ARTHRITIS & RHEUMATISM
Vol. 42, No. 7, July 1999, pp 1361–1369

EFFECTIVENESS OF EXERCISE THERAPY IN PATIENTS WITH OSTEOARTHRITIS OF THE HIP OR KNEE

A Systematic Review of Randomized Clinical Trials

MARGRIET E. VAN BAAR, WILLEM J. J. ASSENDELFT, JOOST DEKKER, ROB A. B. OOSTENDORP, and JOHANNES W. J. BIJLSMA

Objective. **To review the effectiveness of exercise therapy in patients with osteoarthritis (OA) of the hip or knee.**

Methods. **A computerized literature search of Medline, Embase, and Cinahl was carried out. Randomized clinical trials on exercise therapy for OA of the hip or knee were selected if treatment had been randomly allocated and if pain, self-reported disability, observed disability, or patient's global assessment of effect had been used as outcome measures. The validity of trials was systematically assessed by independent reviewers. Effect sizes and power estimates were calculated. A best evidence synthesis was conducted, weighting the studies with respect to their validity and power.**

Results. **Six of the 11 assessed trials satisfied at least 50% of the validity criteria. Two trials had sufficient power to detect medium-sized effects. Effect sizes indicated small-to-moderate beneficial effects of exercise therapy on pain, small beneficial effects on both disability outcome measures, and moderate-to-great beneficial effects according to patient's global assessment of effect.**

Conclusion. **There is evidence of beneficial effects of exercise therapy in patients with OA of the hip or knee. However, the small number of good studies restricts drawing firm conclusions.**

Osteoarthritis (OA) is a relatively common musculoskeletal disorder. A population-based incidence study from the US showed incident rates of 0.5/1,000 per year for hip OA (presented for medical consultation) and 2/1,000 per year for knee OA (1). Prevalence increases with age (2,3). In recent treatment guidelines for OA of the hip and knee (4,5), exercise therapy is considered to be an important nonpharmacologic treatment approach. The goal of exercise therapy in OA patients is to reduce pain and disability. In order to achieve this, exercise therapy aims at the improvement of muscle strength, stability of joints, range of motion, and aerobic fitness. These functions are frequently impaired in patients with OA, presumably contributing to pain and disability (6). Improving these functions is assumed to result in a reduction of pain and disability. In addition, exercise therapy aims directly at reduction of disability, e.g., through corrections of the walking pattern (7).

Since the publication of the treatment guidelines mentioned above, several new randomized clinical trials (RCTs) on exercise therapy in OA patients have been published (8–12). These newly published RCTs were not included in previous reviews (7,13–19). Furthermore, the methodology for reviews of the literature has evolved considerably (20–22). Current methodology requires several explicit and systematic steps to be made in conducting reviews of the literature. These steps are a systematic search of the literature (23), selection of studies based on explicit inclusion and exclusion criteria (24), assessment of methodologic quality (25), and a

Supported by a grant from the Dutch Fund of Investigative Medicine of the Dutch Health Insurance Council (OG92.066).

Margriet E. van Baar, PhD: Netherlands Institute of Primary Health Care, Utrecht, The Netherlands; Willem J. J. Assendelft, MD, PhD: Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam, The Netherlands; Joost Dekker, PhD: Netherlands Institute of Primary Health Care, Utrecht, and Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam, The Netherlands; Rob A. B. Oostendorp, PhD, PT: Dutch National Institute for Allied Health Professions, Amersfoort, and Free University of Brussels, Brussels, Belgium; Johannes W. J. Bijlsma, MD, PhD: University Hospital, Utrecht, The Netherlands.

Address reprint requests to Joost Dekker, PhD, Netherlands Institute of Primary Health Care, PO Box 1568, 3500 BN Utrecht, The Netherlands.

Submitted for publication April 22, 1998; accepted in revised form February 23, 1999.

systematic approach toward data extraction and data analysis (26). Neither the treatment guidelines nor the previously published reviews satisfy the current methodologic requirements for literature reviews. Therefore, considerable improvement can be made by summarizing the evidence available on exercise therapy in OA of the hip or knee by 1) including newly published RCTs, and 2) applying presently required review methodology.

Our objective was to determine the effectiveness of exercise therapy in patients with OA of the knee or hip, based on a systematic review of the evidence from RCTs. We focused on the effects of exercise therapy on pain, self-reported disability, observed disability, and patient's global assessment of effect. These outcome measures are recommended for RCTs that involve patients with OA of the hip and knee (27).

MATERIALS AND METHODS

Literature search. A comprehensive computer-aided search of the Medline (January 1966 to September 1997), Embase (January 1988 to September 1997), and Cinahl (January 1982 to September 1997) databases was carried out by 1 reviewer (WJJA). A highly sensitive search strategy for RCTs (28) and systematic reviews (29) was used. The Cochrane Controlled Trial Register (30) was also searched. References of relevant review articles and trials were screened.

Criteria for studies considered for inclusion. Trial reports that met the following criteria were eligible. 1) The trial concerned patients with OA of the knee or hip, and this was assessed using either clinical or radiologic criteria (or a combination) for OA. 2) Treatment had been allocated using a random procedure (31). 3) At least 1 of the treatments had included exercise therapy. Exercise therapy is defined as a range of activities intended to improve strength, range of motion, endurance, balance, coordination, posture, motor function, or motor development. Exercise therapy can be performed actively, passively, or against resistance (32). No restrictions were made as to type of supervision and group size. Additional interventions were allowed. 4) At least 1 of the following outcome measures had been included: pain, self-reported disability, observed disability, and patient's global assessment of effect. 5) Results had been published as a full report.

Trial reports were excluded if 1) they concerned perioperative exercise therapy, or 2) intervention groups received identical exercise therapy and therefore no contrast existed between the intervention groups. No restrictions were made concerning the language of publication (33).

Assessment of methodologic quality. A list of specific criteria for the methodologic quality assessment was used, consisting of internal validity criteria, descriptive criteria, and statistical criteria (see Table 1). The internal validity criteria refer to requirements for the design and conduct of intervention research. The descriptive and statistical criteria refer to the external validity of a study. This list of criteria, known as the Maastricht-Amsterdam consensus list (22), is an adaptation of a list that has already been used in a great number of systematic reviews and also in the field of physical therapy (24). It includes all of the criteria of Schultz et al (31), Jadad et al (34), and Verhagen et al (35). For the present review, we elaborated on some of the criteria (V9, D1, D2, and D6 in Table 1) for application to OA and exercise therapy (for more details, see the Netherlands Institute of Primary Health Care web site at http://www.nivel.nl).

The information about each criterion from the study reports was analyzed. If sufficient information was given, the design and conduct of the study were assessed. If bias was unlikely, the criterion was rated positive. If bias was likely, the criterion was rated negative. In the case of insufficient or missing information, the criterion was rated inconclusive ("don't know"). A total score for internal validity was calculated by summing up the number of positive criteria. Equal weights were applied, resulting in an internal validity score with a range of 0–12.

The methodologic quality of the study reports was assessed by 2 reviewers (MEvB and WJJA) independently. One trial report that was written by 1 of the reviewers (MEvB) was assessed by the other initial reviewer (WJJA) and by another uninvolved assessor (D. A. W. M. van der Windt). In the case of disagreement, a consensus method was used to discuss and resolve the disagreement between the reviewers. Another reviewer (JD) was available to arbitrate in persisting disagreement.

Data extraction and presentation. Quantitative data were extracted by 1 reviewer (MEvB). Effect sizes (ES) and their variances were computed (36). For differences in continuous outcome measurements, Hedge's g statistic was computed; for differences in proportions, Cohen's h was computed. Computation of Hedge's g required calculation of the mean and SD for each group. In the absence of these data, the ES was calculated from Z scores and sample sizes (36). If possible, the ES was based on change scores (posttreatment minus pretreatment). In the absence of change scores, posttreatment scores were used. The ES was interpreted as described by Cohen (37); i.e., an ES of 0.2 was considered a small beneficial effect, 0.5 a medium effect, and 0.8 a large effect of exercise therapy. Power estimates for an ES of 0.2 and 0.5 were made, to study whether trials could detect an existing difference between interventions using a significance level of 0.05 (37).

Best evidence synthesis. Studies were weighted as to 1) their validity, and 2) their power level. Studies that satisfied at least 50% (n = 6) of the validity criteria were classified as studies with "acceptable validity," versus studies with "low validity." In addition, studies with a sufficient power of at least 0.80 (based on an ES of 0.5) were distinguished from studies with a lower power. A power of 0.80 is generally considered sufficient to detect medium significant differences between interventions (37). Conclusions were mainly based on studies that had both acceptable validity and sufficient power.

RESULTS

Selection of the studies. We initially identified 19 publications concerning 17 trials that met our inclusion criteria. Seven studies were excluded from the review; 4

Table 1. Criteria for the methodologic assessment of randomized clinical trials*

Criteria type, number	Description of criteria
Validity criteria	
V1	Randomization: adequate procedure for generation of a random number list
V2	Randomization: concealed random allocation of treatments, by an independent person not responsible for determining eligibility of patients
V3	Cointerventions: control for cointerventions in design
V4	Cointerventions: reported for each group
V5	Adherence to interventions: >70% in index group, and also in reference groups in placebo-controlled trials
V6	Care provider blinded
V7	Patient blinded
V8	Outcome assessment blinded
V9	Relevant outcome measures: ≥2 of the following outcome measures: pain, disability, and patient global assessment
V10	Withdrawals and dropouts: ≤20% for short-term followup and ≤30% for long-term followup, and no substantial bias (inequality between groups; reason for withdrawal/dropout)
V11	Identical timing of outcome assessment for all intervention groups
V12	Intention-to-treat analysis
Descriptive criteria	
D1	Specification of eligibility criteria, including explicit classification criteria for OA; both an established set of criteria (i.e., American College of Rheumatology) and clinical criteria, including symptoms, were up to standard
D2	Baseline similarity regarding age, radiologic OA, duration of disease, location of OA, and baseline main outcome measure(s)
D3	Description of interventions: adequate description of type, modality, application technique, intensity, duration, and number (or frequency) of sessions for both the index intervention and reference groups
D4	Adverse effects described and attributed to allocated treatment, or explicit report of no adverse effects
D5	Short-term followup: outcome assessment at the end of the intervention period
D6	Long-term followup: outcome assessment ≥6 months after randomization
Statistical criteria	
S1	Sample size: to be presented at randomization and for most important outcome assessments
S2	Presentation of point estimates and distribution measures, for each important outcome measure separately

* For details on the operationalization of criteria and assessment forms, see http://www.nivel.nl. OA = osteoarthritis.

of these concerned perioperative exercise therapy (38–41) and 3 did not have a contrast for exercise therapy between the intervention groups (42–44). Consequently, 12 publications concerning 10 trials were included in our systematic review (8–12,45–51). The information was combined for 2 trials that were reported twice in the literature (ref. 11 combined with ref. 49, and ref. 47 with ref. 48). In addition, 1 report that had been accepted for publication was included (52).

Methodologic quality of the studies. There was initial disagreement between the 2 independent reviewers on 42 (21%) of the 200 (10 × 20) items scored. Disagreement mainly concerned the following criteria: control for cointerventions in trial design (V3), baseline similarity (D2), adequate description of interventions (D3), and intention-to-treat analysis (V12). Nearly all disagreements were due to reading errors or a difference in interpretation of the methodologic criteria. After the consensus meeting, no disagreement persisted.

Table 2 presents, for each trial, the criteria for which bias was considered likely or for which incomplete information hampered the methodologic assessment.

Table 2. Methodologic assessment of trials of exercise therapy for osteoarthritis of the knee or hip, ranked in order of validity score*

First author (reference)	Validity score†	Bias considered likely	Incomplete information for validity assessment	Incomplete information for description and data extraction
Van Baar (52)	9	V6,7	V2	D6
Ettinger (11), Messier (49)	8	V3,4,6,7		D2,4
Callaghan (10)	7	V4,6,7	V2,5	D1,4,6; S2
Börjesson (9)	6	V6,7	V2,4,5,8	D4,6
Minor (50)	6	V3,6,7,12	V2,8	D2,4,6; S1
Sylvester (51)	6	V4,5,6,7	V2,3	D1,3,4,6
Kovar (47), Peterson (48)	5	V4,6,7,8	V2,3,12	D3,6
Schilke (12)	5	V3,4,6,7	V2,5,8	D1,2,3,4,6
Bautch (8)	4	V4,6,7	V2,3,5,8,12	D2,4,6
Chamberlain (45)	4	V4,6,7,10,12	V2,3,5	D2,6; S2
Jan (46)	0	V1,2,4,6,7,9,11	V3,5,8,10,12	D1,3,4,6; S1

* Equally ranked trials are ordered alphabetically. See Table 1 for definition of criteria numbers. V = validity; D = descriptive; S = statistical.
† The validity score is calculated as the sum of all items with bias unlikely. Each item is given equal weight (range 0–12). Incomplete information for the validity assessment is considered as "bias likely," thus having a score of 0.

Validity criteria. Six trials satisfied at least 6 of the 12 validity criteria (50% of the criteria) (9–11,50–52). One trial did not satisfy any of the validity criteria (46).

As a consequence of the nature of exercise therapy, neither care providers nor patients can be blinded to the exercise therapy. Thus, the criteria blinding of care providers (V6) and blinding of patients (V7) were not met in any of the trials studied. The most prevalent shortcomings concerned cointerventions: the design of 3 trials did not control for cointerventions concerning physical therapy strategies or medication (V3), and in 8 trials, there was no report of these cointerventions for each group (V4). In 2 trials, bias was likely due to the absence of an intention-to-treat analysis (V12).

Many trials lacked sufficient information on several validity criteria: concealment of treatment allocation (V2), the level of compliance (V5), control for cointerventions in the design (V3), and blinding of outcome assessment (V8).

Informativeness of the study. Information on adverse effects of exercise therapy (D4) and long-term (≥6 months after randomization) outcome assessment (D6) was often missing in trial reports. In 2 trial reports (50,52), long-term followup was mentioned, but no results were presented. Other frequent deficiencies were in reporting on specification of eligibility criteria (D1) and description of the interventions (D3).

Sample size and power. The sample size and power of the trials varied widely (Table 3). Five trials compared groups of <25 patients, while 2 trials compared ≥100 patients (median group size 34). Two studies (11,52) were designed with sufficient power (≥0.80)

Table 3. Power calculations for all included trials*

First author (reference)	Validity score	Mean number of patients per group	Power with effect size of 0.2	Power with effect size of 0.5
Van Baar (52)	9	100	0.29	0.94
Ettinger (11), Messier (49)	8	146	0.40	0.99
Callaghan (10)	7	9	0.07	0.16
Börjesson (9)	6	34	0.13	0.53
Minor (50)†	6	38	0.14	0.57
Sylvester (51)	6	7	0.07	0.14
Kovar (47), Peterson (48)	5	51	0.17	0.71
Schilke (12)	5	10	0.07	0.18
Bautch (8)	4	17	0.09	0.29
Chamberlain (45)	4	21	0.10	0.35
Jan (46)	0	47	0.16	0.67

* Based on a *t* test for differences between the means of 2 independent samples of equal size and equal variances (see ref. 37); a power ≥0.80 is generally considered sufficient.
† Data extraction was confined to data from osteoarthritis patients.

Table 4. Best evidence synthesis*

First author (reference)	Pain	Self-reported disability	Observed disability in walking	Patient's global assessment of effect
Acceptable validity, sufficient power				
Van Baar (52)	0.58 (0.54, 0.62)	0.26 (0.22,0.30)	0.28 (0.24, 0.32)	0.64 (0.60, 0.68)
Ettinger (11; aerobic exercise)	0.47 (0.44, 0.50)	0.41 (0.38, 0.44)	0.89 (0.85, 0.93)	NM
Ettinger (11; resistance exercise)	0.31 (0.28, 0.34)	0.36 (0.33, 0.39)	0.31 (0.28, 0.34)	NM
Acceptable validity, low power				
Callaghan (10)	NA	NM	NA	NM
Börjesson (9)	0.20 (0.08, 0.32)	NM	−0.11 (−0.17, −0.05)	1.40 (1.28, 1.52)
Low validity, low power				
Kovar (47)	0.52 (0.43, 0.61)	0.88 (0.78, 0.98)	0.92 (0.82, 1.02)	NM
Schilke (12)	0.07 (−0.32, 0.46)	NA	NA	NM
Bautch (8)	0.25 (−0.01, 0.51)	−0.65 (−0.93, −0.37)	NM	NM
Jan (46)	NM	1.01 (0.91, 1.10)	NM	NM

* Values are the effect sizes (95% confidence intervals) for each outcome measure.
NM = not measured; NA = not able to calculate effect sizes due to insufficient data presentation.

to detect medium-sized effects (ES of 0.5). Both of these studies had acceptable validity (≥50% criteria positive). Two studies (46,47) were designed with a nearly sufficient power (0.67 and 0.71, respectively) to detect medium-sized effects; these latter studies had a low validity score.

Effectiveness of exercise therapy in comparison with placebo treatment or no treatment. The majority of the trials included in this review were designed to study the differences between exercise therapy and placebo treatment or no treatment (8–12,47,52; see also http://www.nivel.nl). One of these trials also aimed to study differences between different exercise therapy interventions (10). In Table 4, the ES and the 95% confidence intervals (95% CI) are presented for all studied outcome parameters.

Pain. Pain was used as an outcome measure in 7 trials. In these trials, 4 different outcome measures were used to assess pain. No information was available on timing of pain assessment in relation to the days of exercise. In 1 trial (10), data presentation was insufficient to calculate the ES. One trial (11) included 2 comparisons between exercise therapy interventions (aerobic exercise and resistance exercise) and a placebo treatment. Therefore, these ES values are presented separately in Table 4.

In the 2 trials with acceptable validity and sufficient power, the lower limit of the 95% CI exceeded 0.2 (small effect) (11,52), and in 1 of the 2 trials, the lower limit exceeded 0.5 (medium effect) (52). These trials, however, differed in terms of participants and content of intervention. The participants had either hip or knee OA (52) or knee OA only (11). In both trials, radiographic evidence indicated a mild-to-moderate stage of disease. In both trials, patients were recruited through physicians, and in 1 trial (11), this was supplemented with community-based recruitment. The intervention in 1 trial (52) concerned supervised individual therapy, including strengthening exercises, range of motion exercises, and functional training. The other trial (11) concerned supervised group therapy followed by a home-based program. Exercises included aerobic exercises or resistance exercises (11). In both trials, the supervised parts of the interventions took 12 weeks to complete.

The only trial with acceptable validity, but low power, was borderline significant (9). This study concerned patients with knee OA who had radiographic evidence and symptoms (both not specified) and were recruited from a clinical setting. Two 4-week exercise programs were compared: individual weight-bearing exercises and supervised group therapy consisting of non–weight-bearing exercises.

Among the 3 trials with low validity and low power (8,12,47), the lower limit of the 95% CI exceeded 0.2 in 1 of them (47). This study concerned patients with knee OA for a mean duration of >10 years, and participants were recruited from the community and the clinic. The intervention concerned an 8-week supervised group therapy that mainly consisted of "fitness walking" (47). The other studies concerned patients with knee OA according to the criteria of the American College of Rheumatology who were recruited from both the community and the clinic (8), and patients with knee OA (not specified) who were recruited in the clinic (12). The exercise interventions consisted of a 12-week walking program (8) or an 8-week strength training program monitored on a dynamometer (12).

Thus, the evidence indicates a small-to-moderate

beneficial effect of exercise therapy on pain in knee OA and, to a lesser extent, in hip OA. This effect was found in participants with minimal-to-moderate OA who were recruited from both the community and the clinic and were being treated with various types of exercise therapy.

Self-reported disability. Self-reported disability was used as an outcome measure in 6 trials. Three different outcome measures were used. In 1 trial (12), data presentation was insufficient to calculate the ES. In 2 trials with acceptable validity and sufficient power (11,52), the lower limits of the 95% CI exceeded an ES of 0.2, indicating a small effect.

Among the 3 trials with low validity and low power (8,46,47), the 95% CI indicates a large effect in 2 of the 3 trials (46,47). The 95% CI of the third trial (8) included an extreme value of −0.9. This is probably a biased estimate, due to the forced use of posttreatment scores in combination with a significant baseline difference for this outcome parameter.

It can be concluded that there is evidence for a small beneficial effect of exercise therapy on self-reported disability. This evidence is based on participants with knee OA and, to a lesser extent, those with hip OA. This effect was found in participants with minimal-to-moderate OA who were recruited from both the community and the clinic and were being treated with various types of exercise therapy.

Walking. Walking, the most frequently used outcome parameter for observed disability, was assessed in 6 trials. In these trials, 4 different assessments were used. In 2 trials (10,12), data presentation was insufficient to calculate the ES.

In the 2 trials with acceptable validity and sufficient power (11,52), the lower limit of the 95% CI exceeded 0.2, indicating a small beneficial effect of exercise therapy on walking performance. In the Ettinger et al trial (11), the 95% CI for aerobic exercise exceeded 0.8. The 95% CI of the trial that had acceptable validity, but low power, ranged from −0.17 to −0.05 (9). Again, this is probably a biased estimate, due to the forced use of posttreatment scores in combination with a significant baseline difference for walking. The trial with a low validity score and low power resulted in a 95% CI exceeding 0.8 (47).

In conclusion, the evidence indicates a small beneficial effect of exercise therapy on walking performance.

Patient's global assessment of effect. In only 2 trials, a global assessment of effect by the patient was used as the outcome parameter. In the trial with acceptable validity and sufficient power (52), the lower limit of the 95% CI was 0.6. In the trial with acceptable validity and low power (9), the lower limit of the 95% CI was 1.28.

These data indicate a medium-to-great beneficial effect of exercise therapy according to the patient's global assessment.

Comparison between different exercise therapy programs. Four trials (10,45,50,51) explicitly studied the differences between different exercise therapy interventions (for details, see http://www.nivel.nl).

Pain was assessed in all 4 trials. Three outcome measures were used. In 2 studies, information was given concerning the timing of pain assessment in relation to the days of exercise. In 1 study (45), outcome assessment preceded treatment, while in another study (51), pain was assessed the week following the completion of treatment. Self-reported disability was assessed in 3 trials (45,50,51), and walking in 2 trials (10,50).

Three trials (10,50,51) had an acceptable validity score; however, power was insufficient (0.80) in all trials. The ES could be calculated for 2 studies, 1 with acceptable validity (3 outcome measures) (50) and 1 with low validity (2 outcome measures) (45). All but 1 of the calculated 95% CI included 0. The exception was the 95% CI for pain in the low-validity study (45), which exceeded 0.2 with its lower limit, indicating a small beneficial effect on pain in favor of home exercises. In this study, participants with knee OA were recruited from the clinic, and a 4-week supervised hospital-based exercise regimen was compared with a 4-week home-based exercise regimen. The other study for which the ES could be calculated concerned both OA and rheumatoid arthritis in the hip, knee, and tarsal joints with specified criteria (50). Subjects were recruited from the community and outpatient clinics. Data extraction for this review was confined to OA patients. Three exercise interventions were compared: an aerobic walking program, aerobic hydrotherapy, and a nonaerobic program directed to range of motion. The calculated 95% CI of this study included 0.

In conclusion, no evidence is available in favor of one particular type of exercise therapy program.

DISCUSSION

In this systematic review, we have summarized the available evidence on the effectiveness of exercise therapy in OA of the hip or knee. We assessed the methodologic quality and the power of 11 RCTs. It can be concluded that exercise therapy is effective in patients with OA of the hip or knee. Available evidence indicates beneficial effects on all studied outcome parameters:

pain, self-reported disability, observed disability in walking, and patient's global assessment of effect. The ES values indicated small effects on both disability outcome measures, a small-to-moderate effect on pain, and a moderate-to-great effect according to the patient's global assessment of effect. Since pain and disability are the main symptoms in patients with OA, exercise therapy seems indicated. However, the size of the effects is modest and needs to be enlarged.

Some critical remarks have to be made. These conclusions are based on a small number of good studies. Only 2 RCTs had an acceptable validity score as well as sufficient power (11,52). Another 2 studies had an acceptable validity score, but low power (9,10). Furthermore, trials frequently did not include all relevant outcome measures as required in our criteria. Therefore, for some outcome measures, evidence is based on a limited number of studies, especially with regard to observed disability (i.e., walking) and patient's global assessment of effect. In addition, a number of different instruments have been used for assessment of specific outcome measures. This complicates the comparison of ES, because of possible underlying differences in validity, reliability, and responsiveness. The recently published list of candidate instruments provided by Bellamy (53) can be seen as a first step in the accomplishment of standardization of assessment.

Moreover, hardly any information is available on long-term effects of exercise therapy. In only 2 publications describing the same trial, long-term effects were reported, and beneficial effects were reported for pain and disability (11,49). However, in this trial, exercise therapy was continued to some extent during the entire followup period. Therefore, no insight was gained into duration of effects after completing exercise therapy. This lack of information concerning long-term effects is a remarkable omission, since the clinical impression is that effects disappear over time.

Finally, the effectiveness of exercise therapy in patients with hip OA has hardly been studied. Patients with hip OA were included in only 1 trial (52). In that trial, both hip and knee OA patients were studied. Therefore, there is only limited insight into the effectiveness of exercise therapy in OA of the hip.

There is insufficient evidence to draw conclusions on the optimal content of an exercise therapy intervention. The 2 trials with acceptable validity and sufficient power showed beneficial effects of different types of exercise therapy: aerobic exercises, resistance exercises, or a mixture of several types of exercise therapy (11,52). The ES of different exercise therapy interventions were comparable. Trials comparing effects of different exercise therapy programs remained inconclusive (10,45, 50,51).

The methodologic assessment revealed some major threats to the validity of clinical trials concerning exercise therapy. Blinding of providers and patients was absent in all studies. As a consequence of the nature of exercise therapy, blinding of both providers and patients is not possible. Therefore, blinding of outcome measurement is vital. However, in only half of the trial reports, blinded outcome assessment was explicitly reported. Another potential source of bias was the frequently occurring absence of information on adherence to the intervention. This hampers the interpretation of a study with negative results. It remains unclear whether the exercise therapy intervention was ineffective due to the intervention itself or due to the participants' failure to adhere to the therapy.

We tried to satisfy the current requirements for systematic review (20–22). We included a methodologic quality assessment to elucidate sources of bias in included trials. There is a multiplicity of lists available for methodologic quality assessment, with a variable number of items (varying between 3 and 35). The choice of a specific instrument has been found to affect the outcome of assessment (54). We have used the Maastricht-Amsterdam consensus list on quality assessment (22), a comprehensive list including all items from formal validated lists (31,34,35). The added value of this list is that it addresses a great variety of items considered relevant in rehabilitation and physical therapy research. In addition, we studied the power of the included trials, i.e., whether a trial could detect an existing difference between interventions. Methodologic quality and power were used to weight the level of evidence of a study (55).

Outcome assessment was focused on effects on pain, self-reported disability, observed disability, and patient's global assessment of effect. Within the category of observed disability, we chose to focus on walking, because data on walking were the most frequently presented. Trials were included if at least 1 relevant outcome measure was assessed. In addition, as part of the methodologic quality assessment, trials were graded according to the number of relevant outcome measures used, with the criterion of at least 2 relevant outcome measures. No further criteria were applied, although studies varied considerably in the number of outcome measures. The quality of outcome assessment instruments was not assessed. Our criteria should be viewed as the minimum criteria for outcome measures necessary

to significantly contribute to our knowledge of effectiveness of exercise therapy in OA.

We examined studies for their control for cointerventions concerning physical therapy strategies and medication. Recent research, however, suggests that control for health education and social interaction should have been included as well (56).

We included the results from our own study to provide an overview of all available evidence. This study was reviewed by an independent assessor who was experienced in reviewing musculoskeletal trials; similar review procedures were used. Exclusion of our study, however, would not have changed the conclusions of our review with regard to the effectiveness of exercise therapy.

To enable direct comparison between trials, we calculated the ES for the same outcome measures in different trials. However, in our calculations, we were hampered by insufficient data presentation in trial reports. First, in several trials, we had to use posttreatment data as a basis for ES calculations, instead of the preferred change scores and their standard deviations. ES based on posttreatment scores proved to be less adequate, especially in trials with small sample sizes. In these trials, treatment groups sometimes differed at baseline for an outcome measure. As a consequence, posttreatment outcomes were not informative, as were (pooled) ES based on these data. Second, in 3 trials, no ES could be calculated.

In conclusion, the available evidence indicates beneficial short-term effects of exercise therapy in patients with OA of the knee and, to a lesser extent (less evidence available), in those with OA of the hip. Given the limited number of studies available, this conclusion applies to patients with mild-to-moderate OA who are recruited in outpatient settings and the community. Beneficial effects have been found for various types of exercise therapy. Exercise therapy may be recommended for patients with OA of the knee and also for patients with OA of the hip with a mild-to-moderate stage of disease.

Further research could expand on this recommendation. In particular, additional clinical trials are needed to study the long-term effectiveness of exercise therapy and effectiveness of exercise therapy in patients with hip OA. In the design and conduct of these trials, specific attention should be paid to a sufficient sample size, adherence to exercise therapy, controls for cointerventions, blinded outcome assessment, and an adequate data analysis including an intention-to-treat analysis. The incorporation of a standard set of outcome measures (27) in combination with the adoption of a standard for reporting results (57) will greatly enhance evidence synthesis in this area.

ACKNOWLEDGMENT

We wish to express our gratitude to D. A. W. M. van der Windt, PhD, for her contribution to the methodologic assessment.

REFERENCES

1. Wilson MG, Michet CJ, Illstrup DM, Melton LJ. Idiopathic symptomatic osteoarthritis of the hip and knee; a population based incidence study. Mayo Clin Proc 1990;65:1214–21.
2. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly: the Framingham Osteoarthritis Study. Arthritis Rheum 1987;30: 914–8.
3. Miedema H. Reuma-onderzoek meerdere echelons (ROME): basisrapport. Leiden (The Netherlands): NIPG-TNO; 1994.
4. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. Arthritis Rheum 1995;38:1535–40.
5. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995;38:1541–6.
6. Dekker J, Boot B, van der Woude L, Bijlsma JWJ. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. J Behav Med 1992;15:189–214.
7. Dekker J, Mulder PH, Bijlsma JWJ, Oostendorp RAB. Exercise therapy in patients with rheumatoid arthritis and osteoarthritis: a review. Adv Behav Res Ther 1993;15:211–38.
8. Bautch JC, Malone DG, Vailas AC. Effects of exercise on knee joints with osteoarthritis: a pilot study of biologic markers. Arthritis Care Res 1997;10:48–55.
9. Börjesson M, Robertson E, Weidenhielm L, Mattson E, Olsson E. Physiotherapy in knee osteoarthrosis: effect on pain and walking. Physiother Res Int 1996;1:89–97.
10. Callaghan MJ, Oldham JA, Hunt J. An evaluation of exercise regimes for patients with osteoarthritis. Clin Rehabil 1995;9: 213–8.
11. 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. JAMA 1997;277:25–31.
12. 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. Nurs Res 1996;45:68–72.
13. Semble EL, Loeser RF, Wise CM. Therapeutic exercise for rheumatoid arthritis and osteoarthritis. Semin Arthritis Rheum 1990;20:32–40.
14. Minor MA. Physical activity and management of arthritis. Ann Behav Med 1991;13:117–24.
15. Marks R. Quadriceps strength training for osteoarthritis of the knee: a literature review and analysis. Physiotherapy 1993;79:13–8.
16. Puett DW, Griffin MR. Published trials of non-medical and non-invasive therapies for hip and knee osteoarthritis. Ann Intern Med 1994;121:133–40.
17. Ytterberg SR, Mahowald ML, Krug HE. Exercise for arthritis. Baillieres Clin Rheumatol 1994;8:161–89.
18. La Mantia K, Marks R. The efficacy of aerobic exercises for treating osteoarthritis of the knee. N Z J Physiother 1995;23–30.
19. Hoving JL, van der Heijden GJMG. Fysiotherapie bij heup-

klachten: systematische review van klinisch effectonderzoek. Ned Tijdschr Fysiother 1997;107:2–9.
20. Mulrow CD, Oxman AD. Cochrane collaboration handbook. Oxford: Update Software; 1997.
21. Oxman AD, Cook DJ, Guyatt GH. Users' guide to the medical literature. VI. How to use an overview: evidence-based medicine working group. JAMA 1994;272:1367–71.
22. Van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane collaboration back review group for spinal disorders. Spine 1997;22:2323–30.
23. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997;127: 380–7.
24. Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. Ann Intern Med 1997;127:531–7.
25. De Vet HCW, de Bie RA, van der Heijden GJMG, Verhagen AP, Sijpkes P, Knipschild PG. Systematic reviews on the basis of methodological criteria. Physiotherapy 1997;83:284–9.
26. Lau J, Ioannidis JPA, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820–6.
27. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures in future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. J Rheumatol 1997;24: 799–802.
28. Greenhalgh T. How to read a paper—Medline database. BMJ 1997;315:180–3.
29. Hunt DL, McKibbon KA. Locating and appraising systematic reviews. Ann Intern Med 1997;126:532–8.
30. Cochrane Collaboration, The Cochrane Library. Cochrane controlled trial register. Oxford: Update Software; 1997.
31. Schultz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA 1994;272: 125–8.
32. American Physical Therapy Association. Guide to physical therapist practice. Phys Ther 1997;77:1163–650.
33. Gregoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias? J Clin Epidemiol 1995;48:158–63.
34. Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17: 1–12.
35. Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM, et al. The Delphi List: a criteria list for quality assessment of randomized clinical trials developed by Delphi consensus. J Clin Epidemiol 1998;51:1235–41.
36. Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges HV, editors. The handbook of research synthesis. New York: Sage Foundation; 1994. p. 231–44.
37. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
38. Kumar PJ, McPherson EJ, Dorr LD, Wan Z, Baldwin K. Rehabilitation after total knee arthroplasty: a comparison of 2 rehabilitation techniques. Clin Orthop 1996;331:93–101.
39. D'Lima DD, Colwell CW Jr, Morris BA, Hardwick ME, Kozin F. The effect of preoperative exercise on total knee replacement outcomes. Clin Orthop 1996;326:174–82.
40. Patterson AJ, Murphy NM, Nugent AM, Finlay OE, Nicholls DP, Boreham CA, et al. The effect of minimal exercise on fitness in elderly women after hip surgery. Ulster Med J 1995;64:118–25.
41. Wijgman AJ, Dekkers GH, Waltje E, Krekels T, Arens HJ. Geen positief effect van preoperatieve oefentherapie en instructie bij patienten die heupartroplastiek zullen ondergaan. Ned Tijdschr Geneeskd 1994;138:949–52.
42. 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.
43. Grigor'eva VD, Suzdal'nitskii DV, Strel'tsova EN, Nikolaeva TG. Vliianie krio- i krioelektroterapii na regional'nuiu gemodinamiku u bol'nykh koksartrozom. Vopr Kurortol Fizioter Lech Fiz Kult 1992;49–54.
44. Singer F, Schieler K. Vergleich der Kurzwellendiathermie, pulsierender hoch -und niederfrequenter elektro-magnetischer Energie und Heilgymnastik bei der Behandlung der Coxarthrose. Z Phys Med 1977:172–7.
45. Chamberlain MA, Care G, Harfield B. Physiotherapy in osteoarthritis of the knees. Int J Rehabil Med 1982;4:101–6.
46. Jan MH, Lai JS. The effects of physiotherapy on osteoarthritic knees of females. J Formos Med Assoc 1991;90:1008–13.
47. Kovar PA, Allegrante JP, MacKenzie R, Peterson MGE, Gutin B, Charlson ME. Supervised fitness walking in patients with osteoarthritis of the knee. Ann Intern Med 1992;116:529–34.
48. Peterson MGE, Kovar-Toledano PA, Otis JC, Allegrante JP, Mackenzie CR, Gutin B, et al. Effect of a walking program on gait characteristics in patients with osteoarthritis. Arthritis Care Res 1993;6:11–6.
49. 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.
50. 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.
51. Sylvester KL. Pilot study: investigation of the effect of hydrotherapy in the treatment of osteoarthritic hips. Clin Rehabil 1989;4: 223–8.
52. 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 hip or knee: a randomised clinical trial. J Rheumatol 1998;25:2432–9.
53. Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. J Rheumatol 1997;24:768–78.
54. Moher D, Jadad JR, Tugwell P. Assessing the quality of randomized controlled trials—current issues and future directions. Int J Technol Assess Health Care 1996;12:195–208.
55. Cook, DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 1992;102 Suppl:305–11.
56. Mazzuca SA, Brandt KD, Katz BP, Chambers M, Byrd D, Hanna M. Effects of self-care education on the health status of inner-city patients with osteoarthritis of the knee. Arthritis Rheum 1997;40: 1466–74.
57. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized clinical trials: the CONSORT statement. JAMA 1996;276:637–9.