Therapeutic Exercise for People with Osteoarthritis of the Hip or Knee. A Systematic Review

MARLENE FRANSEN, SARA McCONNELL, and MARY BELL

ABSTRACT. ***Objective.*** To determine whether land based therapeutic exercise is beneficial for people with osteoarthritis (OA) of the hip or knee in terms of reduced joint pain, improved physical function, and/or the patient's global assessment of therapeutic effectiveness.

Methods. Five databases were searched for randomized clinical trials. Standardized mean differences (SMD) with their 95% confidence intervals (CI) were calculated for each study and then combined using a fixed effects model.

Results. Only 2 studies, totaling about 100 participants, could potentially provide data on people with hip OA. Fourteen studies provided data on 1633 participants with knee OA. Nine of these studies were considered of high methodological quality. For pain, combining the results revealed a mean moderate beneficial effect (SMD 0.46, 95% CI 0.35, 0.57), while for self-reported physical function a mean small beneficial effect (SMD 0.33, 95% CI 0.23, 0.43) was found. These results appeared to be sensitive to blinding of outcome assessor and choice of control group.

Conclusion. Land based therapeutic exercise was shown to reduce pain and improve physical function for people with OA of the knee. (J Rheumatol 2002;29:1737–45)

Key Indexing Terms:
OSTEOARTHRITIS KNEE HIP
PHYSICAL THERAPY EXERCISE

Osteoarthritis (OA), the most common rheumatic disease, primarily affects the articular cartilage and subchondral bone of a synovial joint and results in joint failure. People with symptomatic OA of the hip or knee will complain of deep, aching pain and will experience increasing difficulty with daily functional activities and then particularly walking, standing up from a chair, stair climbing, and housekeeping[1,2]. Ultimately chronic OA, particularly when involving the lower limbs, is associated with reduced physical fitness and increased risk of cardiovascular comorbidity[3,4].

Currently there is no cure for OA; however, disease related factors, such as impaired muscle function and fitness, are potentially amenable to exercise intervention[5,6]. To date, there have been no large definitive randomized controlled trials to provide conclusive scientific evidence for this prescribed treatment mode in people with OA of the hip or knee. Without strategies designed to limit selection, performance, attrition, and detection bias, treatment effects are often overestimated[7-10]. To date the only "systematic" review of randomized controlled trials, which combined the results of 6 studies investigating the effectiveness of exercise for OA of the hip or knee published up to September 1997, concluded that "the small number of good studies restricts drawing firm conclusions"[11]. However, several randomized controlled trials have been published since September 1997.

The aim of this systematic review of randomized clinical trials was to include the recent publications into the body of knowledge to determine whether land based therapeutic exercise is beneficial for people with OA of either the hip or knee in terms of reduced joint pain and/or improved physical function. In contrast to the previous systematic review[11], the results for people with OA of the hip will be analyzed separately from the results for people with OA of the knee. We also wished to analyze if treatment effectiveness was associated with mode of treatment delivery (individual treatments, group format programs, home programs), type of control group (active intervention versus inactive or waiting list), and study methodological quality.

MATERIALS AND METHODS

Inclusion criteria. To be included in this review, the clinical trial had to be randomized (or quasi-randomised), to compare some form of therapeutic exercise with a non-exercise or sham exercise group, and to be published as a full length article. Participants needed to be adults with an established diagnosis of OA of either the hip or knee according to accepted criteria[12] or self-reporting OA of the hip or knee on the basis of chronic joint pain. Further, to be included, the study needed to provide data on at least one of the following core set of outcome measures[13]: self-reported pain; self-reported physical function; or patient global assessment. Studies investigating presurgery exercise programs or hydrotherapy based programs were not included.

From the Institute for International Health, University of Sydney, Sydney, Australia; and the University of Toronto, Toronto, Ontario, Canada.

Dr. Fransen was supported by the National Health and Medical Research Council of Australia.

M. Fransen, PhD, Institute for International Health, University of Sydney; S. McConnell, MSc; M. Bell, MD, Department of Physical Therapy, University of Toronto.

Address reprint requests to Dr. M. Fransen, Institute for International Health, University of Sydney, PO Box 576, Newtown, NSW 2042, Australia. E-mail: mfransen@iih.usyd.edu.au

Submitted July 31, 2001; revision accepted January 30, 2002.

Search and evaluation strategy. Five databases were searched: the Cochrane Controlled Trials Register, the Cochrane Musculoskeletal Group Trials Register, Medline (January 1966 to March 2001), CINAHL (January 1982 to March 2001), and PEDro (Physiotherapy Evidence Database). The Medical Subject Heading key words used were "osteoarthritis, knee" and "osteoarthritis, hip"; each exploded and coupled with "exercise" or "exercise therapy"; "physical therapy" or "physiotherapy"; and "rehabilitation." In addition, reference lists of retrieved articles and reviews were examined. Only English-language articles were reviewed due to limited resources for translation. Two reviewers (MF, SM) independently screened retrieved clinical studies for inclusion, extracted data from all included studies using standardized forms, and scored methodological quality. Authors were contacted if the data could not be extrapolated in the desired form from the published article. Study methodological quality was scored using the *Instrument to Measure the Likelihood of Bias* (Appendix)[14]. This is a short scale with proven high interrater reliability assessing 3 aspects of methodology: randomization, double blinding and withdrawals/dropouts, and results in a score range from 0 to 5. If agreement was not achieved at any stage, a third reviewer (MB), blinded to the results of the previous reviewers' decisions, adjudicated.

Statistical analysis. The primary outcome measures chosen for this review were continuous level aggregate data on self-reported pain, self-reported physical function, and/or the patient's global assessment. Mean change scores (posttreatment – baseline) were used, as it was anticipated that many of the studies would have relatively small sample sizes with resultant possible differences in baseline outcome scores between the allocation groups. Standardized mean differences (SMD) with their 95% confidence intervals (CI), calculated from the change score and baseline standard deviation, for the effects of exercise intervention above control intervention were estimated for each study. The estimates were combined using a fixed effects model[15] (Metaview 4.1©, 1999, Update Software, Oxford, UK). The combined results were tested for homogeneity using chi-square tests.

RESULTS

Due to insufficient number of studies on people with OA of the hip or studies measuring patient global assessment of therapeutic effectiveness, this review is limited to studies assessing the effectiveness of exercise for people with OA of the knee in terms of self-reported pain and physical function. There were only 4 studies able to provide data on people with OA of the hip[16-19]. However, 2 of these studies[16,19] had to be excluded as hydrotherapy was either the active treatment or the control treatment. This left 2 studies, with a total of roughly 100 participants. Only 3 of the retrieved studies provided data on the patient's global assessment of therapeutic effectiveness[19-21]; one of these studies was later excluded as the participants were all presurgery patients[20].

Included studies. Of the 31 randomized clinical trials identified from the literature search, 16 studies were excluded (Table 1)[16,18,20,22-34]. Fifteen met the inclusion criteria for the current systematic review[17,19,21,35-46]. One study had 2 clearly different exercise intervention groups, aerobic and resistance[37], and was treated as such, with the sample size of the control group being equally divided between the 2 exercise intervention groups to avoid double counting.

Eight of the 15 studies used either the Arthritis Impact Measurement scale (AIMS)[39,40,42] or the Western Ontario and McMaster Universities Arthritis Index (WOMAC)[21,36,38,41,43] to score self-reported pain and physical function. In the other studies, measures of pain ranged from a crude one or 2 item score[19,35] to a more detailed 6 item score[37]. All studies included an immediate posttreatment assessment of outcomes. Eight studies reported data from a followup assessment[17,21,36-38,40,41,44], with one study only reporting one-year followup data[46]. Due to the variety in followup assessment timing (one month to one year), only immediate posttreatment results have been assessed. This decision excluded one more study from the review[46], leaving 14 studies.

Several attempts were made to contact 6 authors for additional data. Three responded, with 2 able to provide the requested results stratified for location of OA (hip or knee)[17,19], and one able to provide WOMAC scores disaggregated for pain and physical function[36]. No contact could be established with the other 3 authors. Therefore for one study a misprint assumption was made on one "impossible" standard error of the mean score[35]. For another study, 2 baseline standard deviations needed to be extrapolated from a similar size study using the same self-report questionnaires[21]. For the third study, the posttreatment results for the control group were used as the baseline for the active treatment groups (2-group analysis)[37].

The 14 included studies provided data on 936 participants allocated to therapeutic exercise and 697 participants allocated to a control group.

Participants (Table 2). There was variability between the studies in participants. Sample recruitment varied, some studies recruiting community volunteers exclusively[17,37,41,42], or a mix of community volunteers and specialist clinic patients[35,40,43], or general physician referrals[19] and physiotherapy waiting lists[36,38]. About 50% of the sample in one study reported symptom duration of less than one year, while other studies reported mean symptom duration of more than 10 years[21,40]. Inclusion criteria varied from only "knee pain in the past week"[41] to full American College of Rheumatology criteria with at least Kellgren-Lawrence grade III radiographic disease[44]. Exclusion criteria were also inconsistent between studies. For example, one study excluded people taking nonsteroidal antiinflammatory drugs (NSAID)[35], whereas another study only included people currently taking NSAID at least twice a week[39].

Interventions (Table 2). The therapeutic exercise interventions we evaluated included programs provided individually to each participant[19,21,36,38,45], group format programs[17,35,37-40,42,44], and home programs[41,43]. Treatment content varied from unilateral quadriceps muscle strengthening[21] and aerobic walking programs[37,39,40] to very comprehensive programs including manual therapy, upper limb and/or truncal muscle strengthening, and balance coordination[19,36,42,44], in addition to the more usual lower limb muscle strengthening. The duration of each exercise session ranged from 30 minutes[19,21,35] to 90 minutes[39]. Total treatment duration ranged from one month[36] to 3 months[35,37,40,42,44].

Table 1. Excluded studies.

Study	Reason for Exclusion
Borjesson, 1996	Exclusively presurgery patients
Callaghan, 1995	No baseline scores, meaningless median % improvements
Chamberlain, 1995	No appropriate control group
Green, 1993	No appropriate control group. No report on selected outcomes
Hurley, 1998	Not even quasi-randomised
Kreindler, 1989	Only outcome measure is muscle strength
Jan, 1991	Not even quasi-randomised
Lankhorst, 1982	No control group in analysis and no report on selected outcomes
Mangione, 1999	No appropriate control group
Messier, 1997	Gait assessment (same sample as Ettinger, 1997)
Messier, 2000 (1)	No appropriate control group
Messier, 2000 (2)	Balance assessment (same sample as Ettinger, 1997)
Peterson, 1993	Gait assessment (same sample as Kovar, 1992)
Quirk, 1985	No appropriate control group
Rejeski, 1998	Stair time/health perceptions (same sample as Ettinger, 1997)
Sullivan, 1998	1 year followup only (same sample as Kovar, 1992)
Sylvester, 1989	No appropriate control group

Appropriate control: non-exercise group. Selected outcomes: self-reported pain/physical function/patient global assessment.

Table 2. Included studies, treatment and methodology.

Study Sample	Treatment Mode/ Content	Control	Outcome Assessment	Analysis
Bautch, 1997 34 patients/volunteers	Individual/ROM, walking	Education	Unblind	Efficacy
Deyle, 2000 83 patients	Individual/Manual, resistance, ROM	US	Blinded	Efficacy
Ettinger, 1997 293 volunteers	Group/ROM, walking	Education	Blinded	Intention-to-treat
Ettinger, 1997 295 volunteers	Group/Resistance	Education	Blinded	Intention-to-treat
Fransen, 2001 126 patients	Individual or group/resistance, aerobic	Waiting list	Unblind	Intention-to-treat
Hopman-Rock, 2000 91 volunteers	Group/resistance	No intervention	Blinded	Efficacy
Kovar, 1992 103 patients	Group/walking, resistance	Telephone calls	Unblind	Efficacy
Maurer, 1999 113 patients	Individual/unilateral quadriceps	Education	Blind	Efficacy
Minor, 1989 80 patients/ volunteers	Group/walking, aquatics	ROM and relaxation	Blinded	Efficacy
O'Reilly, 1999 180 volunteers	Home/resistance, lifestyle advice	Lifestyle advice	Unblind	Intention-to-treat
Peloquin, 1999 137 volunteers	Group/resistance, aerobic	Education	Blinded	Efficacy
Petrella, 2000 179 patients/volunteers	Home/resistance, ROM, NSAID	ROM and NSAID	Blinded	Intention-to-treat
Rogind, 1998 25 patients	Group/complex mix	No intervention	Blinded	Efficacy
Schilke, 1996 20 patients	Individual/resistance	No intervention	Unblind	Efficacy
Van Baar, 1998 200 patients	Individual/physiotherapy, education	GP education	Blinded	Intention-to-treat

Volunteers: community sample, ROM: range of motion exercises, US: ultrasound, SWD: short wave diathermy, NSAID: nonsteroidal antiinflammatory drugs, GP: general practitioner.

Methodological quality (Table 3). Only one of the 14 studies attained the maximum score of 5 on the *Instrument to Measure the Likelihood of Bias*[43], while a further 8 studies attained a score of 3 out of 5. Examining other methodological criteria, 8 of the 14 studies reported that the assessor was blinded to group allocation[17,19,21,36,37,40,42-44], but only 4 of the 14 studies provided results as per intention-to-treat analysis[19,37,38,41].

Study power. Most of the studies were clearly underpowered for even a moderate treatment effect. Three of the 14 studies were very small, comparing groups of less than 20 participants[35,44,45]. Only 6 of the 14 studies, with a total of 1162 participants, had at least 80% power to detect a moderate treatment effect (0.5) at a significance level of 0.05[19,37,38,41-43]. Furthermore these 6 studies were also rated as having acceptable study methodology (Table 3).

Self-reported pain (Table 4). The pain outcome measure for one study[43] was not included in this systematic review as all participants were required to take daily NSAID, unfairly attenuating any pain-relieving benefit attributable to the exercise program. Combining the results of the other 13 studies gave a mean effect size for land based therapeutic exercise over control of 0.46 (95% CI 0.35, 0.57). This effect size would be rated as moderate[47]. The test for heterogeneity was insignificant (chi-square = 14.9, p = 0.32). Combining the results only of the 6 studies with sufficient power to detect a moderate treatment effect gave a comparable mean effect size of 0.45 (95% CI 0.32, 0.58), again with an insignificant test for heterogeneity (chi-square = 2.5, p = 0.78).

Self-reported physical function (Table 5). For self-reported physical function in people with OA of the knee, combining the results of the 14 studies gave a mean effect for land based therapeutic exercise over control of 0.33 (95% CI 0.23, 0.43). This effect size would be considered small[47]. However, the test for heterogeneity was significant (chi-square = 27.1, p = 0.02). Combining the results of only those 6 studies with sufficient power to detect a moderate treatment effect, the mean effect was a comparable 0.31 (95% CI 0.19, 0.43), with an insignificant test for heterogeneity (chi-square = 1.7, p = 0.95).

Sensitivity analysis (Table 6). Study methodology: For both self-reported pain and self-reported physical function, studies with a high methodological quality score (score 3+) appeared to have a larger, but statistically insignificant, mean effect size compared with studies arguably more vulnerable to effect size bias (score 0–2).

Treatment mode: One study[38] randomly allocated participants to 2 different treatment delivery modes, individual treatments or group format program. As the participants in this study were patients referred for treatment, the control group was randomly allocated to one of the active interventions after 8 week assessment, accounting for the increased number of participants in the delivery mode sensitivity analysis. When studies were grouped by delivery mode, for both self-reported pain and physical function there was no clear difference in mean effect size between participants allocated to individual treatments compared with participants allocated to group format programs. There was only one study assessing a home program in terms of self-reported pain[41], and only 2 studies assessing home programs in terms of self-reported physical function[41,43].

Control group: For both self-reported pain and physical function, studies that allocated the control group to attend education classes showed a considerably smaller mean effect compared with studies that allocated the control group to telephone calls, electrophysical agents, or no change in routine treatment.

DISCUSSION

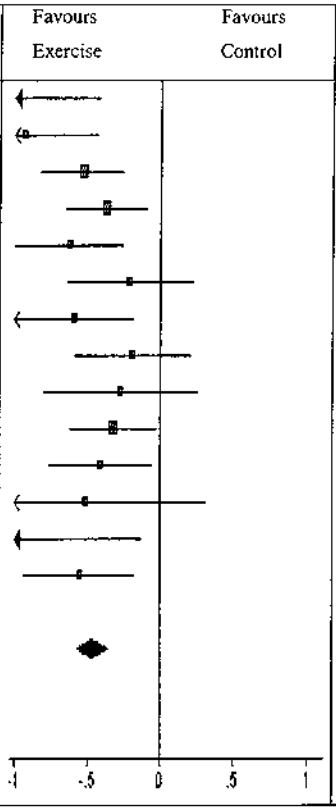
This systematic review, with 14 randomized controlled trials and a total of 1633 participants, demonstrates that land

Table 3. Instrument to Measure the Likelihood of Bias (Appendix).

Study	Randomization	Double Blinding	Withdrawals/ Dropouts	Add/Deduct Point	Score (0–5)
Bautch, 1997	Yes	No	No	No/No	1
Deyle, 2000	Yes	No	Yes	Yes/No	3
Ettinger, 1997	Yes	No	Yes	Yes/No	3
Fransen, 2001	Yes	No	Yes	Yes/No	3
Hopman-Rock, 2000	Yes	No	Yes	No/No	2
Kovar, 1992	Yes	No	Yes	Yes/No	3
Maurer, 1999	Yes	No	No	Yes/No	2
Minor, 1989	Yes	No	Yes	No/No	2
O'Reilly, 1999	Yes	No	Yes	Yes/No	3
Peloquin, 1999	Yes	No	Yes	Yes/No	3
Petrella, 2000	Yes	Yes	Yes	Yes/Yes	5
Rogind, 1998	Yes	No	Yes	Yes/No	3
Schilke, 1996	Yes	No	No	Yes/No	2
Van Baar, 1998	Yes	No	Yes	Yes/No	3

Table 4. Self-reported pain: standardized mean difference (SMD) and 95% confidence intervals (95% CI).

Study	Exercise (n)	Control (n)	SMD (95% CI)
Bautch, 1997	15	15	-1.20 (-0.41, -1.98)
Deyle, 2000	33	36	-0.93 (-0.43, -1.43)
Ettinger (A), 1997	144	75	-0.53 (-0.24, -0.81)
Ettinger (R), 1997	146	75	-0.36 (-0.08, -0.64)
Fransen, 2001	83	43	-0.62 (-0.24, -0.99)
Hopman-Rock, 2000	45	37	-0.20 (0.23, -0.64)
Kovar, 1992	47	45	-0.59 (-0.17, -1.01)
Maurer, 1999	49	49	-0.19 (0.21, -0.58)
Minor, 1989	49	19	-0.27 (0.27, -0.80)
O'Reilly, 1999	108	72	-0.32 (-0.02, -0.62)
Peloquin, 1999	59	65	-0.40 (-0.04, -0.76)
Rogind, 1998	11	12	-0.50 (0.33, -1.34)
Schilke, 1996	10	10	-1.06 (-0.11, -2.01)
Van Baar, 1998	54	59	-0.55 (-0.17, -0.92)
Overall	853	612	-0.46 (-0.35, -0.57)

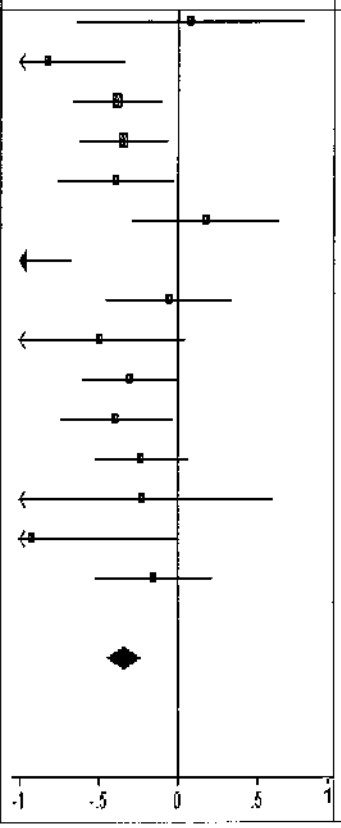


based therapeutic exercise has at least immediate moderate benefit in terms of reducing lower limb pain and immediate small effect in terms of improved physical function for people with symptomatic OA of the knee. There were insufficient studies with similar followup assessments in terms of time to evaluate treatment carry-over. The only previous quantitative systematic review[11] provided data on about 850 participants from 6 studies, but in contrast to the current review, included a study using presurgery patients[20] and did not disaggregate results for participants with OA of the hip and OA of the knee[19]. Our systematic review was able to synthesize data from an additional 8 randomized controlled trials. Further, while the previous systematic review was only able to identify 2 randomized trials with adequate methodology and sufficient power, the current systematic review was able to identify 6 such studies.

The *Instrument to Measure the Likelihood of Bias*[14] was chosen to evaluate methodological quality, as it is a simple instrument with a high interrater reliability, and evaluation is mostly confined to factors (adequate randomization methods and blinding of the assessor to subject allocation) considered most influential in restricting effect size biases[8]. However, the current metaanalysis did not reveal significant effects of study methodological quality as scored by this instrument. It is probable that the effect of study methodological quality was attenuated by the observed difference between the studies in participants recruited, type of active and control intervention used, and treatment intensity and duration (Table 2). On the other hand, this instrument may not be sufficiently sensitive for evaluating studies of physical interventions. Only one study achieved a methodology score of 5 out of 5[43]. This scoring was based on the assumption that the "sham program" provided to the control group (nonprogressive joint unloading and stretches) was not recognized as such by the participants. For the 8 studies with a score of 3 out of 5 (Table 3), the universal reason for not achieving the maximum score of 5 on this scale was a negative response to the question, "Was the study described as double-blind?" A negative response immediately results in a loss of 2 of the available 5 points. With the unavoidable difficulty (due to the nature of the intervention) of masking either the participant or the therapist to group allocation, it would seem essential to provide blinded outcomes assessment in these studies. Nine of the 14 studies reported blinded outcomes assessment[17,19,21,36,37,40,42-44]. For self-reported pain, studies using blinded outcome assessment appeared to show a smaller mean effect (0.43, 95% CI 0.30, 0.55) compared with studies reporting uncertain or unblinded outcome assessment (0.54, 95% CI 0.35, 0.73). For physical function, the difference in effect size, although still not reaching statistical significance, was more evident, with the mean effect size for studies reporting blinded

Table 5. Self-reported physical function: standardized mean difference (SMD) and 95% confidence intervals (95% CI).

Study	Exercise (n)	Control (n)	Favours Exercise	Favours Control	SMD (95% CI)
Bautch, 1997	15	15			0.08 (0.80, -0.63)
Doyle, 2000	33	36			-0.82 (-0.32, -1.31)
Ettinger, 1997	144	75			-0.37 (-0.09, -0.66)
Ettinger, 1997	146	75			-0.33 (-0.05, -0.61)
Fransen, 2001	83	43			-0.39 (-0.01, -0.76)
Hopman-Rock, 2000	37	34			0.18 (0.65, -0.28)
Kovar, 1992	47	45			-1.10 (-0.66, -1.54)
Maurer, 1999	49	49			-0.05 (0.35, -0.44)
Minor, 1989	49	19			-0.48 (0.05, -1.02)
O'Reilly, 1999	108	72			-0.29 (0.01, -0.59)
Peloquin, 1999	59	65			-0.38 (-0.02, -0.74)
Petrella, 2000	91	88			-0.22 (0.07, -0.52)
Rogind, 1998	11	12			-0.22 (0.60, -1.04)
Schilke, 1996	10	10			-0.91 (0.02, -1.84)
Van Baar, 1998	54	59			-0.14 (0.23, -0.51)
Overall	936	697			-0.33 (-0.23, -0.43)



outcome assessment markedly smaller (0.28, 95% CI 0.16, 0.40) compared with studies with unblinded outcome assessment (0.47, 95% CI 0.28, 0.67). It would appear, therefore, that the *Instrument to Measure the Likelihood of Bias* may not allow sufficient methodological differentiation among studies investigating physical interventions, as no credit is given to studies at least blinding outcomes assessment.

Unfortunately, only 4 studies[19,37,38,41] used the more rigorous "per intention-to-treat" or the results of all participants as per randomized allocation data analysis method. Unexpectedly, for self-reported pain, "per intention-to-treat" analysis resulted in a mean effect size of 0.46 (95% CI 0.31, 0.60), while efficacy analysis, or only the results of treatment completers, resulted in a comparable mean effect size of 0.47 (95% CI 0.30, 0.63). Similarly for physical function, an intention-to-treat analysis resulted in a mean effect size of 0.30 (95% CI 0.17, 0.42), while efficacy analysis resulted in a comparable mean effect size of 0.36 (95% CI 0.21, 0.50). Again the effect of data analysis method may have been attenuated by the demonstrated clinical heterogeneity between the studies.

Many studies in this systematic review included an alternative active treatment, such as education classes, electrophysical agents, or telephone monitoring calls, in an attempt to provide similar attention to all study participants. However, education classes have proven efficacy for people with OA[48] and may also have led to contamination of the control group through uptake of the usually recommended lower limb exercises. When a control intervention has treatment efficacy, the comparative effect size of the active intervention will be smaller than would be the case in a study using a truly inactive control. The smaller mean effect size observed in this systematic review for the 4 studies using education classes controls[21,35,37,42] would suggest education classes were effective in terms of reducing self-reported pain and improving physical function. However, physical interventions such as exercise, education classes, dietary regimens, hydrotherapy, and modified footwear each address a different risk factor for OA. The issue is not whether one physical intervention modality can replace another, but rather, how clinically effective each can prove to be. Therefore, if the research question is whether exercise is an effective treatment for OA of the hip or knee, the esti-

Table 6. Sensitivity analysis: standardized mean differences (SMD) and 95% confidence intervals (95% CI); chi-square test of heterogeneity.

	Treatment, n	Control, n	SMD (95%CI)	Chi-square (p)
Methodological quality (Table 3)				
Pain				
Jadad score 3+	685	482	–0.49 (–0.37, –0.61)	6.06 (0.64)
Jadad score 0–2	168	130	–0.35 (–0.12, –0.59)	7.18 (0.10)
Physical function				
Jadad score 3+	776	570	–0.38 (–0.27, –0.49)	16.73 (0.05)
Jadad score 0–2	123	93	–0.12 (0.12, –0.35)	6.57 (0.16)
Treatment mode (Table 2)				
Pain				
Individual treatments	208	176	–0.52 (–0.31, –0.73)	6.65 (0.16)
Group program	575	364	–0.46 (–0.32, –0.60)	7.13 (0.52)
Physical function				
Individual treatments	208	176	–0.31 (–0.10, –0.51)	8.15 (0.09)
Group program	567	361	–0.38 (–0.24, –0.51)	18.01 (0.02)
Home program	199	160	–0.26 (–0.05, –0.47)	0.10 (0.75)
Control group (Table 2)				
Pain				
Educational class control	413	279	–0.43 (–0.27, –0.58)	5.83 (0.21)
Other control	440	333	–0.49 (–0.34, –0.64)	8.70 (0.37)
Physical function				
Education class control	413	279	–0.29 (–0.14, –0.45)	3.15 (0.53)
Other control	523	418	–0.37 (–0.23, –0.50)	23.53 (0.01)

mates of mean effect sizes including studies using control allocations providing effective interventions may be unduly conservative.

Apart from the above discussion on the influence of study methodology and choice of control group intervention on effect sizes, some comments about general limitations noted in these studies may help direct future research.

First, the provision of exercise as a therapeutic intervention is very operator-dependent for both individual treatments and group format programs. The use of only one or a markedly restricted number of treatment providers together with insufficient sample sizes to detect even a moderate treatment effect, and the common use of efficacy analysis instead of intention-to-treat analysis in most studies must have serious implications for the generalizability of the results, and should be 3 aspects of study design worth serious consideration in the future.

Second, some of the larger studies in this systematic review had mostly participants with early or mild symptomatic disease[19,37,41]. Although people with early disease frequently produce reduced physical performance measures, such as muscle strength and aerobic capacity, these physiological impairments often are not sufficient yet to translate into reportable difficulties on simple questionnaires on the performance of daily functional activities. This ceiling effect would considerably reduce the responsiveness of self-report physical function measures in people with early or mild disease and attenuate the real benefit of treatment, which would include increasing an individual's physiological reserve and thereby reducing vulnerability to future physical disability. Including objective measures of physical performance, such as muscle strength, aerobic capacity, or gait variables, not only strengthens the methodological quality of a study where masking of the participant to treatment allocation is unattainable, but also potentially provides data better able to discriminate between people with early disease, where disease related impairments have not yet developed into self-reported functional limitations or disability.

Third, only one study compared different treatment delivery modes within a clinical trial[38] and only one study attempted to compare treatment content (aerobic or resistance program)[37]. Unfortunately, lack of statistical power to compare 2 active treatments led to inconclusive results in both studies. Further, synthesis of the results in our systematic review could not establish a significant difference in the mean effect on self-reported pain or physical function between studies assessing individual treatments compared with those assessing group format programs. It could, however, be argued that the group format potentially provides a cost-effective alternative, and could be more regularly accessed by older people when introduced to community centers or gymnasiums, and that the social contact with peers, particularly those experiencing similar disease related symptoms, is highly likely to encourage treatment attendance and adherence. Longterm adherence to exercise is required to maintain the benefits of improved lower limb muscle function and fitness

achieved by formal therapeutic exercise sessions. Longterm adherence, however, usually requires the stimulus of regular supervision or monitoring. Unfortunately, most individuals or healthcare systems do not have sufficient resources to allow ongoing unrestricted access to individually provided treatments for chronic musculoskeletal conditions. There were insufficient studies to provide our review with sufficient power to establish a comparative analysis of treatment effectiveness on the basis of treatment delivery mode or program content. Hopefully, increasing interest in nonpharmacological interventions with potential to compress morbidity in the aging population will result in resources becoming available for large, thorough, randomized trials with sufficient power to investigate optimal treatment delivery mode or exercise program content.

APPENDIX

Instrument to Measure the Likelihood of Bias[14]

1. Was the study described as randomised (this includes the use of words such as randomly, random and randomisation)?
2. Was the study described as double blind?
3. Was there a description of withdrawals and dropouts?

Scoring the items:

Either give a score of 1 for each "yes" or 0 points for each "no".

There are no in-between marks.

Give 1 additional point if:

For question 1, the method to generate the sequence of randomisation was described and it was appropriate (random numbers table, computer generated, etc.) AND/OR

For question 2, the method of double blinding was described and it was appropriate (identical placebo, active placebo, dummy, etc.).

Deduct 1 point if:

For question 1, the method to generate the sequence of randomisation was described and it was inappropriate (patients were allocated alternately, or according to date of birth, hospital number, etc.) AND/OR

For question 2, the study was described as double blind but the method of blinding was inappropriate (comparison of tablets vs injection with no double dummy).

REFERENCES

1. Davis MA, Ettinger WH, Neuhaus JM, Mallon KP. Knee osteoarthritis and physical functioning: evidence from the NHANES 1 epidemiologic followup study. J Rheumatol 1991;18:591-8.
2. Guccione AA. Arthritis and the process of disablement. Phys Ther 1994;74:408-14.
3. Minor MA, Hewett JE, Webel RR, Dreisinger TE, Kay DR. Exercise tolerance and disease related measures in patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 1988;15:905-11.
4. Philbin EF, Ries MD, Groff G, Sheesley KA, French TS, Pearson TA. Osteoarthritis as a determinant of an adverse coronary heart disease risk profile. J Cardiovasc Risk 1996;3:529-33.
5. Buchner DM, Beresford SA, Larson EB, LaCroix AZ, Wagner EH. Effects of physical activity on health status in older adults II: Intervention studies. Annu Rev Pub Health 1992;13:469-88.
6. Fiatarone MA, Evans WJ. The etiology and reversibility of muscle dysfunction in the aged. J Gerontol 1993;48:77-83.
7. Guyatt GH, Sackett DL, Cook DJ. Users' guide to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993;270:2598-601.
8. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.
9. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998;317:1185-90.
10. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.
11. van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee. Arthritis Rheum 1999;42:1361-9.
12. Altman RD. Criteria for the classification of clinical osteoarthritis. J Rheumatol 1991;18 Suppl 27:10-2.
13. Bellamy N, Kirwan J, Boers M, 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. J Rheumatol 1997;24:799-802.
14. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1-12.
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.
16. Green J, McKenna F, Redfern EJ, Chamberlain MA. Home exercises are as effective as outpatient hydrotherapy for osteoarthritis of the hip. Br J Rheumatol 1993;32:812-5.
17. Hopman-Rock M, Westhoff M. The effects of a health educational and exercise program for older adults with osteoarthritis of the hip or knee. J Rheumatol 2000;27:1947-54.
18. Sylvester KL. Investigation of the effect of hydrotherapy in the treatment of osteoarthritis hips. Clin Rehabil 1989;4:223-8.
19. van Baar ME, Dekker J, Oostendorp RA, et al. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. J Rheumatol 1998;25:2432-9.
20. Borjesson M, Robertson E, Weidenhielm L, Mattsson E, Olsson E. Physiotherapy in knee osteoarthrosis: effect on pain and walking. Physiother Res Int 1996;1:89-97.
21. Maurer BT, Stern AG, Kinossian B, Cook KD, Schumacher HR. Osteoarthritis of the knee: isokinetic quadriceps exercise versus an educational intervention. Arch Phys Med Rehabil 1999;80:1293-9.
22. Callaghan MJ, Oldham JA, Hunt J. An evaluation of exercise regimes for patients with osteoarthritis of the knee: a single blind randomized controlled trial. Clin Rehabil 1995;9:213-8.
23. Chamberlain MA, Care G, Harfield B. Physiotherapy in osteoarthrosis of the knees. A controlled trial of hospital versus home exercises. Int Rehabil Med 1982;4:101-6.
24. Hurley MV, Scott DL. Improvements in quadriceps sensorimotor function and disability of patients with knee osteoarthritis following a clinically practicable exercise regime. Br J Rheumatol 1998;37:1181-7.
25. Jan M-H, Lai J-S. The effects of physiotherapy on osteoarthritic knees of females. J Formosan Med Assoc 1991;90:1008-13.
26. Kreindler H, Lewis CB, Rush S, Schaefer K. Effects of three exercise protocols on strength of persons with osteoarthritis of the knee. Top Geriatr Rehabil 1989;4:32-9.
27. Lankhorst GJ, van der Stadt RJ, van der Korst JK, Hinlopen-Bonrath FM, Griffioen FM, de Boer W. Relationship of isometric

knee extension torque and functional variables in osteoarthrosis of the knee. Scand J Rehab Med 1982;14:7-10.
28. 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. J Gerontol 1999;54A:M184-90.
29. Messier SP, Thompson CD, Ettinger WH. Effects of long-term aerobic or weight training regimens on gait in an older, osteoarthritic population. J Appl Biomech 1997;13:205-25.
30. Messier SP, Loesser RE, Mitchell MN, et al. Exercise and weight loss in obese older adults with knee osteoarthritis: a preliminary study. J Am Geriatr Soc 2000;48:1062-72.
31. 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). J Am Geriatr Soc 2000;48:131-8.
32. Peterson MG, Kovar-Toledano PA, Otis JC, et al. Effect of a walking programme on gait characteristics in patients with osteoarthritis. Arthritis Care Res 1993;6:11-6.
33. 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.
34. Rejeski WJ, Ettinger WH, Martin K, Morgan T. Treating disability in knee osteoarthritis with exercise therapy: a central role for self-effiicacy and pain. Arthritis Care Res 1998;11:94-101.
35. Bautch J, Malone D, Vailas A. Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. Arthritis Care Res 1997;10:48-55.
36. Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB, Allison SC. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. A randomized, controlled trial. Ann Intern Med 2000;132:173-81.
37. Ettinger WH, Burns R, Messier SP, 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.
38. Fransen M, Crosbie J, Edmonds J. Physical therapy is effective for patients with osteoarthritis of the knee: a randomized controlled clinical trial. J Rheumatol 2001;28:156-64.
39. Kovar PA, Allegrante JP, MacKenzie CR, Peterson MG, Gutin B, Charlson ME. Supervised fitness walking in patients with osteoarthritis of the knee. A randomized, controlled trial. Ann Intern Med 1992;116:529-34.
40. Minor MA, Hewett JE, Webel RR, Anderson SK, Kay DR. Efficacy of physical conditioning exercise in patients with rheumatoid arthritis and osteoarthritis. Arthritis Rheum 1989;32:1396-405.
41. O'Reilly SC, Muir KR, Doherty M. Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomised controlled trial. Ann Rheum Dis 1999;58:15-9.
42. 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. J Clin Rheumatol 1999;5:126-36.
43. Petrella RJ, Bartha C. Home based exercise therapy for older patients with knee osteoarthritis: a randomized clinical trial. J Rheumatol 2000;27:2215-21.
44. 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. Arch Phys Med Rehabil 1998;79:1421-7.
45. 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. Nurs Res 1996;45:68-72.
46. 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 Res 1998;11:228-33.
47. Cohen J. Statistical power analysis for the behavioural sciences. New York: Academic Press; 1977.
48. Lorig KR, Mazonson PD, Holman HR. Evidence suggesting that health education for self-management in patients with chronic arthritis has sustained health benefits while reducing health care costs. Arthritis Rheum 1993;36:439-46.