarthritis. *Journal of Alternative & Complementary Medicine* 2010;**16**(3):227-33.

Soni 2012 *{published data only}*

Soni A, Joshi A, Mudge N, Wyatt M, Williamson L. Supervised exercise plus acupuncture for moderate to severe knee osteoarthritis: a small randomised controlled trial. *Acupuncture in Medicine* 2012;**30**(3):176-81.

Stitik 2007 *{published data only}*

Stitik TP, Blacksin MF, Stiskal DM, Kim JH, Foye PM, Schoenheer L, et al. Efficacy and safety of hyaluronan treatment in combination therapy with home exercise for knee osteoarthritis pain. *Archives of Physical and Medical Rehabiltation* 2007;**88**:135-41.

Stitik 2007a *{published data only}*

Stitik TP, Blacksin MF, Stiskal DM, Kim JH, Foye PM, Schoenherr L, et al. Efficacy and safety of hyaluronan treatment in combination therapy with home exercise for knee osteoarthritis pain. *Archives of Physical Medicine & Rehabilitation* 2007;**88**(2):135-41.

Sullivan 1998 *{published data only}*

Sullivan T, Allegrante JP, Peterson MG, Kovar PA, MacKenzie CR. One-year followup of patients with osteoarthritis of the knee who participated in a program of supervised fitness walking and supportive patient education. *Arthritis Care & Research* 1998;**11**:228-33.

Swank 2011 *{published data only}*

Swank AM, Kachelman JB, Bibeau W, Quesada PM, Nyland J, Malkani A, et al. Prehabilitation before total knee arthroplasty increases strength and function in older adults with severe osteoarthritis. *Journal of Strength & Conditioning Research* 2011;**25**(2):318-25.

Sylvester 1989 *{published data only}*

Sylvester KL. Investigation of the effect of hydrotherapy in the treatment of osteoarthritic hips. *Clinical Rehabilitation* 1989;**4**:223-8.

Teixeira 2011 *{published data only}*

Teixeira PE, Piva SR, Fitzgerald GK. Effects of impairment-based exercise on performance of specific self-reported functional tasks in individuals with knee osteoarthritis. *Physical Therapy* 2011;**91**(12):1752-65.

Thiengwittayaporn 2009 *{published data only}*

Thiengwittayaporn S, Wetpiryakul P, Foosakun Y, Ngamsom T, Vathanavit P, Pintongtun J. Comparison of the accuracy of quadriceps isometric exercise between using quadriceps education device (QED) and not using QED for osteoarthritic knee patients: a randomized controlled trial. *Journal of the Medical Association of Thailand* 2009;**92 Suppl 6**:S33-8.

Toda 2001 *{published data only}*

Toda Y. The effect of energy restriction, walking and exercise on lower extremity lean body mass in obese women with osteoarthritis of the knee. *Journal of Orthopaedic Science* 2001;**6**:148-54.

Tok 2011 *{published data only}*

Tok F, Aydemir K, Peker F, Safaz I, Taskaynatan MA, Ozgul A. The effects of electrical stimulation combined with continuous passive motion versus isometric exercise on symptoms, functional capacity, quality of life and balance in knee osteoarthritis: randomized clinical trial. *Rheumatology International* 2011;**31**(2):177-81.

Topp 2009 *{published data only}*

Topp R, Swank AM, Quesada PM, Nyland J, Malkani A. The effect of prehabilitation exercise on strength and functioning after total knee arthroplasty. *PM&R* 2009;**1**(8):729-35.

Tsauo 2008 *{published data only}*

Tsauo J, Cheng P, Yang R. The effects of sensorimotor training on knee proprioception and function for patients with knee osteoarthritis: a preliminary report. *Clinical Rehabilitation* 2008;**22**(5):448-57.

Tunay 2010 *{published data only}*

Tunay VB, Baltaci G, Atay AO. Hospital-based versus home-based proprioceptive and strengthening exercise programs in knee osteoarthritis. *Acta Orthopaedica et Traumatologica Turcica* 2010;**44**(4):270-7.

Tuzun 2004 *{published data only}*

Tuzun EH, Aytar A, Eker L, Daskapan A. Effectiveness of two different physical therapy programmes in the treatment of knee osteoarthritis. *Pain Clinic* 2004;**16**:379-87.

van Baar 2001 *{published data only}*

van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. *Annals of the Rheumatic Diseases* 2001;**60**:1123-30.

Van Gool 2005 *{published data only}*

van Gool CH, Penninx BW, Kempen GI, Rejeski WJ, Miller GD, van Eijk JT, et al. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. *Arthritis Care & Research* 2005;**53**:24-32.

Veenhof 2007 *{published data only}*

Veenhof C, Dekker J, Köke AJA, Oostendorp RA, Bijlsma JWJ. Which patients with osteoarthritis of hip and/or knee benefit most from behavioral graded activity?. *International Journal of Behavioral Medicine* 2007;**14**(2):86-91.

Walls 2010 *{published data only}*

Walls RJ, McHugh G, O'Gorman DJ, Moyna N, O'Byrne JM. Effects of preoperative neuromuscular electrical stimulation on quadriceps strength and functional recovery in total knee arthroplasty. A pilot study. *BMC Musculoskeletal Disorders* 2010;**11**:119.

Wang 2006 *{published data only}*

Wang T-J, Belza B, Thompson FE, Whitney JD, Bennett K. Effects of aquatic exercise on flexibility, strength and aerobic fitness in adults with osteoarthritis of the hip or knee. *Journal of Advanced Nursing* 2006;**57**:141-52.

Wang 2007 *{published data only}*

Wang T, Belza B, Elaine Thompson F, Whitney JD, Bennett K. Effects of aquatic exercise on flexibility, strength and aerobic fitness in adults with osteoarthritis of the hip or knee. *Journal of Advanced Nursing* 2007;**57**(2):141-52.

Wang 2007a *{published data only}*

Wang X, Miller GD, Messier SP, Nicklas BJ. Knee strength maintained despite loss of lean body mass during weight loss in older obese adults with knee osteoarthritis. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2007;**62**(8):866-71.

Wang 2009 *{published data only}*

Wang C, Schmid CH, Hibberd PL, Kalish R, Roubenoff R, Rones R, et al. Tai Chi is effective in treating knee osteoarthritis: a randomized controlled trial. *Arthritis & Rheumatism* 2009;**61**(11):1545-53.

Weng 2009 *{published data only}*

Weng MC, Lee CL, Chen CH, Hsu JJ, Lee WD, Huang MH, et al. Effects of different stretching techniques on the outcomes of isokinetic exercise in patients with knee osteoarthritis. *The Kaohsiung Journal of Medical Sciences* 2009;**25**(6):306-15.

Whitehurst 2011 *{published data only}*

Whitehurst DG, Bryan S, Hay EM, Thomas E, Young J, Foster NE. Cost-effectiveness of acupuncture care as an adjunct to exercise-based physical therapy for osteoarthritis of the knee. *Physical Therapy* 2011;**91**(5):630-41.

Williamson 2007 *{published data only}*

Williamson L, Wyatt MR, Yein K, Melton JT. Severe knee osteoarthritis: a randomized controlled trial of acupuncture, physiotherapy (supervised exercise) and standard management for patients awaiting knee replacement. *Rheumatology* 2007;**46**:1445-9.

Williamson 2007a *{published data only}*

Williamson L, Wyatt MR, Yein K, Melton JT. Severe knee osteoarthritis: a randomized controlled trial of acupuncture, physiotherapy (supervised exercise) and standard management for patients awaiting knee replacement. *Rheumatology (Oxford, England)* 2007;**46**(9):1445-9.

Wyatt 2001 *{published data only}*

Wyatt FB, Milam S, Manske RC, Deere R. The effects of aquatic and traditional exercise programs on persons with knee osteoarthritis. *Journal of Strengthening Research* 2001;**15**:337-40.

Yilmaz 2010 *{published data only}*

Yilmaz OO, Senocak O, Sahin E, Baydar M, Gulbahar S, Bircan C, et al. Efficacy of EMG-biofeedback in knee osteoarthritis. *Rheumatology International* 2010;**30**(7):887-92.

Yip 2007a *{published data only}*

Yip YB, Sit JW, Fung KK, Wong DY, Chong SY, Chung LH, et al. Effects of a self-management arthritis programme with an added exercise component for osteoarthritic knee: randomized controlled trial. *Journal of Advanced Nursing* 2007;**59**(1):20-8.

Yip 2008 *{published data only}*

Yip Y, Sit JW, Wong DYS, Chong SYC, Chung L. A 1-year follow-up of an experimental study of a self-management arthritis programme with an added exercise component of clients with osteoarthritis of the knee. *Psychology, Health & Medicine* 2008;**13**(4):402-14.

References to studies awaiting assessment

Carlos 2012 *{published data only}*

Carlos KP, Belli BdS, Alfredo PP. Effect of pulsed ultrasound and continuous ultrasound linked to exercise in patients with knee osteoarthritis: pilot study [Portuguese]. *Fisioterapia e Pesquisa* 2012;**19**(3):275-81.

Eungpinichpong 1997 *{published data only}*

Eungpinichpong W. The efficacy of physical exercise programmes for patients with osteoarthritis of the knee as determined by clinical and gait parameters. *New Zealand Journal of Physiotherapy* 1998;**26**:5.

Ghroubi 2008 *{published data only}*

Ghroubi S, Elleuch H, Kaffel N, Echikh T, Abid M, Elleuch MH. [Contribution of exercise and diet in the management of knee osteoarthritis in the obese]. *Annales de Readaptation et de Medecine Physique* 2008;**51**(8):663-70.

Keogan 2007 *{published data only}*

Keogan F, Gilsenan C, Hussey J, O'Connell P. Open or closed chain quadriceps exercises in treatment of osteoarthritis of the knee; which is more effective? A blinded randomised controlled trial. *Physiotherapy Ireland* 2007;**28**(1):47-8.

Oida 2008 *{published data only}*

Oida Y, Morozumi K, Nakamura N, Kitabatake Y, Shiozawa S, Sato S, et al. [Effectiveness of a community health service program using exercise intervention for elderly people with osteoarthritis of the knees: a randomized controlled trial]. *Nippon Koshu Eisei Zasshi - Japanese Journal of Public Health* 2008;**55**(4):228-37.

Rosa 2012 *{published data only}*

Rosa UH, Velasquez Tlapanco J, Lara Maya C, Villarreal Rios E, Martinez Gonzalez L, Vargas Daza ER, et al. [Comparison of the effectiveness of isokinetic vs isometric therapeutic exercise in patients with osteoarthritis of knee]. *Reumatologia Clinica* 2012;**8**(1):10-4.

Additional references

Altman 1991

Altman RD. Criteria for the classification of clinical osteoarthritis. *Journal of Rheumatology* 1991;**18 Suppl 27**:10-2.

Bellamy 1997

Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. *Journal of Rheumatology* 1997;**24**:799-802.

Bennell 2008

Bennell KL, Hunt MA, Wrigley TV, Lim BW, Hinman RS. Role of muscle in the genesis and management of knee osteoarthritis. *Rheumatic Disease Clinics of North America* 2008;**34**:731-54.

Brazier 1999

Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999;**38**:870-7.

Buchner 1992

Buchner DM, Beresford SA, Larson EB, LaCroix AZ, Wagner EH. Effects of physical activity on health status in older adults II: Intervention studies. *Annual Review of Public Health* 1992;**13**:469-88.

Busija 2010

Busija L, Bridgett L, Williams S, Osborne R, Buchbinder R, March L, et al. Burden of musculoskeletal conditions: osteoarthritis. *Best Practice & Research Clinical Rheumatology* 2010;**24**:757-69.

Cohen 1977

Cohen J. Statistical Power Analysis for the Behavioural Sciences. New York: Academic, 1977.

Cooper 1995

Cooper C. Occupational activity and the risk of osteoarthritis. *Journal of Rheumatology* 1995;**22**:10-2.

Davis 1991

Davis MA, Ettinger WH, Neuhaus JM, Mallon KP. Knee osteoarthritis and physical functioning: evidence from the NHANES 1 epidemiologic followup study. *Journal of Rheumatology* 1991;**18**:591-8.

Deeks 2011

Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2: The Cochrane Collaboration. www.cochrane-handbook.org. The Cochrane Collaboration, March 2011.

Dekker 2013

Dekker J. Exercise and Physical Functioning in Osteoarthritis; Medical, Neuromuscular and Behavioral Perspectives. Berlin: Springer-Verlag, 2013.

Felson 1995

Felson DT. Weight and osteoarthritis. *Journal of Rheumatology* 1995;**22**:7-9.

Fiatarone 1993

Fiatarone MA, Evans WJ. The etiology and reversibility of muscle dysfunction in the aged. *Journal of Gerontology* 1993;**48**:77-83.

Guccione 1994

Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham study. *American Journal of Public Health* 1994;**84**:351-8.

Hochberg 2008

Hochberg MC. Mortality in osteoarthritis. *Clinical and Experimental Rheumatology* 2008;**26**:S120-4.

Hochberg 2012

Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care & Research* 2012;**64**:465-74.

Juni 2006

Juni P, Reichenbach S, Dieppe P. Osteoarthritis: rational approach to treating the individual. *Best Practice & Research in Clinical Rheumatology* 2006;**20**:721-40.

Kujala 1995

Kujala UM, Kettunen J, Paananen H, Aalto T, Battie MC, Impivaara O, et al. Knee osteoarthritis in former runners, soccer players, weight lifters, and shooters. *Arthritis & Rheumatism* 1995;**38**:539-46.

Liang 1990

Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Medical Care* 1990;**28**:632-42.

McAlindon 1999

McAlindon TE, Wilson PW, Aliabadi P, Weissman B, Felson DT. Level of physical activity and the risk of radiographic and symptomatic knee osteoarthritis in the elderly: the Framingham Study. *American Journal of Medicine* 1999;**106**:151-7.

McAlindon 2014

McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis and Cartilage* 2014;**22**:363-88.

Minor 1988

Minor MA, Hewett JE, Webel RR, Dreisinger TE, Kay DR. Exercise tolerance and disease related measures in patients with rheumatoid arthritis and osteoarthritis. *Journal of Rheumatology* 1988;**15**:905-11.

Monticone 2013

Monticone M, Ferrante S, Salvaderi S, Motta L, Cerri C. Responsiveness and minimal important changes for the Knee Injury and Osteoarthritis Outcome score in subjects undergoing rehabilitation after total knee arthroplasty. *American Journal of Physical Medicine & Rehabilitation* 2013;**92**:864-70.

Nelson 2013

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. *Seminars in Arthritis and Rheumatism* 2014;**43**:701-12.

Nielen 2012

Nielen MM, van Sijl AM, Peters MJ, Verheij RA, Schellevis FG, Nurmohamed MT. Cardiovascular disease prevalence in patients with inflammatory arthritis, diabetes mellitus and osteoarthritis: a cross-sectional study in primary care. *BMC Musculoskeletal Disorders* 2013;**13**:150.

Philbin 1995

Philbin EF, Groff GD, Ries MD, Miller TE. Cardiovascular fitness and health in patients with end-stage osteoarthritis. *Arthritis & Rheumatism* 1995;**38**:799-805.

Rangger 1995

Rangger C, Klestil T, Gloetzer W, Kemmler G, Benedetto KP. Osteoarthritis after arthroscopic partial meniscectomy. *American Journal of Sports Medicine* 1995;**23**:240-4.

Reichenbach 2007

Reichenbach S, Sterchi R, Scherer M, Trelle S, Burgi E, Burgi U, et al. Meta-analysis: chondroitin for osteoarthritis of the knee/hip. *Annals of Internal Medicine* 2007;**146**:580-90.

Schünemann 2011a

Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and 'Summary of findings' tables. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org.

Schünemann 2011b

Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. www.cochrane-handbook.org.

Scott 1993

Scott WW, Lethbridge-Cejku M, Reichle R, Wigley FM, Tobin JD, Hochberg MC. Reliability of grading scales for individual radiographic features of osteoarthritis of the knee. The Baltimore Longitudinal Study of Aging Atlas of Knee Osteoarthritis. *Investigative Radiology* 1993;**28**:497-501.

Slemenda 1997

Slemenda C, Brandt KD, Heilman DK, Mazzuca S, Braunstein EM, Katz BP, et al. Quadriceps weakness and osteoarthritis of the knee. *Annals of Internal Medicine* 1997;**127**:97-104.

Sterne 2011

Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2: The Cochrane Collaboration. www.cochrane-handbook.org.

van Dijk 2006

van Dijk GM, Dekker J, Vennhof C, van den Ende CH. Course of functional status and pain in osteoarthritis of the hip or knee: a systematic review of the literature. *Arthritis & Rheumatology* 2006;**55**:779-85.

Zhang 1996

Zhang Y, Glynn RJ, Felson D. Musculoskeletal disease research: should we analyze the joint or the person?. *Journal of Rheumatology* 1996;**23**:1130-4.

Zhang 2010

Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: Part III: changes in evidence following systematic cumulative update of research published through January. *Osteoarthritis and Cartilage* 2010;**18**(4):476-99.

References to other published versions of this review

Fransen 2002

Fransen M, et al. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. *Journal of Rheumatology* 2002;**29**:1737-45.

Fransen 2008

Fransen M, McConnell S. Exercise for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews* 2008, Issue 4. [DOI: 10.1002/14651858.CD004376.pub2]

Fransen 2009

Fransen M, McConnell S. Land-based exercise for osteoarthritis of the knee: a metaanalysis of randomized controlled trials. *Journal of Rheumatology* 2009;**36**:1109-17.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Abbott 2013

Methods	Low risk of bias
Participants	Hip and knee OA recruitment

Abbott 2013 *(Continued)*

	116 community volunteers with knee OA Mean age 66 years, 55% female ACR criteria
Interventions	Clinic, individual: 1. Manual therapy: 9 sessions × 50 minutes (over 16 weeks) plus home programme (3 × per week) 2. Exercise (aerobic plus strengthening plus neuromuscular control): 9 sessions × 50 minutes plus home programme (3 × per week) 3. Exercise plus manual therapy: 9 sessions × 50 minutes plus home programme (3 × per week) 4. Usual care alone
Outcomes	At 1 year: Pain (WOMAC) Physical function (WOMAC) No quality of life measure
Notes	Compared only allocation 2 with allocation 4 Outcomes measured only at 1 year

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Online randomisation service, stratified for hip or knee OA
Allocation concealment (selection bias)	Low risk	Varied block size randomisation, randomisation service kept schedule
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded participants/therapists
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participant self-reported pain and physical function
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded outcome assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Loss to follow-up balanced between allocation groups, intention-to-treat analysis
Selective reporting (reporting bias)	Low risk	Registered trial

An 2008

Methods	Moderate to high risk of bias
Participants	28 community volunteers, ACR clinical criteria Mean age 65 years, all female, mean BMI 25
Interventions	Clinic, classes: 1. Baduanjin (type of Qigong, less physically demanding than Tai Chi), low-level aerobics and strength, 8 weeks, 5 × 30 minutes 2. No intervention
Outcomes	At 8 weeks: 1. Pain (WOMAC) 2. Physical function (WOMAC) No quality of life scale
Notes	Poor comparability at baseline for WOMAC pain and physical function. Post-treatment scores indicate non-normal distribution, i.e. mean (SD) not appropriate

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Nothing other than 'patients were randomised'
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Not disclosed
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Not disclosed
Incomplete outcome data (attrition bias) All outcomes	High risk	HIgh loss to follow-up: 3 (21%) and 4 (29%). No ITT analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Baker 2001

Methods	Unblinded assessor Intention-to-treat analysis Nutrition education control
Participants	46 volunteers, knee OA 74% female Mean age 69 years ACR criteria
Interventions	1. Home muscle strengthening programme (+ 12 visits) 2. Control: 7 × home visits, nutrition education
Outcomes	At 16 weeks: Pain (WOMAC) Function (WOMAC) No QoL
Notes	Very closely monitored intensive strengthening programme with 12 home visits over 16 weeks (ankle weights, squats, etc)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Generated by independent statistician
Allocation concealment (selection bias)	Low risk	Central allocation
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participant
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Bautch 1997

Methods	Moderate to high risk of bias Unblinded assessor Efficacy analysis Education control
Participants	34 participants/volunteers, knee OA Mean age 68 years ACR criteria
Interventions	Individual programme 1. 12 weeks: providing 36 sessions ROM/walking and education classes 2. Control: 12 weekly education classes
Outcomes	At 12 weeks: Pain (VAS × 2) Function (AIMS) No QoL
Notes	Allocation groups very incomparable base pain/BMI/x-ray with active treatment allocation, demonstrating more severe disease Low-intensity walking

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	Efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Bennell 2005

Methods	Low risk of bias
Participants	140 community volunteers, knee OA ACR criteria, pain > 3/10 68% female, mean age 68 years
Interventions	Individual programme: 1. Taping, knee massage, thoracic mobs and hip muscle strengthening; 12 weeks, 8 sessions 2. Control: 8 × sham ultrasound
Outcomes	At 12 weeks and 24 weeks: VAS pain WOMAC function QoL
Notes	Novel intervention with little attention to knee strengthening. Taping, knee massage, thoracic mobs and hip strengthening

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated random numbers
Allocation concealment (selection bias)	Low risk	Central allocation
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants blinded (sham US control)
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	Low risk	Participants blinded
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Bennell 2010

Methods	Low risk of bias
Participants	89 community volunteers, mean age 65 years, mean BMI 28

Bennell 2010 *(Continued)*

	ACR criteria, KL Grade II+, medial tibiofemoral compartment disease 50% female, 33% KL Grade IV
Interventions	Clinic, individual: 1. Muscle strengthening (targeting hip abductors and adductors), 7 sessions of 15-30 minutes over 2 months plus home exercise programme with cuff weights/Theraband (5× per week) 2. Waiting list
Outcomes	At 12 weeks: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) No quality of life measure
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated random numbers, permuted block 4-6
Allocation concealment (selection bias)	Low risk	Independent investigator, sealed opaque envelopes, central location
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participant
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	SBalanced loss to follow-up (13% vs 16%), ITT analysis
Selective reporting (reporting bias)	Low risk	Registered trial

Bezalel 2010

Methods	Moderate risk of bias
Participants	50 community volunteers 65 years of age and over

Bezalel 2010 *(Continued)*

	70% female, mean age 75 years
Interventions	Clinic, classes: 1. Education + exercises, 4 weeks 1 × 45 minutes clinic classes, home-based exercise programme strengthening and stretches 2. Short-wave diathermy 6 sessions 20 minutes
Outcomes	At 4 weeks and 8 weeks: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) No quality of life
Notes	Scores estimated from graphs

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated random numbers
Allocation concealment (selection bias)	Unclear risk	Not stated
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	24% dropout each allocation, intention-to-treat analysis LOCF
Selective reporting (reporting bias)	Unclear risk	Not registered

Brismée 2007

Methods	Moderate risk of bias
Participants	41 community volunteers 50 years of age and older ACR criteria

Brismée 2007 *(Continued)*

	85% female, mean age 70 years, mean BMI 28
Interventions	Clinic, classes: 1. Tai Chi (simplified Yang style) 6 weeks 3 × 40 minutes followed by 6-week home programme (video-tape) 2. Education programme, 6 weeks 3 × 40 minutes
Outcomes	At 6 weeks and 12 weeks: 1. Pain (WOMAC 7-35) 2. Physical function (WOMAC 17-85) No quality of life measurement
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation table stratified by age and sex
Allocation concealment (selection bias)	Unclear risk	Not stated
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Loss to follow-up 9% intervention, 27% control; apparent intention-to-treat (data from all participants who did not drop out in the first week)
Selective reporting (reporting bias)	Unclear risk	Not registered

Bruce-Brand 2012

Methods	Moderate to high risk of bias (efficacy analysis)
Participants	41 community volunteers 55-75 years of age

Bruce-Brand 2012 *(Continued)*

	KL Grade III+
Interventions	Individual, home-based: 1. Resistance training lower limb, 6 weeks 2 × 30 minutes supervised plus 1 × 30 minutes unsupervised 2. Neuromuscular electrical stimulation (not included in meta-analysis) 3. Standard care
Outcomes	At week 8 and week 14: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) 3. SF-36 MCS
Notes	Standard care included OA education, weight loss, pharmacological therapy and physical therapy (no reporting of participation in any of these interventions) Assumed available to the intervention group

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated random numbers, stratified for age and gender
Allocation concealment (selection bias)	Unclear risk	Investigator with no clinical role in the study
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unlinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	High loss to follow-up (29% and 54%), no intention-to-treat analysis
Selective reporting (reporting bias)	Low risk	Registered trial

Chang 2012

Methods	Moderate to high risk of bias

Chang 2012 *(Continued)*

Participants	41 women, KL Grade II or III, knee flexion > 90 degrees Mean age 67 years, mean BMI 25
Interventions	Clinic, individual: 1. General physiotherapy (SWD, hot packs, TENS, IFC, etc) plus muscle strengthening (Theraband), 8 weeks 2 × 60 minutes 2. General physiotherapy alone, 8 weeks 2 × 30 minutes
Outcomes	At 8 weeks: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) No quality of life
Notes	All participants 'prohibited' from using Chinese medicine/alternative therapies and non-habitual exercise during the study

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Randomly assigned but no description of procedure provided
Allocation concealment (selection bias)	Unclear risk	No indication
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants blinded/therapist unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	Low risk	Participants blinded
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Assessed by lead study author, not clear whether lead study author was also a therapist or was blinded
Incomplete outcome data (attrition bias) All outcomes	High risk	High loss to follow-up, unbalanced (20% and 44% controls), no ITT analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Deyle 2000

Methods	Moderate to high risk of bias

Deyle 2000 *(Continued)*

	Blinded assessor Efficacy analysis Subtherapeutic US control
Participants	83 military care patients, knee OA 60% female Mean age 61 years ACR criteria
Interventions	Individual, clinic programme: 1. Manual therapy/strengthening exercises/aerobic exercise, 4 weeks 2 × 60 minutes Control: ultrasound (subtherapeutic)
Outcomes	At 8 weeks (delayed): Pain + Function (WOMAC) No QoL
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers generator
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants/personnel unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Particpants unblinded
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	Imbalance in missing data between allocation groups, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Doi 2008

Methods	Moderate risk of bias

Doi 2008 *(Continued)*

Participants	142 participants with symptomatic knee OA, 50 years of age and older, osteophytes on x-ray 76% female, mean age 70 years, mean BMI 25
Interventions	Home-based (1 visit for instruction, no monitoring): 1. Quadriceps exercises in sitting or supine, 4 sets of 20 reps (knee extension in sitting) daily. Sandbags for weight, but almost all used just body weight. 2. NSAIDs 3× daily until 'no longer required'
Outcomes	At 8 weeks: 1. Pain (VAS 0-100) 2. Physical function (total WOMAC score) No quality of life. 1 undefined score reported for SF-36 (PCS? MCS? One of the 8 domains?)
Notes	Both allocations could use analgesic patches

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation
Allocation concealment (selection bias)	Low risk	Performed by off-site administrative office in Department of Public Health
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Participants unblinded
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Unclear
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	13%-17% loss to follow-up at 8 weeks, no intention-to-treat
Selective reporting (reporting bias)	Low risk	Registered with Japanese Orthopaedic Association

Ettinger 1997a/b

Methods	Low risk of bias Blinded assessor

Ettinger 1997a/b *(Continued)*

	Intention-to-treat Education control
Participants	293 volunteers, knee OA 69% female Mean age 69 years ACR criteria
Interventions	Class-based programme Ettinger a: aerobic walking, 12 weeks 3 × 1 hour Ettinger b: strengthening upper and lower limbs, 12 weeks 3 × 1 hour Control: 3× monthly education classes, then monthly telephone calls
Outcomes	Mean score at 3, 9, 18 months: Pain (FAST × 6) Function (FAST × 23) No quality of life
Notes	Large classes (10-15 participants) After cessation of classes, high level of regular telephone monitoring (monthly in past 9 months)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated
Allocation concealment (selection bias)	Low risk	Web-based, central
Blinding of participants and personnel (performance bias) All outcomes	High risk	Particpants/personnel unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Participants unblinded
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis, missing data balanced between allocation groups
Selective reporting (reporting bias)	Unclear risk	Not registered

Foley 2003

Methods	Low risk of bias
Participants	*Hip and knee OA 70 patients, most from the clinic Mean age 70 years Radiographic criteria
Interventions	Class-based programme (6 weeks); 18 sessions of muscle strengthening, range of motion Control: waiting list, fortnightly telephone call
Outcomes	At 6 weeks: WOMAC pain WOMAC function SF-12 MCS
Notes	Separate analysis per knee OA only or hip OA, gym-based group versus controls About 40% on orthopaedic waiting list

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated
Allocation concealment (selection bias)	Low risk	Central allocation
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants/personnel unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Participants unblinded
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Small numbers lost to follow-up, balanced between allocation groups, intention-to-treat
Selective reporting (reporting bias)	Unclear risk	Not registered

Foroughi 2011

Methods	Moderate to high risk of bias
Participants	54 women 40 years of age and older, OA confirmed by MRI

Foroughi 2011 *(Continued)*

	Mean age 66 years, mean BMI 32
Interventions	Clinic, classes: 1. Progressive resistance training lower limb muscles using pneumatic Keiser machines, progressive to 80% 1RM (15-18 Borg scale), 3 sets of 8 reps, 24 weeks 3 × 60 minutes 2, Sham exercise: as above, but minimal resistance and no progression, only 2 sets of 8 reps, no hip abduction/adduction, 24 weeks 3 × 60 minutes
Outcomes	At 6 months: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) No quality of life assessment
Notes	Exercise and sham exercise group trained together at same location

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation stratified for glucosamine/chondroitin use and WOMAC physical function subscale score
Allocation concealment (selection bias)	Low risk	Conducted by co-investigator not involved in testing
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Most participants blinded/exercise physiologist supervising treatment unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	Low risk	Participants blinded
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Loss to follow-up: exercise (23%), sham exercise (11%), no intention-to-treat analysis
Selective reporting (reporting bias)	Low risk	Registered study

Fransen 2001

Methods	Low risk of bias
Participants	126 participants, knee OA, ACR criteria

Fransen 2001 *(Continued)*

	70% female, mean age 66 years
Interventions	Individual or class-based allocation (8 weeks), 16 sessions with muscle strengthening and aerobic components Control: waiting list
Outcomes	At 8 weeks: WOMAC pain WOMAC function SF-36 MCS
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Low risk	Sealed opaque envelopes, sequentially numbered
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants/personnel unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Participants unblinded
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Fransen 2007

Methods	Low risk of bias
Participants	*Hip and knee OA, ACR criteria 97 community volunteers, 75% female, mean age 70 years
Interventions	Class-based programme: 1. Tai Chi classes, modified style for OA: 12 weeks 2 × 60 minutes

Fransen 2007 *(Continued)*

	2. Waiting list control
Outcomes	At 12 weeks: 1. WOMAC pain 2. WOMAC function 3. SF-12 MCS
Notes	Disaggregated analysis (hip or knee OA) according to identified signal (most painful) joint

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Low risk	Central allocation by administrator
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants/personnel unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Low risk	Trial registered NCT00123994

Gur 2002

Methods	Moderate to high risk of bias
Participants	23 volunteers, knee OA Gender?, mean age 56 years Radiographic, bilateral KL Grade II-III Sedentary past 10 years, cardiovascular clearance
Interventions	Individual programme (8 weeks), 24 sessions of strengthening extensors/flexors (Cybex) Control: no treatment, but 2 additional testing sessions during 8-week period
Outcomes	At 8 weeks: Pain (VAS: 7 items)

Gur 2002 *(Continued)*

	Fx (VAS: 5 items)
Notes	No medications allowed. Young sample High intensity, maximal effort

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants/personnel unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Participants unblinded
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Assessor unblinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	No dropouts
Selective reporting (reporting bias)	Unclear risk	Not registered

Hay 2006

Methods	Moderate risk of bias
Participants	217 participants referred from general practice presenting with persistent knee pain and 55 years of age and older
Interventions	Exercise advice and access to 3-6 sessions with physiotherapist over a 10-week period Control: advice/education leaflets with 1 follow-up telephone call
Outcomes	At 3 months and 6 months: 1. WOMAC pain 2. WOMAC physical function 3. Hospital Anxiety and Depression Scale
Notes	Proportion with knee OA unknown

Hay 2006 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers generator
Allocation concealment (selection bias)	Unclear risk	Small blocks of 6 per practice
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded particpants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Loss to follow-up minimal and balanced, intention-to-treat analysis
Selective reporting (reporting bias)	Low risk	Registered trial

Hopman-Rock 2000

Methods	Moderate to high risk of bias
Participants	*Hip and knee 91 volunteers with OA knee, 80% female, mean age 65 years
Interventions	Class, clinic: 1. Education + exercise, 6 weeks 1 × 60 minutes 2. Waiting list control
Outcomes	At 6 weeks: 1. VAS pain (2) 2. IRGL mobility No quality of life measure
Notes	Only 6 treatment occasions Separate analysis for OA knee provided

Risk of bias

Hopman-Rock 2000 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No details
Allocation concealment (selection bias)	Unclear risk	No details
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	Efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Huang 2003

Methods	Moderate to high risk of bias
Participants	132 participants, bilateral knee OA
Interventions	Individual clinic: 1. Muscle strengthening (KinCom) extensor/flexor + hotpack/ROM, 8 weeks 3 × 60 minutes 2. Control: hotpack/ROM
Outcomes	At 8 weeks, 1 year: 1. VAS pain 2. Lequesne function No quality of life measure
Notes	Combined the 3 muscle strengthening groups for meta-analysis

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Sequential numbers I-IV (representing treatment allocation)

Huang 2003 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Uncertain
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	8% loss to follow-up, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Huang 2005

Methods	Moderate to high risk of bias
Participants	70 participants, bilateral moderate knee OA Lequesne score < 7, mean age 65 years, 80% female
Interventions	Individual, clinic: 1. Muscle strengthening (KinCom) + hotpack/ROM, 8 weeks 3 × 60 minutes 2. Control: hotpack/ROM
Outcomes	At 8 weeks, 1 year: VAS pain Lequesne fx No quality of life measure
Notes	Analysed group 1 (exercise only) vs group 4 (control), allocation groups 2 and 3 received US and IA hyaluronan

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	High risk	Sequential numbers I-IV (representing treatment allocation)
Allocation concealment (selection bias)	Unclear risk	No information

Huang 2005 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded particpants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	9% loss to follow-up, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Hughes 2004

Methods	Moderate to high risk of bias
Participants	*Hip and knee (combined) 150 community volunteers, ACR criteria, mean age 74 years, 83% female
Interventions	Class, clinic: 1. Muscle strengthening plus aerobic walking (1 hour) plus education/discussion (30 minutes), 8 weeks 3 × 1.5 hours 2. Control: arthritis help book and list of available community exercise programmes
Outcomes	8 weeks, 6 months WOMAC pain WOMAC function No quality of life measure
Notes	Large loss to follow-up at 2 months in controls (40%) Only simple exercise equipment used. Proportion of participants with knee vs hip OA unknown

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	Not reported

Hughes 2004 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded particpants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessors
Incomplete outcome data (attrition bias) All outcomes	High risk	Imbalance in missing data, efficacy analysis, 40% loss to follow-up
Selective reporting (reporting bias)	Unclear risk	Not registered

Hurley 2007

Methods	Moderate risk of bias
Participants	418 participants > 50 years of age who consulted a primary care physician for knee pain of > 6 months' duration 70% female, mean age 67 years, mean BMI 30
Interventions	Clinic, individual or classes (results combined): 1. Strengthening, balance, aerobic and motor control exercises, 6 weeks 2 × 45 minutes 2. Usual primary care (most given analgesics, very few participants referred for other interventions)
Outcomes	At 6 weeks and 6 months: 1. Pain (WOMAC 0-20): only 6 months 2. Physical function (WOMAC 0-68) 3. Quality of life (EQ5D 0-1): only 6 months
Notes	Combined results of 2 exercise-based interventions: individual and class-based for all meta-analyses apart from sensitivity analysis according to delivery mode (individual, class, home) for immediate post-treatment physical function WOMAC physical function was declared main outcome, with results provided for 6 weeks and 6 months

Risk of bias

Bias	Authors' judgement	Support for judgement

Hurley 2007 *(Continued)*

Random sequence generation (selection bias)	Low risk	Cluster randomisation: 54 primary care practices were randomly assigned, not participants. Randomisation list was generated by a study co-author at an external location
Allocation concealment (selection bias)	Low risk	Randomisation list was generated by a study co-author not involved in execution of the study
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded particpants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	18% lost to follow-up, balanced between allocation groups, no intention-to-treat but effect of withdrawal was assessed
Selective reporting (reporting bias)	Low risk	Registered study

Jan 2008

Methods	Moderate to high risk of bias
Participants	102 participants, bilateral knee pain > 6 months, ACR criteria, KL Grade < IV 80% female, mean age 62 years, mean weight 62 kg
Interventions	Clinic, individual: 1. High resistance training (knee extensors and flexors), 60% 1RM, 3 × 8 reps, 8 weeks 3 × 30 minutes; 10 minutes cycling warmup, 10 minutes cold pack knee post session 2. Low resistance training (knee extensors and flexors), 10% 1RM, 10 × 15 reps, 8 weeks 3 × 50 minutes; 10 minutes cycling warmup, 10 minutes cold pack knee post session 3. Health education
Outcomes	At 8 weeks: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) No quality of life assessment
Notes	Participants did not take NSAIDs during study

Jan 2008 *(Continued)*

	Results for high resistance training and low resistance training identical, so combined in meta-analysis

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random integer generator used
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	Unbalanced loss to follow-up: 4 (13%) in control group, 0 in exercise group, no intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Jan 2009

Methods	Low risk of bias
Participants	106 participants 50 years of age and older (not stated whether clinic or community-based recruitment) ACR criteria, KL Grade < IV (most KL II) Bilateral knee pain > 6 months 70% female, mean age 62 years, mean weight 63 kg (BMI around 25 calculated)
Interventions	Clinic, individual: 1. Progressive weight bearing quadriceps strengthening (sitting, using EN-Dynamic resistance device), 4 × 6 reps commencing at 50% RM, increasing to 70% RM, 8 weeks 3 × 30 minutes 2. Progressive non-weight bearing quadriceps strengthening (sitting, using EN-Tree resistance device), 4 × 6reps commencing at 50% RM, increasing to 70% RM, 8 weeks 3 × 30 minutes 3. No intervention control
Outcomes	At 8 weeks:

Jan 2009 *(Continued)*

	1. No pain assessment 2. Physical function (WOMAC 0-68) No quality of life assessment
Notes	Mean of physical function score taken for the 2 quadriceps strengthening allocations in the meta-analysis, as no significant difference in physical function at 8 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	States random number tables
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	No loss to follow-up (5 discontinued treatment in the 2 exercise allocations)
Selective reporting (reporting bias)	Unclear risk	NCT 9100002377 not found

Jenkinson 2009

Methods	Low risk of bias
Participants	389 participants from 5 GP practices in Nottingham, 45 years of age and older BMI > 28, knee pain on most days past month 66% female, mean age 61 years, mean BMI 34, 47% KL Grade II+
Interventions	Most at home, unmonitored (exercise/control) 1. Diet and exercise 2. Diet

Jenkinson 2009 *(Continued)*

	3. Exercise: unsupervised home programme, predominantly strengthening with functional exercises introduced after 2 months and aerobic exercises (walking/stepping up) introduced after 6 months. 2 exercises/d, reps 5 (up to 20) daily for 24 months. Visited every 4 months by a dietitician and received a support telephone call between visits, but the calls were NOT used to reinforce the exercise programme 4. Control: education leaflet (but no information about weight loss or exercise)
Outcomes	At 24 months (delayed): 1. Pain (WOMAC 0-20) 2, Physical function (WOMAC 0-68), not estimable as no data for control group No quality of life measure
Notes	Meta-analysis included data from only 2 allocation groups: Exercise and Control

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer random sequence generator, 2 × 2 factorial design, blocks of 10 stratified by sex, age and BMI
Allocation concealment (selection bias)	Low risk	Prepared by trial researcher, kept in locked drawer, opened by co-ordinator in sequential order
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded particpants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor (mailed questionnaires)
Incomplete outcome data (attrition bias) All outcomes	Low risk	High loss to follow-up at 24 months (26% exercise, 14% control), but intention-to-treat analysis using multiple imputation methods
Selective reporting (reporting bias)	Unclear risk	Study registered, but outcome measures not provided at time of registration. No justification for why only selected domains of SF-36 (physical function, bodily pain)?

Kao 2012

Methods	Moderate to high risk of bias
Participants	259 community volunteers 50 years of age and older, morning stiffness < 30 minutes or crepitus, osteophytes on x-ray

Kao 2012 *(Continued)*

	75% female (81% intervention group, 71% control group), mean age 68 years
Interventions	Clinic, classes: 1. Classes 10-15 participants, education/discussion plus exercise. Stretching and strengthening 'whole body muscles, especially lower limbs,' 4 weeks 1 × 20 minutes 2. Control, no intervention
Outcomes	At 4 and 8 weeks: 1. No pain (only SF-36 bodily pain) 2. Physical function (T-WOMAC 0-170) 3. Quality of life: SF-36 MCS
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Cluster randomisation of 4 districts: 2 to intervention, 2 to control
Allocation concealment (selection bias)	Unclear risk	District allocation would have been known at time of participant screening/recruitment?
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Not reported
Incomplete outcome data (attrition bias) All outcomes	High risk	Unbalanced loss to follow-up: 15% intervention, 27% controls. No intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Study not registered

Keefe 2004

Methods	Moderate to high risk of bias
Participants	34 volunteers and participants. Married Persistent knee pain, mean age 59 years, 50% female

Keefe 2004 *(Continued)*

Interventions	Class, clinic: 1. 36 aerobic sessions, 24 strengthening sessions, 12 weeks 3 × 1 hour 2. Standard care
Outcomes	At 12 weeks: Pain: AIMS pain subscales QoL: AIMS psychological
Notes	Analysed group 3 (exercise only) vs standard care (no spouse intervention groups)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Only 'randomly allocated'
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded particpants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Only 6% loss to follow-up, balanced between allocations. Efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Kovar 1992

Methods	Moderate to high risk of bias
Participants	103 participants, knee OA, 84% female, mean age 69 years Pain and +x-ray
Interventions	Class-based, clinic: 1. Fitness walking/stretch/education, 8 weeks 3 × 60 minutes

Kovar 1992 *(Continued)*

	2. Control: weekly telephone call regarding ADL function
Outcomes	At 8 weeks: Pain (AIMS) Function (AIMS) No quality of life measure
Notes	Large classes (20-30 participants)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Missing data balanced between allocation groups, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Lee 2009

Methods	Low risk of bias
Participants	Participants and community volunteers, KL Grade II+, assessed at least 6 months before study entry, 50-80 years of age 93% female, mean age 69 years, mean BMI 26, most KL Grade II-III
Interventions	Clinic, classes: 1. Tai Chi Qigong (18 movements). Movements of mixed nature (motor control, ROM). Movements involved gentle body stretches, 8 weeks 2 × 45 minutes 2. No intervention control

Lee 2009 *(Continued)*

Outcomes	At 8 weeks: 1. Pain (WOMAC 0-35) 2. Physical function (WOMAC 0-85) 3. Quality of life (SF-36 MCS)
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated balanced block randomisation (2:1)
Allocation concealment (selection bias)	Low risk	Sealed envelopes with identification number. Opened in order
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblilnded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Loss to follow-up low (n = 3), included in analysis
Selective reporting (reporting bias)	Unclear risk	Study not registered

Lim 2008

Methods	Low risk of bias
Participants	107 community volunteers, tibiofemoral knee OA, ACR criteria, medial knee pain, medial compartment osteophytes and medial joint space narrowing > lateral joint space narrowing. < 5 degrees valgus malalignment on x-ray 55% female, mean age 66 years, mean BMI 29
Interventions	Most in home programme: 1. Quadriceps strengthening in varus knee alignment group, 2 × 10 reps (weeks 1-2), 3 × 10 reps (weeks 3-12), 5 days a week. Exercise loads progressed frequently, monitored by 7 home visits by physiotherapist

Lim 2008 *(Continued)*

	2. Quadriceps strengthening exercise in neutral knee alignment group, as above 3. No intervention control varus knee alignment group 4. No intervention control neutral knee alignment group
Outcomes	At week 13: 1. Pain (WOMAC 0-100) 2. Physical function (WOMAC 0-100) No quality of life measure
Notes	Average results for exercise allocations (1,2) vs average results for control allocations (3,4) used in meta-analysis. Study did demonstrate clearly that effects of quadriceps strengthening greater in neutral alignment group

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table used, stratified by alignment in blocks of 6
Allocation concealment (selection bias)	Low risk	Independent researcher randomly assigned participants
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Loss to follow-up 4%-18%, but intention-to-treat analysis using last observation carried forward
Selective reporting (reporting bias)	Low risk	Registered study

Lin 2009

Methods	Low risk of bias
Participants	108 participants 50 years of age and older, KL Grade < IV, history of knee pain > 6 months 70% female, mean age 62 years, mean weight 62 kg

Lin 2009 *(Continued)*

Interventions	Clinic, individual: 1. Proprioception exercises, stepping in multiple directions at various speeds, ROM exercises, 8 weeks 3 × 50 minutes 2. Quadriceps strengthening, 50% 1RM 4 × 6 reps. 1RM tested every 2 weeks and a 5% increase in 1RM implemented to training weight, 8 weeks 3 × 50 minutes 3. No intervention control
Outcomes	At 8 weeks: 1. Pain (WOMAC 0-20) 2. Physical function (WOMAC 0-68) No quality of life measure
Notes	Only allocations 2 and 3 included in meta-analysis

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Low risk	Sealed envelopes
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not blinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	6%-8% loss to follow-up, intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Study not registered

Lund 2008

Methods	Low risk of bias
Participants	79 community volunteers, ACR criteria

Lund 2008 *(Continued)*

	75% female, mean age 69 years, mean weight 68-77 kg
Interventions	Clinic, classes: 1. Aquatic exercise 2. Land-based exercise, mixed strengthening, endurance, balance, stretching, 8 weeks 2 × 50 minutes 3. No intervention control
Outcomes	At 8 weeks and 20 weeks: 1. Pain (KOOS, 100-0) 2. Physical function (KOOS ADL, 100-0) 3. Quality of life (KOOS QoL, 100-0)
Notes	Needed to reverse score KOOS pain and physical function outcomes (KOOS lower score is worse score) and to calculate SD from provided SE

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Envelope method with blocks of 18
Allocation concealment (selection bias)	Low risk	Envelope method, so screener unaware which will be chosen
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	7%-20% loss to follow-up at 20 weeks; however intention-to-treat analysis using LOCF
Selective reporting (reporting bias)	Unclear risk	Study not registered

Maurer 1999

Methods	Moderate risk of bias
Participants	113 participants, knee OA, ACR criteria

Maurer 1999 *(Continued)*

	42% female, mean age 64 years
Interventions	Individual, clinic: 1. Unilateral quadriceps strengthening only, 8 weeks 3 × 30 minutes 2. 4 education classes
Outcomes	At 8 weeks: Pain (WOMAC) Function (WOMAC) No quality of life measure
Notes	Only unilateral exercise but many(?) with bilateral symptoms

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers generator
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Missing data balanced between groups, not study-related, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Messier 2004

Methods	Low risk of bias
Participants	158 obese community volunteers 70% female, mean age 69 years
Interventions	Class, clinic (4 months + optional additional 2 months clinic or home)

Messier 2004 *(Continued)*

	Four allocations: 1. Exercise 2. Exercise + diet 3. Diet 4. Control Exercise: strengthening and aerobic walking (4 months), then telephone monitored home programme (with weights) Control: healthy lifestyle: 3-monthly education meetings re weight loss and exercise with follow-up telephone monitoring (about 8 calls)
Outcomes	At 6 and 18 months (delayed): 1. WOMAC pain 2. WOMAC function No quality of life measure
Notes	Analysis of exercise only vs healthy lifestyle control group Physical function assessed at 18 months, SD of baseline used. Very obese sample

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated
Allocation concealment (selection bias)	Low risk	Web-based
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Mikesky 2006

Methods	Moderate to high risk of bias
Participants	37 community volunteers in OA/pain strata 60% female Mean age 69 years Pain and +x-ray
Interventions	Clinic (0-12 months), then home programme thereafter (12-30 months): 1. Lower and upper limb strengthening (KinCom) 0-12 months, 45 clinic sessions: 12-30 months, home programme strengthening (Theraband) 2. ROM control
Outcomes	30 months (delayed): 1. WOMAC pain 2. WOMAC function 3. SF-36 MCS
Notes	Analysis only of participants with knee OA/pain. Twice-weekly clinic-based classes in the first 12 weeks

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information apart from 'randomized'
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	21% loss to follow-up at 30 months but intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Minor 1989

Methods	Moderate to high risk of bias

Minor 1989 *(Continued)*

Participants	80 participants/volunteers, knee OA, 80% female, mean age 64 years Pain and +x-ray
Interventions	Class, clinic: 1. Aerobic walking, 12 weeks 3 × 1 hour 2. Control: ROM/relaxation, 12 weeks 3 × 1 hour
Outcomes	At 12 weeks and 1 year: Pain (AIMS) QoL (AIMS depression)
Notes	Large classes (max 12 participants) Aim of treatment was to increase aerobic capacity without exacerbating symptoms

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information provided
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Uncertain blinding assessors
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Efficacy analysis, 7% loss to follow-up
Selective reporting (reporting bias)	Unclear risk	Not registered

O'Reilly 1999

Methods	Moderate to high risk of bias
Participants	180 volunteers, knee OA, 66% female, mean age 62 years Knee pain past week
Interventions	Home programme:

O'Reilly 1999 *(Continued)*

	1. Lower limb strengthening (4 home visits to monitor) + lifestyle advice 2. Lifestyle advice only
Outcomes	At 6 months (delayed): 1, Pain (WOMAC) 2. Function (WOMAC) 3. Quality of life (HADS depression)
Notes	Community sample, most with mild radiographic/symptomatic disease (only 41% > KL Grade I)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	Not reported whether sealed envelopes were opaque with sequential numbers for audit trail
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded partcipants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessors
Incomplete outcome data (attrition bias) All outcomes	Low risk	6% loss to follow-up, intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Peloquin 1999

Methods	Moderate to high risk of bias
Participants	137 volunteers, knee OA, 70% female, mean age 66 years +x-ray (< Grade IV)
Interventions	Class-based, clinic: 1. Aerobic and strengthening/stretching exercise, 12 weeks 3 × 1 hour 2. Control: 12× education classes
Outcomes	At 12 weeks:

Peloquin 1999 *(Continued)*

	Pain (AIMS) Function (AIMS) No quality of life measure
Notes	Excluded people with severe disease: > 10 degrees varum, KL Grade IV, > 15 degrees of flexion deformity

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	Imbalance in missing data between allocation groups, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Quilty 2003

Methods	Low risk of bias
Participants	87 community volunteers with patellofemoral pain Mean age 67 years
Interventions	1. 9 physiotherapy sessions over 10 weeks 2. Standard care
Outcomes	At 5 and 12 months: VAS pain WOMAC function No quality of life measure

Quilty 2003 *(Continued)*

Notes	Zelen randomisation Treatment directed at patellofemoral joint

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated
Allocation concealment (selection bias)	Low risk	Sealed opaque envelopes, sequentially numbered for audit trail
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Unblinded to intervention, but Zelen randomisation resulted in blinded control group
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	Low risk	Blinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Rogind 1998

Methods	Moderate to high risk of bias
Participants	25 participants, knee OA ACR criteria and +x-ray (> KL Grade II) 92% female, mean age 72 years
Interventions	Class-based programme: 1. Complex mix of exercises, 12 weeks 2 × 1 hour 2. Control: no intervention
Outcomes	At 12 weeks and 1 year: Pain (VAS × 3) Function (AFI × 10) No quality of life measure
Notes	Moderate to severe disease. Only median (IQR) provided. Baseline differences in pain scores Very complex exercise programme (including venous, truncal muscles, balance)

Rogind 1998 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Missing data minimal and balanced between allocation groups, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Salacinski 2012

Methods	Moderate to high risk of bias
Participants	37 participants/community volunteers, KL Grades I-III > 90 degrees knee flexion and exclude patellofemoral pain precluding stationary cycling 77%-60% female, mean age 53-61 years, mean BMI 22-27, experimental/control
Interventions	Clinic, classes: 1. Aerobic cycling (modified 'spinning,' 70% maximum heart rate) 12 weeks 2 × 60 minutes 2. Wait list control
Outcomes	At 12 weeks: Pain (WOMAC): reverse score Physical function (WOMAC): reverse score KOOS QoL
Notes	Baseline incomparability between groups for BMI and age

Salacinski 2012 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Preset randomisation scheme with computer assignment
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	17% (control)-32%(intervention) lost to follow-up and not included in analysis
Selective reporting (reporting bias)	Low risk	Registered

Salli 2010

Methods	Moderate to high risk of bias
Participants	75 community volunteers 45-65 years of age, leading sedentary life ACR criteria, KL Grade I or II, mean age 57 years, 83% female, mean BMI 32
Interventions	Clinic, individual: 1. Concentric-eccentric exercise programme (8 weeks 3 × 60 minutes individual, used isokinetic dynamometer) + PRN paracetamol to max 2 grams per day 2. Isometric exercise programme (8 weeks 3 × 60 minutes individual, used isokinetic dynamometer) + PRN paracetamol to max 2 grams per day 3. Control (PRN paracetamol to max 2 grams per day)
Outcomes	At week 8 and week 20: Pain (VAS motion) Physical function (WOMAC) No quality of life

Salli 2010 *(Continued)*

Notes	Early radiographic disease, SF-36 MCS scores appear extremely high (70.1) Participants assigned to concentric-eccentric experienced a short period of difficulty in adaptation, but no later adverse effects were observed

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information provided, major differences between allocation groups in physical function/SF-36 MCS at baseline
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Only 4/75 lost fo follow-up (5%), balanced between allocation groups
Selective reporting (reporting bias)	Unclear risk	Not registered

Schilke 1996

Methods	Moderate to high risk of bias
Participants	20 participants, knee OA, 85% female, mean age 66 years Rheumatology clinic attendees
Interventions	Individual, clinic: 1. Strengthening bilateral knee extensors and flexors, 8 weeks 3 × 1 hour 2. No intervention control
Outcomes	At 8 weeks: Pain (OASI) Function (OASI) No quality of life measure
Notes	All training on Cybex

Schilke 1996 *(Continued)*

	Intensive, maximal effort

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Uncertain
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing data
Selective reporting (reporting bias)	Unclear risk	Not registered

Simao 2012

Methods	Moderate risk of bias
Participants	35 participants, ACR criteria, KL Grade II+ 90% female, mean age 70 years, mean BMI 27-30
Interventions	Clinic classes: 1. Squat exercises on a vibratory platform 2. Cycle (70% maximum heart rate) and squatting exercises (progressive 20 × 6 reps), 12 weeks 3 × 30 minutes 3. Telephone calls to confirm adherence to routine activities, i.e. not starting exercise programme (control)
Outcomes	At 12 weeks (median, IQR provided): 1. Pain (WOMAC 0-500) 2. Physical function (WOMAC 0-1700)

Simao 2012 *(Continued)*

	No QoL
Notes	The 2 allocation groups were incomparable at baseline for BMI (27 vs 30) WOMAC pain and function Large proportion same KL 4 (27%-40%)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No procedure described
Allocation concealment (selection bias)	Low risk	Serial numbered opaque envelopes
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Only 1 participant in each allocation lost to follow-up at 12 weeks
Selective reporting (reporting bias)	Low risk	Registered

Song 2003

Methods	Moderate risk of bias
Participants	72 sedentary female participants, knee OA (confirmed by email), clinical and radiographic criteria Mean age 65 years
Interventions	Class-based programme, clinic: 1. Tai Chi classes, 16 1-hour sessions 2. Control: weekly telephone call
Outcomes	At 12 weeks: Pain and function: Korean WOMAC No quality of life measure
Notes	About 40% dropout

Song 2003 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Low risk	Central allocation
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	High risk	43% missing data, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Talbot 2003

Methods	Moderate to high risk of bias
Participants	34 participants, knee OA, ACR criteria Mean age 70 years, 78% female
Interventions	Home programme: 1. 12 ASMP classes plus home-based pedometer walking programme 2. Control: 12 weekly ASMP classes
Outcomes	At 12 week and 24 weeks: Pain (McGill Pain Questionnaire) No physical function measure No quality of life measure
Notes	Evaluating the addition of a home-based pedometer monitored walking programme to the Arthritis Self-Management Programme (ASMP)

Risk of bias

Talbot 2003 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	High risk	Unblinded assessor
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Minimal missing data, balanced between allocation groups, efficacy analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Thomas 2002

Methods	Low risk of bias
Participants	786 participants, knee pain 65% female, mean age 62 years
Interventions	Home programme: 1. Daily muscle strength training, bilateral, with Theraband plus 4 home visits during first 2 months, then 1 visit per 6 months (8, 14, 20 months?) 2. Control: short (2-minute) monthly telephone call
Outcomes	At 24 months (delayed): Pain (WOMAC) Function (WOMAC) No quality of life measure
Notes	Participants with knee pain, all may not be OA

Risk of bias

Bias	Authors' judgement	Support for judgement

Thomas 2002 *(Continued)*

Random sequence generation (selection bias)	Low risk	Computer-generated
Allocation concealment (selection bias)	Low risk	Central administration, sequential list audit trail
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Thorstensson 2005

Methods	Low risk of bias
Participants	65 participants (identified by radiologists/orthopaedic surgeons) with radiographic knee OA (KL Grade III or higher) and long-standing knee pain Between 35 and 65 years of age
Interventions	Clinic-based classes: 1. Intensive muscle strengthening programme, 6 weeks 2 × 1 hour 2. Control: waiting list for 6 months
Outcomes	At 6 weeks and 6 months: 1. KOOS pain 2. KOOS ADL 3. SF-36 MCS
Notes	Younger sample and more severe radiographic disease than most RCTs evaluating exercise for OA

Risk of bias

Bias	Authors' judgement	Support for judgement

Thorstensson 2005 *(Continued)*

Random sequence generation (selection bias)	Low risk	No sequence generation, sealed envelopes produced before randomisation
Allocation concealment (selection bias)	Low risk	Participants selected sealed envelope
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Uncertain
Incomplete outcome data (attrition bias) All outcomes	Low risk	Only 7%-10% loss to follow-up, no intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Topp 2002

Methods	Moderate to high risk of bias
Participants	102 volunteers, ACR clinical criteria 74% female, mean age 63 years
Interventions	Class-based, clinic: 1. Muscle strengthening (dynamic or isometric) with Theraband, 15 weeks 1 × 1 hour (clinic), home 16 weeks 2 × 1 hour 2. Control: no intervention
Outcomes	At 16 weeks: Pain (WOMAC) Function (WOMAC) No quality of life measure
Notes	Clinic-based classes 1× per week Home programme 2× per week

Risk of bias

Bias	Authors' judgement	Support for judgement

Topp 2002 *(Continued)*

Random sequence generation (selection bias)	Unclear risk	No information
Allocation concealment (selection bias)	Unclear risk	No information
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	Uncertain
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	No information
Selective reporting (reporting bias)	Unclear risk	Not registered

van Baar 1998

Methods	Low risk of bias
Participants	113 participants, knee OA, ACR criteria 79% female, mean age 68 years
Interventions	Individual, clinic: 1. Physiotherapy + GP education, 12 weeks, 17 sessions total 2. GP education
Outcomes	At 12 weeks: Pain (VAS × 1) Function IRGL No quality of life measure
Notes	Recruited participants with hip and knee OA. Separate results provided for knee OA. Most with early disease, as approximately 50% of sample had symptom duration < 1 year

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table

van Baar 1998 *(Continued)*

Allocation concealment (selection bias)	Low risk	Sealed opaque envelopes, sequential numbering for audit trail
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Wang 2011

Methods	Low risk of bias
Participants	84 community volunteers, 55 years of age and older Physician diagnosis of OA Not currently exercising > 60 minutes per week, past 2 months
Interventions	Clinic, classes: 1. Land-based exercise, PACE programme (flexibility and aerobic), 12 weeks 3 × 60 minutes 2. Aquatic exercise programme 3. Control (no intervention)
Outcomes	At 12 weeks: 1. KOOS pain (0-100), reverse scored 2. KOOS ADL (0-100), reverse scored 3. KOOS quality of life (0-100)
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement

Wang 2011 *(Continued)*

Random sequence generation (selection bias)	Low risk	Computer-generated
Allocation concealment (selection bias)	Low risk	External, researcher not recruiting participants
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Low risk	Blinded assessor
Incomplete outcome data (attrition bias) All outcomes	Low risk	Only 7% loss to follow-up, 2/28 in each allocation, no intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	Not registered

Yip 2007

Methods	Moderate to high risk of bias
Participants	182 participants 50 years of age and older, ACR clinical criteria Mean age 65 years, 84% female
Interventions	Clinic, classes: 1. ASMP + stretching/walking/Tai Chi (8 movements), 6 weeks 1 × 120 minutes (15 minutes for exercise) 2. No intervention
Outcomes	At week 7 and week 23: 1. Current pain (0-100) No physical function (HAQ score inappropriate: most upper limb function; scores inaccurate: outside 0-3 range) No quality of life
Notes	Large loss to follow-up due to SARS (Hong Kong): discouraged from attending hospital clinics Health Assessment Questionnaire for rheumatoid arthritis developed (not specific to lower limb disability)
Risk of bias	

Yip 2007 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random number table
Allocation concealment (selection bias)	Unclear risk	No information provided
Blinding of participants and personnel (performance bias) All outcomes	High risk	Unblinded
Blinding of outcome assessment (detection bias) - subjective self-reported outcomes (pain, function, quality of life)	High risk	Unblinded participants
Blinding of outcome assessment (detection bias) - other outcomes	Unclear risk	No indication that outcomes assessors blinded
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	High loss to follow-up post treatment due to SARS: 25% control, 10% intervention; 16 weeks: 44% control, 24% intervention. ITT analysis conducted but method not clarified
Selective reporting (reporting bias)	Unclear risk	Not registered

1RM: One-repetition maximum.
ACR: American College of Rheumatology.
ADL: Activity of daily living.
AFI: Arthritis Function Index.
AIMS: Arthritis Impact Measurement Scales.
ASMP: Arthritis Self-Management Programme.
BMI: Body mass index.
EQ5D: Standardised measure of health outcome.
FAST: Fitness Arthritis and Seniors Trial.
GP: General practitioner.
HADS: Hospital Anxiety Depression Scale.
IFC: International Functional Classification.
IQR: Interquartile range.
IRGL: Influence of Rheumatic Disease on Health and Lifestyle scale.
ITT: Intention-to-treat.
KL: Kellgren and Lawrence.
KOOS: Knee Osteoarthritis Outcome Scale.
LOCF: Last observation carried forward.
MCS: Mental Component Summary.
MRI: Magnetic resonance imaging.
NSAIDs: Non-steroidal anti-inflammatory drugs.
OA: Osteoarthritis.
OASI: Osteoarthritis Screening Index.
PACE: Patient-centered Assessment and Counseling for Exercise.
PCS: Physical Component Summary.
QoL: Quality of life.
RCT: Randomised controlled trial.
ROM: Range of motion.

SARS: Severe acute respiratory syndrome.
SD: Standard deviation.
SF: Short Form.
SWD: Short Wave Diathermy.
TENS: Transcutaneous electrical nerve stimulation.
US: Ultrasound.
VAS: Visual analogue scale.
WOMAC: Western Ontario and McMasters Universities Osteoarthritis Index.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Ageberg 2010	No non-exercise control, not randomised
Aglamis 2008	Large baseline differences between 2 small comparator groups in pain and physical function scores
Aglamis 2009	Secondary analysis (Aglamis 2009)
Akyol 2010	No non-exercise control
Alfredo 2012	No non-exercise control
Anwer 2011	No non-exercise control
Aoki 2009	Prehabilitation (home stretching programme)
Atamaz 2006	Physical therapy did not include an exercise programme (IR, short-wave diathermy, interferential)
Atamaz 2012	No non-exercise control
Boocock 2009	No non-exercise control, not randomised, no self-report measures
Borjesson 1996	Patients scheduled for joint replacement surgery
Brosseau 2012	No pain/physical function/quality of life measures
Bulthuis 2007	All non-arthroplasty patients had RA
Bulthuis 2008	Secondary analysis (Bulthuis 2007)
Callaghan 1995	Unable to ascertain effect size, as only provided with median % improvements without baseline scores and with extremely wide confidence intervals because of small sample size
Cetin 2008	No non-exercise control
Chaipinyo 2009	No non-exercise control
Chamberlain 1982	No appropriate control. Assessed benefit of SWD added to exercise
Cheing 2002	No control group. Control group used extremely effective sham TENS
Cheing 2004	Secondary analysis of Cheing 2002. Only gait and muscle strength evaluated
Ciolac 2011	No non-exercise control, not randomised
Coupe 2007	Secondary analysis (Veenhof 2007)

Study	Reason for exclusion
Crotty 2009	Prehabilitation
Deyle 2005	No non-exercise control
Dias 2003	Unable to extract change (SD) or post-treatment (SD) scores from published manuscript. Unusually, published manuscript provided only median/test statistic/degrees of freedom data
Diracoglu 2005	No non-exercise control
Duman 2012	All study patients taking fixed-dose NSAIDs (meloxicam 15 mg daily)
Durmus 2007	No non-exercise control
Durmus 2012	No non-exercise control
Ebnezar 2012	No non-exercise control
Ebnezar 2012a	No non-exercise control
Evcik 2002	Not a randomised trial. Patients were 'separated' into 3 groups
Evgeniadis 2008	Prehabilitation
Eyigor 2004	No non-exercise control
Farr 2010	No self-reported pain/physical function/quality of life
Feinglass 2012	No non-exercise control
Fitzgerald 2011	No non-exercise control
Forestier 2010	Aquatic exercise
Foroughi 2011a	Secondary analysis (Foroughi 2011)
Foster 2007	No non-exercise control
Gaal 2008	Aquatic exercise
Gaudreault 2011	No randomly assigned allocation
Gill 2009	No non-exercise control
Green 1993	No appropriate control. Assessed benefit of hydrotherapy added to home exercise
Gremion 2009	Inappropriate control group (biomagnetic therapy)
Haslam 2001	Advice and exercise given in control group. Evaluated treatment was acupuncture
Helmark 2010	No self-reported pain/function outcomes
Helmark 2012	No randomly assigned allocation
Hinman 2007	Aquatic exercise

Study	Reason for exclusion
Hiyama 2012	No non-exercise control
Hoeksma 2004	No non-exercise control. Manual therapy vs exercise
Huang 2005b	Earlier version of Huang 2005 (1)
Hughes 2010	Secondary analysis
Hurley 1998	Not even quasi-randomised
Hurley 2007a	Secondary analysis
Hurley 2012	18- and 30-month outcomes for a 6-week intervention (Hurley 2007). Already submitted 6-month outcomes for sustainability evaluation
Jan 1991	Not even quasi-randomised
Jan 2008a	Preliminary analysis for Lin 2009
Jessep 2009	No non-exercise control
Karagulle 2007	Aquatic exercise
Kawasaki 2008	No non-exercise control
Kawasaki 2009	Inappropriate control: weekly intra-articular hyaluronate injections
King 2008	No randomly assigned allocation
Konishi 2009	No randomly assigned allocation
Kreindler 1989	No pain/function/patient global outcome assessment. Only outcome is muscle strength
Kuptniratsaikul 2002	Cluster random sampling
Lankhorst 1982	No control group in analysis of results. No pain/function/patient global outcome assessment
Lim 2002	No non-exercise control
Lim 2010	No non-exercise control
Lin 2004	Water exercise programme
Lin 2007	Preliminary analysis (Lin 2009)
Liu 2008	No non-exercise control
Mangione 1999	No appropriate control group. Both allocations on stationary cycling, high vs low intensity
Marra 2012	Exercise only a small component of the experimental allocation (pharmacist-led education programme)
Mascarin 2012	No non-exercise control
McCarthy 2004	No non-exercise control

Study	Reason for exclusion
McKnight 2010	No appropriate control group (comprehensive and well-monitored self-management programme, including exercise component)
McQuade 2011	No randomly assigned allocation
Messier 1997	Secondary analysis (Ettinger 1997). Gait assessment
Messier 2000a	No appropriate control group. Assessed benefit of dietary therapy added to an exercise programme
Messier 2000b	Secondary analysis (Ettinger 1997a/b). Balance assessment
Messier 2007	No non-exercise control
Messier 2008	No pain, physical function, quality of life outcomes
Miller 2012	Secondary analysis (ADAPT study)
Moss 2007	No exercise group; patients passive for mobilisation
Murphy 2008	No non-exercise control
Neves 2011	No non-exercise control
Ng 2010	No non-exercise control
Nicklas 2004	Secondary analysis (Messier 2004). Outcomes limited to markers of chronic inflammation
Ozdincler 2005	No non-exercise control
Penninx 2001	Secondary analysis (Ettinger 1997a/b)
Penninx 2002	Secondary analysis (Ettinger 1997a/b)
Pereira, 2011	No non-exercise allocation
Petersen 2010	No non-exercise allocation, no pain/physical function measures
Petersen 2011	No non-exercise allocation
Peterson 1993	Secondary analysis (Kovar 1992). Gait assessment
Petrella 2000	All study patients taking fixed-dose NSAIDs (oxaprozin 1200 mg daily)
Pietrosimone 2010	No non-exercise allocation
Pietrosimone 2012	No non-exercise allocation, no pain/function/quality of life outcomes
Pisters 2010	No non-exercise allocation
Pisters 2010a	Secondary analyses
Piva 2011	Secondary analysis
Piyakhachornrot 2011	No non-exercise allocation

Study	Reason for exclusion
Quirk 1985	No appropriate control group. Assessed benefit of interferential therapy or SWD added to exercise
Rattanachaiyanont 2008	No non-exercise allocation
Ravaud 2004	Cluster-randomised trial
Reid 2010	No non-exercise allocation
Reid 2011	No self-report pain/physical function/quality of life outcomes
Rejeski 1998	Secondary analysis (Ettinger 1997a/b)
Sayers 2012	No non-exercise control
Schlenk 2011	Not randomised
Scopaz 2009	Not randomised
Selfe 2008	No non-exercise allocation
Sen 2004	No non-exercise control
Sevick 2009	Secondary analysis ADAPT study
Shakoor 2007	No non-exercise allocation
Shakoor 2010	Not randomised, no specific pain/function/quality of life outcomes
Shen 2008	Not randomised
Silva 2008	No non-exercise allocation
Sled 2010	Not randomised
Song 2010	No pain/function/quality of life outcomes
Soni 2012	No non-exercise allocation
Stitik 2007	Not randomised or quasi-randomised—sequentially assigned. In addition, all patients received hyaluronan (5 or 3 weekly injections)
Stitik 2007a	Not randomised
Sullivan 1998	Secondary analysis (1-year follow-up) (Kovar 1992)
Swank 2011	Secondary analysis (Topp)
Sylvester 1989	No appropriate control. Hydrotherapy compared with exercise plus SWD (N = 14)
Teixeira 2011	Secondary analysis (Fitzgerald 2011), no non-exercise control
Thiengwittayaporn 2009	No non-exercise allocation
Toda 2001	Not randomised

Study	Reason for exclusion
Tok 2011	No non-exercise allocation
Topp 2009	Prehabilitation
Tsauo 2008	No non-exercise allocation
Tunay 2010	No non-exercise allocation
Tuzun 2004	No non-exercise control
van Baar 2001	Secondary analysis (van Baar 1998) (follow-up study)
Van Gool 2005	Secondary analysis ADAPT study
Veenhof 2007	No non-exercise allocation
Walls 2010	Prehabilitation
Wang 2006	No land-based exercise group
Wang 2007	Aquatic exercise only
Wang 2007a	No pain/function/quality of life outcomes
Wang 2009	No non-exercise control
Weng 2009	No non-exercise control
Whitehurst 2011	No non-exercise allocation
Williamson 2007	Patients awaiting knee replacement surgery
Williamson 2007a	Prehabilitation
Wyatt 2001	No non-exercise control
Yilmaz 2010	No non-exercise allocation
Yip 2007a	Secondary analysis
Yip 2008	Secondary analysis

ADAPT: Arthritis Diet and Activity Promotion Trial
IR: Infra-Red
NSAIDs: Non-steroidal anti-inflammatory drugsav
RA: Rheumatoid Arthritis
SD: standard deviation
SWD: Short Wave Diathermy
TENS: Transcutaneous electrical nerve stimulation

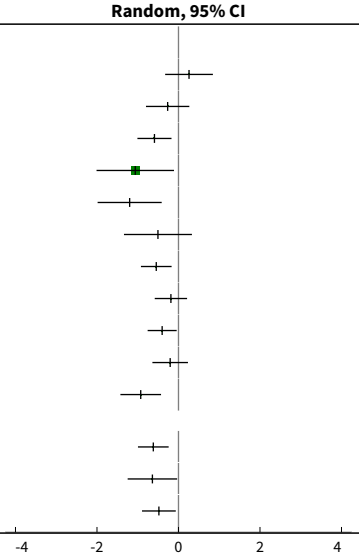
DATA AND ANALYSES

Comparison 1. Post treatment

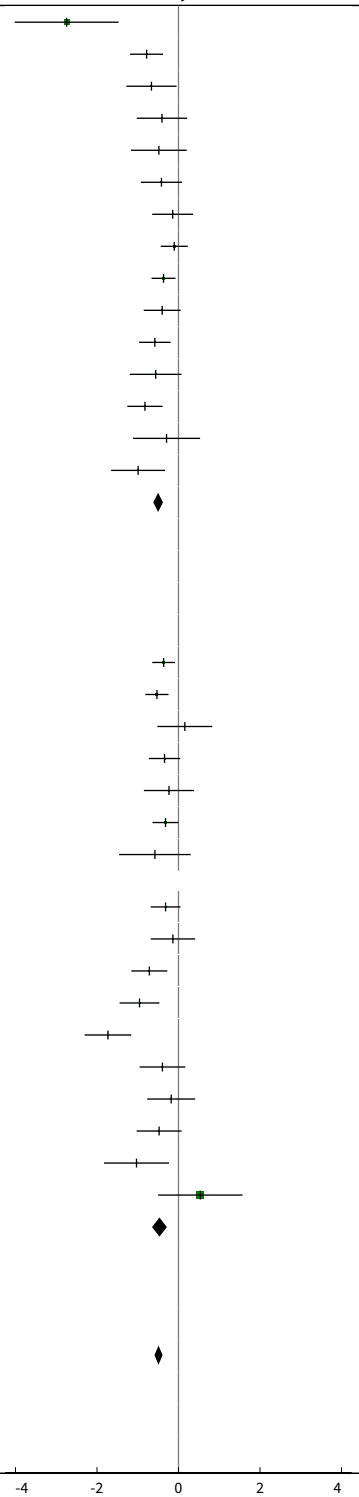
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain	44	3537	Std. Mean Difference (IV, Random, 95% CI)	-0.49 [-0.59, -0.39]
1.1 Change scores	28	2136	Std. Mean Difference (IV, Random, 95% CI)	-0.50 [-0.62, -0.38]
1.2 End of treatment scores	16	1401	Std. Mean Difference (IV, Random, 95% CI)	-0.47 [-0.65, -0.29]
2 Physical function	44	3913	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.64, -0.39]
2.1 Change scores	28	2253	Std. Mean Difference (IV, Random, 95% CI)	-0.47 [-0.63, -0.31]
2.2 End of treatment scores	16	1660	Std. Mean Difference (IV, Random, 95% CI)	-0.59 [-0.78, -0.40]
3 Quality of Life	13	1073	Std. Mean Difference (IV, Random, 95% CI)	0.28 [0.15, 0.40]
3.1 Change scores	8	848	Std. Mean Difference (IV, Random, 95% CI)	0.27 [0.13, 0.42]
3.2 End of treatment scores	5	225	Std. Mean Difference (IV, Random, 95% CI)	0.30 [0.04, 0.57]
4 Study withdrawals	45	4607	Odds Ratio (M-H, Random, 95% CI)	0.93 [0.75, 1.15]

Analysis 1.1. Comparison 1 Post treatment, Outcome 1 Pain.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
1.1.1 Change scores							
Minor 1989	26	-0.6 (1.9)	20	-1.1 (1.9)		1.85%	0.26[-0.33,0.84]
Minor 1989	49	-0.8 (1.7)	19	-0.3 (1.6)		2.08%	-0.27[-0.8,0.27]
Kovar 1992	47	-1.4 (2)	45	-0.1 (2.3)		2.69%	-0.59[-1.01,-0.17]
Schilke 1996	10	-6.1 (4.9)	10	0.4 (6.7)		0.9%	-1.06[-2.01,-0.11]
Bautch 1997	15	-1.4 (2.3)	15	1 (1.6)		1.22%	-1.2[-1.98,-0.41]
Rogind 1998	11	-3 (3.9)	12	-0.1 (6.7)		1.12%	-0.5[-1.34,0.33]
van Baar 1998	54	-27.4 (28.7)	59	-11.7 (28.5)		2.95%	-0.55[-0.92,-0.17]
Maurer 1999	49	-43.5 (80.3)	49	-28.5 (80.3)		2.82%	-0.19[-0.58,0.21]
Peloquin 1999	59	-1.4 (2)	65	-0.6 (2.2)		3.08%	-0.4[-0.76,-0.04]
Hopman-Rock 2000	45	-0.7 (24.1)	37	4 (21.2)		2.58%	-0.2[-0.64,0.23]
Deyle 2000	33	-129.6 (91)	36	-33.8 (111.5)		2.24%	-0.93[-1.43,-0.43]
Fransen 2001	83	-10.6 (19.5)	43	1.5 (19.4)		2.95%	-0.62[-0.99,-0.24]
Baker 2001	22	-79 (88)	22	-20 (93)		1.76%	-0.64[-1.25,-0.03]
Topp 2002	67	-1.5 (3.2)	35	0 (3.2)		2.71%	-0.48[-0.9,-0.07]

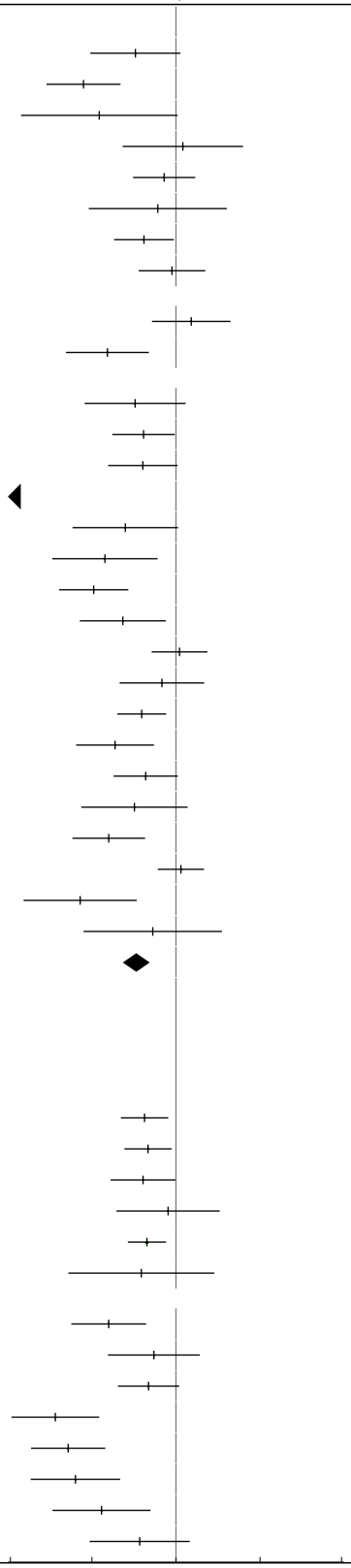


Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Gur 2002	17	-20.9 (8.3)	6	0.7 (4.6)		0.55%	-2.74[-4.02,-1.47]
Huang 2003	99	-1.6 (1.5)	33	-0.4 (1.6)		2.76%	-0.78[-1.19,-0.38]
Song 2003	22	-2.4 (3.9)	21	0.6 (5.1)		1.73%	-0.66[-1.28,-0.05]
Foley 2003	21	-1.2 (2.9)	20	-0 (2.6)		1.72%	-0.41[-1.02,0.21]
Keefe 2004	16	-0.7 (1.7)	18	0 (1.3)		1.5%	-0.48[-1.17,0.2]
Huang 2005	30	-1.2 (1.6)	32	-0.5 (1.7)		2.21%	-0.42[-0.92,0.09]
Thorstensson 2005	30	-1.8 (14)	31	0.3 (15)		2.22%	-0.14[-0.65,0.36]
Bennell 2005	73	-2.2 (1.7)	67	-2 (2.1)		3.25%	-0.1[-0.44,0.23]
Hay 2006	93	-1.6 (3.4)	89	-0.4 (2.8)		3.53%	-0.37[-0.66,-0.07]
Fransen 2007	41	-1.7 (3.3)	36	-0.5 (2.4)		2.48%	-0.4[-0.85,0.05]
Lim 2008	53	-9 (12)	54	-1.7 (12.8)		2.88%	-0.58[-0.97,-0.19]
Lee 2009	29	-2.2 (4.1)	15	-0.2 (1.8)		1.66%	-0.56[-1.2,0.07]
Bennell 2010	45	-2.6 (2.6)	44	-0.4 (2.7)		2.59%	-0.82[-1.26,-0.39]
Simao 2012	11	-62.5 (296)	12	0 (35)		1.14%	-0.29[-1.12,0.53]
Chang 2012	24	-2.3 (1.3)	17	-0.9 (1.5)		1.57%	-0.99[-1.65,-0.33]
Subtotal ***	**1174**		**962**			**62.72%**	**-0.5[-0.62,-0.38]**
Heterogeneity: Tau^2=0.04; Chi^2=46.73, df=28(P=0.01); I^2=40.08%							
Test for overall effect: Z=8.22(P<0.0001)							
1.1.2 End of treatment scores							
Ettinger 1997a/b	146	2.2 (0.7)	75	2.5 (0.6)		3.63%	-0.36[-0.64,-0.08]
Ettinger 1997a/b	144	2.1 (0.6)	75	2.5 (0.6)		3.6%	-0.53[-0.81,-0.24]
Talbot 2003	17	1.4 (0.9)	17	1.2 (1)		1.53%	0.16[-0.52,0.83]
Hughes 2004	68	4.9 (3.4)	43	6.2 (4.3)		2.89%	-0.34[-0.73,0.04]
Brismée 2007	22	15.4 (5.7)	19	16.6 (4.7)		1.73%	-0.23[-0.85,0.38]
Yip 2007	79	37.3 (21.1)	74	44.4 (23.2)		3.34%	-0.32[-0.64,0]
An 2008	11	71.1 (110.1)	10	138.2 (112.6)		1.03%	-0.58[-1.46,0.3]
Doi 2008	61	22.6 (20.7)	56	29.6 (23.4)		3.02%	-0.32[-0.68,0.05]
Lund 2008	25	38 (12.5)	27	39.7 (12)		2.02%	-0.14[-0.68,0.41]
Jan 2008	68	4.8 (3.1)	30	7.1 (3.4)		2.55%	-0.71[-1.16,-0.27]
Lin 2009	36	4.2 (3)	36	7.3 (3.4)		2.29%	-0.96[-1.45,-0.47]
Salli 2010	47	3.4 (1.8)	24	6.5 (1.8)		1.9%	-1.73[-2.3,-1.16]
Bezalel 2010	25	7 (7.5)	25	10 (7.5)		1.95%	-0.39[-0.95,0.17]
Foroughi 2011	20	3.8 (2.7)	25	4.4 (3.7)		1.83%	-0.18[-0.77,0.41]
Wang 2011	26	24 (15)	26	32 (18)		1.99%	-0.48[-1.03,0.08]
Salacinski 2012	13	18.6 (13.4)	15	34.3 (15.9)		1.19%	-1.03[-1.83,-0.23]
Bruce-Brand 2012	10	10.8 (4.3)	6	8.3 (4.4)		0.79%	0.54[-0.5,1.57]
Subtotal ***	**818**		**583**			**37.28%**	**-0.47[-0.65,-0.29]**
Heterogeneity: Tau^2=0.08; Chi^2=38.16, df=16(P=0); I^2=58.07%							
Test for overall effect: Z=5.03(P<0.0001)							
Total ***	**1992**		**1545**			**100%**	**-0.49[-0.59,-0.39]**
Heterogeneity: Tau^2=0.05; Chi^2=84.97, df=45(P=0); I^2=47.04%							
Test for overall effect: Z=9.64(P<0.0001)							
Test for subgroup differences: Chi^2=0.08, df=1 (P=0.77), I^2=0%							



Analysis 1.2. Comparison 1 Post treatment, Outcome 2 Physical function.

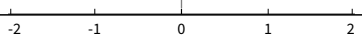
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
1.2.1 Change scores							
Minor 1989	49	-0.9 (2.5)	19	0.3 (2.5)		2.15%	-0.48[-1.02,0.05]
Kovar 1992	47	-2.4 (2.3)	45	0.2 (2.5)		2.49%	-1.1[-1.54,-0.66]
Schilke 1996	10	-3.7 (3.3)	10	-0.4 (3.5)		1.18%	-0.91[-1.84,0.02]
Bautch 1997	15	-2.8 (7.8)	15	-3.5 (8.2)		1.63%	0.08[-0.63,0.8]
van Baar 1998	54	-1.3 (5.7)	59	-0.5 (5.6)		2.75%	-0.14[-0.51,0.23]
Rogind 1998	11	-3 (3.3)	12	-2 (5.3)		1.39%	-0.22[-1.04,0.6]
Peloquin 1999	59	-1.5 (2.4)	65	-0.5 (2.6)		2.8%	-0.38[-0.74,-0.02]
Maurer 1999	49	-106.9 (390.1)	49	-88.3 (390.1)		2.65%	-0.05[-0.44,0.35]
Hopman-Rock 2000	37	-0.8 (4.6)	34	-1.7 (5.2)		2.39%	0.18[-0.28,0.65]
Deyle 2000	33	-402.5 (339.6)	36	-98.2 (393.9)		2.3%	-0.82[-1.31,-0.32]
Baker 2001	22	-272 (295)	22	-119 (323)		1.95%	-0.49[-1.09,0.11]
Fransen 2001	83	-7.7 (19.9)	43	0.1 (20.5)		2.74%	-0.39[-0.76,-0.01]
Topp 2002	67	-4.2 (10.9)	35	0.2 (10.9)		2.59%	-0.39[-0.81,0.02]
Gur 2002	17	-13.8 (4.1)	6	1 (2.5)		0.55%	-3.77[-5.29,-2.26]
Foley 2003	21	-2.8 (7.9)	20	2.1 (8.1)		1.87%	-0.6[-1.23,0.03]
Song 2003	22	-11.1 (12)	21	-1.3 (10.6)		1.87%	-0.84[-1.47,-0.22]
Huang 2003	99	-2 (1.6)	33	-0.4 (1.7)		2.59%	-0.98[-1.39,-0.57]
Huang 2005	30	-1.5 (1.4)	32	-0.5 (1.7)		2.24%	-0.63[-1.14,-0.12]
Bennell 2005	73	-7.8 (8.7)	67	-8.2 (10)		2.89%	0.04[-0.29,0.37]
Thorstensson 2005	30	-2 (12)	31	0.6 (18)		2.27%	-0.17[-0.67,0.34]
Hay 2006	95	-4.8 (10.8)	90	-0.8 (8.5)		3.03%	-0.41[-0.7,-0.12]
Fransen 2007	41	-5 (10.3)	36	2.1 (9.1)		2.41%	-0.72[-1.19,-0.26]
Lim 2008	53	-6.5 (10.6)	54	-2.6 (10.9)		2.7%	-0.36[-0.74,0.02]
Lee 2009	29	-9.4 (14.4)	15	-2.7 (10.8)		1.85%	-0.49[-1.13,0.14]
Bennell 2010	45	-8.1 (7.7)	44	-1.9 (7.6)		2.52%	-0.8[-1.23,-0.37]
Kao 2012	114	3.2 (34)	91	1.5 (20.3)		3.09%	0.06[-0.22,0.33]
Chang 2012	24	-10.7 (5.9)	17	-4.5 (4.4)		1.74%	-1.14[-1.81,-0.47]
Simao 2012	11	-100 (740)	12	75 (463)		1.38%	-0.28[-1.1,0.55]
Subtotal ***	**1240**		**1013**			**61.99%**	**-0.47[-0.63,-0.31]**
Heterogeneity: Tau²=0.12; Chi²=84.69, df=27(P<0.0001); I²=68.12%							
Test for overall effect: Z=5.78(P<0.0001)							
1.2.2 End of treatment scores							
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.37[-0.66,-0.09]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.33[-0.61,-0.05]
Hughes 2004	68	17.3 (12.6)	43	22.3 (12.8)		2.69%	-0.39[-0.78,-0.01]
Brismée 2007	22	39.5 (13)	19	40.7 (11.9)		1.91%	-0.09[-0.71,0.52]
Hurley 2007	229	20 (18.5)	113	25.9 (13.6)		3.25%	-0.35[-0.57,-0.12]
An 2008	11	347.5 (383.8)	10	511.8 (381.6)		1.29%	-0.41[-1.28,0.46]
Jan 2008	68	14.8 (8.9)	30	22.5 (10.9)		2.47%	-0.8[-1.24,-0.36]
Lund 2008	25	35.9 (11.5)	27	38.9 (11)		2.12%	-0.26[-0.81,0.28]
Doi 2008	61	13.7 (13.5)	56	18.6 (16.4)		2.76%	-0.33[-0.69,0.04]
Lin 2009	36	10.1 (8.3)	36	24.9 (11.8)		2.2%	-1.44[-1.96,-0.91]
Jan 2009	71	11.2 (10.1)	35	25 (11.8)		2.48%	-1.28[-1.72,-0.84]
Salli 2010	47	20.7 (8.9)	24	32.6 (11.6)		2.17%	-1.2[-1.73,-0.66]
Bezalel 2010	25	25 (10)	25	34 (10)		2%	-0.89[-1.47,-0.3]
Foroughi 2011	20	13.3 (9.4)	25	18.1 (12)		1.96%	-0.43[-1.03,0.16]



Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Wang 2011	26	18 (14)	26	31 (18)		2.06%	-0.79[-1.36,-0.23]
Bruce-Brand 2012	10	33.9 (12.9)	6	26.1 (15.3)		1.02%	0.53[-0.5,1.57]
Salacinski 2012	13	15.8 (13.9)	15	28.9 (16.2)		1.48%	-0.84[-1.62,-0.06]
Subtotal ***	**1020**		**640**			**38.01%**	**-0.59[-0.78,-0.4]**
Heterogeneity: Tau²=0.1; Chi²=47.46, df=16(P<0.0001); I²=66.29%							
Test for overall effect: Z=6(P<0.0001)							
Total ***	**2260**		**1653**			**100%**	**-0.52[-0.64,-0.39]**
Heterogeneity: Tau²=0.11; Chi²=135.5, df=44(P<0.0001); I²=67.53%							
Test for overall effect: Z=8.23(P<0.0001)							
Test for subgroup differences: Chi²=0.83, df=1 (P=0.36), I²=0%							

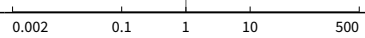
Analysis 1.3. Comparison 1 Post treatment, Outcome 3 Quality of Life.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
1.3.1 Change scores							
Minor 1989	28	-1.7 (1.3)	28	-2.4 (1.7)		5.26%	0.46[-0.07,0.99]
Fransen 2001	83	2 (6.4)	43	-0.7 (3.7)		10.66%	0.48[0.1,0.85]
Keefe 2004	16	0.4 (1.2)	18	0.1 (0.3)		3.21%	0.37[-0.31,1.05]
Bennell 2005	73	0.5 (0.1)	67	0.5 (0.2)		13.48%	-0.07[-0.4,0.27]
Thorstensson 2005	30	4 (13)	31	-0.7 (14)		5.8%	0.34[-0.16,0.85]
Hay 2006	93	0.1 (2)	89	-0.3 (2)		17.46%	0.21[-0.08,0.5]
Lee 2009	29	19.2 (15.9)	15	9.1 (10.3)		3.6%	0.69[0.05,1.34]
Kao 2012	114	2.1 (9.3)	91	-0.3 (7.9)		19.35%	0.28[0,0.55]
Subtotal ***	**466**		**382**			**78.83%**	**0.27[0.13,0.42]**
Heterogeneity: Tau²=0; Chi²=7.61, df=7(P=0.37); I²=8.02%							
Test for overall effect: Z=3.7(P=0)							
1.3.2 End of treatment scores							
Fransen 2007	41	49.6 (8.8)	36	47.6 (8.2)		7.35%	0.23[-0.22,0.68]
Lund 2008	25	43.8 (12.5)	27	43.1 (11.5)		5.01%	0.06[-0.49,0.6]
Wang 2011	26	74 (11)	26	67 (13)		4.81%	0.57[0.02,1.13]
Bruce-Brand 2012	10	66.6 (20.4)	6	65 (27.8)		1.45%	0.07[-0.95,1.08]
Salacinski 2012	13	59.2 (17.5)	15	46.7 (22.6)		2.56%	0.59[-0.17,1.36]
Subtotal ***	**115**		**110**			**21.17%**	**0.3[0.04,0.57]**
Heterogeneity: Tau²=0; Chi²=2.55, df=4(P=0.64); I²=0%							
Test for overall effect: Z=2.23(P=0.03)							
Total ***	**581**		**492**			**100%**	**0.28[0.15,0.4]**
Heterogeneity: Tau²=0; Chi²=10.2, df=12(P=0.6); I²=0%							
Test for overall effect: Z=4.45(P<0.0001)							
Test for subgroup differences: Chi²=0.03, df=1 (P=0.86), I²=0%							



Analysis 1.4. Comparison 1 Post treatment, Outcome 4 Study withdrawals.

Study or subgroup	Exercise n/N	Control n/N	Odds Ratio M-H, Random, 95% CI	Weight	Odds Ratio M-H, Random, 95% CI
An 2008	3/14	4/14		1.37%	0.68[0.12,3.83]
Baker 2001	1/23	1/23		0.55%	1[0.06,17.02]
Bautch 1997	2/17	2/17		0.97%	1[0.12,8.06]
Bennell 2005	13/73	2/67		1.69%	7.04[1.53,32.5]
Bennell 2010	6/45	7/44		2.61%	0.81[0.25,2.65]
Bezalel 2010	6/25	6/25		2.24%	1[0.27,3.66]
Brismée 2007	4/22	6/19		1.85%	0.48[0.11,2.06]
Bruce-Brand 2012	8/28	7/13		2.06%	0.34[0.09,1.34]
Chang 2012	6/30	13/30		2.72%	0.33[0.1,1.03]
Deyle 2000	9/42	5/41		2.57%	1.96[0.6,6.46]
Doi 2008	11/72	14/70		4.09%	0.72[0.3,1.72]
Ettinger 1997a/b	26/146	23/149		6.17%	1.19[0.64,2.19]
Foley 2003	3/35	3/35		1.45%	1[0.19,5.33]
Foroughi 2011	6/20	3/28		1.69%	3.57[0.77,16.54]
Fransen 2001	3/43	2/43		1.22%	1.54[0.24,9.69]
Fransen 2007	4/56	0/41		0.51%	7.11[0.37,135.92]
Gur 2002	0/17	2/8		0.44%	0.07[0,1.76]
Hay 2006	16/109	19/108		5.13%	0.81[0.39,1.67]
Huang 2003	8/99	0/33		0.53%	6.22[0.35,110.82]
Huang 2005	5/35	3/35		1.72%	1.78[0.39,8.09]
Hughes 2004	12/80	10/70		3.85%	1.06[0.43,2.63]
Hurley 2007	49/278	27/140		7.19%	0.9[0.53,1.51]
Jan 2008	3/68	4/34		1.64%	0.35[0.07,1.64]
Jan 2009	5/71	4/35		2.01%	0.59[0.15,2.34]
Jenkinson 2009	21/82	11/76		4.49%	2.03[0.91,4.57]
Kao 2012	20/134	34/125		6.13%	0.47[0.25,0.87]
Keefe 2004	0/16	2/18		0.46%	0.2[0.01,4.49]
Kovar 1992	5/52	6/51		2.36%	0.8[0.23,2.8]
Lee 2009	1/29	2/15		0.7%	0.23[0.02,2.8]
Lim 2008	4/53	6/54		2.16%	0.65[0.17,2.46]
Lin 2009	2/72	3/36		1.23%	0.31[0.05,1.97]
Lund 2008	5/25	3/27		1.66%	2[0.42,9.42]
Maurer 1999	8/57	7/56		2.95%	1.14[0.38,3.4]
O'Reilly 1999	5/113	6/78		2.46%	0.56[0.16,1.89]
Peloquin 1999	10/69	3/68		2.13%	3.67[0.96,13.99]
Rogind 1998	1/12	1/13		0.53%	1.09[0.06,19.63]
Salacinski 2012	6/19	3/18		1.62%	2.31[0.48,11.12]
Salli 2010	3/50	1/25		0.8%	1.53[0.15,15.53]
Simao 2012	1/11	1/12		0.52%	1.1[0.06,20.01]
Song 2003	6/10	10/31		1.81%	3.15[0.72,13.73]
Talbot 2003	2/17	4/17		1.21%	0.43[0.07,2.76]
Thorstensson 2005	2/30	3/31		1.19%	0.67[0.1,4.3]
van Baar 1998	6/99	4/102		2.24%	1.58[0.43,5.78]
Wang 2011	2/26	2/26		1.01%	1[0.13,7.69]
Yip 2007	24/88	41/94		6.09%	0.48[0.26,0.9]
Total (95% CI)	**2512**	**2095**		**100%**	**0.93[0.75,1.15]**
Total events: 343 (Exercise), 320 (Control)					
Heterogeneity: Tau²=0.1; Chi²=55.21, df=44(P=0.12); I²=20.3%					
Test for overall effect: Z=0.65(P=0.51)					



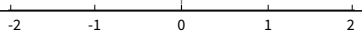
Favours exercise — Favours control

Comparison 2. Treatment sustainability 2-6 months

Outcome or sub-group title	No. of studies	No. of partici-pants	Statistical method	Effect size
1 Pain	12	1468	Std. Mean Difference (IV, Random, 95% CI)	-0.24 [-0.35, -0.14]
1.1 Change	4	563	Std. Mean Difference (IV, Random, 95% CI)	-0.19 [-0.36, -0.02]
1.2 End of follow-up	8	905	Std. Mean Difference (IV, Random, 95% CI)	-0.28 [-0.42, -0.15]
2 Physical function	10	1279	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.26, -0.04]
2.1 Change scores	4	566	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.31, 0.02]
2.2 End of follow-up	6	713	Std. Mean Difference (IV, Random, 95% CI)	-0.15 [-0.32, 0.02]

Analysis 2.1. Comparison 2 Treatment sustainability 2-6 months, Outcome 1 Pain.

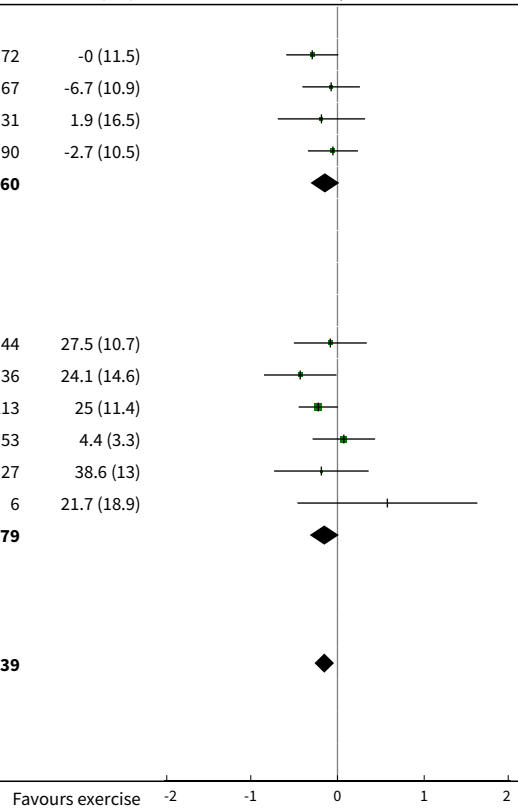
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
2.1.1 Change							
O'Reilly 1999	108	-1.4 (3.5)	72	-0.4 (2.8)		12.21%	-0.32[-0.62,-0.02]
Thorstensson 2005	30	-3.1 (14.3)	31	1.1 (15.6)		4.32%	-0.28[-0.78,0.23]
Bennell 2005	73	-2.1 (1.7)	67	-1.6 (2.5)		9.92%	-0.23[-0.57,0.1]
Hay 2006	93	-1.2 (3.9)	89	-1 (34)		13.01%	-0.01[-0.3,0.28]
Subtotal ***	**304**		**259**			**39.46%**	**-0.19[-0.36,-0.02]**
Heterogeneity: Tau^2=0; Chi^2=2.41, df=3(P=0.49); I^2=0%							
Test for overall effect: Z=2.22(P=0.03)							
2.1.2 End of follow-up							
Quilty 2003	43	42.8 (25.1)	44	50.5 (25.6)		6.15%	-0.3[-0.72,0.12]
Talbot 2003	17	1.1 (0.8)	17	1.6 (1.1)		2.35%	-0.5[-1.19,0.18]
Hughes 2004	60	5.1 (3.7)	36	6.7 (3.9)		6.3%	-0.42[-0.84,-0]
Hurley 2007	229	5.7 (3.1)	113	6.7 (3.3)		21.41%	-0.32[-0.54,-0.09]
Yip 2007	67	38.6 (22)	53	42.5 (23.7)		8.43%	-0.17[-0.53,0.19]
Lund 2008	25	38 (13)	27	37.4 (12.5)		3.71%	0.05[-0.5,0.59]
Jenkinson 2009	82	5.7 (4)	76	7 (4.2)		11.13%	-0.33[-0.64,-0.01]
Bruce-Brand 2012	10	9.6 (4.1)	6	8.3 (4.1)		1.06%	0.29[-0.73,1.31]
Subtotal ***	**533**		**372**			**60.54%**	**-0.28[-0.42,-0.15]**
Heterogeneity: Tau^2=0; Chi^2=3.96, df=7(P=0.78); I^2=0%							
Test for overall effect: Z=4.09(P<0.0001)							
Total ***	**837**		**631**			**100%**	**-0.24[-0.35,-0.14]**
Heterogeneity: Tau^2=0; Chi^2=7.08, df=11(P=0.79); I^2=0%							
Test for overall effect: Z=4.58(P<0.0001)							
Test for subgroup differences: Chi^2=0.71, df=1 (P=0.4), I^2=0%							



Favours exercise — Favours control

Analysis 2.2. Comparison 2 Treatment sustainability 2-6 months, Outcome 2 Physical function.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
2.2.1 Change scores							
O'Reilly 1999	108	-3.5 (12.5)	72	-0 (11.5)		14.06%	-0.29[-0.59,0.01]
Bennell 2005	73	-7.5 (10.9)	67	-6.7 (10.9)		11.48%	-0.07[-0.4,0.26]
Thorstensson 2005	30	-0.9 (13.1)	31	1.9 (16.5)		4.99%	-0.19[-0.69,0.32]
Hay 2006	95	-3.3 (12.2)	90	-2.7 (10.5)		15.19%	-0.05[-0.34,0.24]
Subtotal ***	**306**		**260**			**45.73%**	**-0.15[-0.31,0.02]**
Heterogeneity: Tau^2=0; Chi^2=1.51, df=3(P=0.68); I^2=0%							
Test for overall effect: Z=1.72(P=0.09)							
2.2.2 End of follow-up							
Quilty 2003	43	26.5 (13.2)	44	27.5 (10.7)		7.15%	-0.08[-0.5,0.34]
Hughes 2004	60	18.3 (12.6)	36	24.1 (14.6)		7.24%	-0.43[-0.85,-0.01]
Hurley 2007	229	21.7 (16.6)	113	25 (11.4)		24.75%	-0.22[-0.45,0]
Yip 2007	67	4.7 (3.7)	53	4.4 (3.3)		9.72%	0.07[-0.29,0.43]
Lund 2008	25	36.1 (13.5)	27	38.6 (13)		4.25%	-0.19[-0.73,0.36]
Bruce-Brand 2012	10	31.5 (14.4)	6	21.7 (18.9)		1.17%	0.58[-0.46,1.61]
Subtotal ***	**434**		**279**			**54.27%**	**-0.15[-0.32,0.02]**
Heterogeneity: Tau^2=0; Chi^2=5.57, df=5(P=0.35); I^2=10.16%							
Test for overall effect: Z=1.78(P=0.07)							
Total ***	**740**		**539**			**100%**	**-0.15[-0.26,-0.04]**
Heterogeneity: Tau^2=0; Chi^2=7.09, df=9(P=0.63); I^2=0%							
Test for overall effect: Z=2.66(P=0.01)							
Test for subgroup differences: Chi^2=0, df=1 (P=0.95), I^2=0%							

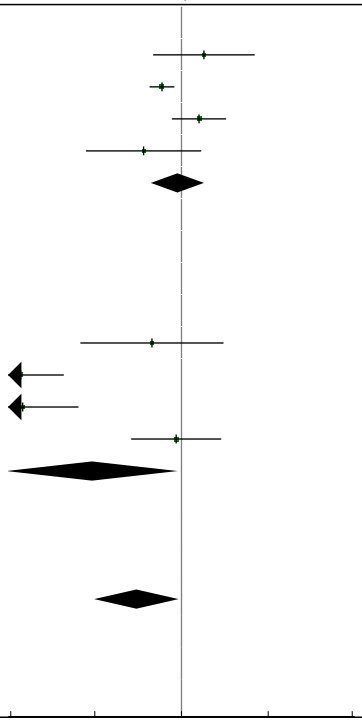


Comparison 3. Treatment sustainability > 6 months

Outcome or sub-group title	No. of studies	No. of partici-pants	Statistical method	Effect size
1 Pain	8	1272	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-1.01, -0.03]
1.1 Change	4	1024	Std. Mean Difference (IV, Random, 95% CI)	-0.05 [-0.35, 0.26]
1.2 End of follow-up	4	248	Std. Mean Difference (IV, Random, 95% CI)	-1.03 [-2.02, -0.04]
2 Physical function	8	1266	Std. Mean Difference (IV, Random, 95% CI)	-0.57 [-1.05, -0.10]
2.1 Change	4	1024	Std. Mean Difference (IV, Random, 95% CI)	-0.20 [-0.32, -0.07]
2.2 End of follow-up	4	242	Std. Mean Difference (IV, Random, 95% CI)	-1.03 [-2.07, 0.02]

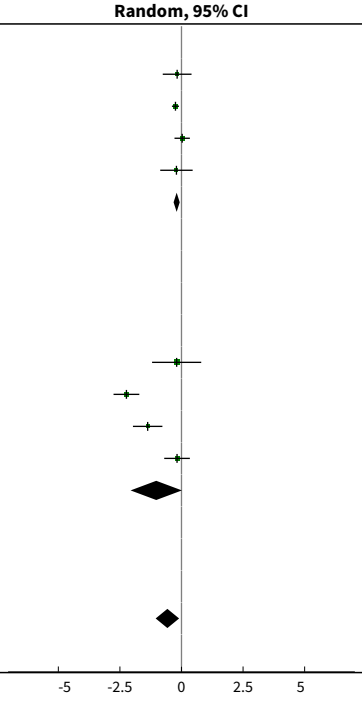
Analysis 3.1. Comparison 3 Treatment sustainability > 6 months, Outcome 1 Pain.

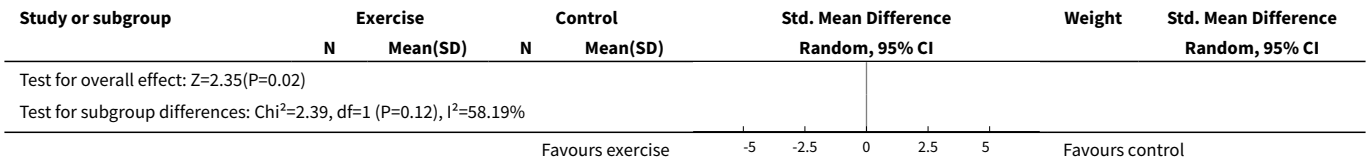
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
3.1.1 Change							
Minor 1989	26	-0.6 (1.9)	20	-1.1 (1.9)		12.19%	0.26[-0.33,0.84]
Thomas 2002	467	-1.3 (3.6)	316	-0.5 (3.6)		14.62%	-0.22[-0.37,-0.08]
Messier 2004	80	-0.4 (4.2)	78	-1.2 (4)		13.95%	0.2[-0.11,0.51]
Mikesky 2006	15	-1.6 (5.5)	22	0.4 (3.4)		11.6%	-0.44[-1.1,0.23]
Subtotal ***	**588**		**436**			**52.36%**	**-0.05[-0.35,0.26]**
Heterogeneity: Tau²=0.06; Chi²=8.49, df=3(P=0.04); I²=64.66%							
Test for overall effect: Z=0.3(P=0.76)							
3.1.2 End of follow-up							
Rogind 1998	11	3 (2.2)	12	4 (3.3)		10.38%	-0.34[-1.17,0.48]
Huang 2003	87	4.8 (1.5)	27	7.6 (1.5)		12.83%	-1.85[-2.35,-1.36]
Huang 2005	26	3.9 (1.4)	28	6.6 (1.5)		11.76%	-1.83[-2.47,-1.19]
Abbott 2013	29	14.9 (11.2)	28	15.6 (11.4)		12.67%	-0.06[-0.58,0.46]
Subtotal ***	**153**		**95**			**47.64%**	**-1.03[-2.02,-0.04]**
Heterogeneity: Tau²=0.92; Chi²=32.45, df=3(P<0.0001); I²=90.76%							
Test for overall effect: Z=2.05(P=0.04)							
Total ***	**741**		**531**			**100%**	**-0.52[-1.01,-0.03]**
Heterogeneity: Tau²=0.42; Chi²=74.84, df=7(P<0.0001); I²=90.65%							
Test for overall effect: Z=2.1(P=0.04)							
Test for subgroup differences: Chi²=3.48, df=1 (P=0.06), I²=71.26%							



Analysis 3.2. Comparison 3 Treatment sustainability > 6 months, Outcome 2 Physical function.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
3.2.1 Change							
Minor 1989	26	-1.1 (2)	20	-0.7 (2.5)		12.31%	-0.18[-0.76,0.41]
Thomas 2002	467	-2.6 (10.5)	316	-0 (10.5)		14.9%	-0.24[-0.39,-0.1]
Messier 2004	80	-3.1 (11.6)	78	-3.4 (11.5)		14.19%	0.03[-0.28,0.34]
Mikesky 2006	15	-0.2 (11.6)	22	1.9 (9.1)		11.73%	-0.2[-0.86,0.45]
Subtotal ***	**588**		**436**			**53.13%**	**-0.2[-0.32,-0.07]**
Heterogeneity: Tau²=0; Chi²=2.43, df=3(P=0.49); I²=0%							
Test for overall effect: Z=3.08(P=0)							
3.2.2 End of follow-up							
Rogind 1998	11	10 (2.6)	6	11.5 (12)		9.09%	-0.2[-1.19,0.8]
Huang 2003	87	3 (1.4)	27	6.1 (1.3)		12.78%	-2.24[-2.76,-1.71]
Huang 2005	26	5.8 (1.8)	28	8.1 (1.5)		12.2%	-1.37[-1.97,-0.78]
Abbott 2013	29	50 (34.2)	28	56.8 (40)		12.8%	-0.18[-0.7,0.34]
Subtotal ***	**153**		**89**			**46.87%**	**-1.03[-2.07,0.02]**
Heterogeneity: Tau²=1.01; Chi²=33.92, df=3(P<0.0001); I²=91.15%							
Test for overall effect: Z=1.93(P=0.05)							
Total ***	**741**		**525**			**100%**	**-0.57[-1.05,-0.1]**
Heterogeneity: Tau²=0.39; Chi²=69.88, df=7(P<0.0001); I²=89.98%							



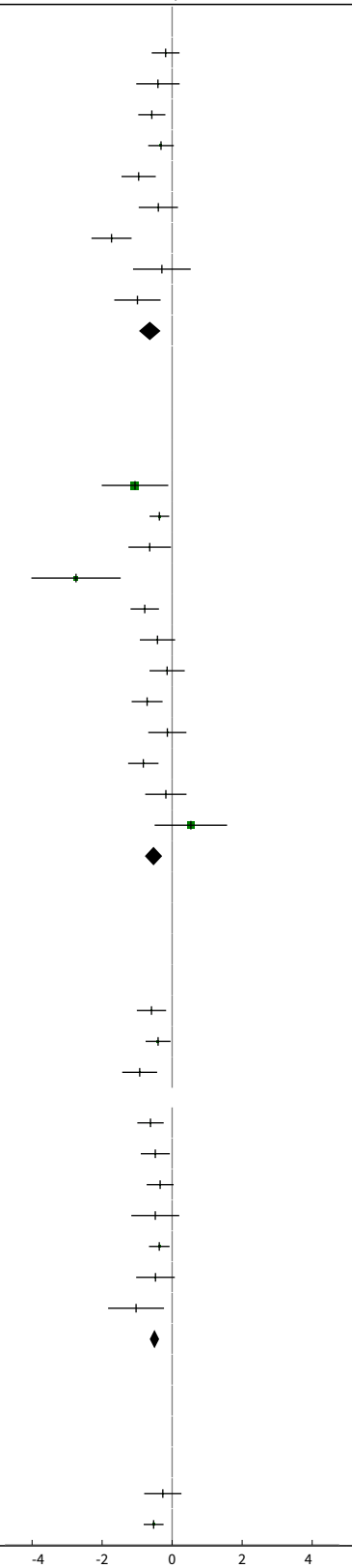


Comparison 4. Treatment content

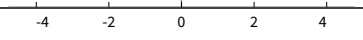
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain	44	3487	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.60, -0.41]
1.1 Quads strengthening only	9	620	Std. Mean Difference (IV, Random, 95% CI)	-0.64 [-0.95, -0.33]
1.2 Lower limb strengthening	12	863	Std. Mean Difference (IV, Random, 95% CI)	-0.53 [-0.78, -0.28]
1.3 Strengthening and aerobics	10	920	Std. Mean Difference (IV, Random, 95% CI)	-0.50 [-0.64, -0.37]
1.4 Walking programmes	4	351	Std. Mean Difference (IV, Random, 95% CI)	-0.48 [-0.83, -0.13]
1.5 Other programmes	10	733	Std. Mean Difference (IV, Random, 95% CI)	-0.35 [-0.49, -0.20]
2 Physical function	44	4255	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.62, -0.39]
2.1 Quadriceps strengthening only	10	726	Std. Mean Difference (IV, Random, 95% CI)	-0.74 [-1.07, -0.41]
2.2 Lower limb strengthening	13	1066	Std. Mean Difference (IV, Random, 95% CI)	-0.54 [-0.83, -0.26]
2.3 Strengthening and aerobics	10	1231	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.67, -0.36]
2.4 Walking programmes	3	317	Std. Mean Difference (IV, Random, 95% CI)	-0.35 [-0.58, -0.11]
2.5 Other programmes	10	915	Std. Mean Difference (IV, Random, 95% CI)	-0.27 [-0.47, -0.07]

Analysis 4.1. Comparison 4 Treatment content, Outcome 1 Pain.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
4.1.1 Quads strengthening only							
Maurer 1999	49	-43.5 (80.3)	49	-28.5 (80.3)		2.9%	-0.19[-0.58,0.21]
Foley 2003	21	-1.2 (2.9)	20	-0 (2.6)		1.7%	-0.41[-1.02,0.21]
Lim 2008	53	-9 (12)	54	-1.7 (12.8)		2.97%	-0.58[-0.97,-0.19]
Doi 2008	61	22.6 (20.7)	56	29.6 (23.4)		3.14%	-0.32[-0.68,0.05]
Lin 2009	36	4.2 (3)	36	7.3 (3.4)		2.31%	-0.96[-1.45,-0.47]
Bezalel 2010	25	7 (7.5)	25	10 (7.5)		1.95%	-0.39[-0.95,0.17]
Salli 2010	47	3.4 (1.8)	24	6.5 (1.8)		1.9%	-1.73[-2.3,-1.16]
Simao 2012	11	-62.5 (296)	12	0 (35)		1.11%	-0.29[-1.12,0.53]
Chang 2012	24	-2.3 (1.3)	17	-0.9 (1.5)		1.55%	-0.99[-1.65,-0.33]
Subtotal ***	**327**		**293**			**19.54%**	**-0.64[-0.95,-0.33]**
Heterogeneity: Tau^2=0.15; Chi^2=26.27, df=8(P=0); I^2=69.54%							
Test for overall effect: Z=4.06(P<0.0001)							
4.1.2 Lower limb strengthening							
Schilke 1996	10	-6.1 (4.9)	10	0.4 (6.7)		0.87%	-1.06[-2.01,-0.11]
Ettinger 1997a/b	146	2.2 (0.7)	75	2.5 (0.6)		3.85%	-0.36[-0.64,-0.08]
Baker 2001	22	-79 (88)	22	-20 (93)		1.75%	-0.64[-1.25,-0.03]
Gur 2002	17	-20.9 (8.3)	6	0.7 (4.6)		0.52%	-2.74[-4.02,-1.47]
Huang 2003	99	-1.6 (1.5)	33	-0.4 (1.6)		2.84%	-0.78[-1.19,-0.38]
Huang 2005	30	-1.2 (1.6)	32	-0.5 (1.7)		2.23%	-0.42[-0.92,0.09]
Thorstensson 2005	30	-1.8 (14)	31	0.3 (15)		2.24%	-0.14[-0.65,0.36]
Jan 2008	68	4.8 (3.1)	30	7.1 (3.4)		2.6%	-0.71[-1.16,-0.27]
Lund 2008	25	38 (12.5)	27	39.7 (12)		2.02%	-0.14[-0.68,0.41]
Bennell 2010	45	-2.6 (2.6)	44	-0.4 (2.7)		2.65%	-0.82[-1.26,-0.39]
Foroughi 2011	20	3.8 (2.7)	25	4.4 (3.7)		1.82%	-0.18[-0.77,0.41]
Bruce-Brand 2012	10	10.8 (4.3)	6	8.3 (4.4)		0.75%	0.54[-0.5,1.57]
Subtotal ***	**522**		**341**			**24.16%**	**-0.53[-0.78,-0.28]**
Heterogeneity: Tau^2=0.1; Chi^2=27.99, df=11(P=0); I^2=60.7%							
Test for overall effect: Z=4.21(P<0.0001)							
4.1.3 Strengthening and aerobics							
Kovar 1992	47	-1.4 (2)	45	-0.1 (2.3)		2.76%	-0.59[-1.01,-0.17]
Peloquin 1999	59	-1.4 (2)	65	-0.6 (2.2)		3.21%	-0.4[-0.76,-0.04]
Deyle 2000	33	-129.6 (91)	36	-33.8 (111.5)		2.26%	-0.93[-1.43,-0.43]
Fransen 2001	83	-10.6 (19.5)	43	1.5 (19.4)		3.06%	-0.62[-0.99,-0.24]
Topp 2002	67	-1.5 (3.2)	35	0 (3.2)		2.78%	-0.48[-0.9,-0.07]
Hughes 2004	68	4.9 (3.4)	43	6.2 (4.3)		2.99%	-0.34[-0.73,0.04]
Keefe 2004	16	-0.7 (1.7)	18	0 (1.3)		1.47%	-0.48[-1.17,0.2]
Hay 2006	93	-1.6 (3.4)	89	-0.4 (2.8)		3.74%	-0.37[-0.66,-0.07]
Wang 2011	26	24 (15)	26	32 (18)		1.99%	-0.48[-1.03,0.08]
Salacinski 2012	13	18.6 (13.4)	15	34.3 (15.9)		1.16%	-1.03[-1.83,-0.23]
Subtotal ***	**505**		**415**			**25.43%**	**-0.5[-0.64,-0.37]**
Heterogeneity: Tau^2=0; Chi^2=6.81, df=9(P=0.66); I^2=0%							
Test for overall effect: Z=7.4(P<0.0001)							
4.1.4 Walking programmes							
Minor 1989	49	-0.8 (1.7)	19	-0.3 (1.6)		2.09%	-0.27[-0.8,0.27]
Ettinger 1997a/b	144	2.1 (0.6)	75	2.5 (0.6)		3.82%	-0.53[-0.81,-0.24]

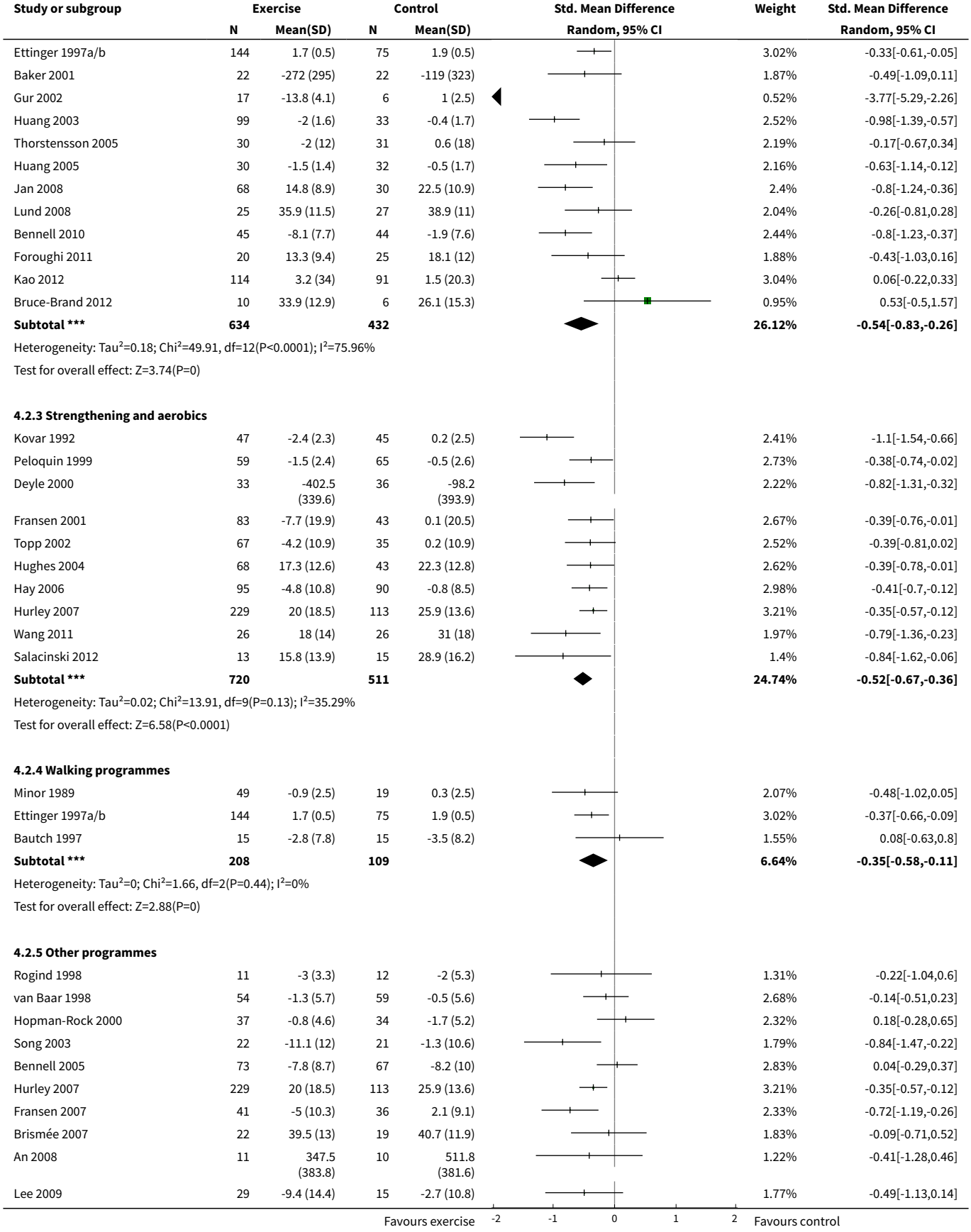


Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Bautch 1997	15	-1.4 (2.3)	15	1 (1.6)		1.19%	-1.2[-1.98,-0.41]
Talbot 2003	17	-0.1 (1)	17	0 (1.7)		1.51%	-0.09[-0.76,0.59]
Subtotal ***	**225**		**126**			**8.62%**	**-0.48[-0.83,-0.13]**
Heterogeneity: Tau^2=0.05; Chi^2=5.26, df=3(P=0.15); I^2=42.99%							
Test for overall effect: Z=2.71(P=0.01)							
4.1.5 Other programmes							
Rogind 1998	11	-3 (3.9)	12	-0.1 (6.7)		1.08%	-0.5[-1.34,0.33]
van Baar 1998	54	-27.4 (28.7)	59	-11.7 (28.5)		3.06%	-0.55[-0.92,-0.17]
Hopman-Rock 2000	45	-0.7 (24.1)	37	4 (21.2)		2.63%	-0.2[-0.64,0.23]
Song 2003	22	-2.4 (3.9)	21	0.6 (5.1)		1.72%	-0.66[-1.28,-0.05]
Bennell 2005	73	-2.2 (1.7)	67	-2 (2.1)		3.41%	-0.1[-0.44,0.23]
Fransen 2007	41	-1.7 (3.3)	36	-0.5 (2.4)		2.53%	-0.4[-0.85,0.05]
Brismée 2007	22	15.4 (5.7)	19	16.6 (4.7)		1.71%	-0.23[-0.85,0.38]
Yip 2007	79	37.3 (21.1)	70	44.4 (23.2)		3.47%	-0.32[-0.64,0.01]
An 2008	11	71.1 (110.1)	10	138.2 (112.6)		1%	-0.58[-1.46,0.3]
Lee 2009	29	-2.2 (4.1)	15	-0.2 (1.8)		1.64%	-0.56[-1.2,0.07]
Subtotal ***	**387**		**346**			**22.26%**	**-0.35[-0.49,-0.2]**
Heterogeneity: Tau^2=0; Chi^2=5.6, df=9(P=0.78); I^2=0%							
Test for overall effect: Z=4.64(P<0.0001)							
Total ***	**1966**		**1521**			**100%**	**-0.51[-0.6,-0.41]**
Heterogeneity: Tau^2=0.04; Chi^2=76.81, df=44(P=0); I^2=42.71%							
Test for overall effect: Z=10.27(P<0.0001)							
Test for subgroup differences: Chi^2=4.3, df=1 (P=0.37), I^2=6.91%							

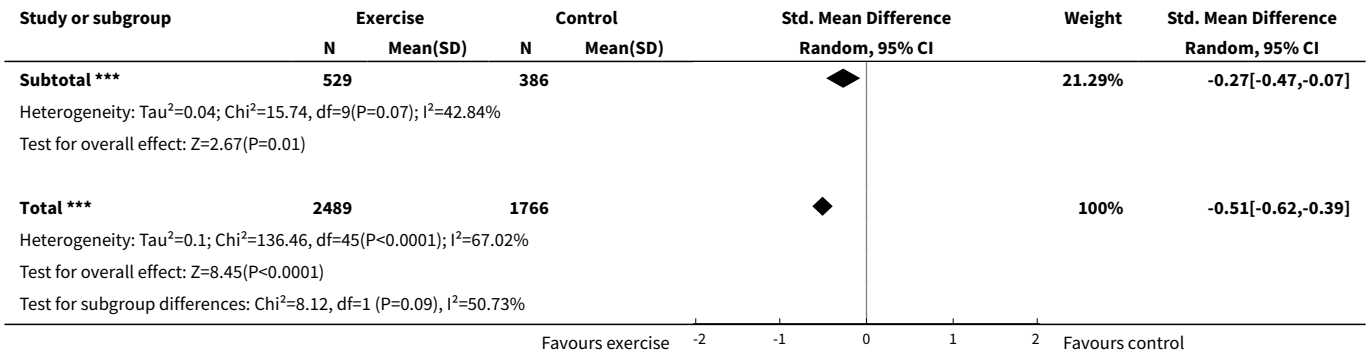


Analysis 4.2. Comparison 4 Treatment content, Outcome 2 Physical function.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
4.2.1 Quadriceps strengthening only							
Maurer 1999	49	-106.9 (390.1)	49	-88.3 (390.1)		2.58%	-0.05[-0.44,0.35]
Foley 2003	21	-2.8 (7.9)	20	2.1 (8.1)		1.79%	-0.6[-1.23,0.03]
Doi 2008	61	13.7 (13.5)	56	18.6 (16.4)		2.7%	-0.33[-0.69,0.04]
Lim 2008	53	-6 (10.6)	54	-2.6 (10.9)		2.63%	-0.32[-0.7,0.06]
Lin 2009	36	10.1 (8.3)	36	24.9 (11.8)		2.12%	-1.44[-1.96,-0.91]
Jan 2009	71	11.2 (10.1)	35	25 (11.8)		2.41%	-1.28[-1.72,-0.84]
Bezalel 2010	25	25 (10)	25	34 (10)		1.92%	-0.89[-1.47,-0.3]
Salli 2010	47	20.7 (8.9)	24	32.6 (11.6)		2.09%	-1.2[-1.73,-0.66]
Chang 2012	24	-10.7 (5.9)	17	-4.5 (4.4)		1.66%	-1.14[-1.81,-0.47]
Simao 2012	11	-100 (740)	12	75 (463)		1.31%	-0.28[-1.1,0.55]
Subtotal ***	**398**		**328**			**21.21%**	**-0.74[-1.07,-0.41]**
Heterogeneity: Tau^2=0.21; Chi^2=38.91, df=9(P<0.0001); I^2=76.87%							
Test for overall effect: Z=4.4(P<0.0001)							
4.2.2 Lower limb strengthening							
Schilke 1996	10	-3.7 (3.3)	10	-0.4 (3.5)		1.11%	-0.91[-1.84,0.02]



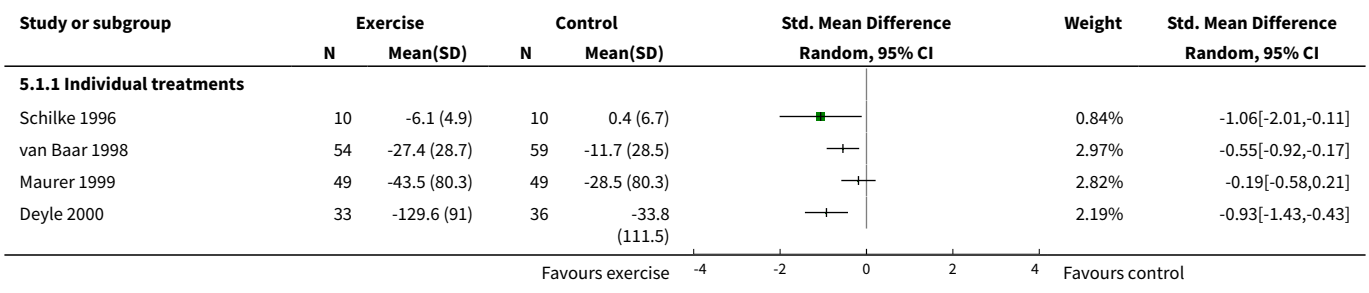
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.02%	-0.33[-0.61,-0.05]
Baker 2001	22	-272 (295)	22	-119 (323)		1.87%	-0.49[-1.09,0.11]
Gur 2002	17	-13.8 (4.1)	6	1 (2.5)		0.52%	-3.77[-5.29,-2.26]
Huang 2003	99	-2 (1.6)	33	-0.4 (1.7)		2.52%	-0.98[-1.39,-0.57]
Thorstensson 2005	30	-2 (12)	31	0.6 (18)		2.19%	-0.17[-0.67,0.34]
Huang 2005	30	-1.5 (1.4)	32	-0.5 (1.7)		2.16%	-0.63[-1.14,-0.12]
Jan 2008	68	14.8 (8.9)	30	22.5 (10.9)		2.4%	-0.8[-1.24,-0.36]
Lund 2008	25	35.9 (11.5)	27	38.9 (11)		2.04%	-0.26[-0.81,0.28]
Bennell 2010	45	-8.1 (7.7)	44	-1.9 (7.6)		2.44%	-0.8[-1.23,-0.37]
Foroughi 2011	20	13.3 (9.4)	25	18.1 (12)		1.88%	-0.43[-1.03,0.16]
Kao 2012	114	3.2 (34)	91	1.5 (20.3)		3.04%	0.06[-0.22,0.33]
Bruce-Brand 2012	10	33.9 (12.9)	6	26.1 (15.3)		0.95%	0.53[-0.5,1.57]
Subtotal ***	**634**		**432**			**26.12%**	**-0.54[-0.83,-0.26]**
Heterogeneity: Tau^2=0.18; Chi^2=49.91, df=12(P<0.0001); I^2=75.96%							
Test for overall effect: Z=3.74(P=0)							
4.2.3 Strengthening and aerobics							
Kovar 1992	47	-2.4 (2.3)	45	0.2 (2.5)		2.41%	-1.1[-1.54,-0.66]
Peloquin 1999	59	-1.5 (2.4)	65	-0.5 (2.6)		2.73%	-0.38[-0.74,-0.02]
Deyle 2000	33	-402.5 (339.6)	36	-98.2 (393.9)		2.22%	-0.82[-1.31,-0.32]
Fransen 2001	83	-7.7 (19.9)	43	0.1 (20.5)		2.67%	-0.39[-0.76,-0.01]
Topp 2002	67	-4.2 (10.9)	35	0.2 (10.9)		2.52%	-0.39[-0.81,0.02]
Hughes 2004	68	17.3 (12.6)	43	22.3 (12.8)		2.62%	-0.39[-0.78,-0.01]
Hay 2006	95	-4.8 (10.8)	90	-0.8 (8.5)		2.98%	-0.41[-0.7,-0.12]
Hurley 2007	229	20 (18.5)	113	25.9 (13.6)		3.21%	-0.35[-0.57,-0.12]
Wang 2011	26	18 (14)	26	31 (18)		1.97%	-0.79[-1.36,-0.23]
Salacinski 2012	13	15.8 (13.9)	15	28.9 (16.2)		1.4%	-0.84[-1.62,-0.06]
Subtotal ***	**720**		**511**			**24.74%**	**-0.52[-0.67,-0.36]**
Heterogeneity: Tau^2=0.02; Chi^2=13.91, df=9(P=0.13); I^2=35.29%							
Test for overall effect: Z=6.58(P<0.0001)							
4.2.4 Walking programmes							
Minor 1989	49	-0.9 (2.5)	19	0.3 (2.5)		2.07%	-0.48[-1.02,0.05]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.02%	-0.37[-0.66,-0.09]
Bautch 1997	15	-2.8 (7.8)	15	-3.5 (8.2)		1.55%	0.08[-0.63,0.8]
Subtotal ***	**208**		**109**			**6.64%**	**-0.35[-0.58,-0.11]**
Heterogeneity: Tau^2=0; Chi^2=1.66, df=2(P=0.44); I^2=0%							
Test for overall effect: Z=2.88(P=0)							
4.2.5 Other programmes							
Rogind 1998	11	-3 (3.3)	12	-2 (5.3)		1.31%	-0.22[-1.04,0.6]
van Baar 1998	54	-1.3 (5.7)	59	-0.5 (5.6)		2.68%	-0.14[-0.51,0.23]
Hopman-Rock 2000	37	-0.8 (4.6)	34	-1.7 (5.2)		2.32%	0.18[-0.28,0.65]
Song 2003	22	-11.1 (12)	21	-1.3 (10.6)		1.79%	-0.84[-1.47,-0.22]
Bennell 2005	73	-7.8 (8.7)	67	-8.2 (10)		2.83%	0.04[-0.29,0.37]
Hurley 2007	229	20 (18.5)	113	25.9 (13.6)		3.21%	-0.35[-0.57,-0.12]
Fransen 2007	41	-5 (10.3)	36	2.1 (9.1)		2.33%	-0.72[-1.19,-0.26]
Brismée 2007	22	39.5 (13)	19	40.7 (11.9)		1.83%	-0.09[-0.71,0.52]
An 2008	11	347.5 (383.8)	10	511.8 (381.6)		1.22%	-0.41[-1.28,0.46]
Lee 2009	29	-9.4 (14.4)	15	-2.7 (10.8)		1.77%	-0.49[-1.13,0.14]

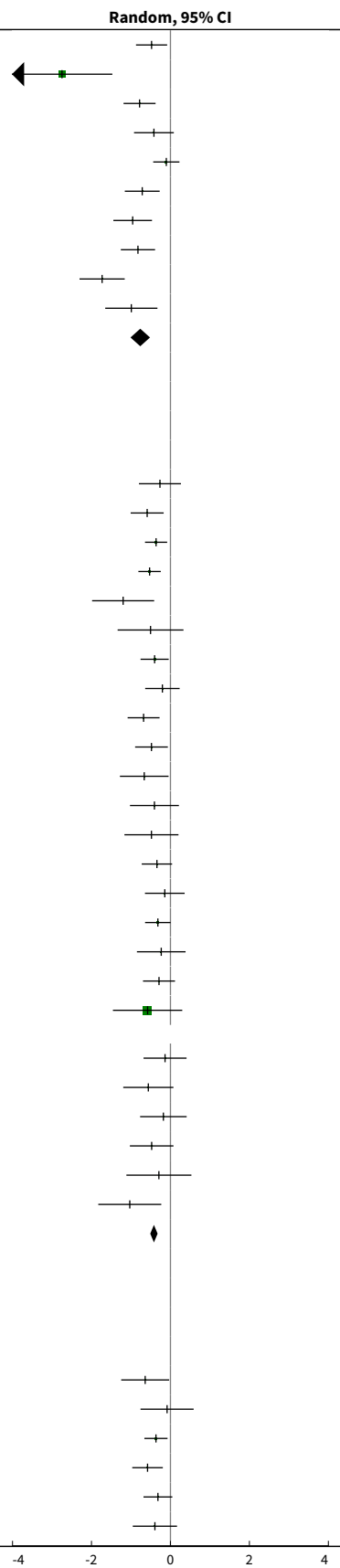


Comparison 5. Treatment delivery mode

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain	44	3588	Std. Mean Difference (IV, Random, 95% CI)	-0.50 [-0.60, -0.41]
1.1 Individual treatments	14	1133	Std. Mean Difference (IV, Random, 95% CI)	-0.76 [-1.01, -0.52]
1.2 Class-based programmes	24	1905	Std. Mean Difference (IV, Random, 95% CI)	-0.42 [-0.51, -0.33]
1.3 Home programmes	7	550	Std. Mean Difference (IV, Random, 95% CI)	-0.38 [-0.55, -0.21]
2 Physical Function	45	4344	Std. Mean Difference (IV, Random, 95% CI)	-0.49 [-0.61, -0.38]
2.1 Individual treatments	16	1493	Std. Mean Difference (IV, Random, 95% CI)	-0.76 [-1.03, -0.50]
2.2 Class-based programmes	24	2152	Std. Mean Difference (IV, Random, 95% CI)	-0.38 [-0.49, -0.26]
2.3 Home programmes	7	699	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.53, -0.21]

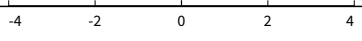
Analysis 5.1. Comparison 5 Treatment delivery mode, Outcome 1 Pain.





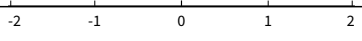
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Fransen 2001	62	-7.7 (18.9)	43	1.5 (19.4)		2.84%	-0.48[-0.87,-0.08]
Gur 2002	17	-20.9 (8.3)	6	0.7 (4.6)		0.5%	-2.75[-4.02,-1.47]
Huang 2003	99	-1.6 (1.5)	33	-0.4 (1.6)		2.76%	-0.78[-1.19,-0.38]
Huang 2005	30	-1.2 (1.6)	32	-0.5 (1.7)		2.16%	-0.42[-0.92,0.09]
Bennell 2005	73	-2.2 (1.7)	67	-2 (2.1)		3.32%	-0.1[-0.44,0.23]
Jan 2008	68	4.8 (3.1)	30	7.1 (3.4)		2.52%	-0.71[-1.16,-0.27]
Lin 2009	36	4.2 (3)	36	7.3 (3.4)		2.24%	-0.96[-1.45,-0.47]
Bennell 2010	45	-2.6 (2.6)	44	-0.4 (2.7)		2.57%	-0.82[-1.26,-0.39]
Salli 2010	47	3.4 (1.8)	24	6.5 (1.8)		1.83%	-1.73[-2.3,-1.16]
Chang 2012	24	-2.3 (1.3)	17	-0.9 (1.5)		1.5%	-0.99[-1.65,-0.33]
Subtotal ***	**647**		**486**			**31.09%**	**-0.76[-1.01,-0.52]**
Heterogeneity: Tau²=0.15; Chi²=47.09, df=13(P<0.0001); I²=72.39%							
Test for overall effect: Z=6.06(P<0.0001)							
5.1.2 Class-based programmes							
Minor 1989	49	-0.8 (1.7)	19	-0.3 (1.6)		2.02%	-0.27[-0.8,0.27]
Kovar 1992	47	-1.4 (2)	45	-0.1 (2.3)		2.68%	-0.59[-1.01,-0.17]
Ettinger 1997a/b	146	2.2 (0.7)	75	2.5 (0.6)		3.76%	-0.36[-0.64,-0.08]
Ettinger 1997a/b	144	2.1 (0.6)	75	2.5 (0.6)		3.73%	-0.53[-0.81,-0.24]
Bautch 1997	15	-1.4 (2.3)	15	1 (1.6)		1.15%	-1.2[-1.98,-0.41]
Rogind 1998	11	-3 (3.9)	12	-0.1 (6.7)		1.05%	-0.5[-1.34,0.33]
Peloquin 1999	59	-1.4 (2)	65	-0.6 (2.2)		3.13%	-0.4[-0.76,-0.04]
Hopman-Rock 2000	45	-0.7 (24.1)	37	4 (21.2)		2.56%	-0.2[-0.64,0.23]
Fransen 2001	59	-11.4 (18.4)	43	1.5 (19.4)		2.77%	-0.68[-1.08,-0.28]
Topp 2002	67	-1.5 (3.2)	35	0 (3.2)		2.7%	-0.48[-0.9,-0.07]
Song 2003	22	-2.4 (3.9)	21	0.6 (5.1)		1.66%	-0.66[-1.28,-0.05]
Foley 2003	21	-1.2 (2.9)	20	-0 (2.6)		1.65%	-0.41[-1.02,0.21]
Keefe 2004	16	-0.7 (1.7)	18	0 (1.3)		1.42%	-0.48[-1.17,0.2]
Hughes 2004	68	4.9 (3.4)	43	6.2 (4.3)		2.91%	-0.34[-0.73,0.04]
Thorstensson 2005	30	-1.8 (14)	31	0.3 (15)		2.17%	-0.14[-0.65,0.36]
Yip 2007	79	37.3 (21.1)	70	44.4 (23.2)		3.39%	-0.32[-0.64,0.01]
Brismée 2007	22	15.4 (5.7)	19	16.6 (4.7)		1.66%	-0.23[-0.85,0.38]
Fransen 2007	56	-9.6 (18.9)	41	-4.4 (16.2)		2.77%	-0.29[-0.69,0.12]
An 2008	11	71.1 (110.1)	10	138.2 (112.6)		0.96%	-0.58[-1.46,0.3]
Lund 2008	25	38 (12.5)	27	39.7 (12)		1.96%	-0.14[-0.68,0.41]
Lee 2009	29	-2.2 (4.1)	15	-0.2 (1.8)		1.59%	-0.56[-1.2,0.07]
Foroughi 2011	20	3.8 (2.7)	25	4.4 (3.7)		1.76%	-0.18[-0.77,0.41]
Wang 2011	26	24 (15)	26	32 (18)		1.93%	-0.48[-1.03,0.08]
Simao 2012	11	-62.5 (296)	12	0 (35)		1.07%	-0.29[-1.12,0.53]
Salacinski 2012	13	18.6 (13.4)	15	34.3 (15.9)		1.12%	-1.03[-1.83,-0.23]
Subtotal ***	**1091**		**814**			**53.55%**	**-0.42[-0.51,-0.33]**
Heterogeneity: Tau²=0; Chi²=15.56, df=24(P=0.9); I²=0%							
Test for overall effect: Z=8.8(P<0.0001)							
5.1.3 Home programmes							
Baker 2001	22	-79 (88)	22	-20 (93)		1.69%	-0.64[-1.25,-0.03]
Talbot 2003	17	-0.1 (1)	17	0 (1.7)		1.46%	-0.09[-0.76,0.59]
Hay 2006	93	-1.6 (3.4)	89	-0.4 (2.8)		3.65%	-0.37[-0.66,-0.07]
Lim 2008	53	-9 (12)	54	-1.7 (12.8)		2.89%	-0.58[-0.97,-0.19]
Doi 2008	61	22.6 (20.7)	56	29.6 (23.4)		3.06%	-0.32[-0.68,0.05]
Bezalel 2010	25	7 (7.5)	25	10 (7.5)		1.89%	-0.39[-0.95,0.17]

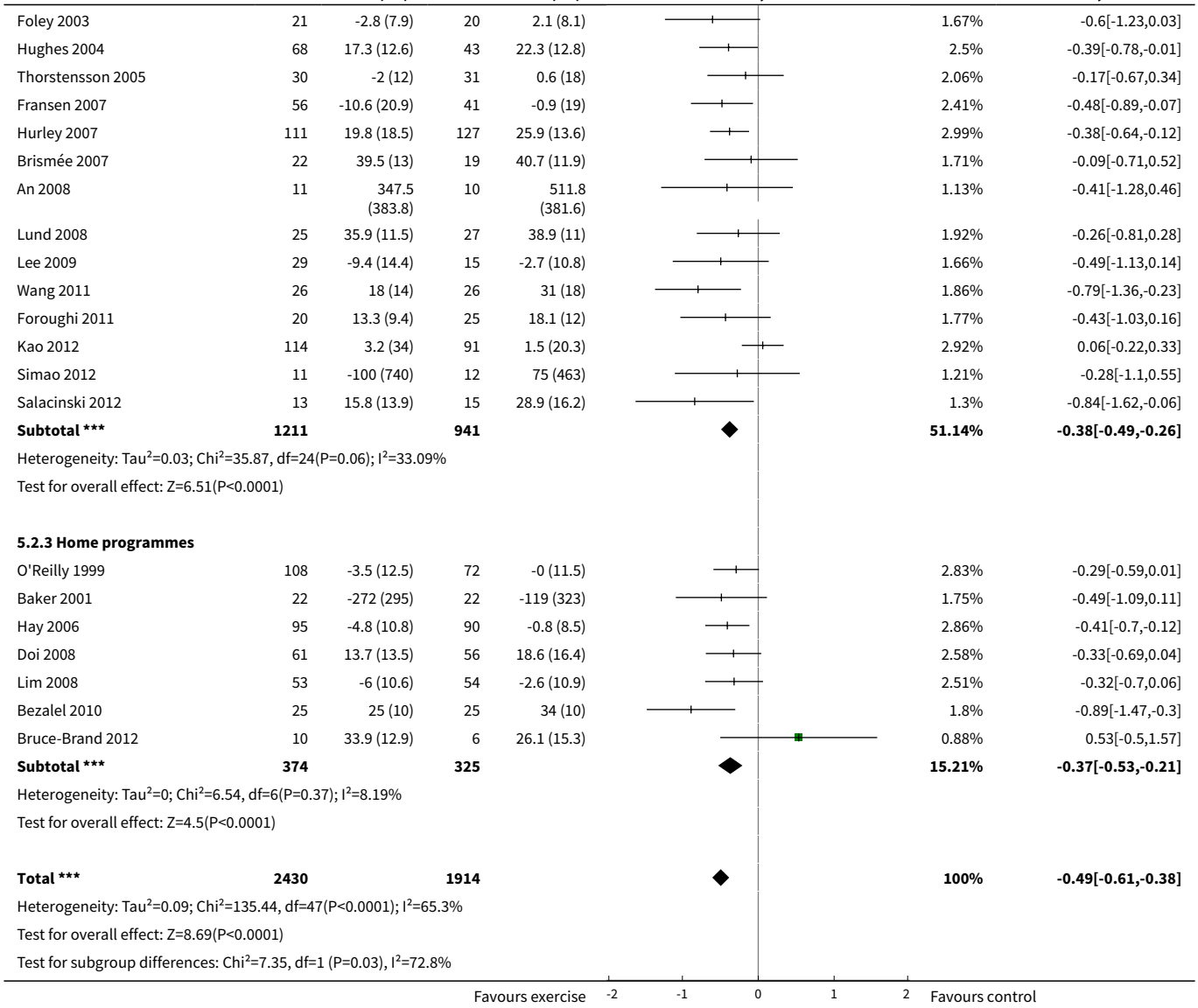
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Bruce-Brand 2012	10	10.8 (4.3)	6	8.3 (4.4)		0.73%	0.54[-0.5,1.57]
Subtotal ***	**281**		**269**			**15.37%**	**-0.38[-0.55,-0.21]**
Heterogeneity: Tau^2=0; Chi^2=5.6, df=6(P=0.47); I^2=0%							
Test for overall effect: Z=4.38(P<0.0001)							
Total ***	**2019**		**1569**			**100%**	**-0.5[-0.6,-0.41]**
Heterogeneity: Tau^2=0.04; Chi^2=77.99, df=45(P=0); I^2=42.3%							
Test for overall effect: Z=10.43(P<0.0001)							
Test for subgroup differences: Chi^2=7.31, df=1 (P=0.03), I^2=72.66%							



Analysis 5.2. Comparison 5 Treatment delivery mode, Outcome 2 Physical Function.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
5.2.1 Individual treatments							
Schilke 1996	10	-3.7 (3.3)	10	-0.4 (3.5)		1.02%	-0.91[-1.84,0.02]
van Baar 1998	54	-1.3 (5.7)	59	-0.5 (5.6)		2.56%	-0.14[-0.51,0.23]
Maurer 1999	49	-106.9 (390.1)	49	-88.3 (390.1)		2.46%	-0.05[-0.44,0.35]
Deyle 2000	33	-402.5 (339.6)	36	-98.2 (393.9)		2.1%	-0.82[-1.31,-0.32]
Fransen 2001	62	-6.6 (15.7)	43	0.1 (20.5)		2.47%	-0.37[-0.77,0.02]
Gur 2002	17	-13.8 (4.1)	6	1 (2.5)		0.47%	-3.76[-5.27,-2.24]
Huang 2003	99	-2 (1.6)	33	-0.4 (1.7)		2.4%	-0.98[-1.39,-0.57]
Huang 2005	30	-1.5 (1.4)	32	-0.5 (1.7)		2.04%	-0.63[-1.14,-0.12]
Bennell 2005	73	-7.8 (8.7)	67	-8.2 (10)		2.71%	0.04[-0.29,0.37]
Hurley 2007	127	20.2 (18.5)	127	25.9 (13.6)		3.03%	-0.35[-0.6,-0.1]
Jan 2008	68	14.8 (8.9)	30	22.5 (10.9)		2.27%	-0.8[-1.24,-0.36]
Lin 2009	36	10.1 (8.3)	36	24.9 (11.8)		2%	-1.44[-1.96,-0.91]
Jan 2009	71	11.2 (10.1)	35	25 (11.8)		2.29%	-1.28[-1.72,-0.84]
Salli 2010	47	20.7 (8.9)	24	32.6 (11.6)		1.97%	-1.2[-1.73,-0.66]
Bennell 2010	45	-8.1 (7.7)	44	-1.9 (7.6)		2.32%	-0.8[-1.23,-0.37]
Chang 2012	24	-10.7 (5.9)	17	-4.5 (4.4)		1.55%	-1.14[-1.81,-0.47]
Subtotal ***	**845**		**648**			**33.64%**	**-0.76[-1.03,-0.5]**
Heterogeneity: Tau^2=0.22; Chi^2=82.04, df=15(P<0.0001); I^2=81.72%							
Test for overall effect: Z=5.63(P<0.0001)							
5.2.2 Class-based programmes							
Minor 1989	49	-0.9 (2.5)	19	0.3 (2.5)		1.95%	-0.48[-1.02,0.05]
Kovar 1992	47	-2.4 (2.3)	45	0.2 (2.5)		2.29%	-1.1[-1.54,-0.66]
Bautch 1997	15	-2.8 (7.8)	15	-3.5 (8.2)		1.44%	0.08[-0.63,0.8]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		2.9%	-0.33[-0.61,-0.05]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		2.9%	-0.37[-0.66,-0.09]
Rogind 1998	11	-3 (3.3)	12	-2 (5.3)		1.22%	-0.22[-1.04,0.6]
Peloquin 1999	59	-1.5 (2.4)	65	-0.5 (2.6)		2.61%	-0.38[-0.74,-0.02]
Hopman-Rock 2000	37	-0.8 (4.6)	34	-1.7 (5.2)		2.19%	0.18[-0.28,0.65]
Fransen 2001	59	-8.5 (15.7)	43	0.1 (20.5)		2.45%	-0.48[-0.88,-0.08]
Topp 2002	67	-4.2 (10.9)	35	0.2 (10.9)		2.39%	-0.39[-0.81,0.02]
Song 2003	22	-11.1 (12)	21	-1.3 (10.6)		1.68%	-0.84[-1.47,-0.22]





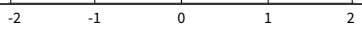
Comparison 6. Number of contact occasions

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain	44	3487	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.60, -0.41]
1.1 Fewer than 12 occasions	10	1019	Std. Mean Difference (IV, Random, 95% CI)	-0.40 [-0.56, -0.24]
1.2 12 or more occasions	34	2468	Std. Mean Difference (IV, Random, 95% CI)	-0.55 [-0.66, -0.43]

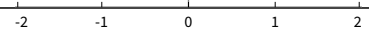
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2 Physical function	44	3913	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.64, -0.39]
2.1 Fewer than 12 occasions	9	1033	Std. Mean Difference (IV, Random, 95% CI)	-0.33 [-0.57, -0.09]
2.2 12 or more occasions	35	2880	Std. Mean Difference (IV, Random, 95% CI)	-0.57 [-0.71, -0.43]

Analysis 6.1. Comparison 6 Number of contact occasions, Outcome 1 Pain.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
6.1.1 Fewer than 12 occasions							
Deyle 2000	33	-129.6 (91)	36	-33.8 (111.5)		2.26%	-0.93[-1.43,-0.43]
Hopman-Rock 2000	45	-0.7 (24.1)	37	4 (21.2)		2.63%	-0.2[-0.64,0.23]
Talbot 2003	17	-0.1 (1)	17	0 (1.7)		1.51%	-0.09[-0.76,0.59]
Bennell 2005	73	-2.2 (1.7)	67	-2 (2.1)		3.41%	-0.1[-0.44,0.23]
Hay 2006	93	-1.6 (3.4)	89	-0.4 (2.8)		3.74%	-0.37[-0.66,-0.07]
Yip 2007	79	37.3 (21.1)	70	44.4 (23.2)		3.47%	-0.32[-0.64,0.01]
Lim 2008	53	-9 (12)	54	-1.7 (12.8)		2.97%	-0.58[-0.97,-0.19]
Doi 2008	61	22.6 (20.7)	56	29.6 (23.4)		3.14%	-0.32[-0.68,0.05]
Bezalel 2010	25	7 (7.5)	25	10 (7.5)		1.95%	-0.39[-0.95,0.17]
Bennell 2010	45	-2.6 (2.6)	44	-0.4 (2.7)		2.65%	-0.82[-1.26,-0.39]
Subtotal ***	**524**		**495**			**27.75%**	**-0.4[-0.56,-0.24]**
Heterogeneity: Tau^2=0.02; Chi^2=13.9, df=9(P=0.13); I^2=35.27%							
Test for overall effect: Z=4.95(P<0.0001)							
6.1.2 12 or more occasions							
Minor 1989	49	-0.8 (1.7)	19	-0.3 (1.6)		2.09%	-0.27[-0.8,0.27]
Kovar 1992	47	-1.4 (2)	45	-0.1 (2.3)		2.76%	-0.59[-1.01,-0.17]
Schilke 1996	10	-6.1 (4.9)	10	0.4 (6.7)		0.87%	-1.06[-2.01,-0.11]
Ettinger 1997a/b	146	2.2 (0.7)	75	2.5 (0.6)		3.85%	-0.36[-0.64,-0.08]
Bautch 1997	15	-1.4 (2.3)	15	1 (1.6)		1.19%	-1.2[-1.98,-0.41]
Ettinger 1997a/b	144	2.1 (0.6)	75	2.5 (0.6)		3.82%	-0.53[-0.81,-0.24]
Rogind 1998	11	-3 (3.9)	12	-0.1 (6.7)		1.08%	-0.5[-1.34,0.33]
van Baar 1998	54	-27.4 (28.7)	59	-11.7 (28.5)		3.06%	-0.55[-0.92,-0.17]
Maurer 1999	49	-43.5 (80.3)	49	-28.5 (80.3)		2.9%	-0.19[-0.58,0.21]
Peloquin 1999	59	-1.4 (2)	65	-0.6 (2.2)		3.21%	-0.4[-0.76,-0.04]
Fransen 2001	83	-10.6 (19.5)	43	1.5 (19.4)		3.06%	-0.62[-0.99,-0.24]
Baker 2001	22	-79 (88)	22	-20 (93)		1.75%	-0.64[-1.25,-0.03]
Gur 2002	17	-20.9 (8.3)	6	0.7 (4.6)		0.52%	-2.74[-4.02,-1.47]
Topp 2002	67	-1.5 (3.2)	35	0 (3.2)		2.78%	-0.48[-0.9,-0.07]
Huang 2003	99	-1.6 (1.5)	33	-0.4 (1.6)		2.84%	-0.78[-1.19,-0.38]
Foley 2003	21	-1.2 (2.9)	20	-0 (2.6)		1.7%	-0.41[-1.02,0.21]
Song 2003	22	-2.4 (3.9)	21	0.6 (5.1)		1.72%	-0.66[-1.28,-0.05]
Keefe 2004	16	-0.7 (1.7)	18	0 (1.3)		1.47%	-0.48[-1.17,0.2]
Hughes 2004	68	4.9 (3.4)	43	6.2 (4.3)		2.99%	-0.34[-0.73,0.04]



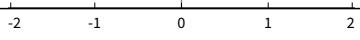
Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Thorstensson 2005	30	-1.8 (14)	31	0.3 (15)		2.24%	-0.14[-0.65,0.36]
Huang 2005	30	-1.2 (1.6)	32	-0.5 (1.7)		2.23%	-0.42[-0.92,0.09]
Brismée 2007	22	15.4 (5.7)	19	16.6 (4.7)		1.71%	-0.23[-0.85,0.38]
Fransen 2007	41	-1.7 (3.3)	36	-0.5 (2.4)		2.53%	-0.4[-0.85,0.05]
Jan 2008	68	4.8 (3.1)	30	7.1 (3.4)		2.6%	-0.71[-1.16,-0.27]
Lund 2008	25	38 (12.5)	27	39.7 (12)		2.02%	-0.14[-0.68,0.41]
An 2008	11	71.1 (110.1)	10	138.2 (112.6)		1%	-0.58[-1.46,0.3]
Lee 2009	29	-2.2 (4.1)	15	-0.2 (1.8)		1.64%	-0.56[-1.2,0.07]
Lin 2009	36	4.2 (3)	36	7.3 (3.4)		2.31%	-0.96[-1.45,-0.47]
Salli 2010	47	3.4 (1.8)	24	6.5 (1.8)		1.9%	-1.73[-2.3,-1.16]
Wang 2011	26	24 (15)	26	32 (18)		1.99%	-0.48[-1.03,0.08]
Foroughi 2011	20	3.8 (2.7)	25	4.4 (3.7)		1.82%	-0.18[-0.77,0.41]
Bruce-Brand 2012	10	10.8 (4.3)	6	8.3 (4.4)		0.75%	0.54[-0.5,1.57]
Chang 2012	24	-2.3 (1.3)	17	-0.9 (1.5)		1.55%	-0.99[-1.65,-0.33]
Simao 2012	11	-62.5 (296)	12	0 (35)		1.11%	-0.29[-1.12,0.53]
Salacinski 2012	13	18.6 (13.4)	15	34.3 (15.9)		1.16%	-1.03[-1.83,-0.23]
Subtotal * **	**1442**		**1026**			**72.25%**	**-0.55[-0.66,-0.43]**
Heterogeneity: Tau²=0.05; Chi²=59.84, df=34(P=0); I²=43.19%							
Test for overall effect: Z=9.13(P<0.0001)							
Total * **	**1966**		**1521**			**100%**	**-0.51[-0.6,-0.41]**
Heterogeneity: Tau²=0.04; Chi²=76.81, df=44(P=0); I²=42.71%							
Test for overall effect: Z=10.27(P<0.0001)							
Test for subgroup differences: Chi²=2.03, df=1 (P=0.15), I²=50.74%							



Analysis 6.2. Comparison 6 Number of contact occasions, Outcome 2 Physical function.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
6.2.1 Fewer than 12 occasions							
Hopman-Rock 2000	37	-0.8 (4.6)	34	-1.7 (5.2)		2.39%	0.18[-0.28,0.65]
Deyle 2000	33	-402.5 (339.6)	36	-98.2 (393.9)		2.3%	-0.82[-1.31,-0.32]
Bennell 2005	73	-7.8 (8.7)	67	-8.2 (10)		2.89%	0.04[-0.29,0.37]
Hay 2006	95	-4.8 (10.8)	90	-0.8 (8.5)		3.03%	-0.41[-0.7,-0.12]
Doi 2008	61	13.7 (13.5)	56	18.6 (16.4)		2.76%	-0.33[-0.69,0.04]
Lim 2008	53	-6 (10.6)	54	-2.6 (10.9)		2.7%	-0.32[-0.7,0.06]
Bennell 2010	45	-8.1 (7.7)	44	-1.9 (7.6)		2.51%	-0.8[-1.23,-0.37]
Bezalel 2010	25	25 (10)	25	34 (10)		2%	-0.89[-1.47,-0.3]
Kao 2012	114	3.2 (34)	91	1.5 (20.3)		3.09%	0.06[-0.22,0.33]
Subtotal * **	**536**		**497**			**23.68%**	**-0.33[-0.57,-0.09]**
Heterogeneity: Tau²=0.09; Chi²=28.25, df=8(P=0); I²=71.68%							
Test for overall effect: Z=2.71(P=0.01)							
6.2.2 12 or more occasions							
Minor 1989	49	-0.9 (2.5)	19	0.3 (2.5)		2.15%	-0.48[-1.02,0.05]
Kovar 1992	47	-2.4 (2.3)	45	0.2 (2.5)		2.49%	-1.1[-1.54,-0.66]
Schilke 1996	10	-3.7 (3.3)	10	-0.4 (3.5)		1.18%	-0.91[-1.84,0.02]

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.33[-0.61,-0.05]
Bautch 1997	15	-2.8 (7.8)	15	-3.5 (8.2)		1.63%	0.08[-0.63,0.8]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.37[-0.66,-0.09]
van Baar 1998	54	-1.3 (5.7)	59	-0.5 (5.6)		2.75%	-0.14[-0.51,0.23]
Rogind 1998	11	-3 (3.3)	12	-2 (5.3)		1.39%	-0.22[-1.04,0.6]
Peloquin 1999	59	-1.5 (2.4)	65	-0.5 (2.6)		2.8%	-0.38[-0.74,-0.02]
Maurer 1999	49	-106.9 (390.1)	49	-88.3 (390.1)		2.65%	-0.05[-0.44,0.35]
Fransen 2001	83	-7.7 (19.9)	43	0.1 (20.5)		2.74%	-0.39[-0.76,-0.01]
Baker 2001	22	-272 (295)	22	-119 (323)		1.95%	-0.49[-1.09,0.11]
Gur 2002	17	-13.8 (4.1)	6	1 (2.5)		0.56%	-3.77[-5.29,-2.26]
Topp 2002	67	-4.2 (10.9)	35	0.2 (10.9)		2.59%	-0.39[-0.81,0.02]
Song 2003	22	-11.1 (12)	21	-1.3 (10.6)		1.87%	-0.84[-1.47,-0.22]
Huang 2003	99	-2 (1.6)	33	-0.4 (1.7)		2.59%	-0.98[-1.39,-0.57]
Foley 2003	21	-2.8 (7.9)	20	2.1 (8.1)		1.87%	-0.6[-1.23,0.03]
Hughes 2004	68	17.3 (12.6)	43	22.3 (12.8)		2.69%	-0.39[-0.78,-0.01]
Thorstensson 2005	30	-2 (12)	31	0.6 (18)		2.27%	-0.17[-0.67,0.34]
Huang 2005	30	-1.5 (1.4)	32	-0.5 (1.7)		2.24%	-0.63[-1.14,-0.12]
Brismée 2007	22	39.5 (13)	19	40.7 (11.9)		1.91%	-0.09[-0.71,0.52]
Fransen 2007	41	-5 (10.3)	36	2.1 (9.1)		2.41%	-0.72[-1.19,-0.26]
Hurley 2007	229	20 (18.5)	113	25.9 (13.6)		3.25%	-0.35[-0.57,-0.12]
Lund 2008	25	35.9 (11.5)	27	38.9 (11)		2.12%	-0.26[-0.81,0.28]
An 2008	11	347.5 (383.8)	10	511.8 (381.6)		1.29%	-0.41[-1.28,0.46]
Jan 2008	68	14.8 (8.9)	30	22.5 (10.9)		2.47%	-0.8[-1.24,-0.36]
Jan 2009	71	11.2 (10.1)	35	25 (11.8)		2.48%	-1.28[-1.72,-0.84]
Lin 2009	36	10.1 (8.3)	36	24.9 (11.8)		2.2%	-1.44[-1.96,-0.91]
Lee 2009	29	-9.4 (14.4)	15	-2.7 (10.8)		1.85%	-0.49[-1.13,0.14]
Salli 2010	47	20.7 (8.9)	24	32.6 (11.6)		2.17%	-1.2[-1.73,-0.66]
Foroughi 2011	20	13.3 (9.4)	25	18.1 (12)		1.97%	-0.43[-1.03,0.16]
Wang 2011	26	18 (14)	26	31 (18)		2.06%	-0.79[-1.36,-0.23]
Chang 2012	24	-10.7 (5.9)	17	-4.5 (4.4)		1.74%	-1.14[-1.81,-0.47]
Salacinski 2012	13	15.8 (13.9)	15	28.9 (16.2)		1.48%	-0.84[-1.62,-0.06]
Bruce-Brand 2012	10	33.9 (12.9)	6	26.1 (15.3)		1.02%	0.53[-0.5,1.57]
Simao 2012	11	-100 (740)	12	75 (463)		1.38%	-0.28[-1.1,0.55]
Subtotal * **	**1724**		**1156**			**76.32%**	**-0.57[-0.71,-0.43]**
Heterogeneity: Tau^2=0.1; Chi^2=96.42, df=35(P<0.0001); I^2=63.7%							
Test for overall effect: Z=8.08(P<0.0001)							
Total * **	**2260**		**1653**			**100%**	**-0.51[-0.64,-0.39]**
Heterogeneity: Tau^2=0.11; Chi^2=135.74, df=44(P<0.0001); I^2=67.58%							
Test for overall effect: Z=8.21(P<0.0001)							
Test for subgroup differences: Chi^2=2.83, df=1 (P=0.09), I^2=64.63%							

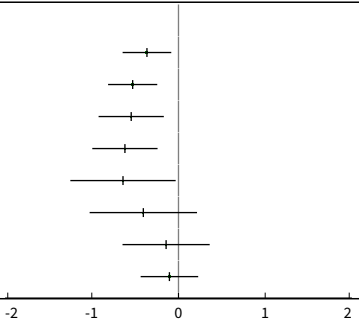


Comparison 7. Sensitivity Analyses

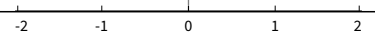
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Selection and attrition bias: pain	44	3487	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.60, -0.41]
1.1 Low risk	14	1458	Std. Mean Difference (IV, Random, 95% CI)	-0.47 [-0.59, -0.36]
1.2 Unclear or high risk	30	2029	Std. Mean Difference (IV, Random, 95% CI)	-0.53 [-0.67, -0.39]
2 Selection and attrition bias: physical function	44	3913	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.64, -0.39]
2.1 Low risk	14	1456	Std. Mean Difference (IV, Random, 95% CI)	-0.45 [-0.63, -0.28]
2.2 Unclear or high risk	30	2457	Std. Mean Difference (IV, Random, 95% CI)	-0.55 [-0.72, -0.38]
3 Detection bias: pain	44	3487	Std. Mean Difference (IV, Random, 95% CI)	-0.51 [-0.60, -0.41]
3.1 Low risk	3	226	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.87, 0.13]
3.2 Unclear or high risk	41	3261	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.61, -0.42]
4 Detection bias: physical function	44	3913	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.64, -0.39]
4.1 Low risk	3	226	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [-1.14, 0.22]
4.2 Unclear or high risk	41	3687	Std. Mean Difference (IV, Random, 95% CI)	-0.52 [-0.65, -0.40]

Analysis 7.1. Comparison 7 Sensitivity Analyses, Outcome 1 Selection and attrition bias: pain.

Study or subgroup	Favours exercise N	Favours exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
7.1.1 Low risk							
Ettinger 1997a/b	146	2.2 (0.7)	75	2.5 (0.6)		3.85%	-0.36[-0.64,-0.08]
Ettinger 1997a/b	144	2.1 (0.6)	75	2.5 (0.6)		3.82%	-0.53[-0.81,-0.24]
van Baar 1998	54	-27.4 (28.7)	59	-11.7 (28.5)		3.06%	-0.55[-0.92,-0.17]
Fransen 2001	83	-10.6 (19.5)	43	1.5 (19.4)		3.06%	-0.62[-0.99,-0.24]
Baker 2001	22	-79 (88)	22	-20 (93)		1.75%	-0.64[-1.25,-0.03]
Foley 2003	21	-1.2 (2.9)	20	-0 (2.6)		1.7%	-0.41[-1.02,0.21]
Thorstensson 2005	30	-1.8 (14)	31	0.3 (15)		2.24%	-0.14[-0.65,0.36]
Bennell 2005	73	-2.2 (1.7)	67	-2 (2.1)		3.41%	-0.1[-0.44,0.23]



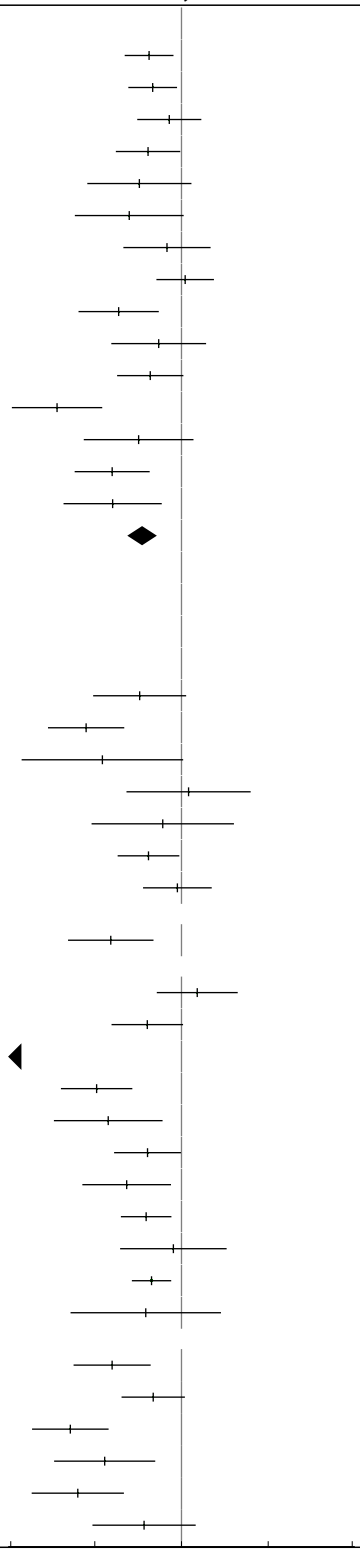
Study or subgroup	Favours exercise N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Fransen 2007	41	-1.7 (3.3)	36	-0.5 (2.4)		2.53%	-0.4[-0.85,0.05]
Lund 2008	25	38 (12.5)	27	39.7 (12)		2.02%	-0.14[-0.68,0.41]
Lim 2008	53	-9 (12)	54	-1.7 (12.8)		2.97%	-0.58[-0.97,-0.19]
Lee 2009	29	-2.2 (4.1)	15	-0.2 (1.8)		1.64%	-0.56[-1.2,0.07]
Lin 2009	36	4.2 (3)	36	7.3 (3.4)		2.31%	-0.96[-1.45,-0.47]
Bennell 2010	45	-2.6 (2.6)	44	-0.4 (2.7)		2.65%	-0.82[-1.26,-0.39]
Wang 2011	26	24 (15)	26	32 (18)		1.99%	-0.48[-1.03,0.08]
Subtotal ***	**828**		**630**			**39.01%**	**-0.47[-0.59,-0.36]**
Heterogeneity: Tau²=0.01; Chi²=16.35, df=14(P=0.29); I²=14.39%							
Test for overall effect: Z=7.9(P<0.0001)							
7.1.2 Unclear or high risk							
Minor 1989	49	-0.8 (1.7)	19	-0.3 (1.6)		2.09%	-0.27[-0.8,0.27]
Kovar 1992	47	-1.4 (2)	45	-0.1 (2.3)		2.76%	-0.59[-1.01,-0.17]
Schilke 1996	10	-6.1 (4.9)	10	0.4 (6.7)		0.87%	-1.06[-2.01,-0.11]
Bautch 1997	15	-1.4 (2.3)	15	1 (1.6)		1.19%	-1.2[-1.98,-0.41]
Rogind 1998	11	-3 (3.9)	12	-0.1 (6.7)		1.08%	-0.5[-1.34,0.33]
Peloquin 1999	59	-1.4 (2)	65	-0.6 (2.2)		3.21%	-0.4[-0.76,-0.04]
Maurer 1999	49	-43.5 (80.3)	49	-28.5 (80.3)		2.9%	-0.19[-0.58,0.21]
Deyle 2000	33	-129.6 (91)	36	-33.8 (111.5)		2.26%	-0.93[-1.43,-0.43]
Hopman-Rock 2000	45	-0.7 (24.1)	37	4 (21.2)		2.63%	-0.2[-0.64,0.23]
Gur 2002	17	-20.9 (8.3)	6	0.7 (4.6)		0.52%	-2.74[-4.02,-1.47]
Topp 2002	67	-1.5 (3.2)	35	0 (3.2)		2.78%	-0.48[-0.9,-0.07]
Song 2003	22	-2.4 (3.9)	21	0.6 (5.1)		1.72%	-0.66[-1.28,-0.05]
Talbot 2003	17	-0.1 (1)	17	0 (1.7)		1.51%	-0.09[-0.76,0.59]
Huang 2003	99	-1.6 (1.5)	33	-0.4 (1.6)		2.84%	-0.78[-1.19,-0.38]
Keefe 2004	16	-0.7 (1.7)	18	0 (1.3)		1.47%	-0.48[-1.17,0.2]
Hughes 2004	68	4.9 (3.4)	43	6.2 (4.3)		2.99%	-0.34[-0.73,0.04]
Huang 2005	30	-1.2 (1.6)	32	-0.5 (1.7)		2.23%	-0.42[-0.92,0.09]
Hay 2006	93	-1.6 (3.4)	89	-0.4 (2.8)		3.74%	-0.37[-0.66,-0.07]
Yip 2007	79	37.3 (21.1)	70	44.4 (23.2)		3.47%	-0.32[-0.64,0.01]
Brismée 2007	22	15.4 (5.7)	19	16.6 (4.7)		1.71%	-0.23[-0.85,0.38]
An 2008	11	71.1 (110.1)	10	138.2 (112.6)		1%	-0.58[-1.46,0.3]
Doi 2008	61	22.6 (20.7)	56	29.6 (23.4)		3.14%	-0.32[-0.68,0.05]
Jan 2008	68	4.8 (3.1)	30	7.1 (3.4)		2.6%	-0.71[-1.16,-0.27]
Bezalel 2010	25	7 (7.5)	25	10 (7.5)		1.95%	-0.39[-0.95,0.17]
Salli 2010	47	3.4 (1.8)	24	6.5 (1.8)		1.9%	-1.73[-2.3,-1.16]
Foroughi 2011	20	3.8 (2.7)	25	4.4 (3.7)		1.82%	-0.18[-0.77,0.41]
Salacinski 2012	13	18.6 (13.4)	15	34.3 (15.9)		1.16%	-1.03[-1.83,-0.23]
Chang 2012	24	-2.3 (1.3)	17	-0.9 (1.5)		1.55%	-0.99[-1.65,-0.33]
Bruce-Brand 2012	10	10.8 (4.3)	6	8.3 (4.4)		0.75%	0.54[-0.5,1.57]
Simao 2012	11	-62.5 (296)	12	0 (35)		1.11%	-0.29[-1.12,0.53]
Subtotal ***	**1138**		**891**			**60.99%**	**-0.53[-0.67,-0.39]**
Heterogeneity: Tau²=0.07; Chi²=60.37, df=29(P=0); I²=51.97%							
Test for overall effect: Z=7.46(P<0.0001)							
Total ***	**1966**		**1521**			**100%**	**-0.51[-0.6,-0.41]**
Heterogeneity: Tau²=0.04; Chi²=76.81, df=44(P=0); I²=42.71%							
Test for overall effect: Z=10.27(P<0.0001)							
Test for subgroup differences: Chi²=0.38, df=1 (P=0.54), I²=0%							



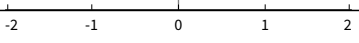
Favours exercise Favours control

Analysis 7.2. Comparison 7 Sensitivity Analyses, Outcome 2 Selection and attrition bias: physical function.

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
7.2.1 Low risk							
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.37[-0.66,-0.09]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.33[-0.61,-0.05]
van Baar 1998	54	-1.3 (5.7)	59	-0.5 (5.6)		2.75%	-0.14[-0.51,0.23]
Fransen 2001	83	-7.7 (19.9)	43	0.1 (20.5)		2.74%	-0.39[-0.76,-0.01]
Baker 2001	22	-272 (295)	22	-119 (323)		1.95%	-0.49[-1.09,0.11]
Foley 2003	21	-2.8 (7.9)	20	2.1 (8.1)		1.87%	-0.6[-1.23,0.03]
Thorstensson 2005	30	-2 (12)	31	0.6 (18)		2.27%	-0.17[-0.67,0.34]
Bennell 2005	73	-7.8 (8.7)	67	-8.2 (10)		2.89%	0.04[-0.29,0.37]
Fransen 2007	41	-5 (10.3)	36	2.1 (9.1)		2.41%	-0.72[-1.19,-0.26]
Lund 2008	25	35.9 (11.5)	27	38.9 (11)		2.12%	-0.26[-0.81,0.28]
Lim 2008	53	-6.5 (10.6)	54	-2.6 (10.9)		2.7%	-0.36[-0.74,0.02]
Lin 2009	36	10.1 (8.3)	36	24.9 (11.8)		2.2%	-1.44[-1.96,-0.91]
Lee 2009	29	-9.4 (14.4)	15	-2.7 (10.8)		1.85%	-0.49[-1.13,0.14]
Bennell 2010	45	-8.1 (7.7)	44	-1.9 (7.6)		2.52%	-0.8[-1.23,-0.37]
Wang 2011	26	18 (14)	26	31 (18)		2.06%	-0.79[-1.36,-0.23]
Subtotal ***	**826**		**630**			**36.45%**	**-0.45[-0.63,-0.28]**
Heterogeneity: Tau^2=0.06; Chi^2=32.82, df=14(P=0); I^2=57.35%							
Test for overall effect: Z=5.22(P<0.0001)							
7.2.2 Unclear or high risk							
Minor 1989	49	-0.9 (2.5)	19	0.3 (2.5)		2.15%	-0.48[-1.02,0.05]
Kovar 1992	47	-2.4 (2.3)	45	0.2 (2.5)		2.49%	-1.1[-1.54,-0.66]
Schilke 1996	10	-3.7 (3.3)	10	-0.4 (3.5)		1.18%	-0.91[-1.84,0.02]
Bautch 1997	15	-2.8 (7.8)	15	-3.5 (8.2)		1.63%	0.08[-0.63,0.8]
Rogind 1998	11	-3 (3.3)	12	-2 (5.3)		1.39%	-0.22[-1.04,0.6]
Peloquin 1999	59	-1.5 (2.4)	65	-0.5 (2.6)		2.8%	-0.38[-0.74,-0.02]
Maurer 1999	49	-106.9 (390.1)	49	-88.3 (390.1)		2.65%	-0.05[-0.44,0.35]
Deyle 2000	33	-402.5 (339.6)	36	-98.2 (393.9)		2.3%	-0.82[-1.31,-0.32]
Hopman-Rock 2000	37	-0.8 (4.6)	34	-1.7 (5.2)		2.39%	0.18[-0.28,0.65]
Topp 2002	67	-4.2 (10.9)	35	0.2 (10.9)		2.59%	-0.39[-0.81,0.02]
Gur 2002	17	-13.8 (4.1)	6	1 (2.5)		0.55%	-3.77[-5.29,-2.26]
Huang 2003	99	-2 (1.6)	33	-0.4 (1.7)		2.59%	-0.98[-1.39,-0.57]
Song 2003	22	-11.1 (12)	21	-1.3 (10.6)		1.87%	-0.84[-1.47,-0.22]
Hughes 2004	68	17.3 (12.6)	43	22.3 (12.8)		2.69%	-0.39[-0.78,-0.01]
Huang 2005	30	-1.5 (1.4)	32	-0.5 (1.7)		2.24%	-0.63[-1.14,-0.12]
Hay 2006	95	-4.8 (10.8)	90	-0.8 (8.5)		3.03%	-0.41[-0.7,-0.12]
Brismée 2007	22	39.5 (13)	19	40.7 (11.9)		1.91%	-0.09[-0.71,0.52]
Hurley 2007	229	20 (18.5)	113	25.9 (13.6)		3.25%	-0.35[-0.57,-0.12]
An 2008	11	347.5 (383.8)	10	511.8 (381.6)		1.29%	-0.41[-1.28,0.46]
Jan 2008	68	14.8 (8.9)	30	22.5 (10.9)		2.47%	-0.8[-1.24,-0.36]
Doi 2008	61	13.7 (13.5)	56	18.6 (16.4)		2.76%	-0.33[-0.69,0.04]
Jan 2009	71	11.2 (10.1)	35	25 (11.8)		2.48%	-1.28[-1.72,-0.84]
Bezalel 2010	25	25 (10)	25	34 (10)		2%	-0.89[-1.47,-0.3]
Salli 2010	47	20.7 (8.9)	24	32.6 (11.6)		2.17%	-1.2[-1.73,-0.66]
Foroughi 2011	20	13.3 (9.4)	25	18.1 (12)		1.96%	-0.43[-1.03,0.16]

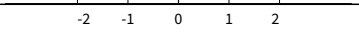


Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Salacinski 2012	13	15.8 (13.9)	15	28.9 (16.2)		1.48%	-0.84[-1.62,-0.06]
Simao 2012	11	-100 (740)	12	75 (463)		1.38%	-0.28[-1.1,0.55]
Kao 2012	114	3.2 (34)	91	1.5 (20.3)		3.09%	0.06[-0.22,0.33]
Chang 2012	24	-10.7 (5.9)	17	-4.5 (4.4)		1.74%	-1.14[-1.81,-0.47]
Bruce-Brand 2012	10	33.9 (12.9)	6	26.1 (15.3)		1.02%	0.53[-0.5,1.57]
Subtotal ***	**1434**		**1023**			**63.55%**	**-0.55[-0.72,-0.38]**
Heterogeneity: Tau^2=0.14; Chi^2=101.8, df=29(P<0.0001); I^2=71.51%							
Test for overall effect: Z=6.4(P<0.0001)							
Total ***	**2260**		**1653**			**100%**	**-0.52[-0.64,-0.39]**
Heterogeneity: Tau^2=0.11; Chi^2=135.5, df=44(P<0.0001); I^2=67.53%							
Test for overall effect: Z=8.23(P<0.0001)							
Test for subgroup differences: Chi^2=0.6, df=1 (P=0.44), I^2=0%							

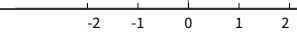


Analysis 7.3. Comparison 7 Sensitivity Analyses, Outcome 3 Detection bias: pain.

Study or subgroup	Favours exercise N	Favours exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
7.3.1 Low risk							
Bennell 2005	73	-2.2 (1.7)	67	-2 (2.1)		3.41%	-0.1[-0.44,0.23]
Foroughi 2011	20	3.8 (2.7)	25	4.4 (3.7)		1.82%	-0.18[-0.77,0.41]
Chang 2012	24	-2.3 (1.3)	17	-0.9 (1.5)		1.55%	-0.99[-1.65,-0.33]
Subtotal ***	**117**		**109**			**6.78%**	**-0.37[-0.87,0.13]**
Heterogeneity: Tau^2=0.13; Chi^2=5.61, df=2(P=0.06); I^2=64.36%							
Test for overall effect: Z=1.45(P=0.15)							
7.3.2 Unclear or high risk							
Minor 1989	49	-0.8 (1.7)	19	-0.3 (1.6)		2.09%	-0.27[-0.8,0.27]
Kovar 1992	47	-1.4 (2)	45	-0.1 (2.3)		2.76%	-0.59[-1.01,-0.17]
Schilke 1996	10	-6.1 (4.9)	10	0.4 (6.7)		0.87%	-1.06[-2.01,-0.11]
Ettinger 1997a/b	144	2.1 (0.6)	75	2.5 (0.6)		3.82%	-0.53[-0.81,-0.24]
Ettinger 1997a/b	146	2.2 (0.7)	75	2.5 (0.6)		3.85%	-0.36[-0.64,-0.08]
Bautch 1997	15	-1.4 (2.3)	15	1 (1.6)		1.19%	-1.2[-1.98,-0.41]
van Baar 1998	54	-27.4 (28.7)	59	-11.7 (28.5)		3.06%	-0.55[-0.92,-0.17]
Rogind 1998	11	-3 (3.9)	12	-0.1 (6.7)		1.08%	-0.5[-1.34,0.33]
Maurer 1999	49	-43.5 (80.3)	49	-28.5 (80.3)		2.9%	-0.19[-0.58,0.21]
Peloquin 1999	59	-1.4 (2)	65	-0.6 (2.2)		3.21%	-0.4[-0.76,-0.04]
Hopman-Rock 2000	45	-0.7 (24.1)	37	4 (21.2)		2.63%	-0.2[-0.64,0.23]
Deyle 2000	33	-129.6 (91)	36	-33.8 (111.5)		2.26%	-0.93[-1.43,-0.43]
Baker 2001	22	-79 (88)	22	-20 (93)		1.75%	-0.64[-1.25,-0.03]
Fransen 2001	83	-10.6 (19.5)	43	1.5 (19.4)		3.06%	-0.62[-0.99,-0.24]
Gur 2002	17	-20.9 (8.3)	6	0.7 (4.6)		0.52%	-2.74[-4.02,-1.47]
Topp 2002	67	-1.5 (3.2)	35	0 (3.2)		2.78%	-0.48[-0.9,-0.07]
Foley 2003	21	-1.2 (2.9)	20	-0 (2.6)		1.7%	-0.41[-1.02,0.21]
Song 2003	22	-2.4 (3.9)	21	0.6 (5.1)		1.72%	-0.66[-1.28,-0.05]
Talbot 2003	17	-0.1 (1)	17	0 (1.7)		1.51%	-0.09[-0.76,0.59]
Huang 2003	99	-1.6 (1.5)	33	-0.4 (1.6)		2.84%	-0.78[-1.19,-0.38]



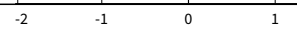
Study or subgroup	Favours exercise N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Keefe 2004	16	-0.7 (1.7)	18	0 (1.3)		1.47%	-0.48[-1.17,0.2]
Hughes 2004	68	4.9 (3.4)	43	6.2 (4.3)		2.99%	-0.34[-0.73,0.04]
Huang 2005	30	-1.2 (1.6)	32	-0.5 (1.7)		2.23%	-0.42[-0.92,0.09]
Thorstensson 2005	30	-1.8 (14)	31	0.3 (15)		2.24%	-0.14[-0.65,0.36]
Hay 2006	93	-1.6 (3.4)	89	-0.4 (2.8)		3.74%	-0.37[-0.66,-0.07]
Fransen 2007	41	-1.7 (3.3)	36	-0.5 (2.4)		2.53%	-0.4[-0.85,0.05]
Yip 2007	79	37.3 (21.1)	70	44.4 (23.2)		3.47%	-0.32[-0.64,0.01]
Brismée 2007	22	15.4 (5.7)	19	16.6 (4.7)		1.71%	-0.23[-0.85,0.38]
Lim 2008	53	-9 (12)	54	-1.7 (12.8)		2.97%	-0.58[-0.97,-0.19]
Lund 2008	25	38 (12.5)	27	39.7 (12)		2.02%	-0.14[-0.68,0.41]
An 2008	11	71.1 (110.1)	10	138.2 (112.6)		1%	-0.58[-1.46,0.3]
Jan 2008	68	4.8 (3.1)	30	7.1 (3.4)		2.6%	-0.71[-1.16,-0.27]
Doi 2008	61	22.6 (20.7)	56	29.6 (23.4)		3.14%	-0.32[-0.68,0.05]
Lin 2009	36	4.2 (3)	36	7.3 (3.4)		2.31%	-0.96[-1.45,-0.47]
Lee 2009	29	-2.2 (4.1)	15	-0.2 (1.8)		1.64%	-0.56[-1.2,0.07]
Bezalel 2010	25	7 (7.5)	25	10 (7.5)		1.95%	-0.39[-0.95,0.17]
Salli 2010	47	3.4 (1.8)	24	6.5 (1.8)		1.9%	-1.73[-2.3,-1.16]
Bennell 2010	45	-2.6 (2.6)	44	-0.4 (2.7)		2.65%	-0.82[-1.26,-0.39]
Wang 2011	26	24 (15)	26	32 (18)		1.99%	-0.48[-1.03,0.08]
Simao 2012	11	-62.5 (296)	12	0 (35)		1.11%	-0.29[-1.12,0.53]
Bruce-Brand 2012	10	10.8 (4.3)	6	8.3 (4.4)		0.75%	0.54[-0.5,1.57]
Salacinski 2012	13	18.6 (13.4)	15	34.3 (15.9)		1.16%	-1.03[-1.83,-0.23]
Subtotal * **	**1849**		**1412**			**93.22%**	**-0.52[-0.61,-0.42]**
Heterogeneity: Tau²=0.04; Chi²=68.34, df=41(P=0); I²=40.01%							
Test for overall effect: Z=10.39(P<0.0001)							
Total * **	**1966**		**1521**			**100%**	**-0.51[-0.6,-0.41]**
Heterogeneity: Tau²=0.04; Chi²=76.81, df=44(P=0); I²=42.71%							
Test for overall effect: Z=10.27(P<0.0001)							
Test for subgroup differences: Chi²=0.32, df=1 (P=0.57), I²=0%							



Favours exercise — Favours control

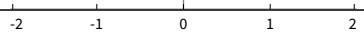
Analysis 7.4. Comparison 7 Sensitivity Analyses, Outcome 4 Detection bias: physical function.

Study or subgroup	Exercise N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
7.4.1 Low risk							
Bennell 2005	73	-7.8 (8.7)	67	-8.2 (10)		2.89%	0.04[-0.29,0.37]
Foroughi 2011	20	13.3 (9.4)	25	18.1 (12)		1.96%	-0.43[-1.03,0.16]
Chang 2012	24	-10.7 (5.9)	17	-4.5 (4.4)		1.74%	-1.14[-1.81,-0.47]
Subtotal * **	**117**		**109**			**6.59%**	**-0.46[-1.14,0.22]**
Heterogeneity: Tau²=0.28; Chi²=10.04, df=2(P=0.01); I²=80.08%							
Test for overall effect: Z=1.33(P=0.18)							
7.4.2 Unclear or high risk							
Minor 1989	49	-0.9 (2.5)	19	0.3 (2.5)		2.15%	-0.48[-1.02,0.05]
Kovar 1992	47	-2.4 (2.3)	45	0.2 (2.5)		2.49%	-1.1[-1.54,-0.66]
Schilke 1996	10	-3.7 (3.3)	10	-0.4 (3.5)		1.18%	-0.91[-1.84,0.02]



Favours exercise — Favours control

Study or subgroup	Exercise N	Exercise Mean(SD)	Control N	Control Mean(SD)	Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
Bautch 1997	15	-2.8 (7.8)	15	-3.5 (8.2)		1.63%	0.08[-0.63,0.8]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.33[-0.61,-0.05]
Ettinger 1997a/b	144	1.7 (0.5)	75	1.9 (0.5)		3.07%	-0.37[-0.66,-0.09]
van Baar 1998	54	-1.3 (5.7)	59	-0.5 (5.6)		2.75%	-0.14[-0.51,0.23]
Rogind 1998	11	-3 (3.3)	12	-2 (5.3)		1.39%	-0.22[-1.04,0.6]
Peloquin 1999	59	-1.5 (2.4)	65	-0.5 (2.6)		2.8%	-0.38[-0.74,-0.02]
Maurer 1999	49	-106.9 (390.1)	49	-88.3 (390.1)		2.65%	-0.05[-0.44,0.35]
Deyle 2000	33	-402.5 (339.6)	36	-98.2 (393.9)		2.3%	-0.82[-1.31,-0.32]
Hopman-Rock 2000	37	-0.8 (4.6)	34	-1.7 (5.2)		2.39%	0.18[-0.28,0.65]
Baker 2001	22	-272 (295)	22	-119 (323)		1.95%	-0.49[-1.09,0.11]
Fransen 2001	83	-7.7 (19.9)	43	0.1 (20.5)		2.74%	-0.39[-0.76,-0.01]
Gur 2002	17	-13.8 (4.1)	6	1 (2.5)		0.55%	-3.77[-5.29,-2.26]
Topp 2002	67	-4.2 (10.9)	35	0.2 (10.9)		2.59%	-0.39[-0.81,0.02]
Song 2003	22	-11.1 (12)	21	-1.3 (10.6)		1.87%	-0.84[-1.47,-0.22]
Huang 2003	99	-2 (1.6)	33	-0.4 (1.7)		2.59%	-0.98[-1.39,-0.57]
Foley 2003	21	-2.8 (7.9)	20	2.1 (8.1)		1.87%	-0.6[-1.23,0.03]
Hughes 2004	68	17.3 (12.6)	43	22.3 (12.8)		2.69%	-0.39[-0.78,-0.01]
Thorstensson 2005	30	-2 (12)	31	0.6 (18)		2.27%	-0.17[-0.67,0.34]
Huang 2005	30	-1.5 (1.4)	32	-0.5 (1.7)		2.24%	-0.63[-1.14,-0.12]
Hay 2006	95	-4.8 (10.8)	90	-0.8 (8.5)		3.03%	-0.41[-0.7,-0.12]
Fransen 2007	41	-5 (10.3)	36	2.1 (9.1)		2.41%	-0.72[-1.19,-0.26]
Brismée 2007	22	39.5 (13)	19	40.7 (11.9)		1.91%	-0.09[-0.71,0.52]
Hurley 2007	229	20 (18.5)	113	25.9 (13.6)		3.25%	-0.35[-0.57,-0.12]
Doi 2008	61	13.7 (13.5)	56	18.6 (16.4)		2.76%	-0.33[-0.69,0.04]
An 2008	11	347.5 (383.8)	10	511.8 (381.6)		1.29%	-0.41[-1.28,0.46]
Lim 2008	53	-6.5 (10.6)	54	-2.6 (10.9)		2.7%	-0.36[-0.74,0.02]
Lund 2008	25	35.9 (11.5)	27	38.9 (11)		2.12%	-0.26[-0.81,0.28]
Jan 2008	68	14.8 (8.9)	30	22.5 (10.9)		2.47%	-0.8[-1.24,-0.36]
Lee 2009	29	-9.4 (14.4)	15	-2.7 (10.8)		1.85%	-0.49[-1.13,0.14]
Lin 2009	36	10.1 (8.3)	36	24.9 (11.8)		2.2%	-1.44[-1.96,-0.91]
Jan 2009	71	11.2 (10.1)	35	25 (11.8)		2.48%	-1.28[-1.72,-0.84]
Bennell 2010	45	-8.1 (7.7)	44	-1.9 (7.6)		2.52%	-0.8[-1.23,-0.37]
Bezalel 2010	25	25 (10)	25	34 (10)		2%	-0.89[-1.47,-0.3]
Salli 2010	47	20.7 (8.9)	24	32.6 (11.6)		2.17%	-1.2[-1.73,-0.66]
Wang 2011	26	18 (14)	26	31 (18)		2.06%	-0.79[-1.36,-0.23]
Bruce-Brand 2012	10	33.9 (12.9)	6	26.1 (15.3)		1.02%	0.53[-0.5,1.57]
Simao 2012	11	-100 (740)	12	75 (463)		1.38%	-0.28[-1.1,0.55]
Kao 2012	114	3.2 (34)	91	1.5 (20.3)		3.09%	0.06[-0.22,0.33]
Salacinski 2012	13	15.8 (13.9)	15	28.9 (16.2)		1.48%	-0.84[-1.62,-0.06]
Subtotal * **	**2143**		**1544**			**93.41%**	**-0.52[-0.65,-0.4]**
Heterogeneity: Tau²=0.1; Chi²=122.85, df=41(P<0.0001); I²=66.63%							
Test for overall effect: Z=8.18(P<0.0001)							
Total * **	**2260**		**1653**			**100%**	**-0.52[-0.64,-0.39]**
Heterogeneity: Tau²=0.11; Chi²=135.5, df=44(P<0.0001); I²=67.53%							
Test for overall effect: Z=8.23(P<0.0001)							
Test for subgroup differences: Chi²=0.03, df=1 (P=0.86), I²=0%							



APPENDICES

Appendix 1. MEDLINE search strategy

1. exp osteoarthritis/
2. osteoarthr$.tw.
3. (degenerative adj2 arthritis).tw.
4. arthrosis.tw.
5. or/1-4
6. Knee/
7. exp Knee Joint/
8. knee$.tw.
9. or/6-8
10. exp EXERCISE/
11. exp exertion/
12. exp Physical Fitness/
13. exp Exercise Test/
14. exp Exercise Tolerance/
15. exp Sports/
16. exp PLIABILITY/
17. exp Physical Endurance/
18. exertion$.tw.
19. exercis$.tw.
20. sport$.tw.
21. ((physical or motion) adj5 (fitness or therap$)).tw.
22. (physical$ adj2 endur$).tw.
23. ((strength$ or isometric$ or isotonic$ or isokinetic$ or aerobic$ or endurance or weight$) adj5 (exercis$ or train$)).tw.
24. exp physical therapy modalities/
25. physiotherap$.tw.
26. manipulat$.tw.
27. kinesiotherap$.tw.
28. exp Rehabilitation/
29. rehab$.tw.
30. (skate$ or skating).tw.
31. run$.tw.
32. jog$.tw.
33. treadmill$.tw.
34. swim$.tw.
35. bicycl$.tw.
36. (cycle$ or cycling).tw.
37. walk$.tw.
38. (row or rows or rowing).tw.
39. muscle strength$.tw.
40. or/10-39
41. randomized controlled trial.pt.
42. controlled clinical trial.pt.
43. randomized.ab.
44. placebo.ab.
45. drug therapy.fs.
46. randomly.ab.
47. trial.ab.
48. groups.ab.

49.41 or 42 or 43 or 44 or 45 or 46 or 47 or 48
50.humans.sh.
51.49 and 50
52.and/5,9,40,51

Appendix 2. EMBASE (Ovid) search strategy

1. exp osteoarthritis/
2. osteoarthr$.tw.
3. (degenerative adj2 arthritis).tw.
4. arthrosis.tw.
5. or/1-4
6. Knee/
7. knee$.tw.
8. 6 or 7
9. exp EXERCISE/
10.fitness/
11.exercise test/
12.exercise tolerance/
13.exp Sport/
14.pliability/
15.exp "physical activity, capacity and performance"/
16.exertion$.tw.
17.exercis$.tw.
18.sport$.tw.
19.((physical or motion) adj5 (fitness or therap$)).tw.
20.(physical$ adj2 endur$).tw.
21.((strength$ or isometric$ or isotonic$ or isokinetic$ or aerobic$ or endurance or weight$) adj5 (exercis$ or train$)).tw.
22.exp physiotherapy/
23.physiotherap$.tw.
24.manipulat$.tw.
25.kinesiotherap$.tw.
26.exp REHABILITATION/
27.rehab$.tw.
28.(skate$ or skating).tw.
29.run$.tw.
30.jog$.tw.
31.treadmill$.tw.
32.swim$.tw.
33.bicycl$.tw.
34.(cycle$ or cycling).tw.
35.walk$.tw.
36.(row or rows or rowing).tw.
37.muscle strength$.tw.
38.or/9-37
39.and/5,8,38
40.random$.ti,ab.
41.factorial$.ti,ab.
42.(crossover$ or cross over$ or cross-over$).ti,ab.
43.placebo$.ti,ab.
44.(doubl$ adj blind$).ti,ab.
45.(singl$ adj blind$).ti,ab.
46.assign$.ti,ab.

47.allocat$.ti,ab.
48.volunteer$.ti,ab.
49.crossover procedure.sh.
50.double blind procedure.sh.
51.randomized controlled trial.sh.
52.single blind procedure.sh.
53.or/40-52
54.exp animal/ or nonhuman/ or exp animal experiment/
55.exp human/
56.54 and 55
57.54 not 56
58.53 not 57
59.39 and 58

Appendix 3. *The Cochrane Library* (Wiley Interscience) search strategy

1. MeSH descriptor Osteoarthritis explode all trees
2. osteoarthr*:ti,ab
3. (degenerative next arthritis):ti,ab
4. arthrosis:ti,ab
5. (#1 OR #2 OR #3 OR #4)
6. MeSH descriptor Knee explode all trees
7. MeSH descriptor Knee Joint explode all trees
8. knee*:ti,ab
9. (#6 OR #7 OR #8)
10. MeSH descriptor Exercise explode all trees
11. MeSH descriptor Exertion explode all trees
12. MeSH descriptor Physical Fitness explode all trees
13. MeSH descriptor Exercise Test explode all trees
14. MeSH descriptor Exercise Tolerance explode all trees
15. MeSH descriptor Sports explode all trees
16. MeSH descriptor Pliability explode all trees
17. MeSH descriptor Physical Endurance explode all trees
18. exertion*:ti,ab
19. exercis*:ti,ab
20. sport*:ti,ab
21. ((physical or motion) near/5 (fitness or therap*)):ti,ab
22. (physical* near/2 endur*):ti,ab
23. ((strength* or isometric* or isotonic* or isokinetic* or aerobic* or endurance or weight*) near/5 (exercis* or train*)):ti,ab
24. MeSH descriptor Physical Therapy Modalities explode all trees
25. (physical next therap*):ti,ab
26. physiotherap*:ti,ab
27. manipulat*:ti,ab
28. kinesiotherap*:ti,ab
29. MeSH descriptor Rehabilitation explode all trees
30. rehab*:ti,ab
31. (skate* or skating):ti,ab
32. run*:ti,ab
33. jog*:ti,ab
34. treadmill*:ti,ab
35. swim*:ti,ab
36. bicycl*:ti,ab
37. (cycle* or cycling):ti,ab

38. walk*:ti,ab
39. (row or rows or rowing):ti,ab
40. muscle next strength:ti,ab
41. (#10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40)
42. (#5 AND #9 AND #41)

Appendix 4. CINAHL (EBSCOhost) search strategy

1. S56 S55 and S42
2. S55 S54 or S53 or S52 or S51 or S50 or S49 or S48 or S47 or S46 or S45 or S44 or S43 S54 TI Allocat* random* or AB Allocat* random*
3. S53 (MH "Quantitative Studies")
4. S52 (MH "Placebos")
5. S51 TI Placebo* or AB Placebo*
6. S50 TI Random* allocat* or AB Random* allocat*
7. S49 (MH "Random Assignment")
8. S48 TI Randomi?ed control* trial* or AB Randomi?ed control* trial*
9. S47 TI singl* mask* or TI doubl* mask* or TI treb* mask* or TI tripl* mask* or AB singl* mask* or AB doubl* mask* or AB treb* mask* or AB tripl* mask*
10. S46 TI singl* blind* or TI doubl* blind* or TI treb* blind* or TI tripl* blind* or AB singl* blind* or AB doubl* blind* or AB treb* blind* or AB tripl* blind*
11. S45 TI "clinic* trial*" or AB "clinic* trial*"
12. S44 PT Clinical Trial
13. S43 (MH "Clinical Trials+")
14. S42 S41 and S40 and S5
15. S41 S39 or S38 or S37 or S36 or S35 or S34 or S33 or S32 or S31 or S30 or S29 or S28 or S27 or S26 or S25 or S24 or S23 or S22 or S21 or S20 or S19 or S18 or S17 or S16 or S15 or S14 or S13 or S12 or S11 or S10 or S9 or S8 or S7 or S6
16. S40 S8 or S7 or S6
17. S39 (ti "muscle strength*") or (ab "muscle strength*")
18. S38 (ti row or rows or rowing) or (ab row or rows or rowing)
19. S37 (ti walk*) or (ab walk*)
20. S36 (ti cycle* or cycling) or (ab cycle* or cycling)
21. S35 (ti bicycl*) or (ab bicycl*)
22. S34 (ti swim*) or (ab swim*)
23. S33 (ti swim*) or (ab swim*)
24. S32 (ti treadmill*) or (ab treadmill*)
25. S31 (ti jog*) or (ab jog*)
26. S30 (ti run*) or (ab run*)
27. S29 (ti skate* or skating) or (ab skate* or skating)
28. S28 (ti rehab*) or (ab rehab*)
29. S27 (MH "Rehabilitation+")
30. S26 (ti kinesiotherap*) or (ab kinesiotherap*)
31. S25 (ti manipulat*) or (ab manipulat*)
32. S24 (ti physiotherap*) or (ab physiotherap*)
33. S23 (MH "Physical Therapy+")
34. S22 TI (strength* or isometric* or isotonic* or isokinetic*or aerobic* or endurance or weight*) or AB (strength* or isometric* or isotonic* or isokinetic*or aerobic* or endurance or weight*)
35. S21 TI physical* n2 endur* or AB physical* n2 endur*
36. S20 TI physical N5 fitness or TI physical N5 therap* or AB physical N5 fitness or AB physical N5 therap* or TI motion n5 therap* or AB motion n5 therap*
37. S19 (ti sport*) or (ab sport*)
38. S18 (ti exercis*) or (ab exercis*)
39. S17 (ti exertion*) or (ab exertion*)
40. S16 (MH "Physical Endurance+")

41.S15 (MH "Pliability
42.S14 (MH "Sports+")
43.S13 (MH "Exercise Tolerance+")
44.S12 (MH "Exercise Test+")
45.S11 (MH "Physical Fitness")
46.S10 (MH "Exertion+")
47.S9 (MH "Exercise+")
48.S8 (ti knee*) or (ab knee*)
49.S7 (MH "Knee Joint
50.S6 (MH "Knee")
51.S5 S4 or S3 or S2 or S1
52.S4 (ti arthrosis) or (ab arthrosis)
53.S3 (ti degenerative N2 arthritis) or (ab degenerative N2 arthritis)
54.S2 (ti osteoarthr*) or (ab osteoarthr*)
55.S1 (MH "Osteoarthritis+")

Appendix 5. PEDro search strategy

1. Advanced search
2. Therapy: Fitness training OR Strength training
3. Body Part: Lower leg or knee

WHAT'S NEW

Date	Event	Description
30 September 2014	New citation required but conclusions have not changed	Methods were updated in accordance with current recommendations of The Cochrane Collaboration: 'Risk of bias' assessment and 'Summary of findings' tables were added Quality of life assessment and study withdrawal rates were added to the update Pain and physical function outcomes were further disaggregated into immediate post-treatment effects and sustainability (2 to 6 months and > 6 months post treatment)
29 October 2013	New search has been performed	Twenty-three new studies were added to this update: Brismée 2007; Hurley 2007; Yip 2007; An 2008; Doi 2008; Jan 2008; Lim 2008; Lund 2008; Jan 2009; Jenkinson 2009; Lee 2009; Lin 2009; Bennell 2010; Bezalel 2010; Salli 2010; Foroughi 2011; Wang 2011; Bruce-Brand 2012; Chang 2012; Kao 2012; Salacinski 2012; Simao 2012; Abbott 2013. One study that was included in the original review was excluded from this update: Petrella 2000

HISTORY

Review first published: Issue 3, 2003

Date	Event	Description
12 May 2009	Amended	Minor amendment; see Published notes
13 August 2008	Amended	CMSG ID A007-R

Date	Event	Description
11 August 2008	New citation required but conclusions have not changed	Substantive amendment
3 June 2008	New search has been performed	This updated review is 1 of 2 Cochrane reviews replacing an earlier review, 'Exercise for osteoarthritis of the hip or knee.' Since the time of the original review, the editors decided to subdivide the review into separate conditions The Background section has been revised to provide information on the specific disorder only, and the search strategy has been revised accordingly. The Methods section has been updated to reflect current methods of the Cochrane Musculoskeletal Review Group A total of 15 new studies were added to this updated review: Gur 2002; Foley 2003; Huang 2003; Quilty 2003; Song 2003; Talbot 2003; Hughes 2004; Keefe 2004; Messier 2004; Bennell 2005; Huang 2005a; Thorstensson 2005; Hay 2006; Mikesky 2006; Fransen 2007
3 June 2008	Amended	Converted to new review format

CONTRIBUTIONS OF AUTHORS

M Fransen, S McConnell, A Harmer, M Van der Esch, M Simic and K Bennell conducted the updated review.
M Fransen is the guarantor of the review.

DECLARATIONS OF INTEREST

None known.

SOURCES OF SUPPORT

Internal sources

- National Health and Medical Research Council, Australia.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Methods described in the review have been updated since the original protocol, in accordance with current recommended methods of the Cochrane Musculoskeletal Review Group and The Cochrane Collaboration: 'Risk of bias' assessment and 'Summary of findings' tables were added. Two outcomes—quality of life assessment and study withdrawal rates—were added to the update. Pain and physical function outcomes were further disaggregated into immediate post-treatment effects and sustainability (three to six months post treatment).

The original protocol was prepared for a review entitled "Exercise for osteoarthritis of the hip or knee." Since the time the original review was published, the editors have decided to subdivide the review into two reviews of separate conditions. For this update of the specific review for hip OA, we have included two additional outcomes: quality of life and study withdrawal rates.

NOTES

The original protocol was prepared for a review entitled "Exercise for osteoarthritis of the hip or knee." Since the time the original review was published, the editors have decided to subdivide the review into two reviews of separate conditions. The current review provides a second update of the review "Exercise for osteoarthritis of the knee."

INDEX TERMS

Medical Subject Headings (MeSH)

*Exercise Therapy; Arthralgia [rehabilitation]; Osteoarthritis, Knee [*rehabilitation]; Randomized Controlled Trials as Topic

MeSH check words

Humans