Exercise for osteoarthritis of the knee: a Cochrane systematic review

Marlene Fransen,¹ Sara McConnell,² Alison R Harmer,¹ Martin Van der Esch,³ Milena Simic,¹ Kim L Bennell⁴

Abstract

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

Methods Five electronic databases were searched, up until May 2013. Randomised clinical trials comparing some form of land-based therapeutic exercise with a non-exercise control were selected. Three teams of two review authors independently extracted data and assessed risk of bias for each study. Standardised mean differences immediately after treatment and 2–6 months after cessation of formal treatment were separately pooled using a random effects model.

Results In total, 54 studies were identified. Overall, 19 (35%) studies reported adequate random sequence generation, allocation concealment and adequately accounted for incomplete outcome data. However, research results may be vulnerable to selection, attrition and detection bias. Pooled results from 44 trials indicated that exercise significantly reduced pain (12 points/100; 95% CI 10 to 15) and improved physical function (10 points/100; 95% CI 8 to 13) to a moderate degree immediately after treatment, while evidence from 13 studies revealed that exercise significantly improved quality of life immediately after treatment with small effect (4 points/100; 95% CI 2 to 5). In addition, 12 studies provided 2-month to 6-month post-treatment sustainability data which showed significantly reduced knee pain (6 points/100; 95% CI 3 to 9) and 10 studies which showed improved physical function (3 points/100; 95% CI 1 to 5).

Conclusions Among people with knee osteoarthritis, land-based therapeutic exercise provides short-term benefit that is sustained for at least 2–6 months after cessation of formal treatment.
Cochrane Central Register of Controlled Trials (CENTRAL), the Cumulative Index to Nursing and Allied Health Literature (CINAHL) and the Physiotherapy Evidence Database (PEDro). Also included was a search of ClinicalTrials.gov (http://www.ClinicalTrials.gov) and the WHO trials portal (http://www.who.int/ictrp/en/).

Three teams of two review authors (MF, SMcC, ARH, MVdE, MS, KLB) 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. Those teams 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, data were extracted from the pain scale that was highest on a list according to a previously described hierarchy of pain-related outcomes.\(^17\)\(^18\) Data on more than one physical function scale, when reported in a trial, were extracted according to a hierarchy format (eg, WOMAC or other functional scale). If data on more than one quality of life scale were reported in a trial, data were extracted according to a hierarchy format (eg, SF-12 or other quality of life scale). Risk of bias was assessed in accordance with methods recommended by The Cochrane Collaboration. Each potential source of bias was graded as high, low or unclear. If random sequence generation, allocation concealment and incomplete outcome data domains were adequately met by a study, the overall risk of bias on pain and physical function was judged as ‘low’ for that study. All other studies were categorised as ‘unclear’ or ‘high’ risk of bias. If participants were stated to be blinded to treatment allocation, the study was considered as low risk for detection bias on pain and physical function.

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

Heterogeneity was assessed in a random-effects model, and overall effects were adjusted to include an estimate of the degree of variation between studies, or heterogeneity, in intervention effect ($\tau^2$).\(^19\) The impact of heterogeneity on meta-analysis results was 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.\(^19\) 30–60% probably represents moderate heterogeneity, and >50% is usually considered as representing substantial heterogeneity. For studies published after 1 July 2005, the Clinical Trials Register at the International Clinical Trials Registry Platform of the WHO (http://apps.who.int/ trialssearch) was screened to obtain the a priori trial protocol. The GRADEpro software and the five GRADE (Grades of Recommendation, Assessment, Development and Evaluation) considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) were used to assess the quality of a body of evidence for stated outcomes.\(^20\)\(^21\) Finally, with sensitivity analysis, the effects of (1) potential selection and attrition bias on immediate post-treatment pain and physical function outcomes and (2) the effect of potential detection bias on immediate post-treatment pain and physical function outcomes were assessed.

Subgroup analyses were conducted to determine the sustainability of treatment effects, as well as to determine if the size of treatment effect was mediated by exercise content, number of face-to-face sessions or the method of treatment delivery.

RESULTS

Study characteristics

Of 212 retrieved RCTs identified by the literature search, 54 met the inclusion criteria.\(^12\)\(^13\) Among these, marked variability 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,\(^23\)\(^25\)\(^29\)\(^31\)\(^33\)\(^35\)\(^39\)\(^53\)\(^52\)\(^59\)\(^61\)\(^64\) where five studies recruited more than 200 participants,\(^22\)\(^45\)\(^48\)\(^70\) one of which recruited 750 participants.

Sample recruitment varied widely, with studies recruiting exclusively community volunteers,\(^37\)\(^44\)\(^49\)\(^53\)\(^60\)\(^69\)\(^72\)\(^74\) patients drawn from specialist rheumatology or orthopaedic clinics,\(^28\)\(^30\)\(^33\)\(^35\)\(^46\)\(^47\)\(^54\)\(^66\)\(^68\)\(^71\)\(^75\) a mix of community volunteers and patients from specialist clinics or patients referred by general practitioners.\(^22\)\(^25\)\(^48\)\(^50\)

A wide range of therapeutic exercise programmes were assessed. Variability was found in delivery mode, type of exercise, and treatment ‘dosage’ (duration, frequency, intensity) and many studies did not provide a clear rationale for their choice. With regard to treatment duration, monitored treatment sessions, presented in individual or class-based format, ranged from 20 to 60 min. Exercise frequency for monitored classes or for individual clinic sessions in most studies was two to three times per week but varied between one to five times per week.\(^23\) Intensity achieved during strength training using free or limb weights or Theraband was commonly a 10-repetition maximum with varying numbers of sets.\(^27\)\(^31\)\(^34\)\(^53\) or was at least moderate.\(^30\)\(^72\)\(^74\) Aerobic exercise training, achieved via walking or cycling programmes, ranged from low to moderate.\(^34\)\(^37\)\(^44\)\(^50\)\(^57\)\(^59\)\(^61\)\(^64\)\(^69\)\(^74\) intensity.

According to the methodological quality assessment, a total of 19 of 54 studies (35%) could be considered as achieving ‘low risk of bias’ from the published report.\(^22\)\(^24\)\(^26\)\(^27\)\(^34\)\(^35\)\(^37\)\(^38\)\(^48\)\(^52\)\(^53\)\(^57\)\(^62\)\(^70\)\(^71\)\(^73\)\(^74\) Only 4 of the 54 included studies claimed blinding of study participants.\(^26\)\(^31\)\(^36\)\(^62\)

Pooled results of 44 studies (see online supplementary table S1) demonstrated statistically significant benefit of exercise on pain immediately post-treatment, with an SMD of 0.49 (95% CI 0.39 to 0.59). This effect size would be considered moderate.\(^76\) and was equivalent to a reduction of 10 points (95% CI 10 to 15 points) on a 0 to 100-point pain scale (0 indicating 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 ($p=0.77$; $I^2=0\%$).

Pooled results of 44 studies (see online supplementary table S2) demonstrated statistically significant benefit of exercise on physical function immediately post-treatment, with an SMD of 0.52 (95% CI 0.39 to 0.64). This effect size would be considered moderate and was equivalent to an improvement of 10 points (95% CI 8 to 13 points) on a 0–100-point scale. Between-study heterogeneity was substantial ($I^2=68\%$). No
significant difference was noted between change and end of treatment scores \((p=0.36; \text{I}^2=0\%)\).

Pooled results of 13 studies (see online supplementary table S3) demonstrated statistically significant benefit of exercise on quality of life immediately post-treatment, with an SMD of 0.28 (95% CI 0.15 to 0.40). This effect size would be considered small\(^{26}\) and was equivalent to an improvement of four points (95% CI 2 to 5 points) on a 0–100-point scale. Between-study heterogeneity was negligible (\(\text{I}^2=0\%\)). No significant difference was noted between change scores and end of treatment scores \((p=0.86; \text{I}^2=0\%)\).

Pooled results from 12 studies\(^{26, 30} 40 44 45 48 55 60 62 69 75 77\) demonstrated a statistically significant benefit (SMD 0.24, 95% CI 0.14 to 0.33) of exercise on pain at 2–6 months postexercise training. This effect size would be considered small and was equivalent to a reduction of six (95% CI 3 to 9) points on a 0–100-point scale. There was no between-study heterogeneity (\(\text{I}^2=0\%\)). No significant difference was noted between change scores and end of treatment scores \((p=0.40; \text{I}^2=0\%)\).

In 10 studies\(^{26, 30} 40 44 45 55 60 62 69 75 77\) pooled results demonstrated a statistically significant benefit (SMD 0.15, 95% CI 0.04 to 0.26) of exercise on physical function at 2–6 months postexercise training. This effect size would be considered small and was equivalent to an improvement of 3 (95% CI 1 to 5) points on a 0–100-point scale. There was no between-study heterogeneity (\(\text{I}^2=0\%\)). No significant difference was noted between change scores and end of treatment scores \((p=0.95; \text{I}^2=0\%)\).

In six studies\(^{22, 42} 43 58 63 70\) a non-significant effect (SMD 0.08, 95% CI –0.15 to 0.30) on pain after more than 6 months was found. The between-study heterogeneity was moderate (\(\text{I}^2=43\%\)). No significant difference was noted between change scores and end of treatment scores \((p=0.95; \text{I}^2=0\%)\). Pooled results of seven studies\(^{22, 42} 43 58 59 63 70\) demonstrated statistically significant benefit (SMD 0.20, 95% CI 0.08 to 0.32) on physical function after more than 6 months. Between-study heterogeneity was absent (\(\text{I}^2=0\%\)). No significant difference was noted between change scores and end of treatment scores \((p=0.95; \text{I}^2=0\%)\).

The magnitude of the immediate treatment effect for both pain and physical function increased with the number of face-to-face contact occasions with the healthcare professional supervising or monitoring the exercise programme. However, the difference between fewer than 12 occasions and 12 or more occasions failed to reach statistical significance for pain \((p=0.15)\) and for physical function \((p=0.09)\), respectively. Pooled analysis demonstrated that each of the treatment delivery modes (ie, individual treatments, class-based programmes, and ‘home’ programmes) provided significant reductions in pain and physical function but with no significant difference between the modes for pain \((p=0.14)\) or physical function \((p=0.06)\). While each type of exercise (quadriceps strengthening only, lower limb strengthening, combination strengthening and aerobic exercise, walking programmes and ‘others’) reduced pain and improved physical function, there was no statistical difference between the exercise training types on pain \((p=0.37)\) and physical function \((p=0.09)\).

Only 11 RCTs specifically reported on adverse events\(^{22, 26} 31 35 36 38 45 47 53 55 73\). 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.

Sensitivity analyses showed that the risk of selection and attrition bias was ‘low’ in 14 studies (1458 participants) for pain and for physical function (14 studies, 1456 participants)\(^{24, 26} 27 34 35 37 38 52-55 71 73 74\). All other studies were categorised as ‘unclear’ or ‘high’ risk of bias (30 studies, 2029 and 2457 participants, respectively). The risk of detection bias for pain and physical function was ‘low’ in three studies with 226 participants\(^{26} 31 36\), indicating that participants were stated to be blinded to treatment allocation. All other included studies were categorised as ‘unclear’ or ‘high’ risk of bias (41 studies, 3261 and 3687 participants, respectively).

DISCUSSION

This systematic review is an update of a previous Cochrane review, published in 2008, which included 32 RCTs. An additional 22 RCTs have been included in this update for a total of 54 trials, providing data from 3362 participants for outcomes on pain and from 3222 participants for outcomes on physical function. Overall, meta-analysis demonstrated that land-based therapeutic exercise programmes resulted in an immediate mean treatment benefit for knee pain, physical function and quality of life. These mean immediate treatment benefits, extracted from 44 RCTs involving 3537 participants for pain and 3913 participants for physical function, would be considered moderate. Treatment benefit for quality of life, extracted from 13 trials involving 1073 participants, would be considered small. The benefit for pain is comparable with reported estimates for current simple analgesics and non-steroidal anti-inflammatory drugs taken for knee pain\(^{27}\). 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. The pain-relieving benefit of exercise declined at 2–6 months postexercise but was still significant, as evidenced in 12 studies involving 1468 participants. However, pain benefits were lost longer than 6 months postexercise, as was found in six studies. A small but significant treatment benefit for physical function remained 2–6 months following exercise, as extracted from 10 studies involving 1279 participants, as well as at time points longer than 6 months, as evidenced in seven studies. These results suggest that although the pain-relieving benefit of exercise therapy is not maintained six or more months after treatment, improvements in physical function are better sustained.

This review highlighted the variety in content of exercise programmes. 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\(^{12, 77}\). 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\(^{34}\). Two other studies compared different strengthening regimens: weight bearing quadriceps exercises versus non-weight bearing quadriceps exercises in one study\(^{47}\), and concentric-eccentric strengthening exercises versus isometric strengthening exercises in the other\(^{48}\). Neither study found significant differences in effect 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. 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 and seemed to be less effective than strengthening and aerobic exercise.

The magnitude of the treatment effect for pain and physical function was influenced by the delivery mode such as the...
number of face-to-face contact occasions with the healthcare professional who was supervising or monitoring the exercise programme. However, unlike in the previous Cochrane review,\(^4\) the difference between fewer than 12 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.

Exercise ‘dosage,’ which is a factor of programme duration, frequency and intensity, varied considerably between the studies included in this review. Even if dosages were similar, uncertainty may still be inherent. Prescribed dosage may not be translated into actual dosage as it is dependent on individual effort (and hence intensity) during each session; and prescribed dosage is also dependent on effort during testing and retesting sessions. The influence of programme duration on 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-intensity and low-intensity resistance training of the knee flexor and extensor muscles while controlling for total exercise workload.\(^4\) That study found no difference between high-intensity and low-intensity strength training in improving clinical effects.

Overall quality of the body of evidence was assessed as high when the GRADE approach was applied for pain and quality of life. Although selection, attrition and detection bias may have resulted in overestimation of the 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).

Treatment effect size for many of the studies was modest. Multifaceted interventions that incorporate exercise strategies into patient care may provide greater benefit and should be tested. Future studies are needed to: (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 multiarmed placebo-controlled RCTs to help provide evidence of optimal exercise content and dosage, and (3) assess the long-term effectiveness of exercise for people with knee OA in terms of structural disease progression.

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\(^78\) have not been controlled for in the exercise studies. Thus it is not possible to determine the exact magnitude of beneficial effects directly arising from exercise per se. 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-matched and gender-matched peers without symptomatic OA, these physiological impairments often are not yet large enough to translate into reportable difficulties on simple questionnaires. Lack of reportable difficulties would considerably reduce the potential range of improvement that was possible (ceiling effect) on self-report questionnaires among people with early or mild disease.

Several limitations of this review have been identified. Although we conducted an extensive literature search, because resources were limited, we extracted data only from studies published in the English language, potentially excluding other important evidence. Four studies were published in a language other than English,\(^79\) and we were unable to source full text for two studies.\(^8\) These studies await classification. However, the possibility of publication bias could not be excluded, 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, which is consistent with previously published systematic reviews.\(^83\)\(^8\) Mediating effects of exercise dosage and disease severity on the effectiveness of exercise could not be ascertained because of large variability in reported data.

Conclusions

High-quality evidence suggests that land-based therapeutic exercise provides benefit in terms of reduced knee pain and improved quality of life and moderate-quality evidence of improved physical function among people with knee OA. Healthcare professionals and people with OA can be reassured that any type of exercise programme that is performed regularly and is closely monitored by healthcare professionals can improve pain, physical function and quality of life related to knee OA in the short term.

What are the new findings?

- High-quality evidence suggests that land-based therapeutic exercise provides benefit in terms of reduced knee pain and improved quality of life and moderate-quality evidence of improved physical function among people with knee osteoarthritis.
- It can be assured that any type of exercise programme that is performed regularly and is closely monitored can improve pain, physical function and quality of life related to knee OA in the short term.
- The magnitude of immediate treatment effects of exercise on pain and physical function increases with the number of face-to-face contact occasions with the healthcare professional.
- Mediating effects of exercise dosage and disease severity on the effectiveness of exercise could not be ascertained because of large variability in reported data.

Competing interests None.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES


What are the new findings?

- High-quality evidence suggests that land-based therapeutic exercise provides benefit in terms of reduced knee pain and improved quality of life and moderate-quality evidence of improved physical function among people with knee osteoarthritis.
- It can be assured that any type of exercise programme that is performed regularly and is closely monitored can improve pain, physical function and quality of life related to knee OA in the short term.
- The magnitude of immediate treatment effects of exercise on pain and physical function increases with the number of face-to-face contact occasions with the healthcare professional.
- Mediating effects of exercise dosage and disease severity on the effectiveness of exercise could not be ascertained because of large variability in reported data.

Competing interests None.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES

Review


