National Institute for Health and Care Excellence

Final

Osteoarthritis in over 16s: diagnosis and management

[K] Evidence review for the clinical and cost-effectiveness of treatment packages for the management of osteoarthritis

NICE guideline NG226

Evidence reviews underpinning recommendation 1.3.4 in the NICE guideline

October 2022

Final

Disclaimer

Copyright

ISBN: 978-1-4731-4740-9

Contents

1 Treatment packages

1.1 Review question

What is the clinical and cost-effectiveness of treatment packages (that include combinations of interventions) for the management of osteoarthritis?

1.1.1 Introduction

The management of osteoarthritis involves multiple approaches, for example, exercise, weight control and approaches to reduce pain and improve function. Osteoarthritis is a chronic pain condition, and this can have negative impacts on mental health. The limitation in function can also perpetuate co-existent health problems. Finally, undertaking exercise when movement of an affected joint is painful, can create concern and anxiety about the appropriateness of this intervention. To address these issues, behaviour change and/or education approaches are sometimes used. To date, healthcare professionals have often been good at providing some elements of osteoarthritis treatment but not all the required management approaches. Treatment packages for osteoarthritis have therefore been developed and are defined as any intervention for osteoarthritis (including: exercise, manual therapy, electrotherapy, acupuncture, devices, pharmacological management [including oral, topical, transdermal and intra-articular formulations], arthroscopic procedures) combined with one of the following:

1. Behaviour change interventions (for example: joint protection principles, cognitive-behavioural therapy)

2. An education programme, including those based on behavioural theory (defined as education sessions provided by one or more healthcare professionals over multiple sessions where the study provides clear information about the content included in the education sessions)

Current practice for people with osteoarthritis is to be provided with reactive, symptom based approaches to care. Some healthcare professionals have insufficient expertise or time to deliver the tailored approaches sometimes needed for this population. Referrals can be made for physiotherapy or pain management services to address some of the barriers, however, osteoarthritis treatment packages are not available in a standardised way throughout the country.

This review aims to evaluate the clinical and cost-effectiveness of treatment packages, where combinations of interventions are used together, for the management of osteoarthritis.

1.1.2 Summary of the protocol

Table 1: PICO characteristics of review question

Population	Inclusion: • Adults (age ≥16 years) with osteoarthritis affecting any joint Exclusion: • Children (age <16 years) • People with conditions that may make them susceptible to osteoarthritis or often occur alongside osteoarthritis (including: crystal arthritis, inflammatory arthritis, septic arthritis, hemochromatosis, haemophilic arthropathy, diseases of childhood that may predispose to osteoarthritis, and malignancy).

	• Studies with an unclear population (e,g, proportion of participants with osteoarthritis unclear) • Spinal osteoarthritis
Interventions	Treatment packages (minimum intervention duration 1 week). A treatment package is defined as any intervention for osteoarthritis (including: exercise, manual therapy, electrotherapy, acupuncture, devices, pharmacological management [including oral, topical, transdermal and intra-articular formulations], arthroscopic procedures) combined with one of the following: 1. Behaviour change interventions (for example: joint protection principles, cognitive-behavioural therapy) 2. An education programme, including those based on behavioural theory (defined as education sessions provided by one or more healthcare professionals over multiple sessions where the study provides clear information about the content included in the education sessions)
Comparisons	• Non-combined active treatment for osteoarthritis, started at the time of trial initiation ○ Exercise ○ Manual therapy ○ Electrotherapy ○ Acupuncture ○ Devices ○ Pharmacological management (oral, topical, transdermal or intra-articular therapy) ○ Arthroscopic procedures ○ Other (education programmes, behaviour change interventions) • Standard care (non-organised) or no treatment **No treatment defined as either (1) doing nothing or (2) very low intensity intervention such as advice*
Outcomes	Primary outcomes (critical outcomes): Stratify by ≤/>3 months (longest time-point in each): • Health-related quality of life [validated patient-reported outcomes, continuous data prioritised] • Pain [validated patient-reported outcomes, continuous data prioritised] • Physical function [validated patient-reported outcomes, continuous data prioritised] Secondary outcomes (important outcomes): • Psychological distress [validated patient-reported outcomes, continuous data prioritised] • Osteoarthritis flares [dichotomous data prioritised] • Discontinuation [dichotomous data]
Study design	RCTs or systematic reviews of RCTs

For full details see the review protocol in Appendix A.

1.1.3 Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in Appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.1.4 Effectiveness evidence

1.1.4.1 Included studies

Fifty-six randomised-controlled trial studies (eighty-four papers) were included in the review;[4, 7, 8, 18, 24, 33, 36-38, 43, 45, 46, 56, 75, 80, 87, 89, 93-95, 98, 99, 103, 106, 109, 125-127, 130, 132, 138, 141, 144, 146, 148, 150, 151, 155, 156, 159, 167, 184, 186, 208, 211, 222, 228, 230, 235, 236, 244, 253, 261, 266-268, 279, 288] these are summarised in Table 2 below. Evidence from these studies is summarised in the clinical evidence summary below (Table 3). The majority of studies included people with knee or hip osteoarthritis (with a minority including people with hand osteoarthritis). Three studies[45, 138, 186] reported including people with chronic knee pain, without specifying that they had osteoarthritis. These studies were included but noted to be an indirect population.

The treatment packages included in this review used the following interventions as components:

- Exercise[7, 8, 18, 24, 33, 36-38, 43, 45, 46, 56, 80, 89, 94, 98, 99, 103, 106, 125, 126, 130, 132, 138, 141, 144, 146, 148, 150, 155, 156, 159, 167, 184, 186, 208, 211, 235, 236, 244, 266, 267, 279, 288]
- Manual therapy[228]
- Electrotherapy[109, 268]
- Devices[4]
- Combinations of the above with additional interventions (including acupuncture and pharmacological management)[75, 87, 93, 95, 127, 151, 222, 230, 253, 261]

These were combined with:

- Behaviour change interventions (including joint protection, pain coping skills training, goal setting, weight management counselling, ect.)[7, 18, 36-38, 43, 45, 46, 56, 94, 95, 98, 103, 106, 126, 127, 130, 132, 138, 141, 144, 146, 148, 150, 167, 184, 186, 208, 222, 235, 236, 279]
- Educational programmes[4, 8, 24, 33, 75, 80, 87, 89, 93, 99, 109, 125, 151, 155, 156, 159, 211, 228, 230, 244, 253, 261, 266-268, 288]

The treatment packages varied in length, including studies delivered over less than or equal to 6 weeks and more than 6 weeks.

No relevant clinical studies comparing treatment packages to the following non-combined active treatments were identified:

- Acupuncture
- Pharmacological management (oral, topical, transdermal or intra-articular therapy)
- Arthroscopic procedures

See also the study selection flow chart in Appendix C, study evidence tables in Appendix D, forest plots in Appendix E and GRADE tables in Appendix F.

1.1.4.2 Excluded studies

Two Cochrane reviews were identified and checked [136, 292] but were not included in the review. This was because the reviews did not include treatment packages by the definition used in our protocol.

See the excluded studies list in Appendix J.

1.1.5 Summary of studies included in the effectiveness evidence

1.1.5.1 Treatment packages compared to exercise alone

Table 2: Summary of studies included in the evidence review for the comparison of treatment packages and exercise alone

Study	Intervention and comparison	Population	Outcomes	Comments
Alasfour 2020[7]	**Treatment package - Exercise and behaviour change intervention** (n=20) An app providing a guide for exercise performance, including alerts and a monitoring system controlled by the physical therapist. The app provided automatic recording of exercise adherence, including the time and completed sessions. Length of package ≤ 6 weeks (6 weeks). **Exercise only** (n=20) Exercise program only. **Concomitant therapy:** All participants from both groups had the same exercise program. This was a simple strengthening exercise program for lower-limb muscles (mainly for knee extensor and hip abductor muscles and improve function.	**Knee osteoarthritis** Mean age (SD): 54.4 (4.4). years N = 40 Definition: Diagnosed by the physician with unilateral or bilateral chronic knee osteoarthritis (diagnosis at least 6 months) with mild to moderate pain intensity (score no more than 7 on the Arabic Numeric Pain Rating Scale) Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated/ unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Alferi 2020[8]	**Treatment package - Exercise and education programme** (n=29) Exercise plus lifestyle group. In addition to exercise, there were 8 sessions of lectures and group discussions two times/week on the following topics: nutrition; self-management of the disease: self-care strategies, relationships with family, friends and other social support providers, pain management, and improvement of living conditions and social relations; and health education. Length of package: > 6 weeks (8 weeks). **Exercise only** (n=32) Exercise program only **Concomitant therapy:** A therapeutic exercise program including warm-up, flexibility, active muscle strengthening exercises, balance and proprioception exercises.	**Knee osteoarthritis** Mean age (SD): 64.0 (7.8) years N = 83 Definition: Clinical and radiographic diagnosis of unilateral or bilateral knee osteoarthritis Severity: Kellgren Lawrence grade 1-4 Duration of symptoms: Not stated/unclear Presence of multimorbidities: Not stated/ unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	
Arnold 2010[24]	**Treatment package – Exercise and education programme** (n=28) Aquatic exercise sessions for 45 minutes (delivered twice a week for 11 weeks) with group education sessions for 30	**Hip osteoarthritis** Mean age (SD): 74.4 (6.3) years N = 82 Definition: People with hip pain for at least 6 months who	Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	minutes once a week for 11 weeks. Education included a cognitive behavioural approach to persuade people to change behaviours and adopt positive fall-prevention strategies to motivate them to participate in exercise. Length of package: ≤6 weeks (5 weeks) **Exercise only** (n=27) Exercise component only **Standard care (non-organised) or no treatment** (n=27) People were instructed to not begin an exercise program during the control period and would be offered a treatment after 11 weeks **Concomitant therapy:** People were allowed to start new therapies if necessary	were diagnosed with hip osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: High morbidity score (Number of comorbidities (mean [SD]): 2.1 (1.3)).		
Bennell 2016[37] Subsidiary papers: Bennell 2015[47]	**Treatment package – Exercise and behaviour change intervention** (n=73) Pain coping skills training including 10 weekly sessions. Pain coping skills training included cognitive and behavioural strategies. The exercises included a	**Knee osteoarthritis** Mean age (SD): 63.4 (8.1) years N = 222 Definition: Knee osteoarthritis fulfilling the American College of Rheumatology criteria (pain	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Psychological distress at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	standardised home-based strength exercise program. Length of package: >6 weeks (12 weeks) **Exercise only** (n=75) **Behaviour change intervention only** (n=74) **Concomitant therapy:** No additional information	on most days in the past month and radiographic changes) with knee pain for at least 3 months Severity: Radiographic grade 2-4, median grade 3 Duration of symptoms (median [IQR]): Exercise = 6 (3-10), PCST = 5.5 (4-10), treatment package = 5.5 (2-10). Presence of multimorbidities: Not stated / Unclear	Discontinuation at ≤3 months and >3 months	
Bennell 2017[38] Subsidiary papers: Bennell 2012[39]	**Treatment package – Exercise and behaviour change intervention** (n=84) Exercise programme (with some education) and coaching sessions. Exercise included strengthening exercise three times a week. People were provided with pedometers. Coaching included 6 additional sessions where the coach discussed the person's preference, confidence and success in the exercise to help reinforce desired behavioural change. Delivered in 5 exercise sessions and 6 coaching session over 6 months. Length of package: >6 weeks (6 months)	**Knee osteoarthritis** Mean age (SD): 62.3 (7.5) years N = 168 Definition: American College of Rheumatology clinical criteria for knee osteoarthritis Severity: Not stated Duration of symptoms: <2->10 years, median time 2-10 years Presence of multimorbidities: Not stated / Unclear	Quality of life at >3 months Pain at >3 months Physical function at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Exercise only (n=84) **Concomitant therapy:** No additional information			
Bennell 2018[46] HOPE trial	**Treatment package – Exercise and behaviour change intervention** (n=73) Pain coping skills training (eight 35- to 45-minute modules delivered once per week, including progressive muscle relaxation, brief relaxation practices, activity-rest cycling, pleasant activity scheduling, cognitive restructuring, pleasant imagery, distraction techniques and problem solving) and exercise including strength and flexibility exercises. Length of package: >6 weeks (24 weeks) **Exercise only** (n=71) **Concomitant therapy:** All people received 8 information sheets (covering arthritis, osteoarthritis, managing pain, physical activity, saving energy, health eating, emotions and tips for hip osteoarthritis) produced by Arthritis Australia	Knee osteoarthritis Mean age (SD): 61.3 (7.2) years N = 144 Definition: Hip osteoarthritis with hip pain for at least 3 months on most days of the past month Severity: Not stated Duration of symptoms: <2 years to >10 years, median 2-10 years. Presence of multimorbidities: Not stated / Unclear	Quality of life at >3 months Pain at >3 months Physical function at >3 months Psychological distress at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Bennell 2020[36] Subsidiary paper: Bennell 2020[43]	**Treatment package - Exercise and behaviour change intervention** (n=56) People received an automated, semi-interactive SMS intervention delivered via mobile phone to support adherence to the home exercise program. Length of package: > 6 weeks (24 weeks). **Exercise only** (n=54) No SMS text messaging intervention. **Concomitant therapy:** People continued their previously allocated home exercise program as an unsupervised program for 24 weeks but to reduce the frequency from four times per week to three times per week.	Knee osteoarthritis Mean age (SD): 62.3 (6.8) years N = 110 Definition: Knee pain on most days of the last month with knee pain for at least 3 months, average overall pain severity of at least 4 on an 11-point numeric rating scale and tibiofemoral osteophytes on x-ray Severity: Kellgren Lawrence grade 2-4, median grade 3 Duration of symptoms (mean [SD]): 8.2 (7.5) years Presence of multimorbidities: Not stated/ unclear	Quality of life at >3 months Pain at >3 months Physical function at >3 months Discontinuation at >3 months	The population included people who also had problems in other joints, including the hand, neck, back, hip, foot and shoulder. However, all participants had knee osteoarthritis.
Brosseau 2012[56]	**Treatment package – Exercise and behaviour change intervention** (n=69) Walking and behavioural intervention, including a supervised walking program delivered over a 12 month period three times a week with 45 minute aerobic walking phases achieving approximately 50 to 70% of the subjects' pre-determined maximum heart rate,	Knee osteoarthritis Mean age (SD): 63.4 (8.6) years N = 222 Definition: Mild to moderate unilateral or bilateral osteoarthritis of the knee according to the American College of Rheumatology clinical and	Quality of life at >3 months Pain at >3 months Physical function at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	and a behavioural intervention using the adapted Program for Arthritis Control through Education and Exercise program, discussing the benefits of physical activity, and counselling to provide support and explore barriers. Length of package: >6 weeks (12 months) **Exercise only** (n=79) **Standard care (non-organised) or no treatment** (n=74) Non-organised care (self-directed) **Concomitant therapy:** Everyone was given educational pamphlets and a pedometer as a measurement tool for exercise	radiographic/magnetic resonance imagery criteria Severity: Not stated Duration of symptoms (mean [SD]): 10.3 (9.26) Presence of multimorbidities: Not stated / Unclear		
Dziedzic 2015[94] SMOotH trial Subsidiary papers: Dziedzic 2011[96] Oppong 2014[213]	**Treatment package – Exercise and behaviour change intervention** (n=65) Joint protection instruction and hand exercises. Hand exercises including stretching and strengthening exercises. Joint protection principles included: weight distribution while completing tasks, using as large a grip as possible, avoiding strain and repetitive movements,	**Hand osteoarthritis** Mean age (SD): 65.8 (9.1) years N = 257 Definition: Meeting the American College of Rheumatology criteria for features of hand osteoarthritis, or had	Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	avoiding prolonged grips in one position, reducing the effort needed to do a task and energy conservation. This was delivered over 4 weekly sessions lasting 1.5 hours. Length of package: ≤6 weeks (4 weeks) **Exercise only** (n=65) **Behaviour change intervention only** (n=62) **Standard care (non-organised) or no treatment** (n=65) No additional treatments **Concomitant therapy:** All people were given standardised written information on self-management approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using, and were given advice to consult their general practitioner if symptoms continued to be troublesome.	unilateral or bilateral thumb base osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear		

Study	Intervention and comparison	Population	Outcomes	Comments
Dziedzic 2016[95] Subsidiary papers: Jordan 2017[145] Oppong 2018[212]	**Treatment package – Combination and behaviour change intervention** (n=4*) An enhanced GP consultation to make, give and explain the diagnosis including behaviour change advice and education, advice on weight management, general exercise and physical activity. Length of package: ≤6 weeks (the package was likely delivered for less than 6 weeks, but follow up may continue for a long time after this). **Standard care (non-organised) or no treatment** (n=4*) Usual care (no additional training, guidebooks or dedicated nurse osteoarthritis clinic). **Concomitant therapy:** No additional information	**Mixed (hand, hip, knee, ankle) osteoarthritis** Mean age (SD): 67.3 (10.5) years N = 8* Definition: Joint pain in the past year self-reported in the Health Survey in people 45 years and over Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Psychological distress at ≤3 months and >3 months Discontinuation at >3 months (this outcome uses the number of participants instead of the number of centers)	MOSAIC study *This trial is a cluster randomised study and so the unit of randomisation used for the analysis was the number of centers rather than the number of participants.
Farr 2010[98]	**Treatment package – Exercise and behaviour change intervention** (n=100) Combined resistance training and self-management. Self-management included 12 weekly 90 minute classroom sessions including modules on an overview of osteoarthritis,	**Knee osteoarthritis** Mean age (SD): 55.1 (7.0) years N = 293 Definition: Pain on 4 or more days of the week in one or both knees for at least 4	Pain at ≤3 months and >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	general exercise principles, stress management, foot care, pain management, analgesic and anti-inflammatory medications, nutrition for health, coping mechanisms, communication with health care providers and healthy lifestyle practices. Length of package: >6 weeks (9 months) **Exercise only** (n=95) **Behaviour change intervention only** (n=98) **Concomitant therapy:** No additional information	months during the previous year with radiographic status of grade 2 osteoarthritis in at least one knee. All people met the American College of Rheumatology classification criteria for early osteoarthritis of the knee Severity: Kellgren Lawrence grade 2 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear		
Focht 2005[106] Subsidiary papers: Focht 2004[105] Messier 2004[188] Miller 2003[194] Van gool 2005[275] Shea 2010[252]	**Treatment package – Exercise and behaviour change intervention** (n=76) Diet and exercise. The dietary intervention was conducted by dieticians discussing healthy food selection with portion and dietary fat control, aiming for a weight loss of at least 5%, and exercise including aerobic and strength phases. Length of package: >6 weeks (18 months) **Exercise only** (n=80)	**Knee osteoarthritis** Mean age (SD): 68.7 (6.3) years N = 316 Definition: Knee pain on most days with radiographic evidence of grade 1-3 tibiofemoral or patellofemoral osteoarthritis based on weight-bearing anteroposterior and sunrise view radiographs	Pain at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Behaviour change intervention only (n=82) **Standard care (non-organised) or no treatment** (n=78) No intervention, but regular meetings of participants to provide attention and social interaction with some health education Concomitant therapy: No additional information	Severity: Mean Kellgren Lawrence score: 2.3 (0.7) Duration of symptoms: Not stated Presence of multimorbidities: High morbidity score (70-84% were obese, 53-58% had arthritis in other joints, 44-54% had hypertension, 23-34% had coronary heart disease, 6-12% had diabetes)		
Focht 2014[103] Subsidiary papers: Focht 2017[104]	**Treatment package – Exercise and behaviour change intervention** (n=40) Exercise and cognitive behavioural therapy intervention, delivered as 27, 80-minute center based sessions. Including 60 minutes of exercise (aerobic and strength) and 20 minutes of cognitive behavioural activity counselling in each session. Length of package: >6 weeks (3 months) **Exercise only** (n=40) **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 63.5 (6.9) years N = 80 Definition: Radiographically confirmed, symptomatic knee osteoarthritis Severity: Kellgren Lawrence grade 2-3 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Hsu 2021[126]	**Treatment package - Exercise and behaviour change intervention** (n=22) Both diet control (balanced low-energy diet of 1200 kcal/day) and the elastic band resistance program interventions (seated, open-chain exercises to strengthen the major muscle groups of the lower extremities). Length of package: > 6 weeks (12 weeks). **Exercise only** (n=22) Exercise only **Behaviour change intervention only** (n=22) Dietary advice intervention only **Concomitant therapy:** All people continued their previous therapies	**Knee osteoarthritis** Mean age (SD): 65.3 (4.0) years N = 63 Definition: Knee osteoarthritis diagnosed when x-ray findings indicated a Kellgren and Lawrence grade of no more than 3 and visual analog scale at least 4 out of 10.. Severity: Kellgren Lawrence grade (mean [SD]): 1.73 (0.78) (grades I-III) Duration of symptoms: Not stated/unclear Presence of multimorbidities: Not stated/ unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	
Keefe 2004[150]	**Treatment package – Exercise and behaviour change intervention** (n=20) Spouse assisted coping skills training and exercise. Spouse assisted coping skills training consisted of 12 weekly, 2 hour sessions and discussed: pain being complex; gate control theory; acquiring and maintaining pain coping skills; osteoarthritis being a couples	**Knee osteoarthritis** Mean age (SD): 59.5 (11.4) years N = 72 Definition: Persistent knee pain due to osteoarthritis Severity: Not stated Duration of symptoms: Not stated	Quality of life at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	issue and so everyone's involvement can be useful. Exercise included strength, aerobic and flexibility training. Length of package: >6 weeks (12 weeks) **Exercise only** (n=16) **Behaviour change intervention only** (n=18) **Standard care (non-organised) or no treatment** (n=18) Standard care **Concomitant therapy:** No additional information	Presence of multimorbidities: Not stated / Unclear		
Mcknight 2010[184]	**Treatment package – Exercise and behaviour change intervention** (n=95) Strength training and self-management sessions. Strength training focussed on muscle strengthening, stretching and balance, range or motion and flexibility, with sessions three times a week for 1 hour. Self-management was delivered through 12 weekly 90 minute sessions discussing coping and self-efficacy skills, and then	**Knee osteoarthritis** Mean age (SD): 52.6 (7.2) years N = 273 Definition: Pain on most days in 1 or both knees for less than 4 years with a Kellgren Lawrence score of 2 in one or both knees Severity: Kellgren Lawrence grade of 2	Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	weekly, biweekly, monthly and bimonthly phone calls after this. Length of package: >6 weeks (2 years) **Exercise only** (n=91) **Behaviour change intervention only** (n=87) **Concomitant therapy:** No additional information	Duration of symptoms: Less than 5 years Presence of multimorbidities: Not stated / Unclear		
Quilty 2003[230]	**Treatment package – Combination and education programme** (n=43) Physiotherapy and patellar taping, postural, footwear and weight reduction advice delivered in 9 sessions over 10 weeks lasting half an hour each. Exercises were strengthening in nature. Medial patellar taping was applied. Length of package: >6 weeks (12 weeks) **Exercise only** (n=44) **Concomitant therapy:** All people were given an information sheet and encouraged to continue with the exercises after the formal period of supervised therapy	**Knee osteoarthritis** Mean age (SD): 66.8 (10.4) years N = 87 Definition: Chronic knee or hip pain with radiographic evidence of knee osteoarthritis (Kellgren Lawrence grade less than and equal to 2). Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Pain at >3 months Physical function at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Rejeski 2002[235] ADAPT trial	**Treatment package – Exercise and behaviour change intervention** (n=68) Exercise and dietary weight loss. Exercise 3 days per week (4 months facility based, the remaining 14 months could be home based). Weight loss through group and individual discussion sessions. Length of package: >6 weeks (18 months) **Exercise only** (n=69) **Behaviour change intervention only** (n=73) A fourth group (n=68) was not included in this analysis as it did not fulfil the inclusion criteria (education program only, which was not a component of the treatment package). **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 68.52 (6.30) years N = 278 Definition: Knee pain on most days of the month, limitations in activity and radiographic tibiofemoral osteoarthritis on weight-bearing anteroposterior x-rays Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Low morbidity score (48.73% had hypertension, 15.51% had cardiovascular disease, 9.49% had diabetes).	Quality of life at >3 months	

1.1.5.2 Treatment packages compared to manual therapy alone

Table 3: Summary of studies included in the evidence review for the comparison of treatment packages and manual therapy alone

Study	Intervention and comparison	Population	Outcomes	Comments
Dwyer 2015[93]	**Treatment package – Combination and education programme** (n=28) Manual therapy and rehabilitation (including exercise and education). 6 treatment sessions of manual therapy over 4 weeks including mobilization, manipulation and soft tissue treatment. Education discussed diagnosis and prognosis, and advice on health promotion and lifestyle. Length of package: ≤6 weeks (4 weeks) **Manual therapy only** (n=27) A third group (n=28) was not included as it did not fulfil the inclusion criteria (was a treatment package of exercise and education, with no valid comparator in the others provided). **Concomitant therapy:** Leaflet advice about the diagnosis, prognosis, and lifestyle advice was provided to all participants.	**Knee osteoarthritis** Mean age (SD): 62.2 (11.1) years N = 83 Definition: Mild-moderate knee osteoarthritis based on the American College of Rheumatology and the Kellgren Lawrence grade (suitable grades being grades 0 to 3) Severity: Grade 1-2, median grade 1 Duration of symptoms (mean [SD]): 83.9 (96.1) months Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

1.1.5.3 Treatment packages compared to electrotherapy alone

Table 4: Summary of studies included in the evidence review for the comparison of treatment packages and electrotherapy alone

Study	Intervention and comparison	Population	Outcomes	Comments
Huang 2000[127]	**Treatment package – Combination and behaviour change intervention** (n=42) Weight reduction therapy, with electrotherapy, auricular acupuncture and exercise. Diet control was supported through counselling, advising people to reduce the number of calories taken in per day. Aerobic exercise was achieved through an ergonomic bicycle. Electrotherapy was delivered as ultrasound and TENS. Each course of treatment included 3 treatments per week for 12 weeks . Length of package: > 6 weeks (12 weeks) **Electrotherapy only** (n=42) A third group (n=42) was not included as they did not fulfil the inclusion criteria (was another treatment package with only diet, exercise and acupuncture, with no valid comparator available in the other interventions)	**Knee osteoarthritis** Mean age: 54.8 years N = 126 Definition: People with osteoarthritis stage 2-4 according to the Altman criteria. Severity: Altman grade 2-4 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Concomitant therapy: No additional information			

1.1.5.5 Treatment packages compared to behaviour change interventions alone

Table 5: Summary of studies included in the evidence review for the comparison of treatment packages and behaviour change interventions alone

Study	Intervention and comparison	Population	Outcomes	Comments
Bennell 2016[37] Subsidiary papers: Bennell 2015[47]	**Treatment package – Exercise and behaviour change intervention** (n=73) Pain coping skills training including 10 weekly sessions. Pain coping skills training included cognitive and behavioural strategies. The exercises included a standardised home-based strength exercise program. Length of package: >6 weeks (12 weeks) **Exercise only** (n=75) **Behaviour change intervention only** (n=74) **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 63.4 (8.1) years N = 222 Definition: Knee osteoarthritis fulfilling the American College of Rheumatology criteria (pain on most days in the past month and radiographic changes) with knee pain for at least 3 months Severity: Radiographic grade 2-4, median grade 3 Duration of symptoms (median [IQR]): Exercise = 6 (3-10), PCST = 5.5 (4-10), treatment package = 5.5 (2-10). Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Psychological distress at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Dziedzic 2015[94] SMOotH trial Subsidiary papers: Dziedzic 2011[96] Oppong 2014[213]	**Treatment package – Exercise and behaviour change intervention** (n=65) Joint protection instruction and hand exercises. Hand exercises including stretching and strengthening exercises. Joint protection principles included: weight distribution while completing tasks, using as large a grip as possible, avoiding strain and repetitive movements, avoiding prolonged grips in one position, reducing the effort needed to do a task and energy conservation. This was delivered over 4 weekly sessions lasting 1.5 hours. Length of package: ≤6 weeks (4 weeks) **Exercise only** (n=65) **Behaviour change intervention only** (n=62) **Standard care (non-organised) or no treatment** (n=65) No additional treatments **Concomitant therapy:** All people were given standardised written information on self-management	**Hand osteoarthritis** Mean age (SD): 65.8 (9.1) years N = 257 Definition: Meeting the American College of Rheumatology criteria for features of hand osteoarthritis, or had unilateral or bilateral thumb base osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using, and were given advice to consult their general practitioner if symptoms continued to be troublesome.			
Farr 2010[98]	**Treatment package – Exercise and behaviour change intervention** (n=100) Combined resistance training and self-management. Self-management included 12 weekly 90 minute classroom sessions including modules on an overview of osteoarthritis, general exercise principles, stress management, foot care, pain management, analgesic and anti-inflammatory medications, nutrition for health, coping mechanisms, communication with health care providers and healthy lifestyle practices. Length of package: >6 weeks (9 months) **Exercise only** (n=95) **Behaviour change intervention only** (n=98)	**Knee osteoarthritis** Mean age (SD): 55.1 (7.0) years N = 293 Definition: Pain on 4 or more days of the week in one or both knees for at least 4 months during the previous year with radiographic status of grade 2 osteoarthritis in at least one knee. All people met the American College of Rheumatology classification criteria for early osteoarthritis of the knee Severity: Kellgren Lawrence grade 2 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months and >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Concomitant therapy: No additional information			
Focht 2005[106] Subsidiary papers: Focht 2004[105] Messier 2004[188] Miller 2003[194] Van gool 2005[275] Shea 2010[252]	**Treatment package – Exercise and behaviour change intervention** (n=76) Diet and exercise. The dietary intervention was conducted by dieticians discussing healthy food selection with portion and dietary fat control, aiming for a weight loss of at least 5%, and exercise including aerobic and strength phases. Length of package: >6 weeks (18 months) **Exercise only** (n=80) **Behaviour change intervention only** (n=82) **Standard care (non-organised) or no treatment** (n=78) No intervention, but regular meetings of participants to provide attention and social interaction with some health education **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 68.7 (6.3) years N = 316 Definition: Knee pain on most days with radiographic evidence of grade 1-3 tibiofemoral or patellofemoral osteoarthritis based on weight-bearing anteroposterior and sunrise view radiographs Severity: Mean Kellgren Lawrence score: 2.3 (0.7) Duration of symptoms: Not stated Presence of multimorbidities: High morbidity score (70-84% were obese, 53-58% had arthritis in other joints, 44-54% had hypertension, 23-34% had coronary heart disease, 6-12% had diabetes)	Pain at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Hsu 2021[126]	**Treatment package - Exercise and behaviour change intervention** (n=22) Both diet control (balanced low-energy diet of 1200 kcal/day) and the elastic band resistance program interventions (seated, open-chain exercises to strengthen the major muscle groups of the lower extremities). Length of package: > 6 weeks (12 weeks). **Exercise only** (n=22) Exercise only. **Behaviour change intervention only** (n=22) Dietary advice intervention only. **Concomitant therapy:** All people continued their previous therapies	**Knee osteoarthritis** Mean age (SD): 65.3 (4.0) years N = 63 Definition: Knee osteoarthritis diagnosed when x-ray findings indicated a Kellgren and Lawrence grade of no more than 3 and visual analog scale at least 4 out of 10. Severity: Kellgren Lawrence grade (mean [SD]): 1.73 (0.78) (grades I-III) Duration of symptoms: Not stated/unclear Presence of multimorbidities: Not stated/ unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	
Keefe 2004[150]	**Treatment package – Exercise and behaviour change intervention** (n=20) Spouse assisted coping skills training and exercise. Spouse assisted coping skills training consisted of 12 weekly, 2 hour sessions and discussed: pain being complex; gate control theory; acquiring and maintaining pain coping skills; osteoarthritis being a couples	**Knee osteoarthritis** Mean age (SD): 59.5 (11.4) years N = 72 Definition: Persistent knee pain due to osteoarthritis Severity: Not stated	Quality of life at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	issue and so everyone's involvement can be useful. Exercise included strength, aerobic and flexibility training.. Length of package: >6 weeks (12 weeks) **Exercise only** (n=16) **Behaviour change intervention only** (n=18) **Standard care (non-organised) or no treatment** (n=18) Standard care **Concomitant therapy:** No additional information	Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear		
Mcknight 2010[184]	**Treatment package – Exercise and behaviour change intervention** (n=95) Strength training and self-management sessions. Strength training focussed on muscle strengthening, stretching and balance, range or motion and flexibility, with sessions three times a week for 1 hour. Self-management was delivered through 12 weekly 90 minute sessions discussing coping and self-efficacy skills, and then	**Knee osteoarthritis** Mean age (SD): 52.6 (7.2) years N = 273 Definition: Pain on most days in 1 or both knees for less than 4 years with a Kellgren Lawrence score of 2 in one or both knees Severity: Kellgren Lawrence grade of 2	Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	weekly, biweekly, monthly and bimonthly phone calls after this. Length of package: >6 weeks (2 years) **Exercise only** (n=91) **Behaviour change intervention only** (n=87) **Concomitant therapy:** No additional information	Duration of symptoms: Less than 5 years Presence of multimorbidities: Not stated / Unclear		
Rejeski 2002[235] ADAPT trial	**Treatment package – Exercise and behaviour change intervention** (n=68) Exercise and dietary weight loss. Exercise 3 days per week (4 months facility based, the remaining 14 months could be home based). Weight loss through group and individual discussion sessions. Length of package: >6 weeks (18 months) **Exercise only** (n=69) **Behaviour change intervention only** (n=73) A fourth group (n=68) was not included in this analysis as it did not fulfil the inclusion criteria (education program only, which	**Knee osteoarthritis** Mean age (SD): 68.52 (6.30) years N = 278 Definition: Knee pain on most days of the month, limitations in activity and radiographic tibiofemoral osteoarthritis on weight-bearing anteroposterior x-rays Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Low morbidity score (48.73% had hypertension, 15.51% had cardiovascular disease, 9.49% had diabetes).	Quality of life at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	was not a component of the treatment package). **Concomitant therapy:** No additional information			

1.1.5.4 Treatment packages compared to education programmes alone

Table 6: Summary of studies included in the evidence review for the comparison of treatment packages and education programmes alone

Study	Intervention and comparison	Population	Outcomes	Comments
Adams 2021[4]	**Treatment package - Devices and education programme** (n=116) Included a self-management programme (plus a thumb splint) consisting of 90 minute 1:1 therapist intervention over two hospital visits, plus hand exercises at least 3 times a week for at least 20 minutes each time. Other elements provided included: Arthritis Research UK Osteoarthritis booklet, a discussion with the therapist about the potential facilitators and barriers to engaging with self-management and a self-management contract sheet; a hand exercise diary. Length of package: > 6 weeks (8 weeks)	**Thumb osteoarthritis** Mean age (SD): 62.6 (9.6) years N = 349 Definition: Base of thumb osteoarthritis reporting at least moderate hand pain (>5) and dysfunction (>9) on the Australian Canadian outcome measure. Severity: Not stated/unclear Duration of symptoms (median [IQR]): Between 2 (0,4) and 1 (0,3). Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Education programme only (n=116) Self-management intervention only. **Treatment package - Devices and education programme** (n=117) Self management program and placebo splint. **Concurrent therapy:** No additional information			
Alfieri 2020[8]	**Treatment package – Exercise and education programme** (n=29) Exercise (supervised strength, proprioceptive and balance exercises) plus lifestyle counselling, including 8 lectures discussing nutrition, self management strategies, health education and coping skills. Delivered 2 times a week over 8 weeks. **Exercise only** (n=32) **Concomitant therapy:** All people received a therapeutic exercise program.	**Knee osteoarthritis** Mean age (SD): 64.0 (7.8) years N = 61 Definition: Clinical and radiographic diagnosis of unilateral or bilateral knee osteoarthritis Severity: Kellgren Lawrence grade 1-4 Duration of symptoms: Not stated/unclear Presence of multimorbidities: Not stated/unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Crossley 2015[75] Subsidiary papers: Crossley 2008[76]	**Treatment package – Combination and education programme** (n=44) Exercise, education, manual therapy and taping. Eight treatments (approximately 60 minutes duration) once a week for 4 weeks and then once every 2 weeks for 8 weeks for each group. The package included strengthening exercises, patellar taping, manual therapy and an education program discussing osteoarthritis, physical activity, healthy eating, complementary therapies and coping strategies. Length of package: >6 weeks (8 weeks) **Education programme only** (n=48) **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 54.4 (9.9) years N = 92 Definition: Anterior or retro-patellar pain with lateral patellofemoral osteophytes on weight-bearing skyline radiographs Severity: Kellgren Lawrence grade 0-2, median grade 0 (this study looks at people with patellofemoral osteoarthritis) Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	
Deveza 2021[87]	**Treatment package - Combination and education programme** (n=102) Education, splint, hand exercises and diclofenac sodium 1% gel. Length of package: ≤6 weeks **Education programme only** (n=102)	**Thumb osteoarthritis** Mean age (SD): 65.6 (8.1) years N = 204 Definition: Thumb base pain at least half of the days in the past month, average pain rated at 40 or greater on a 0 to 100mm visual analog scale over the 30 days and in the	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Education only. **Concomitant therapy:** Both groups were provided with education about osteoarthritis and ergonomic principles using a 9-page educational booklet and 2 individual, face-to-face sessions with the study physiotherapist. The educational booklet did not provide information about exercises or splints	48 hours prior to screening, score of 6 or higher on the Functional Index of Hand Osteoarthritis and radiographic evidence of osteoarthritis at the first metacarpal joint, read by a trained rheumatologist. Severity: Kellgren Lawrence grade 2-3, median grade 3. Duration of symptoms: <1 to >5 years. Median 1-5 years. Presence of multimorbidities: Not stated/ unclear		
Dias 2017[89]	**Treatment package – Exercise and education programme** (n=37) Hydrotherapy and education about diagnosis, symptoms, prognosis and basic care of knee osteoarthritis during daily activities (through one lecture and weekly advice on telephone discussions). Length of package: ≤ 6 weeks (6 weeks). **Education programme only** (n=36) **Concomitant therapy:** No additional information.	**Hip osteoarthritis** Mean age (SD): 70.9 (5.1) years N = 73 Definition: People diagnosed with osteoarthritis in at least one knee based on the clinical and radiographic criteria of the American College of Rheumatology diagnosed with hip osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Fernandes 2010[99] Subsidiary papers: Svege 2016[264] Svege 2015[265]	**Treatment package – Exercise and education programme** (n=55) "Hip School" with patient education and supervised exercise. Education included three group based sessions and one individual session. Exercises included strengthening, functional and flexibility exercises. Length of package: >6 weeks (12 weeks) **Education programme only** (n=54) **Concomitant therapy:** No additional information	**Hip osteoarthritis** Mean age (SD): 57.8 (9.9) years N = 109 Definition: Radiographical and symptomatic hip osteoarthritis Severity: Not stated Duration of symptoms (mean [SD]): 48.4 (52.1) months Presence of multimorbidities: Not stated / Unclear	Quality of life at >3 months Pain at >3 months Physical function at >3 months Discontinuation at >3 months	
Gaines 2004[109]	**Treatment package – Electrotherapy and education programme** (n=20) Neuromuscular electrical stimulation (15 minutes per day, 3 days a week for 36 sessions in total) with the Arthritis Self-management program, including 12 hour community-based sessions discussing arthritis, self-management and helping to produce personalised action plans for the management of arthritis. Length of package: > 6 weeks (12 weeks).	**Knee osteoarthritis** Mean age: 70.8 years N = 38 Definition: Radiographic and clinical evidence of knee osteoarthritis Severity: Grades 1-4, median grades 1-2 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months Pain at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Education programme only (n=18) **Concomitant therapy:** No additional information.			
Kemp 2018[151]	**Treatment package – Combination and education programme** (n=10) A semi-standardised program including manual hip joint and soft tissue mobilisation and stretching; hip muscle retraining; trunk muscle retraining; function, proprioceptive and sports- or activity-specific retraining; enhancing physical activity and education. Length of package: > 6 weeks (12 weeks). **Education programme only** (n=7) **Concomitant therapy:** No additional information.	**Hip osteoarthritis** Mean age (SD): 35.7 (9.9) years N = 17 Definition: Early-onset hip osteoarthritis (defined as chondropathy Outerbridge grade at least 1). Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	This study was noted to have serious population indirectness (people were post-hip arthroscopy, being studied on average 8-9 months afterwards)
Oh 2021[211]	**Treatment package - Exercise and education programme** (n=40) Education and a self-directed home-based resistance training program. Length of package: > 6 weeks (5 months).	**Knee osteoarthritis** Mean age (SD): 71.5 (5.8) years N = 60 Definition: Clinically and radiologically defined degenerative osteoarthritis	Pain at >3 months Physical function at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Education programme only (n=20) Education programme only **Concomitant therapy:** Both groups participated in the health education program, which consisted of 50 minutes, once a month for 5 months and was conducted by a multidisciplinary team. It covered 1) the prevention and management of osteoarthritis; 2) lifestyle modification for pain management; 3) self-care strategies for pain relief; 4) nutrition for weight management; 5) ways to improve health-related quality of life.	Severity: Not stated/unclear Duration of symptoms: Not stated/unclear Presence of multimorbidities: Not stated/ unclear		
Poulsen 2013[228] Subsidiary papers: Poulsen 2011[227] Poulsen 2013[229]	**Treatment package – Manual therapy and education programme** (n=43) Hip school and manual therapy. Hip school involved 5 sessions delivered over 6 weeks with two individual sessions and three group sessions. This included information about epidemiology, anatomy of the hip, pain, activity, natural course of the disease and treatment options. Stretching exercises were taught. The manual therapy included a combination of	**Hip osteoarthritis** Mean age (SD): 64.6 (8.6) years N = 118 Definition: Unilateral hip pan for >3 months' duration with radiographic hip osteoarthritis defined as minimal joint space width (JSW) measurement <2.00mm or a side difference in minimal JSW >10% Severity: Not stated	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	manual soft tissue therapy, stretching and joint manipulation. Length of package: ≤ 6 weeks (6 weeks). **Education programme only** (n=39) **Standard care (non-organised) or no treatment** (n=36) Minimal intervention. People were given a leaflet describing the stretching exercises from hip school and received a short 5 minute instruction in self-care immediately after randomisation. People were advised to live as ususal, not to make any changes to use of possible pain medication or to initiate any other treatment during the following 6 weeks. **Concomitant therapy:** No additional information	Duration of symptoms (mean [SD]): 32 (36) months Presence of multimorbidities: Not stated / Unclear		
Stener-victorin 2004[261]	**Treatment package – Combination and education programme** (n=43) Hydrotherapy or electroacupuncture and education about hip anatomy, the disease process, activity cycling, pain relief and total hip	**Hip osteoarthritis** Age range: 42-86 years N = 82 Definition: Radiographic changes consistent with osteoarthritis in the hip and	Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	arthroplasty. Education was delivered in 2 group meetings of 2 hours each, with hydrotherapy delivered 10 times during 5 weeks. Length of package: ≤ 6 weeks (5 weeks). **Education programme only** (n=39) **Concomitant therapy:** No additional information	pain related to motion and/or pain on load and/or ache during rest Severity: Not stated Duration of symptoms (range): 4 months - 15 years Presence of multimorbidities: Not stated / Unclear		
Talbot 2003[267]	**Treatment package – Exercise and education programme** (n=17) Walk+ program and education (self-management) program. Each person's daily steps were modified to increase them by 10% from baseline every 4 weeks. Each person had individual counselling and participated in the Arthritis self-management program to learn techniques around coping and including exercise in management. Length of package: > 6 weeks (12 weeks). **Education programme only** (n=17)	**Hip osteoarthritis** Mean age (SD): 70.2 (5.8) years N = 34 Definition: Pain in one or both knees on most days, difficulty performing at least one functional task because of pain, and radiographic evidence of osteoarthritis Severity: Grades 1-4, median grade 2 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months and >3 months	In forest plots this study is referred to as Talbot 2003A

Study	Intervention and comparison	Population	Outcomes	Comments
	Concomitant therapy: No additional information			
Talbot 2003[268]	**Treatment package – Electrotherapy and education programme** (n=20) Neuromuscular electrical stimulation delivered to the quadriceps femoris muscle completed at home with 3 training sessions per week for 12 weeks. Each person participated in the Arthritis self-management program to learn techniques around coping and including exercise in management. Length of package: > 6 weeks (12 weeks). **Education programme only** (n=18) **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 70.5 (5.3) years N = 38 Definition: Pain in one or both knees; self reported difficulty in walking, stair climbing or rising from a chair; radiographic evidence of knee osteoarthritis (At least grade 1) based on the criteria of Kellgren and Lawrence Severity: Grades 1-4, median grade 2 Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months and >3 months Discontinuation at >3 months	In forest plots this study is referred to as Talbot 2003B

1.1.5.6 Treatment packages compared to standard care (non-organised) or no treatment

Table 7: Summary of studies included in the evidence review for the comparison of treatment packages and standard care (non-organised) or no treatment

Study	Intervention and comparison	Population	Outcomes	Comments
Allen 2021[18] Subsidiary paper: Kaufman 2021[148]	**Treatment package - Exercise and behaviour change intervention** (n=230) STEP-KOA programme. Began with access to an internet-based exercise program for knee osteoarthritis. After 3 months, people not meeting OMERACT-OARSI response criteria progressed to biweekly telephone coaching to address barriers to physical activity. After 3 months, participants still not meeting response criteria went on to in-person physiotherapy visits. Length of package: > 6 weeks (9 months). **Standard care (non-organised) or no treatment**(n=115) People received educational materials via mail every 2 weeks for 9 months. The intervention included a comprehensive set of topics related to osteoarthritis and its management, described previously and based on established treatment guidelines.	**Knee osteoarthritis** Mean age (SD): 60.0 (10.3) years N = 345 Definition: Physician diagnosis of knee osteoarthritis Severity: Not stated/unclear Duration of symptoms (mean [SD]): 16.4 (11.2) years Presence of multimorbidities: Not stated/ unclear	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Concomitant therapy: No additional information			
Arnold 2010[24]	**Treatment package – Exercise and education programme** (n=28) Aquatic exercise sessions for 45 minutes (delivered twice a week for 11 weeks) with group education sessions for 30 minutes once a week for 11 weeks. Education included a cognitive behavioural apporach to persuade people to change behaviours and adopt positive fall-prevention strategies to motivate them to participate in exercise. **Exercise only** (n=27) Exercise component only **Standard care (non-organised) or no treatment** (n=27) People were instructed to not begin an exercise program during the control period and would be offered a treatment after 11 weeks **Concomitant therapy:** People were allowed to start new therapies if necessary	**Hip osteoarthritis** Mean age (SD): 74.4 (6.3) years N = 82 Definition: People with hip pain for at least 6 months who were diagnosed with hip osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: High morbidity score (Number of comorbidities (mean [SD]): 2.1 (1.3)).	Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Bearne 2011[33]	**Treatment package – Exercise and education programme** (n=24) Ten 75 minute group exercise and self-management sessions (up to 8 people per group, twice a week for 5 weeks) including strength, balance and functional exercises and education, coping and self-management discussion sessions facilitated by a physiotherapist. Length of package: ≤ 6 weeks (5 weeks). **Standard care (non-organised) or no treatment** (n=24) Usual care only **Concomitant therapy:** All people were allowed to continue routine management prescribed by their GPs, including referral to secondary care. Medication for co-existent conditions continued as needed.	**Hip osteoarthritis** Mean age (range): 66 (52-78) years N = 48 Definition: People with hip pain for at least 6 months who were diagnosed with hip osteoarthritis Severity: Not stated Duration of symptoms (mean [range]): 5.0 (1-40) years Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Psychological distress at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	
Bennell 2017[45] IMPACT trial Subsidiary papers: Lawford 2018[163] Lin 2003[168]	**Treatment package – Exercise and behaviour change intervention** (n=74) Videoconferencing sessions with a physiotherapist for home exercise and a pain coping skills training program. Pain coping skills training (PainCOACH)	**Knee osteoarthritis** Mean age (SD): 61.2 (7.1) years N = 214 Definition: Chronic knee pain and reduced physical function	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months	This study was noted to have serious population indirectness (no clear statement about the presence of osteoarthritis)

Study	Intervention and comparison	Population	Outcomes	Comments
	included eight 35- to 45- minute modules that were interactive and automated, and advised practicing pain-coping skills daily (including progressive relaxation, activity-rest cycling, scheduling pleasant activities, changing negative thoughts, pleasant imagery and distraction techniques, and problem solving). The physiotherapy sessions were completed in 7 sessions over 12 weeks including strength exercises. Length of package: > 6 weeks (12 weeks) **Standard care (non-organised) or no treatment** (n=140) No additional treatment (just internet educational material). **Concomitant therapy:** All people had access to internet educational material about exercise and physical activity, pain management, emotions, healthy eating, complementary therapies, and medications (www.arthritisaustralia.com.au) that they were encouraged to access at their leisure.	Severity: Not stated Duration of symptoms: <2 years - >10 years, median 2-10 years Presence of multimorbidities: Not stated / Unclear. Site of osteoarthritis: Knee	Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Brosseau 2012[56]	**Treatment package – Exercise and behaviour change intervention** (n=69) Walking and behavioural intervention, including a supervised walking program delivered over a 12 month period three times a week with 45 minute aerobic walking phases achieving approximately 50 to 70% of the subjects' pre-determined maximum heart rate, and a behavioural intervention using the adapted Program for Arthritis Control through Education and Exercise program, discussing the benefits of physical activity, and counselling to provide support and explore barriers. Length of package: >6 weeks (12 months) **Exercise only** (n=79) **Standard care (non-organised) or no treatment** (n=74) Non-organised care (self-directed) **Concomitant therapy:** Everyone was given educational pamphlets and a pedometer as a measurement tool for exercise	**Knee osteoarthritis** Mean age (SD): 63.4 (8.6) years N = 222 Definition: Mild to moderate unilateral or bilateral osteoarthritis of the knee according to the American College of Rheumatology clinical and radiographic/magnetic resonance imagery criteria Severity: Not stated Duration of symptoms (mean [SD]): 10.3 (9.26) Presence of multimorbidities: Not stated / Unclear	Quality of life at >3 months Pain at >3 months Physical function at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Da silva 2015[80]	**Treatment package – Exercise and education programme** (n=19) A group rehabilitation program with 60 minute sessions twice a week for 8 weeks including educational aspects about knee osteoarthritis (including dietary modification, non-pharmacological management of pain, and home exercise techniques) and a strengthening and balance exercise program. Length of package: >6 weeks (8 weeks) **Standard care (non-organised) or no treatment** (n=22) No additional treatment **Concomitant therapy:** Everyone had one self-management class session with a general orientation about osteoarthritis delivered in a 90 minute lecture.	**Knee osteoarthritis** Mean age (SD): 58.5 (7.1) years N = 41 Definition: A clinical diagnosis of chronic knee osteoarthritis (based on the criteria of the American College of Rheumatology) Severity: Not stated Duration of symptoms: Symptoms in the last year on most days for at least 3 months Presence of multimorbidities: Low morbidity score (Diabetes Mellitus: 3, Hypertension: 18, Hypercholesterolemia: 2).	Quality of life at ≤3 months Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	
Dziedzic 2015[94] SMOotH trial Subsidiary papers: Dziedzic 2011[96] Oppong 2014[213]	**Treatment package – Exercise and behaviour change intervention** (n=65) Joint protection instruction and hand exercises. Hand exercises including stretching and strengthening exercises. Joint protection principles included:	**Hand osteoarthritis** Mean age (SD): 65.8 (9.1) years N = 257 Definition: Meeting the American College of	Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	weight distribution while completing tasks, using as large a grip as possible, avoiding strain and repetitive movements, avoiding prolonged grips in one position, reducing the effort needed to do a task and energy conservation. This was delivered over 4 weekly sessions lasting 1.5 hours. Length of package: ≤6 weeks (4 weeks) **Exercise only** (n=65) **Behaviour change intervention only** (n=62) **Standard care (non-organised) or no treatment** (n=65) No additional treatments **Concomitant therapy:** All people were given standardised written information on self-management approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using and were	Rheumatology criteria for features of hand osteoarthritis, or had unilateral or bilateral thumb base osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear		

Study	Intervention and comparison	Population	Outcomes	Comments
	given advice to consult their general practitioner if symptoms continued to be troublesome.			
Focht 2005[106] Subsidiary papers: Focht 2004[105] Messier 2004[188] Miller 2003[194] Van gool 2005[275] Shea 2010[252]	**Treatment package – Exercise and behaviour change intervention** (n=76) Diet and exercise. The dietary intervention was conducted by dieticians discussing healthy food selection with portion and dietary fat control, aiming for a weight loss of at least 5%, and exercise including aerobic and strength phases. Length of package: >6 weeks (18 months) **Exercise only** (n=80) **Behaviour change intervention only** (n=82) **Standard care (non-organised) or no treatment** (n=78) No intervention, but regular meetings of participants to provide attention and social interaction with some health education **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 68.7 (6.3) years N = 316 Definition: Knee pain on most days with radiographic evidence of grade 1-3 tibiofemoral or patellofemoral osteoarthritis based on weight-bearing anteroposterior and sunrise view radiographs Severity: Mean Kellgren Lawrence score: 2.3 (0.7) Duration of symptoms: Not stated Presence of multimorbidities: High morbidity score (70-84% were obese, 53-58% had arthritis in other joints, 44-54% had hypertension, 23-34% had coronary heart disease, 6-12% had diabetes)	Pain at >3 months Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Hopman-rock 2000[125]	**Treatment package – Exercise and education programme** (n=56) A 6 weekly education programme with an exercise component. The first hour of each session was guided by a peer education discussing: pathophysiology, lifestyle and physical activity, pain management, importance of weight reduction and diet, ergonomic aspects and medical aspects of osteoarthritis. The second hour was a strength exercise program directed by a physical therapist. Length of package: ≤ 6 weeks (6 weeks) **Standard care (non-organised) or no treatment** (n=49) No treatment **Concomitant therapy:** No additional information	**Mixed osteoarthritis (hip and/or knee)** Mean age (SD): 65.3 (5.5) years N = 105 Definition: Radiographs of the hips and knees confirming osteoarthritis of Kellgren Grade at least 2. Following the classification criteria of the American College of Rheumatology. Severity: Kellgren Lawrence score of at least 2 in 795 of people Duration of symptoms: <1 year to >20 years, median 3-10 years Presence of multimorbidities: High morbidity score (Number of other chronic conditions (mean [SD]): 2.5 (1.6)).	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months	
Hughes 2004[130] Subsidiary papers: Hughes 2010[131]	**Treatment package – Exercise and behaviour change intervention** (n=80) Fit & Strong intervention. 90 minute sessions held three times per week for 8 weeks. Included resistance training and walking, and a 30 minute group	**Knee osteoarthritis** Mean age (SD): 73.6 (6.6) years N = 150 Definition: Knee osteoarthritis with at least 3 of the following	Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	discussion educational component, discussing how people would achieve tasks at home and information about the efficacy of exercise. Length of package: > 6 weeks (8 weeks) **Standard care (non-organised) or no treatment** (n=70) No treatment **Concomitant therapy:** All people were given a copy of 'The Arthritis Helpbook' including self-care materials and hand-outs	6: age >60 years, morning stiffness with a duration <30 minutes, crepitus on active motion, tenderness of the bony margins of the joint, bony enlargement on examination, a lack of palpable warmth of the synovium. Hip osteoarthritis if pain is present in combination with either: hip internal rotation at least 15 degrees, pain present on internal rotation of the hip, morning stiffness of the hip for a time no more than 60 minutes, and age <60 years or; hip internal rotation <15 degrees, and hip flexion at least 115 degrees. Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Low morbidity score (Unclear. Around 60% had a cardiovascular disease, 5.5% had an asthma, 4% had emphysema, 11% had diabetes, 5% had cancer).		
Hughes 2006[132]	**Treatment package – Exercise and behaviour change intervention** (n=115) Fit & Strong intervention. 90 minute sessions held three times per week for 8 weeks.	**Knee osteoarthritis** Mean age: 73.3 years N = 215	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Included resistance training and walking, and a 30 minute group discussion educational component, discussing how people would achieve tasks at home and information about the efficacy of exercise. Length of package: > 6 weeks (8 weeks) **Standard care (non-organised) or no treatment** (n=100) No treatment **Concomitant therapy:** All people were given a copy of 'The Arthritis Helpbook' including self-care materials and hand-outs	Definition: Osteoarthritis of the hip or knee as per a modified version of the American College of Rheumatology functional classes Severity: American Rheumatism Association classes I-III, median class II Duration of symptoms: Not stated Presence of multimorbidities: Low morbidity score (Around 60% had hypertension, around 44% had cardiovascular disease, around 6.5% had asthma, around 4% had emphysema, around 13% had diabetes, around 4% had cancer).	Discontinuation at ≤3 months and >3 months	
Hurley 2007[138] Subsidiary papers: Hurley 2012[137] Hurley 2007[139]	**Treatment package – Exercise and behaviour change intervention** (n=278) ESCAPE-knee pain program. Integrated patient education, with simple self-management and pain coping strategies, delivered in the first 15-20 minutes of each rehabilitation session followed by 35-45 minutes of individualised progressive exercise programs. The content of self-management, coping and education settings included goal	**Knee osteoarthritis** Mean age (range): 67 (50-91) years N = 418 Definition: People with chronic knee pain of mild, moderate or severe magnitude for more than 6 months Severity: Not stated Duration of symptoms: Not stated	Quality of life at >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Psychological distress at >3 months Discontinuation at ≤3 months and >3 months	This study was noted to have serious population indirectness (the population had chronic knee pain but no clear statement about the presence of osteoarthritis)

Study	Intervention and comparison	Population	Outcomes	Comments
	setting, pacing and activity-rest cycling, drug management and action plan review, diet and healthy eating, intermediate home exercise regimen and program review, pain gate and review of action plans, managing flares in pain, advanced home exercise regimen and reviewing action plans, mini-relaxation and deep breathing techniques and information regarding pursuing activity and exercise in the community. Length of package: ≤ 6 weeks (6 weeks) **Standard care (non-organised) or no treatment** (n=140) Usual primary care **Concomitant therapy:** No additional information	Presence of multimorbidities (median [IQR]): Usual care: 6 (3-15), Individual rehab: 7 (3-15), Group rehab: 5 (2.5-11).		
Isaramalai 2018[141]	**Treatment package – Exercise and behaviour change intervention** (n=63) Exercise by one of two forms: progressive strengthening exercise or non-weight bearing exercise, with a community based education session and behaviour change intervention and follow up home visits for	**Knee osteoarthritis** Mean age (SD): 66.2 (5.2) years N = 108 Definition: Symptomatic knee osteoarthritis, as determined by the clinical and radiographic criteria of the American College of	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	extra support. The behaviour change intervention included: a twenty-minute job hazard analysis, a one-hour health education session, and thirty minute mutual goal setting. Home-based interventions conducted every other week (with self-directed exercise at least 3 days per week for 8 weeks). Length of package: > 6 weeks (8 weeks) **Standard care (non-organised) or no treatment** (n=45) Usual care only **Concomitant therapy:** No additional information	Rheumatology and the Kellgren-Lawrence radiographic grading scale (<4) Severity: Kellgren and Lawrence grade of knee osteoarthritis 1-3, median grade 2 Duration of symptoms (mean [IQR]): PEM-NEW = 3 (2,5), PEM-PRE = 2 (2,3.3), ST = 3 (2,5). Presence of multimorbidities: Not stated / Unclear		
Jessep 2009[144]	**Treatment package – Exercise and behaviour change intervention** (n=29) ESCAPE-knee pain. 10 sessions held twice a week for 5 weeks, with a review session 4 months after completion of the program. Each session began with an informal themed group discussion led by a supervising physiotherapist for 15-20 minutes, followed by a 40-minute self-paced, progressive exercise circuit to improve	**Knee osteoarthritis** Mean age (range): 67 (51 to 81) years N = 64 Definition: Mild, moderate or severe non-specific knee pain lasting more than 6 months with no identifiable recent cause; these people would be diagnosed as having clinical osteoarthritis based on their clinical presentation and history	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months Psychological distress at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	strength, balance, coordination and function. At 4 months messages were reinforced. Length of package: ≤ 6 weeks (5 weeks). **Standard care (non-organised) or no treatment** (n=35) Outpatient physiotherapy (no additional information) **Concomitant therapy:** No additional information	Severity: Not stated Duration of symptoms (mean [range]): 13 (0.5 to 55) years Presence of multimorbidities: Not stated / Unclear		
Kao 2012[146]	**Treatment package – Exercise and behaviour change intervention** (n=134) A treatment package containing a behaviour change component, education component and exercise component delivered as four 80 minute classes held once a week with 10-15 participants. Education discussed healthy lifestyles, seeking support, solving problems and making action plans. This included a self-efficacy promoting strategy where people made their own goals and shared their experiences with others. Length of package: ≤ 6 weeks (4 weeks).	**Knee osteoarthritis** Mean age (SD): 67.7 (10.6) years N = 259 Definition: Diagnosis by medical history and a physical examination (including an x-ray showing osteophytes) Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Low morbidity score (No comorbidities: 76, High blood pressure: 94, Diabetes mellitus: 27, Hyperlipidaemia: 25, Heart disease: 29).	Quality of life at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Standard care (non-organised) or no treatment (n=125) Standard care available to all participants **Concomitant therapy:** No additional information			
Keefe 2004[150]	**Treatment package – Exercise and behaviour change intervention** (n=20) Spouse assisted coping skills training and exercise. Spouse assisted coping skills training consisted of 12 weekly, 2 hour sessions and discussed: pain being complex; gate control theory; acquiring and maintaining pain coping skills; osteoarthritis being a couples issue and so everyone's involvement can be useful. Exercise included strength, aerobic and flexibility training.. Length of package: >6 weeks (12 weeks) **Exercise only** (n=16) **Behaviour change intervention only** (n=18)	**Knee osteoarthritis** Mean age (SD): 59.5 (11.4) years N = 72 Definition: Persistent knee pain due to osteoarthritis Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Standard care (non-organised) or no treatment (n=18) Standard care **Concomitant therapy:** No additional information			
Klassbo 2003[155]	**Treatment package – Exercise and education programme** (n=77) Hip school. Instructions on home based exercises and an education programme, consisting of three education sessions and 1 individual follow up session 2 months after the last session). The sessions discuss the anatomy of the hip, what hip osteoarthritis is, pain, exercise, self-management strategies, non-pharmacological, pharmacological and surgical management. Length of package: > 6 weeks (6 months) **Standard care (non-organised) or no treatment** (n=68) Usual treatment **Concomitant therapy:** No additional information	**Hip osteoarthritis** Mean age (SD): 61.8 (10.4) years N = 145 Definition: All people had to have fulfilled diagnostic tests (radiography) and clinical criteria, defined as pain in the hip region lasting more than 3 months and manifestations of impaired hip joint range of motion and/or muscle function. Severity: Not stated Duration of symptoms: Between <6 months and 10+ years, median time >2 years <5 years Presence of multimorbidities: Not stated / Unclear	Discontinuation at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Kloek 2018[156]	**Treatment package – Exercise and education programme** (n=109) e-Exercise delivered over 12 weeks with a combination of about 5 face-to-face sessions with a physical therapist and an online application focusing on behavioural graded activity, exercises and information. E-exercise included 3 modules: graded activity; strength and stability; and information (osteoarthritis aetiology, pain management, weight management, motivation, medication and social influences on pain). Length of package: > 6 weeks (12 weeks). **Standard care (non-organised) or no treatment** (n=99) Usual physical therapy according to a Dutch Osteoarthritis guideline **Concomitant therapy:** No additional information	**Mixed osteoarthritis (hip and/or knee osteoarthritis)** Mean age (SD): 63.1 (8.7) years N = 208 Definition: People hip/knee osteoarthritis according to the clinical criteria of the American College of Rheumatology Severity: Not stated Duration of symptoms: <1 to at least 5 years, median time 1-5 years Presence of multimorbidities: Low morbidity score (0 comorbidities: 124, 1 comorbidity: 40, at least 2 comorbidities: 44).	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months Physical function at ≤3 months and >3 months	
Kovar 1992[159] Subsidiary papers: Sullivan 1998[263]	**Treatment package – Exercise and education programme** (n=52) Indoor supervised fitness walking and patient education.	**Knee osteoarthritis** Mean age (SD): 69.5 (10.3) years N = 102	Quality of life at ≤3 months and >3 months Pain at >3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	The program included 24 90-minute walking and education sessions, with education discussing the barriers and benefits of walking, how to walk properly and maintain the habit. Length of package: > 6 weeks (8 weeks). **Standard care (non-organised) or no treatment** (n=50) Standard care only **Concomitant therapy:** No additional information	Definition: Clinical and radiographic osteoarthritis Severity: Not stated Duration of symptoms (mean [SD]): 11.5 (11.5) years Presence of multimorbidities: Not stated / Unclear		
Li 2017[167] Subsidiary papers: Clayton 2015[69]	**Treatment package – Exercise and behaviour change intervention** (n=17) A 1.5 hour education session about physical activity, a FitbitFlex to encourage aerobic exercise, and individual weekly activity counselling with a physiotherapist by telephone. People were counselled to make action plans and identify barriers and solutions. Length of package: ≤ 6 weeks (4 weeks). **Standard care (non-organised) or no treatment** (n=17)	**Knee osteoarthritis** Mean age (SD): 55.5 (8.6) years N = 34 Definition: Physician-confirmed diagnosis of knee osteoarthritis, or pass 2 criteria for early osteoarthritis (being age 50 years or older and having experience pain or discomfort in or around the knee during the previous year lasting 28 or more separate or consecutive days). 98% also met the American College of Rheumatology clinical criteria for knee osteoarthritis.	Quality of life at ≤3 months Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Delayed intervention **Concomitant therapy:** No additional information	Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear		
Mecklenburg 2018[186]	**Treatment package – Exercise and behaviour change intervention** (n=101) Hinge Health 12 weeks digital care package for chronic knee pain. People used a tablet computer with an application on it and sensors to complete exercise instructions, read education articles, achieve weight loss and complete cognitive behavioural therapy on specific weeks. Length of package: > 6 weeks (12 weeks). **Standard care (non-organised) or no treatment** (n=61) Usual care and access to three education articles. **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SD): 46 (12) years N = 162 Definition: Knee pain for at least 1 month in the past 12 months. Severity: Not stated Duration of symptoms: At least 1 month in the past 12 weeks Presence of multimorbidities: Not stated / Unclear	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	This study was noted to have serious population indirectness (the population had chronic knee pain but no clear statement about the presence of osteoarthritis)
Nunez 2006[208]	**Treatment package – Exercise and behaviour change intervention** (n=51)	**Knee osteoarthritis** Mean age (SD): 71.1 (6.7) years N = 100	Quality of life at >3 months Pain at >3 months Physical function at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Strengthening exercises (twice a day for knee exercises and once a day for general exercises) and self-management training, including two 30 minute visits at the first week and 3 months, and two group sessions for around 90 miutes at weeks 3 and 4, with a maximum of 12 people. The sessions discussed energy conservation, joint protection, evaluation and control of pain, use of assistive devices, and general exercises. Length of package: > 6 weeks (12 weeks). **Standard care (non-organised) or no treatment** (n=49) Standard care only. **Concomitant therapy:** Both groups received 3-4g/day of paracetamol alone or no more than 2g/day of paracetamol combined with 2400mg/day of ibuprofen or other NSAIDs (the dose of NSAIDs varying according to individual patient needs).	Definition: People with knee osteoarthritis according to the Kellgren and Lawrence criteria. Severity: Not stated Duration of symptoms (mean [SD]): 11.8 (10.6) months Presence of multimorbidities: Not stated / Unclear	Discontinuation at >3 months	
Paterson 2021[222]	**Treatment package - Combination and behaviour change intervention** (n=15)	**Toe osteoarthritis** Mean age (SD): 59.0 (7.83) years N = 30	Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Foot orthoses (a wedged insole worn for >6 hours/day) and a self management program (wearing shoes with adequate depth and width; advice on analgesia, weight management). Length of package: > 6 weeks (3 months). **Standard care (non-organised) or no treatment**(n=15) Usual care only. **Concomitant therapy:** Usual care was provided to all. People in this treatment group attended one 1 5-minute visit with a GP at which they received advice and/or prescription of analgesics and anti-inflammatory medication at the discretion of the GP. In addition, the GP was also provided advice on weight management and physical activity. Participants were permitted additional visits if they experienced an ongoing problem related to the treatment, and this addition was documented by the GP.	Definition: First metatarsophalangeal osteoarthritis defined as radiographic osteoarthritis (a score of at least 2 for osteophytes or joint space narrowing on either the anteroposterior and lateral views, according to a radiographic atlas), self-reported pain at least 4 for an 11-point numerical rating scale in the corresponding first MTP joint region on most days of the previous month. Severity: Osteophyte grade 2-3, joint space narrowing grade 1-3 (median grade for both = 2). Duration of symptoms (mean [SD]): 8.5 (6.5) years Presence of multimorbidities: Not stated/ Unclear		
Poulsen 2013[228] Subsidiary papers: Poulsen 2011[227]	**Treatment package – Manual therapy and education programme** (n=43)	**Hip osteoarthritis** Mean age (SD): 64.6 (8.6) years N = 118	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Poulsen 2013[229]	Hip school and manual therapy. Hip school involved 5 sessions delivered over 6 weeks with two individual sessions and three group sessions. This included information about epidemiology, anatomy of the hip, pain, activity, natural course of the disease and treatment options. Stretching exercises were taught. The manual therapy included a combination of manual soft tissue therapy, stretching and joint manipulation. Length of package: ≤ 6 weeks (6 weeks). **Education programme only** (n=39) **Standard care (non-organised) or no treatment** (n=36) Minimal intervention. People were given a leaflet describing the stretching exercises from hip school and received a short 5 minute instruction in self-care immediately after randomisation. People were advised to live as ususal, not to make any changes to use of possible pain medication or to initiate any other treatment during the following 6 weeks.	Definition: Unilateral hip pan for >3 months' duration with radiographic hip osteoarthritis defined as minimal joint space width (JSW) measurement <2.00mm or a side difference in minimal JSW >10% Severity: Not stated Duration of symptoms (mean [SD]): 32 (36) months Presence of multimorbidities: Not stated / Unclear	Physical function at ≤3 months and >3 months Discontinuation at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Concomitant therapy: No additional information			
Rezende 2021[236]	**Treatment package - Exercise and behaviour change intervention** (n=111) Two days of a structured educational and exercise-based self-management program that were held two months apart. Length of package: > 6 weeks (24 months). **Standard care (non-organised) or no treatment**(n=111) Usual care only. **Concomitant therapy:** People in both groups were seen by the orthopaedic surgeons at inclusion, six, 12 and 24 months. At inclusion people were already receiving diacerein and/or analgesics such as paracetamol, codeine and/or dipyrone that were prescribed by the physicians when people were first seen.	**Knee osteoarthritis** Mean age (SD): 63.5 (9.1) years N = 222 Definition: Knee osteoarthritis according to the American College of Rheumatology clinical and radiological definitions with Kellgren & Lawrence stages 1-3 Severity: Kellgren and Lawrence grace 1-3 (median grade 2). Duration of symptoms: Not stated/unclear Presence of multimorbidities: not stated/ unclear	Pain at >3 months Physical function at >3 months Discontinuation at >3 months	
Saw 2016[244]	**Treatment package – Exercise and education programme** (n=35) Strengthening exercises and light aerobic exercise combined with goal setting and education	**Mixed osteoarthritis (hip and/or knee)** Mean age (SD): 60.72 (5.54) years N = 71	Quality of life at ≤3 months and >3 months Pain at ≤3 months and >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	regarding: knowledge of understanding of osteoarthritis, pain neuroscience, activity, self-management skills, problem solving, goal setting, coping mechanisms, stress management and pacing. Length of package: ≤ 6 weeks (6 weeks). **Standard care (non-organised) or no treatment** (n=36) Usual care. **Concomitant therapy:** No additional information	Definition: People diagnosed with osteoarthritis who had been placed on the waiting list to receive a hip/knee arthroplasty Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Discontinuation at ≤3 months and >3 months	
Skou 2015[253]	**Treatment package – Combination and education programme** (n=50) MEDIC programme. Combination of education, exercise and insoles (for everyone), weight loss advice and pain medication (if indicated). Delivered over 12 weeks. Booster session between 20 weeks and 52 weeks. **Standard care (non-organised) or no treatment** (n=50) Usual care	**Knee osteoarthritis** Mean age (SD): 66.0 (9.0) years N = 100 Definition: Symptomatic and radiographically-confirmed knee osteoarthritis Severity: Kellgren Lawrence grade 1-4, median grade 3 Duration of symptoms: 0 months - more than 10 years, median 2-5 years. Presence of multimorbidities: Low morbidity score	Quality of life at >3 months Pain at >3 months Physical function at >3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
	Concomitant therapy: Both groups received two educational leaflets	(Charlson comorbidity index, 0->3. Median: 1.)		
Tak 2005[266]	**Treatment package – Exercise and education programme** (n=55) Hop with the Hip program, consisting of 8, 1 hour weekly group sessions of strength training using fitness equipment under supervision of physical therapists. People were provided with personal ergonomic advice and dietary advice. Length of package: > 6 weeks (8 weeks). **Standard care (non-organised) or no treatment** (n=54) No additional treatment apart from appointments organised by the individual. **Concomitant therapy:** No additional information	**Hip osteoarthritis** Mean age (SD): Intervention: 67.4 (7.6) years. Control: 68.9 years N = 109 Definition: The diagnosis of osteoarthritis of the hip had been made by the general practitioner and clinical symptoms, evaluated by physical therapists at baseline, meeting criteria for osteoarthritis of the hip of the American College of Rheumatology (pain in the hip together with endorotation of at least 15 degrees, pain present at endorotation of the hip, morning stiffness for no more than 60 minutes after rising, age >50 years. Severity: Not stated Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months Pain at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Wallis 2017[279]	**Treatment package – Exercise and behaviour change intervention** (n=23) A walking program of moderate intensity of 70 minutes per week, for at least 10 minutes per session for 12 weeks. This was combined with planning sessions with physiotherapists to discuss goals and encourage changes to improve activity. Social support was encouraged. Length of package: > 6 weeks (12 weeks). **Standard care (non-organised) or no treatment** (n=23) Usual care was non-operative management to manage pain and symptoms including pharmacological and non-pharmacological interventions. No new physical activity should be started in the 12 week period. **Concomitant therapy:** People continued taking their usual medications and other non-surgical treatments to manage their knee osteoarthritis, and used normal assistive devices such as a cane	**Knee osteoarthritis** Mean age (SD): 67.5 (7.5) years N = 46 Definition: Severe knee osteoarthritis rating grade III or IV affecting at least one of the tibiofemoral compartments determined radiographically. Severity: Radiographic grade III-IV, median grade IV Duration of symptoms: Not stated Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months Pain at ≤3 months Physical function at ≤3 months Discontinuation at ≤3 months	

Study	Intervention and comparison	Population	Outcomes	Comments
Yip 2007[288] Subsidiary papers: Yip 2007[287] Yip 2008[289]	**Treatment package – Exercise and education programme** (n=23) The Arthritis Self Management Program intervention, consisting of 6x 2 hour classes held once a week with 10-15 people to discuss the basic principles of self management, osteoarthritis symptoms, joint protection, available treatments, managing stress, nutrition and communication skills. Exercise consisted of three types: stretching, walking and Tai Chi. People set exercise goals and received positive feedback by a nurse every week. Length of package: ≤ 6 weeks (6 weeks). **Standard care (non-organised) or no treatment** (n=94) Routine orthopaedic treatment (treatment prescribed by orthopaedic doctors or outpatient clinic) with no other treatment. **Concomitant therapy:** No additional information	**Knee osteoarthritis** Mean age (SE): Intervention: 65.60 (1.03) years. Control: 64.02 (1.06) years. N = 182 Definition: Diagnosed based on the clinical criteria of the American College of Rheumatology Severity: Not stated Duration of symptom (mean [SE]): Intervention: 8.31 (0.78). Control: 7.85 (0.65). Presence of multimorbidities: Not stated / Unclear	Quality of life at ≤3 months Pain at ≤3 months and >3 months Discontinuation at ≤3 months	

See Appendix D for full evidence tables.

1.1.6 Summary of the effectiveness evidence

1.1.6.1 Treatment packages compared to exercise alone

Table 8: Clinical evidence summary: treatment packages compared to exercise alone

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with exercise alone	Risk difference with treatment packages	
Quality of life (AQOL II, -0.11-1, high is good, change score) at ≤3 months	148 (1 RCT) follow up: 12 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean quality of life was 0.1	MD **0** (0.05 lower to 0.05 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS pain, 0-10, high is poor, final value) at ≤3 months	36 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 3.19	MD **1.07 higher** (0.04 lower to 2.18 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS psychological disability, 0-10, high is poor, final value) at ≤3 months	36 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 1.88	MD **0.33 higher** (0.35 lower to 1.01 higher)	MID = 0.5 SD (SMD)
Quality of life (AQOL II, -0.11-1, high is good, change score) at >3 months	421 (3 RCTs) follow up: mean 14 months	⊕⊕⊕◯ MODERATE [a]	-	The mean quality of life was 0.04	MD **0** (0.02 lower to 0.02 higher)	MID = 0.05 (0.5 x median baseline SD)
Quality of life (KOOS, 0-100, high is good, change score) at >3 months	110 (1 RCT) follow up: 24 weeks	⊕⊕⊕⊕ HIGH	-	The mean quality of life was -2.3	MD **0.1 higher** (7.31 lower to 7.51 higher)	MID = 0.5 SD (SMD)
Quality of life (SF-36 physical component, 0-100, high is good, final values) at >3 months	223 (2 RCTs) follow up: mean 18 months	⊕◯◯◯ VERY LOW [a,b,c]	-	-	MD **0.76 higher** (3.7 lower to 5.22 higher)	MID = 2 (established value)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with exercise alone	Risk difference with treatment packages	
Quality of life (SF-36 mental component, 0-100, high is good, final values) at >3 months	223 (2 RCTs) follow up: mean 18 months	⊕⊕◯◯ LOW [a]	-	-	MD **0.25 higher** (1.74 lower to 2.25 higher)	MID = 3 (established value)
Pain (WOMAC, 0-20, high is poor, change scores) at ≤3 months	190 (2 RCTs) follow up: mean 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean pain was -2.6	MD **1.07 lower** (1.69 lower to 0.45 lower)	MID = 0.5 SD (SMD)
Pain (WOMAC, NRS [different scale ranges], high is poor, final values) at ≤3 months	274 (3 RCT) follow up: mean 12 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	-	SMD **0.1 SD lower** (0.71 lower to 0.51 higher)	MID = 0.5 SD (SMD)
Pain (KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months	686 (5 RCTs) follow up: mean 57 weeks	⊕⊕⊕◯ MODERATE [a]	-	-	SMD **0.13 SD lower** (0.28 lower to 0.02 higher)	MID = 0.5 SD (SMD)
Pain (WOMAC, VAS [different scale ranges], high is poor, final values) at >3 months	367 (3 RCTs) follow up: mean 13 months	⊕⊕◯◯ LOW [a]	-	-	SMD **0.04 higher** (0.17 lower to 0.24 higher)	MID = 0.5 SD (SMD)
Physical function (WOMAC, 0-68, high is poor, change scores) at ≤3 months	190 (2 RCT) follow up: mean 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean physical function was -10.1	MD **3.8 lower** (5.3 lower to 2.3 lower)	MID = 0.5 SD (SMD)
Physical function (KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months	530 (4 RCTs) follow up: mean 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean physical function was -13.1	SMD **0.09 SD lower** (0.26 lower to 0.08 higher)	MID = 0.5 SD (SMD)
Physical function (WOMAC [different scale ranges], high is poor, final values) at >3 months	172 (2 RCTs)	⊕◯◯◯ VERY LOW [a,b]	-	-	SMD **0.24 SD higher**	MID = 0.5 SD (SMD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with exercise alone	Risk difference with treatment packages	
	follow up: mean 15 months				(0.06 lower to 0.54 higher)	
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at ≤3 months	148 (1 RCT) follow up: 12 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -1.1	MD **0.2 higher** (1.09 lower to 1.49 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at ≤3 months	148 (1 RCT) follow up: 12 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.7	MD **0.2 lower** (1.91 lower to 1.51 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at ≤3 months	148 (1 RCT) follow up: 12 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -1.5	MD **1 higher** (1.15 lower to 3.15 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at >3 months	292 (2 RCTs) follow up: mean 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.4	MD **0.15 lower** (0.54 lower to 0.23 higher)	MID = 2.25 (0.5 x median baseline SD)
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at >3 months	292 (2 RCTs) follow up: mean 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.3	MD **0.15 lower** (0.62 lower to 0.32 higher)	MID = 2.7 (0.5 x median baseline SD)
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at >3 months	292 (2 RCTs) follow up: mean 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.5	MD **0.24 lower** (0.72 lower to 0.24 higher)	MID = 2.8 (0.5 x median baseline SD)
Discontinuation at ≤3 months	706 (8 RCTs) follow up: mean 11 weeks	⊕◯◯◯ VERY LOW [a,b]	RR 0.75 (0.52 to 1.08)	153 per 1,000	**38 fewer per 1,000** (73 fewer to 12 more)	MID (precision) = RR 0.8-1.25.

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with exercise alone	Risk difference with treatment packages	
Discontinuation at >3 months	1472 (10 RCTs) follow up: mean 14 months	⊕⊕⊕◯ MODERATE [a]	RR 1.00 (0.82 to 1.22)	198 per 1,000	**0 fewer per 1,000** (36 fewer to 44 more)	MID (precision) = RR 0.8-1.25.

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs
c. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

1.1.6.2 Treatment packages compared to manual therapy alone

Table 9: Clinical evidence summary: treatment packages compared to manual therapy alone

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with manual therapy alone	Risk difference with treatment packages	
Pain (WOMAC, 0-500, high is poor, final value) at ≤3 months	55 (1 RCT) follow up: 5 weeks	⊕⊕◯◯ LOW [a,b]	-	The mean pain was 102.3	MD **4.6 lower** (51.06 lower to 41.86 higher)	MID = 0.5 SD (SMD)
Physical function (WOMAC, 0-1800, high is poor, final value) at ≤3 months	55 (1 RCT) follow up: 5 weeks	⊕⊕◯◯ LOW [a,b]	-	The mean physical function was 389.7	MD **10.8 lower** (157.76 lower to 136.16 higher)	MID = 0.5 SD (SMD)
Discontinuation at ≤3 months	55 (1 RCT)	⊕⊕◯◯ LOW [b]	RR 1.93 (0.19 to 20.05)	37 per 1,000	**34 more per 1,000** (30 fewer to 706 more)	MID (precision) = RR 0.8-1.25.

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with manual therapy alone	Risk difference with treatment packages	
	follow up: 5 weeks					

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

1.1.6.3 Treatment packages compared to electrotherapy alone

Table 10: Clinical evidence summary: treatment packages compared to electrotherapy alone

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with electrotherapy alone	Risk difference with treatment packages	
Pain (VAS, 0-10, high is poor, change score) at ≤3 months	84 (1 RCT) follow up: 12 weeks	⨁⨁◯◯ LOW [a]	-	The mean pain was -1.9	MD **2.1 lower** (2.89 lower to 1.31 lower)	MID = 0.5 SD (SMD)

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

1.1.6.5 Treatment packages compared to behaviour change interventions alone

Table 11: Clinical evidence summary: treatment packages compared to behaviour change interventions alone

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with behaviour change interventions alone	Risk difference with treatment packages	
Quality of life (AQOL II, -0.04-1, high is good, change score) at ≤3 months	147 (1 RCT) follow up: 12 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean quality of life was 0.1	MD **0** (0.03 lower to 0.03 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS pain, 0-10, high is poor, final value) at ≤3 months	38 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 4	MD **0.26 higher** (0.7 lower to 1.22 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS psychological distress, 0-10, high is poor, final value) at ≤3 months	38 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 2.38	MD **0.17 lower** (1 lower to 0.66 higher)	MID = 0.5 SD (SMD)
Quality of life (AQOL II, -0.04-1, high is good, change score) at >3 months	147 (1 RCT) follow up: 52 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean quality of life was 0.1	MD **0** (0.03 lower to 0.03 higher)	MID = 0.5 SD (SMD)
Quality of life (SF-36 physical composite, 0-100, high is good, final value) at >3 months	141 (1 RCT) follow up: 18 months	⊕◯◯◯ VERY LOW [a,b]	-	-	MD **2.16 higher** (0.16 lower to 4.48 higher)	MID = 2 (established value)
Quality of life (SF-36 mental composite, 0-100, high is good, final value) at >3 months	141 (1 RCT) follow up: 18 months	⊕⊕◯◯ LOW [a]	-	-	MD **0.55 lower** (2.77 lower to 1.67 higher)	MID = 0.5 SD (SMD)
Pain (WOMAC, 0-20, high is poor, change scores) at ≤3 months	189 (2 RCTs) follow up: mean 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean pain was -2.37	MD **1.22 lower** (2.18 lower to 0.27 lower)	MID = 1.2 (0.5 x median baseline SD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with behaviour change interventions alone	Risk difference with treatment packages	
Pain (WOMAC, 0-500, high is poor, final value) at ≤3 months	198 (1 RCT) follow up: 12 weeks	⊕⊕◯◯ LOW [a]	-	The mean pain was 72	MD **4.9 lower** (23.72 lower to 13.92 higher)	MID = 0.5 SD (SMD)
Pain (WOMAC, 0-20, high is poor, change scores) at >3 months	305 (2 RCTs) follow up: mean 15 months	⊕⊕◯◯ LOW [a,b]	-	The mean pain was -1.84	MD **1.17 lower** (2 lower to 0.34 lower)	MID = 1.6 (0.5 x me baseline SD)
Pain (WOMAC, 0-500, high is poor, final value) at >3 months	198 (1 RCT) follow up: 9 months	⊕⊕◯◯ LOW [a]	-	The mean pain was 62.9	MD **6.7 lower** (28.49 lower to 15.09 higher)	MID = 0.5 SD (SMD)
Physical function (WOMAC, 0-68, high is poor, change score) at ≤3 months	189 (2 RCTs) follow up: mean 12 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	The mean physical function was -8.48	MD **5.65 lower** (11.36 lower to 0.07 higher)	MID = 3,2 (0.5 x median baseline SD)
Physical function (WOMAC, 0-68, high is poor, change score) at >3 months	147 (1 RCT) follow up: 12 months	⊕⊕◯◯ LOW [a,b]	-	The mean physical function was -12.3	MD **6.8 lower** (10.16 lower to 3.44 lower)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at ≤3 months	147 (1 RCT) follow up: 12 weeks	⊕⊕◯◯ LOW [a,b]	-	The mean psychological distress was -1.9	MD **1 higher** (0.33 lower to 2.33 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at ≤3 months	147 (1 RCT) follow up: 12 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.6	MD **0.3 lower** (2.11 lower to 1.51 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at ≤3 months	147 (1 RCT)	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.3	MD **0.2 lower** (2.09 lower to 1.69 higher)	MID = 0.5 SD (SMD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with behaviour change interventions alone	Risk difference with treatment packages	
	follow up: 12 weeks					
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at >3 months	147 (1 RCT) follow up: 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -2.1	MD **0.1 higher** (1.35 lower to 1.55 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at >3 months	147 (1 RCT) follow up: 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -0.9	MD **0.5 lower** (2.18 lower to 1.18 higher)	MID = 0.5 SD (SMD)
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at >3 months	147 (1 RCT) follow up: 12 months	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was -1.7	MD **0.4 lower** (2.5 lower to 1.7 higher)	MID = 0.5 SD (SMD)
Discontinuation at ≤3 months	318 (3 RCTs) follow up: mean 12 weeks	⊕⊕◯◯ LOW [b]	RR 0.66 (0.28 to 1.58)	76 per 1,000	**26 fewer per 1,000** (55 fewer to 44 more)	MID (precision) = RR 0.8-1.25.
Discontinuation at >3 months	812 (5 RCTs) follow up: mean 15 months	⊕⊕◯◯ LOW [a,b]	RR 1.15 (0.86 to 1.55)	164 per 1,000	**25 more per 1,000** (23 fewer to 90 more)	MID (precision) = RR 0.8-1.25.

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

c. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

1.1.6.4 Treatment packages compared to education programmes alone

Table 12: Clinical evidence summary: treatment packages compared to education programmes alone

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with education programmes alone	Risk difference with treatment packages	
Quality of life (EQ-5D 5L, -0.11-1, high is good, final value) at ≤3 months	167 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 0.61	MD **0.02 higher** (0.05 lower to 0.09 higher)	MID = 0.03 (established value)
Quality of life (HOOS, KOOS, 0-100, high is good, change scores and final value) at ≤3 months	173 (3 RCTs) follow up: mean 10 weeks	⊕⊕◯◯ LOW [a,b]	-	The mean quality of life was 18.9	MD **13.15 higher** (6.91 higher to 19.39 higher)	MID = 8.1 (0.5 x median baseline SD)
Quality of life (AIMS-2 pain subscale, 0-10, high is good, final value) at ≤3 months	38 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 5.99	MD **0.81 lower** (2.25 lower to 0.63 higher)	MID = 0.5 SD (SMD)
Quality of life (HOOS, KOOS, 0-100, high is good, change score and final value) at >3 months	144 (2 RCT) follow up: mean 11 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 31	MD **2.52 higher** (4.04 lower to 9.08 higher)	MID = 8.1 (0.5 x median baseline SD)
Quality of life (SF-36 physical function, 0-100, high is good, final value) at >3 months	75 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 71.3	MD **4.2 higher** (5.17 lower to 13.57 higher)	MID = 3 (established value)
Quality of life (SF-36 bodily pain, 0-100, high is good, final value) at >3 months	78 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 61.4	MD **9.1 higher** (0.58 lower to 18.78 higher)	MID = 3 (established value)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with education programmes alone	Risk difference with treatment packages	
Quality of life (SF-36 role physical, 0-100, high is good, final value) at >3 months	78 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 75.7	MD **6.6 higher** (5.58 lower to 18.78 higher)	MID = 3 (established value)
Quality of life (SF-36 vitality, 0-100, high is good, final value) at >3 months	78 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 61.7	MD **2.7 lower** (11.94 lower to 6.54 higher)	MID = 2 (established value)
Quality of life (SF-36 general health, 0-100, high is good, final value) at >3 months	74 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 67.6	MD **3.7 higher** (6.07 lower to 13.47 higher)	MID = 2 (established value)
Quality of life (SF-36 mental health, 0-100, high is good, final value) at >3 months	77 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 82.8	MD **1 lower** (7.78 lower to 5.78 higher)	MID = 3 (established value)
Quality of life (SF-36 role emotional, 0-100, high is good, final value) at >3 months	78 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 90.5	MD **0.2 higher** (8.25 lower to 8.65 higher)	MID = 4 (established value)
Quality of life (SF-36 social function, 0-100, high is good, final value) at >3 months	78 (1 RCT) follow up: 16 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean quality of life was 84.1	MD **7.1 higher** (2.84 lower to 17.04 higher)	MID = 3 (established value)
Pain (HOOS, KOOS, WOMAC, VAS, 0-100, high is good, change scores) at ≤3 months	440 (6 RCTs) follow up: mean 10 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	The mean pain was 23.5	MD **11.31 higher** (5.87 higher to 16.74 higher)	MID = 7.9 (0.5 x median baseline SD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with education programmes alone	Risk difference with treatment packages	
Pain (KOOS, AUSCAN, McGill Pain Questionnaire, pain rating index [different scale ranges], high is poor, final values) at ≤3 months	291 (4 RCTs) follow up: mean 12 weeks	⊕⊕⊕◯ MODERATE	-	-	SMD **0.15 SD higher** (0.08 lower to 0.38 higher)	MID = 0.5 SD (SMD)
Pain (HOOS, KOOS, WOMAC, 0-100, high is poor, change score and final values) at >3 months	222 (3 RCTs) follow up: mean 12 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean pain was 35.6	MD **3.81 lower** (8.41 lower to 0.79 higher)	MID = 7.7 (0.5 x median baseline SD)
Pain (WOMAC, McGill Pain Questionnaire, pain rating index [different scale ranges], high is poor, final values) at >3 months	142 (4 RCTs) follow up: mean 21 weeks	⊕◯◯◯ VERY LOW [a,b]	-	-	SMD **0.09 SD higher** (0.66 lower to 0.83 higher)	MID = 0.5 SD (SMD)
Physical function (HOOS, WOMAC, 0-100, high is good, change scores) at ≤3 months	246 (4 RCTs) follow up: mean 9 weeks	⊕⊕◯◯ LOW [a,b]	-	The mean physical function was 22.4	MD **11.08 higher** (7.66 higher to 14.5 higher)	MID = 8.2 (0.5 x median baseline SD)
Physical function (AUSCAN, Functional Index of Hand Osteoarthritis [different scale ranges], high is good, final values) at ≤3 months	363 (2 RCTs) follow up: mean 12 weeks	⊕◯◯◯ VERY LOW [b,c]	-	-	SMD **0.21 SD higher** (0.23 lower to 0.65 higher)	MID = 0.5 SD (SMD)
Physical function (HOOS, WOMAC, 0-100, high is poor, change score and final values) at >3 months	219 (3 RCTs) follow up: mean 12 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean physical function was 36.5	MD **5.59 lower** (10.18 lower to 1 lower)	MID = 7.8 (0.5 x median baseline SD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with education programmes alone	Risk difference with treatment packages	
Physical function (WOMAC, 0-68, high is poor, final value) at >3 months	32 (1 RCT) follow up: 5 months	⊕◯◯◯ VERY LOW [a,b]	-	The mean physical function was 30.89	MD **14.67 lower** (24.11 lower to 5.13 lower)	MID = 0.5 SD (SMD)
Discontinuation at ≤3 months	738 (7 RCTs) follow up: mean 10 weeks	⊕◯◯◯ VERY LOW [d,e]	RD 0.01 (-0.04 to 0.06)	144 per 1,000	**10 fewer per 1,000** (40 fewer to 60 more) [f]	Precision calculated through Optimal Information Size (OIS) due to zero events in some studies (0.8-0.9 = serious, <0.8 = very serious).
Discontinuation at >3 months	419 (6 RCTs) follow up: mean 9 months	⊕◯◯◯ VERY LOW [a,b]	RR 0.68 (0.51 to 0.92)	339 per 1,000	**109 fewer per 1,000** (166 fewer to 27 fewer)	MID (precision) = RR 0.8-1.25.

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

c. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

d. Downgraded for heterogeneity due to conflicting number of events in different studies (zero events in one or more studies)

e. Downgraded by 1 to 2 increments for imprecision due to zero events and small sample size

f. Absolute effect calculated by risk difference due to zero events in at least one arm of one study

1.1.6.6 Treatment packages compared to standard care (non-organised) or no treatment

Table 13: Clinical evidence summary: treatment packages compared to standard care (non-organised) or no treatment

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Quality of life (EQ-5D, AQoL-2, -0.11-1, high is good, change scores and final values) at ≤3 months	589 (6 RCTs) follow up: mean 11 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	-	MD **0.08 higher** (0.02 higher to 0.13 higher)	MID = 0.075 (0.5 x median baseline SD)
Quality of life (KOOS, HOOS, VAS quality of life, health assessment questionnaire, 0-100, high is good, change score and final values) at ≤3 months	569 (5 RCTs) follow up: mean 7 weeks	⊕◯◯◯ VERY LOW [a,b]	-	-	**MD 2.56 higher** (1.86 lower to 6.97 higher)	MID = 7.9 (0.5 x median baseline SD)
Quality of life (Health related quality of life, 7-39, high is good, final value) at ≤3 months	109 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 27.3	MD **1.3 higher** (0.11 higher to 2.49 higher)	MID = 0.5 SD (SMD)
Quality of life (SF-36 physical component, 0-100, high is good, change scores) at ≤3 months	259 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was -0.76	MD **0.95 higher** (1.16 lower to 3.06 higher)	MID = 2 (established value)
Quality of life (SF-36 mental component, 0-100, high is good, change scores) at ≤3 months	259 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was -1.7	MD **2.56 higher** (0.78 higher to 4.34 higher)	MID = 3 (established value)
Quality of life (Geri-AIMS pain subscale, AIMS pain, 0-10, high is good, final values) at ≤3 months	345 (3 RCTs) follow up: mean 9 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	The mean quality of life was 4.5	MD **0.36 higher** (0.3 lower to 1.01 higher)	MID = 0.7 (0.5 x median baseline SD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Quality of life (AIMS psychological disability, 0-10, high is poor, final value) at ≤3 months	38 (1 RCT) follow up: 12 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 1.8	MD **0.41 higher** (0.31 lower to 1.13 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS arthritis impact, 0-10, high is poor, final value) at ≤3 months	92 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 3.06	**MD 0.2 lower** (0.97 lower to 0.57 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS physical activity, 0-10, high is poor, final value) at ≤3 months	92 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 5.96	MD **2.22 lower** (3.25 lower to 1.19 lower)	MID = 0.5 SD (SMD)
Quality of life (AIMS medications use, 0-6, high is good, final value) at ≤3 months	92 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 2.9	MD **0.74 higher** (0.07 lower to 1.55 higher)	MID = 0.5 SD (SMD)
Quality of life (SF-36 physical function, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 51.33	MD **14 higher** (1.76 higher to 26.24 higher)	MID = 3 (established value)
Quality of life (SF-36 bodily pain, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 42.8	MD **14.8 higher** (2.21 higher to 27.39 higher)	MID = 3 (established value)
Quality of life (SF-36 role physical, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕⊕◯◯ LOW [a]	-	The mean quality of life was 35	MD **53.33 higher** (30.56 higher to 76.1 higher)	MID = 3 (established value)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Quality of life (SF-36 vitality, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 58.33	MD **13.67 higher** (2.3 higher to 25.04 higher)	MID = 2 (established value)
Quality of life (SF-36 general health, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 55.27	MD **13.73 higher** (0.68 higher to 26.78 higher)	MID = 2 (established value)
Quality of life (SF-36 mental health, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 61.07	MD **14.13 higher** (0.09 lower to 28.35 higher)	MID = 3 (established value)
Quality of life (SF-36 role emotional, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕⊕◯◯ LOW [a]	-	The mean quality of life was 53.2	MD **33.47 higher** (10.78 higher to 56.16 higher)	MID = 4 (established value)
Quality of life (SF-36 social function, 0-100, high is good, change scores) at ≤3 months	30 (1 RCT) follow up: 8 weeks	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 90.83	MD **0.84 higher** (8.46 lower to 10.14 higher)	MID = 3 (established value)
Quality of life (EQ-5D, AQoL-2, -0.11-1, high is good, change scores and final values) at >3 months	1001 (7 RCTs) follow up: mean 9 months	⊕◯◯◯ VERY LOW [a,b,c]	-	-	MD **0.05 higher** (0.01 higher to 0.1 higher)	MID = 0.076 (0.5 x median baseline SD)
Quality of life (KOOS, HOOS, VAS quality of life, 0-100, high is good,	313 (3 RCTs) follow up:	⊕⊕◯◯ LOW [a]	-	-	MD **1.67 lower** (6.81 lower to 3.46 higher)	MID = 8.1 (0.5 x median baseline SD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
change score and final values) at >3 months	mean 10 months					
Quality of life (SF-36 physical component, 0-100, high is good, change scores) at >3 months	78 (1 RCT) follow up: 18 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 45.149	MD **4.24 lower** (8.67 lower to 0.19 higher)	MID = 2 (established value)
Quality of life (SF-36 mental component, 0-100, high is good, change scores) at >3 months	78 (1 RCT) follow up: 18 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 53.101	MD **0.82 higher** (3.41 lower to 5.06 higher)	MID = 3 (established value)
Quality of life (Geri-AIMS pain subscale, AIMS pain, 0-10, high is good, final values) at >3 months	267 (2 RCTs) follow up: mean 12 months	⊕⊕◯◯ LOW [a]	-	The mean quality of life was 5.1	MD **0.19 higher** (0.04 lower to 0.42 higher)	MID = 0.74 (0.5 x median baseline SD)
Quality of life (AIMS arthritis impact, 0-10, high is poor, final value) at >3 months	52 (1 RCT) follow up: 12 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 3.8	MD **0.55 lower** (1.82 lower to 0.72 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS physical activity, 0-10, high is poor, final value) at >3 months	52 (1 RCT) follow up: 12 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 6.18	MD **0.11 lower** (1.66 lower to 1.44 higher)	MID = 0.5 SD (SMD)
Quality of life (AIMS general health perception, 0-10, high is poor, final value) at >3 months	52 (1 RCT) follow up: 12 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 3.26	MD **0.45 higher** (0.82 lower to 1.72 higher)	MID = 0.5 SD (SMD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Quality of life (AIMS medications use, 1-6, high is good, final value) at >3 months	52 (1 RCT) follow up: 12 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 3.6	MD **0.26 lower** (1.47 lower to 0.95 higher)	MID = 0.5 SD (SMD)
Quality of life (SF-36 physical function, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 23.47	MD **3.73 higher** (3.94 lower to 11.4 higher)	MID = 3 (established value)
Quality of life (SF-36 bodily pain, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 30.33	MD **8.28 higher** (2.01 lower to 18.57 higher)	MID = 3 (established value)
Quality of life (SF-36 role physical, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 49.31	MD **14.55 lower** (33.57 lower to 4.47 higher)	MID = 3 (established value)
Quality of life (SF-36 vitality, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 54.58	MD **3.24 lower** (14.01 lower to 7.53 higher)	MID = 2 (established value)
Quality of life (SF-36 general health, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 56.42	MD **6.3 lower** (15.82 lower to 3.22 higher)	MID = 2 (established value)
Quality of life (SF-36 mental health, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 63.81	MD **6.54 lower** (17.52 lower to 4.44 higher)	MID = 3 (established value)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Quality of life (SF-36 role emotional, 0-100, high is good, final values) at >3 months	80 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 62.03	MD **4.32 lower** (25.07 lower to 16.43 higher)	MID = 4 (established value)
Quality of life (SF-36 social function, 0-100, high is good, final values) at >3 months	88 (1 RCT) follow up: 9 months	⊕◯◯◯ VERY LOW [a,c]	-	The mean quality of life was 62.53	MD **1.29 lower** (14.66 lower to 12.08 higher)	MID = 3 (established value)
Pain (HOOS, WOMAC, Foot Health Status Questionnaire, VAS [different scale ranges], high is poor, change scores) at ≤3 months	704 (6 RCTs) follow up: mean 10 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	-	SMD **0.53 SD lower** (0.93 lower to 0.13 lower)	MID = 0.5 SD (SMD)
Pain (KOOS, WOMAC, Lequesne index pain subscale, Harris Hip score pain subscale, BPI severity, VAS, Arthritis Self Efficacy Pain subscale [different scale ranges], high is poor, final values) at ≤3 months	1756 (15 RCTs) follow up: mean 8 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	-	SMD **0.35 SD lower** (0.6 lower to 0.1 lower)	MID = 0.5 SD (SMD)
Pain (HOOS, KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months	806 (5 RCTs) follow up: mean 12 months	⊕⊕⊕◯ MODERATE [a]	-	-	SMD **0.33 SD lower** (0.47 lower to 0.19 lower)	MID = 0.5 SD (SMD)
Pain (KOOS, WOMAC, BPI severity, VAS, Arthritis Self Efficacy Pain subscale [different scale ranges], high is poor, final values) at >3 months	1418 (13 RCTs) follow up: mean 12 months	⊕⊕◯◯ LOW [a]	-	-	SMD **0.18 SD lower** (0.28 lower to 0.07 lower)	MID = 0.5 SD (SMD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Physical function (HOOS, WOMAC, Foot Health Status Questionnaire Function Domain [different scale ranges], high is poor, change scores) at ≤3 months	630 (5 RCTs) follow up: mean 11 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	-	SMD **0.46 SD lower** (0.77 lower to 0.15 lower)	MID = 0.5 SD (SMD)
Physical function (KOOS, HOOS, WOMAC, Lequesne index function subscale, Physical Activity Scale for the Elderly [different scale ranges], high is poor, final values) at ≤3 months	1298 (11 RCTs) follow up: mean 8 weeks	⊕◯◯◯ VERY LOW [a,b,c]	-	-	SMD **0.46 SD lower** (0.77 lower to 0.15 lower)	MID = 0.5 SD (SMD)
Physical function (HOOS, KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months	655 (4 RCTs) follow up: mean 11 months	⊕◯◯◯ VERY LOW [a,c]	-	-	SMD **0.43 SD lower** (0.59 lower to 0.27 lower)	MID = 0.5 SD (SMD)
Physical function (KOOS, WOMAC, Physical Activity Scale for the Elderly [different scale ranges], high is poor, final values) at >3 months	1196 (10 RCTs) follow up: mean 14 months	⊕⊕◯◯ LOW [a]	-	-	SMD **0.16 SD lower** (0.28 lower to 0.04 lower)	MID = 0.5 SD (SMD)
Psychological distress (HADS anxiety, GAD-7 [different scale ranges], high is poor, final values) at <3 months	120 (3 RCTs) follow up: mean 6 weeks	⊕⊕⊕◯ MODERATE [a]	-	The mean psychological distress was 4.15	SMD **0.1 SD higher** (0.25 lower to 0.46 higher)	MID = 0.5 SD (SMD)
Psychological distress (HADS depression, PHQ-8 [different scale ranges], high is poor, final values) at <3 months	120 (3 RCTs) follow up: mean 6 weeks	⊕⊕◯◯ LOW [a,c]	-	The mean psychological distress was 2.95	SMD **0.25 SD lower** (0.61 lower to 0.11 higher)	MID = 0.5 SD (SMD)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects		Comments
				Risk with standard care (non-organised) or no treatment	Risk difference with treatment packages	
Psychological distress (HADS anxiety, GAD-7 [different scale ranges], high is poor, final values) at >3 months	462 (4 RCTs) follow up: mean 9 months	⊕⊕◯◯ LOW [a]	-	-	SMD **0.17 SD higher** (0.12 lower to 0.26 higher)	MID = 0.5 SD (SMD)
Psychological distress (HADS depression, PHQ-8 [different scale ranges], high is poor, final values) at >3 months	462 (4 RCTs) follow up: mean 9 months	⊕⊕◯◯ LOW [a]	-	-	SMD **0.07 SD higher** (0.12 lower to 0.26 higher)	MID = 0.5 SD (SMD)
Discontinuation at ≤3 months	2794 (21 RCTs) follow up: mean 10 weeks	⊕◯◯◯ VERY LOW [a,b]	RD -0.03 (-0.09 to 0.02)	211 per 1,000	**30 fewer per 1,000** (90 fewer to 20 more) [d]	Precision calculated through Optimal Information Size (OIS) due to zero events in some studies (0.8-0.9 = serious, <0.8 = very serious).
Discontinuation at >3 months	2430 (15 RCTs) follow up: mean 13 months	⊕◯◯◯ VERY LOW [a,b,c]	RR 0.96 (0.79 to 1.17)	278 per 1,000	**11 fewer per 1,000** (58 fewer to 47 more)	MID (precision) = RR 0.8-1.25.

[a]. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

[b]. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

[c]. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

[d]. Absolute effect calculated by risk difference due to zero events in at least one arm of one study

See Appendix F for full GRADE tables.

1.1.7 Economic evidence

1.1.7.1 Included studies

Four health economic studies with relevant comparisons were included in this review: one comparing treatment packages to exercise alone [37]; and three comparing treatment packages to usual care.[118, 144, 179] These are summarised in the health economic evidence profiles below (Table 14 and Table 15) and the health economic evidence tables in Appendix H.

1.1.7.2 Excluded studies

No relevant health economic studies were excluded due to limited applicability or methodological limitations.

See also the health economic study selection flow chart in Appendix G.

1.1.8 Summary of included economic evidence

Table 14: Health economic evidence profile: Treatment packages compared to exercise alone

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Bennell 2016 [37] (Australia)	Partially applicable[a]	Potentially serious limitations[b]	• Within-trial analysis • Cost-utility analysis (QALYs) • Population: Patients with knee osteoarthritis • Comparators: 1. Exercise alone 2. PCST alone 3. PCST with exercise • Time horizon: 52 weeks	(2−1): £133[c] (3−1): £285[c] (3−2): £152[c]	(2−1): 0.01 QALYs (3−1): 0.03 QALYs (3−2): 0.03 QALYs	(2−1): £13,300 per QALY gained (3−1): £9,500 per QALY gained (3−2): £5,067 per QALY gained	None reported

Abbreviations: PCST= pain coping skills training; QALY= quality-adjusted life years

(a) As the study is from an Australian perspective it has been judged as partially applicable.

(b) Within-trial analysis and so may not reflect full body of available evidence for this comparison; 1 of 12 studies included in the clinical review for treatment packages compared to exercise alone. The costs per QALY were not reported in the study and so were instead estimated using the reported incremental costs (converted to UK pounds) and QALYs. The incremental QALYs were reported to one significant figure which means the cost per QALY gained is subject to uncertainty. For example. the cost per QALY for intervention 2 vs intervention 1 could feasibly range between £9,500 and £27,000 with the addition of another decimal place.

(c) Converted using 2012 purchasing power parities[214]. Cost components incorporated: Therapy and other healthcare-related costs, excluding initial fixed cost of physical therapist training and impact on patient incomes or travel/time costs.

Table 15: Health economic evidence profile: Treatment packages compared to usual care alone

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Health Quality (Ontario HTA) 2018[118] (Canada)	Partially applicable[(a)]	Potentially serious limitations[(b)]	• Probabilistic decision analytical model • Cost-utility analysis (QALYs) • Population: Adults with knee OA • Comparators: 1. Usual care 2. Structured education and neuromuscular exercise program • Time horizon: 1 year	£407[(c)]	0.03 QALYs	£13,550 per QALY gained	Probability Intervention 2 cost effective (£28k/56K threshold): 81%/90% Intervention 2 remains cost effective at a 24-month time horizon
Jessep 2009[144] (UK)	Partially applicable[(d)]	Potentially serious limitations[(e)]	• Within-trial analysis of RCT (same paper) • Cost-utility analysis (QALYs) • Population: People with mild, moderate or severe non-specific chronic knee pain. • Comparators: 1. Outpatient physiotherapy (usual care) 2. ESCAPE knee pain (two exercise-based supervised sessions a week lasting 1 hour up	-£263[(f)]	0.08 QALYs	Intervention 2 dominates	None reported

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			to 5 weeks with educational material provided to take home) • Time horizon: 1 year				
Marra 2014[179] (Canada)	Partially applicable(g)	Potentially serious limitations(h)	• Probabilistic decision analytical model • Cost-utility analysis (QALYs) • Population: Patients with previously undiagnosed knee OA • Comparators: 1. Usual care 2. Healthcare professional package(i) • Time horizon: 6 months	Based on HUI3: £5(j) Based on PAT-5D: £3(j)	Based on HUI3: 0.0221 Based on PAT-5D: 0.0236	Based on HUI3): £254 per QALY gained Based on PAT-5D): £137 per QALY gained	Probability Intervention 2 cost effective (£1,200 threshold): 90%

Abbreviations: HUI3= The Health Utilities Index Mark 3; OA= osteoarthritis; PAT-5D= Paper Adaptive Test-5D; QALY= quality-adjusted life years

(a) As the study is from a Canadian perspective it has been judged as partially applicable.

(b) The clinical evidence was derived from a single RCT. The interventional cost estimates were based primarily on assumptions by experts. Costs and resource use for usual care were taken from a paper published in 2004. The incremental QALYs are reported to two decimal places which is subject to uncertainty (the cost per QALY could feasibly range between £12,000 and £16,000 with the addition of another decimal)

(c) Converted using 2017 purchasing power parities[214]. Cost components incorporated: Consultations with health care professionals, diagnostic tests and examinations, and hospitalisation

(d) Group sessions compared to individual sessions.

(e) Small study with only 67 participants were recruited at baseline. No analysis of uncertainty nor sensitivity analysis of results conducted. Health outcomes based on results from a single trial. The immediate cost of intervention 2 was nearly half that of intervention 1 and seems to be driven by the assumption that 6 participants will attend the complete programme in a group. Costs from 2005 may not reflect current UK NHS practice.

(f) Cost components incorporated: Healthcare utilisations costs included A&E, GP, nurse and outpatient visits, other primary care and medication costs.

(g) As the study is from a Canadian perspective it has been judged as partially applicable.

(h) Short time horizon of 6 months. It is unclear how unit costs were assigned to each component of resource utilisation. It is also unclear how the preference weights for utilities were valued and how QALYs were calculated.

(i) Screening questionnaire, education and pain medication management by a pharmacist, physiotherapy-guided exercise, and communication with primary care physician
(j) Converted using 2009 purchasing power parities[214]. Cost components incorporated: Physicians visits, treatments/ medications, laboratory tests and imaging.

1.1.9 Economic model

This area was not prioritised for economic modelling.

1.1.10 Economic considerations: trade-off between net clinical effects and costs

1.1.11 Economic evidence statements

- One cost-utility analysis reported that treatment packages (pain coping skills training) combined with exercise were cost effective versus treatment packages alone (ICER: £5,067) and exercise alone (ICER: £9,500). Treatment packages alone were also cost effective versus exercise alone (ICER: £13,300). This analysis was graded as partially applicable with potentially serious limitations.
- One cost-utility analysis reported that a structured education and neuromuscular programme was cost effective versus usual care (ICER: £13,550). This analysis was graded as partially applicable with potentially serious limitations.
- One cost-utility analysis reported that a group-based supervised exercise programme along with educational material dominated individual outpatient physiotherapy. This analysis was graded as partially applicable with potentially serious limitations.
- One cost-utility analysis reported that a healthcare professional package consisting of a screening questionnaire, education and pain medication management by a pharmacist, physiotherapy-guided exercise, and communication with primary care physician was cost effective versus usual care (ICER: £254 with HUI3 and £137 with PAT-5D). This analysis was graded as partially applicable with potentially serious limitations.

1.1.12 The committee's discussion and interpretation of the evidence

1.1.12.1. The outcomes that matter most

The critical outcomes were health-related quality of life, pain and physical function. These were considered critical due to their relevance to people with osteoarthritis. The Osteoarthritis Research Society International (OARSI) consider that pain and physical function were the most important outcomes for evaluating interventions. Health-related quality of life gives a broader perspective on the person's wellbeing, allowing for examination of the biopsychosocial impact of interventions. The important outcomes were psychological distress, osteoarthritis flare and discontinuation. Discontinuation events were included for this review as a measure of the tolerability of the treatment package compared to the individual components and standard care.

The committee considered osteoarthritis flares to be important in the lived experience and management of osteoarthritis. However, these were also considered difficult to measure with no clear consensus on their definition. The Flares in OA OMERACT working group have proposed an initial definition and domains of OA flares through a consensus exercise; "it is a transient state, different from the usual state of the condition, with a duration of a few days, characterized by onset, worsening of pain, swelling, stiffness, impact on sleep, activity, functioning, and psychological aspects that can resolve spontaneously or lead to a need to adjust therapy." However, this has been considered to have limitations and has not been widely adopted. Therefore, the committee included the outcome accepting any reasonable definition provided by any studies discussing the event.

Mortality was not considered in the outcomes. Osteoarthritis as a disease process is not considered to cause mortality by itself and so any mortality was considered to either be due to the intervention or external factors. Given this, the committee did not feel that mortality required a specific outcome. Additionally, as this intervention included a combination of interventions that were included in other review questions, the committee agreed that there was unlikely to be additional risks of mortality from combining the interventions and so this did not need to be investigated separately. Finally, while mortality is not examined

separately, participants may be included in the discontinuation outcome due to mortality. The committee were informed where this was the case to inform their decision making.

There was no evidence available for osteoarthritis flares. The committee acknowledged this as an important outcome rather than a critical one and agreed that they could make recommendations even though there was limited information for this outcome. While there was evidence available for other outcomes, there was only limited evidence available for psychological distress throughout the literature.

1.1.12.2 The quality of the evidence

Fifty-five randomised-controlled trial studies were included in this review. Evidence was available comparing treatment packages to exercise alone, manual therapy alone, electrotherapy alone, behaviour change interventions alone, educational programmes alone and standard care (non-organised) or no treatment. There was no evidence comparing treatment packages to acupuncture alone, devices alone, pharmacological management alone and arthroscopic procedures alone. The evidence for treatment packages in some far included combinations with all of the previously listed interventions apart from arthroscopic procedures.

The quality of the evidence varied across comparisons and outcomes but was in general between moderate and very low quality. Outcomes were most often downgraded for risk of bias and imprecision. Where downgraded for risk of bias this was often for selection bias and/or performance bias (as it was not possible to blind participants and those delivering the intervention to the allocated treatment in most cases). On occasions outcomes were downgraded for inconsistency when there was heterogeneity in the results. Some studies included indirect populations. However, these studies were often in a minority out of the study included in an outcome, so this only rarely influenced the quality rating.

Treatment packages compared to exercise alone

Seventeen studies included the comparison of treatment packages to exercise alone. Evidence was generally of low quality, ranging from high to very low quality. Outcomes were often downgraded for risk of bias and imprecision. Two outcomes were downgraded for inconsistency that was not explained by subgroup analysis. The evidence was based on a limited number of studies for some outcomes (for example: quality of life, pain, physical function and psychological distress at less than and equal to 3 months).

Treatment packages compared to manual therapy alone

One study included the comparison of treatment packages to manual therapy alone. Evidence was of low quality. Outcomes were downgraded for imprecision and two outcomes were downgraded for risk of bias.

Treatment packages compared to electrotherapy alone

One study included the comparison of treatment packages to electrotherapy alone. Evidence was of low quality. The outcome was downgraded for risk of bias.

Treatment packages compared to behaviour change interventions alone

Eight studies included the comparison of treatment packages to behaviour change interventions alone. Evidence was generally of low quality, ranging from moderate to very low quality. Outcomes were downgraded for risk of bias and imprecision. One outcome was downgraded for inconsistency, with heterogeneity that could not be resolved by subgroup analysis. The evidence for all outcomes apart from discontinuation was based on limited evidence.

Treatment packages compared to education programmes alone

Thirteen studies included the comparison of treatment packages to education programmes alone. Evidence was generally of very low quality, ranging from moderate to very low quality. Outcomes were downgraded for risk of bias and imprecision. One outcome was downgraded for inconsistency, with heterogeneity that could not be resolved by subgroup analysis. Discontinuation at less than and equal to 3 months included zero events in one or more study arms of at least one study with a small sample size, so was downgraded for inconsistency and imprecision.

Treatment packages compared to standard care (non-organised) or no treatment

Thirty studies included the comparison of treatment packages to standard care (non-organised) or no treatment. Evidence was generally of very low quality, ranging from moderate to very low quality. Outcomes were downgraded for risk of bias and imprecision. Ten outcomes were downgraded for inconsistency, with heterogeneity that could not be resolved by subgroup analysis.

1.1.12.3 Benefits and harms

Key uncertainties

The committee noted the limited evidence for some interventions. While programmes with more than two interventions may include other interventions investigated in this guideline (for example: acupuncture, devices) there were no studies investigating them as the only component being combined with an educational programme or behaviour change intervention. The committee decided that the evidence was generalisable to these interventions, and so if an intervention showed a clinically important benefit by itself then it may also gain benefit from being provided in a treatment package as with the interventions investigated in this review.

The committee acknowledged the challenges of comparing interventions when combined. They noted that smaller effect sizes may be significant benefits when comparing treatment packages to active treatments (for example: exercise) and so acknowledged that there may be important benefits seen in the evidence that are difficult to interpret in this context.

Treatment packages compared to exercise alone

The results for this comparison showed, in general, no clinically important difference between the two interventions for all outcomes included (quality of life, pain, physical function, psychological distress and discontinuation) in both less than and more than 3 months. One exception was seen for quality of life, where one subscale of a measure showed a clinically important harm. However, this was based on the evidence from one very low quality outcome including one small study (n=36) and given the consistency in the rest of the evidence the committee did not consider this as strong evidence compared to the other outcomes.

The committee concluded that there was no difference between the active treatment and the treatment packages in the outcomes measured. However, they agreed that there were additional potential benefits to treatment packages in qualitative outcomes that would not have been found in this review (for example: motivation). The committee considered that some people may respond better to treatment packages than to treatment alone.

Treatment packages compared to manual therapy alone

The limited results for this comparison showed no clinically important difference between the two interventions for pain, physical function and discontinuation at less than and equal to 3 months.

The committee concluded that the evidence for this comparison was very limited. However, it was consistent in showing that treatment packages were not inferior to manual therapy alone and so may be useful for some people.

Treatment packages compared to electrotherapy alone

The limited results for this comparison showed a clinically important benefit between the two interventions for pain at less than and equal to 3 months.

The committee concluded that the evidence for this comparison was very limited and so it would be difficult to draw conclusions based on it. However, the evidence is consistent with other evidence that treatment packages may be useful for some people with osteoarthritis.

Treatment packages compared to behaviour change interventions alone

The results for this comparison showed a clinically important benefit in physical function at less than and more than 3 months, unclear effects on quality of life at more than 3 months, and pain at less than and equal to 3 months (with some outcomes showing clinically important benefits and others showing no clinically important difference) and no clinically important difference in quality of life at less than and equal to 3 months, pain at more than 3 months, psychological distress and discontinuation.

Treatment packages compared to education programmes alone

The results for this comparison showed a clinically important benefit in physical function at less than and equal to 3 months and discontinuation at more than 3 months, unclear effects on quality of life at more than 3 months, and pain at less than and equal to 3 months (with some outcomes showing clinically important benefits and others showing no clinically important difference) and no clinically important difference in quality of life at less than and equal to 3 months, pain at more than 3 months, physical function at more than 3 months, and discontinuation at less than and equal to 3 months.

The committee concluded that the evidence showed a possible benefit for treatment packages when compared to education programmes alone. As the committee would not consider providing an education programme alone for people with osteoarthritis (instead offering it as a part of treatment with other interventions, like exercise), this was consistent with clinical practice. They agreed that the evidence showed that treatment packages may have a benefit beyond the education programme itself.

Treatment packages compared to standard care (non-organised) or no treatment

The results for this comparison showed unclear effects on quality of life, pain and physical function at less than and more than 3 months (with some outcomes showing clinically important benefits, some showing no clinically important difference and others showing clinically important harms) and no clinically important difference in pain and physical function at more than 3 months and psychological distress and discontinuation at less than and more than 3 months.

The committee noted that the evidence showed inconsistent changes in quality of life, possible clinically important benefits for pain at less than and equal to 3 months and possible benefits in physical function at less than and more than 3 months. Otherwise, there was no clinically important difference observed in any other outcomes. When examining the quality of life information at less than and equal to 3 months, the committee agreed that benefits were observed in overall quality of life scales with a larger number of studies and participants contributing to the outcomes. While there were other outcomes using overall quality of life scale scores that showed no clinically important difference, they indicated a positive signal from the treatment that did not fulfil the threshold for clinical importance agreed by the committee, but still indicated a positive effect. The results for subscales of quality of life scores (such as SF-36) were more inconsistent. However, this evidence was based on

outcomes from one study with a small number of participants (n=80) and so, given the equal very low quality rating of all of these outcomes, the committee had greater confidence in these results. This was also true of the pain and physical function outcomes at less than and equal to and greater than 3 months, where outcomes in general showed a positive effect from treatment packages. However, these effects were insufficient to achieve a clinically important difference based on the minimally important differences agreed by the committee. The committee agreed that, given the complexities in combining trials that may include heterogenous interventions and comparisons, that this evidence indicated that there may be a benefit to providing care as treatment packages, including education and behaviour change approaches as required for the person. They concluded that this evidence showed that treatment packages could be an effective treatment when compared to standard care or no treatment.

Weighing up the clinical benefits and harms

The committee considered the need for additional research in this area. While they agreed that this was an area of interest, they agreed that due to how specific the programs are (and therefore how heterogenous they are to each other) that making a new research recommendation was unlikely to provide additional information that would change the recommendation in this guideline. Treatment packages should be considered on a case-by-case basis for their potential efficacy. The committee defined a treatment package as any treatment for osteoarthritis which could include exercise, manual therapy, devices and pharmacological treatments combined with any one of the following: behaviour change approaches, including ways to reduce pain and straining when using joints, pain coping skills training (including spouse-assisted coping skills training), goal setting; motivational coaching; weight management counselling and workplace risk counselling; and an education programme given by 1 or more healthcare professionals over multiple sessions, including those based on behavioural theory.

Overall, evidence showed that treatment packages had a clinically important benefit on physical function compared with education or behaviour change interventions alone and non-clinically important but consistent beneficial changes in quality of life, pain and physical function when compared to standard care. Economic evidence summarised in the next section also suggested treatment packages were cost effective. However, they showed no superiority to individual therapies (such as exercise, manual therapy and electrotherapy). The committee agreed that a person-centred approach is important. Additional education or behavioural change approaches may help some people achieve their goals, while others may not need this. Therefore, the committee recommended combining therapeutic exercise as part of a structured treatment package because this may be more suitable for some people and motivate them to continue with therapeutic exercise.

1.1.12.4 Cost effectiveness and resource use

Four economic evaluations were included in the review. All were in people with knee osteoarthritis.

The first study took a UK perspective and was based on a single-blind pragmatic randomised controlled trial. It had a follow-up of 1 year. The treatment package was two supervised exercise sessions a week over 5 weeks while the comparator was outpatient physiotherapy with a maximum of 10 sessions. The study itself was small with only 67 patients recruited at baseline. QALYs were reported using the EQ-5D measure. An important difference between the two arms related to costs. The intervention was delivered in a group setting while the comparator was not. Since costs were reported on a per patient basis, the intervention was calculated to be cheaper than the comparator. Cost per QALY results were presented, however there were no sensitivity analyses nor analysis of uncertainty. The study was graded as partially applicable with potentially serious limitations.

The second study took an Australian perspective with a time horizon of 1 year. Pain coping skills training alone and in combination with exercise were compared to exercise alone. QALYs were captured using the AQoL-6D. The incremental QALYs were reported to one significant figure only and the addition of another significant figure resulted in vast variations in the final cost per QALY. The study did not report final cost per QALYs, so these were calculated from the available data. This study was graded as being partially applicable with potentially serious limitations.

The other two studies took a Canadian perspective. One study compared structured education and neuromuscular exercise to usual care (defined as educational pamphlets about knee osteoarthritis with the option of pain medication) while the other compared treatment management by various healthcare professionals to usual care (defined as an educational pamphlet on knee osteoarthritis by The Arthritis Society). The first study calculated QALYs using the EQ-5D measure while the other study collected this data using both the HUI3 and the PAD-5D. Both studies were graded as partially applicable. The time horizon in the first study was 1 year. The costs for the intervention were estimates based primarily on expert consultation and group-based programmes while costs for the comparator arm were taken from a study published in 2004. The second study had a time horizon of 6 months. While it defined resource use associated with the treatments, it was unclear how unit costs were assigned to each component of resource use. It was also unclear how the second study valued preference weights for utilities and how QALYs were calculated. For these reasons, both studies were deemed to have potentially serious limitations.

The first study reported that treatment packages dominated outpatient physiotherapy, being cheaper and more effective. The other three studies reported that treatment packages were cost effective at a threshold of £20,000 per QALY gained.

1.1.12.5 Other factors the committee took into account

The committee noted that the research identified does not appear to represent the diverse population of people with osteoarthritis. While they did not make a research recommendation for this review, they agreed that any further research should be representative of the population, including people from different family backgrounds, and socioeconomic backgrounds, disabled people, and people of different ages and genders. Future work should be done to consider the different experiences of people from diverse communities to ensure that the approach taken can be made equitable for everyone.

1.1.13 Recommendations supported by this evidence review

This evidence review supports recommendation 1.3.4. Other evidence supporting these recommendations can be found in evidence review K.

1.1.14 References

1. Ackerman IN, Buchbinder R, Osborne RH. Challenges in evaluating an Arthritis Self-Management Program for people with hip and knee osteoarthritis in real-world clinical settings. Journal of Rheumatology. 2012; 39(5):1047-1055

2. Ackerman IN, Buchbinder R, Osborne RH. Factors limiting participation in arthritis self-management programmes: an exploration of barriers and patient preferences within a randomized controlled trial. Rheumatology. 2013; 52(3):472-479

3. Adams J, Barratt P, Arden NK, Barbosa Boucas S, Bradley S, Doherty M et al. The Osteoarthritis Thumb Therapy (OTTER) II Trial: a study protocol for a three-arm multi-centre randomised placebo controlled trial of the clinical effectiveness and efficacy and cost-effectiveness of splints for symptomatic thumb base osteoarthritis. BMJ Open. 2019; 9(10):e028342

4. Adams J, Barratt P, Rombach I, Arden N, Barbosa Boucas S, Bradley S et al. The clinical and cost effectiveness of splints for thumb base osteoarthritis: a randomized controlled clinical trial. Rheumatology. 2021; 60(6):2862-2877

5. Aglamis B, Toraman NF, Yaman H. The effect of a 12-week supervised multicomponent exercise program on knee OA in Turkish women. Journal of Back and Musculoskeletal Rehabilitation. 2008; 21(2):121-128

6. Ahern M, Skyllas J, Wajon A, Hush J. The effectiveness of physical therapies for patients with base of thumb osteoarthritis: Systematic review and meta-analysis. Musculoskeletal Science & Practice. 2018; 35:46-54

7. Alasfour M, Almarwani M. The effect of innovative smartphone application on adherence to a home-based exercise programs for female older adults with knee osteoarthritis in Saudi Arabia: a randomized controlled trial. Disability and Rehabilitation. 2020:1-8

8. Alfieri FM, Lima ARS, Salgueiro M, Andrade EA, Battistella LR, Silva N. Efficacy of an exercise program combined with lifestyle education in patients with knee osteoarthritis. Acta Reumatologica Portuguesa. 2020; 45(3):201-206

9. Alfredo PP, Bjordal JM, Dreyer SH, Meneses SR, Zaguetti G, Ovanessian V et al. Efficacy of low level laser therapy associated with exercises in knee osteoarthritis: a randomized double-blind study. Clinical Rehabilitation. 2012; 26(6):523-533

10. Ali SA, Kokorelias KM, MacDermid JC, Kloseck M. Education and Social Support as Key Factors in Osteoarthritis Management Programs: A Scoping Review. Arthritis. 2018; 2018:2496190

11. Allegrante JP, Kovar PA, MacKenzie CR, Peterson MG, Gutin B. USA impact of a supervised walking program and education on functional status: results from a controlled trial in patients with osteoarthritis of the knee. Patient Education and Counseling. 1991; 18:283-284

12. Allen KD, Arbeeva L, Callahan LF, Golightly YM, Goode AP, Heiderscheit BC et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis and Cartilage. 2018; 26(3):383-396

13. Allen KD, Bongiorni D, Bosworth HB, Coffman CJ, Datta SK, Edelman D et al. Group versus individual physical therapy for veterans with knee osteoarthritis: Randomized clinical trial. Physical Therapy. 2016; 96(5):597-608

14. Allen KD, Bosworth HB, Brock DS, Chapman JG, Chatterjee R, Coffman CJ et al. Patient and provider interventions for managing osteoarthritis in primary care: protocols for two randomized controlled trials. BMC Musculoskeletal Disorders. 2012; 13:60

15. Allen KD, Oddone EZ, Coffman CJ, Datta SK, Juntilla KA, Lindquist JH et al. Telephone-based self-management of osteoarthritis: A randomized trial. Annals of Internal Medicine. 2010; 153(9):570-579

16. Allen KD, Oddone EZ, Coffman CJ, Jeffreys AS, Bosworth HB, Chatterjee R et al. Patient, Provider, and Combined Interventions for Managing Osteoarthritis in Primary Care: A Cluster Randomized Trial. Annals of Internal Medicine. 2017; 166(6):401-411

17. Allen KD, Somers TJ, Campbell LC, Arbeeva L, Coffman CJ, Cene CW et al. Pain coping skills training for African Americans with osteoarthritis: results of a randomized controlled trial. Pain. 2019; 160(6):1297-1307

18. Allen KD, Woolson S, Hoenig HM, Bongiorni D, Byrd J, Caves K et al. Stepped exercise program for patients with knee osteoarthritis : A randomized controlled trial. Annals of Internal Medicine. 2021; 174(3):298-307

19. Allen KD, Yancy WS, Jr., Bosworth HB, Coffman CJ, Jeffreys AS, Datta SK et al. A combined patient and provider intervention for management of osteoarthritis in veterans: A randomized clinical trial. Annals of Internal Medicine. 2016; 164(2):73-83

20. Altmis H, Oskay D, Elbasan B, Duzgun I, Tuna Z. Mobilization with movement and kinesio taping in knee arthritis-evaluation and outcomes. International Orthopaedics. 2018; 42(12):2807-2815

21. Anonymous. Arthritis self-management programme delivered by post is as effective as group teaching. Evidence-Based Healthcare and Public Health. 2004; 8(5):261-262

22. Anwer S, Alghadir A, Brismee JM. Effect of home exercise program in patients with knee osteoarthritis: A systematic review and meta-analysis. Journal of Geriatric Physical Therapy. 2016; 39(1):38-48

23. Arfaei Chitkar SS, Mohaddes Hakkak HR, Saadati H, Hosseini SH, Jafari Y, Ganji R. The effect of mobile-app-based instruction on the physical function of female patients with knee osteoarthritis: a parallel randomized controlled trial. BMC Women's Health. 2021; 21(1):333

24. Arnold CM, Faulkner RA. The effect of aquatic exercise and education on lowering fall risk in older adults with hip osteoarthritis. Journal of Aging & Physical Activity. 2010; 18(3):245-260

25. Aunger JA, Greaves CJ, Davis ET, Asamane EA, Whittaker AC, Greig CA. A novel behavioural INTErvention to REduce Sitting Time in older adults undergoing orthopaedic surgery (INTEREST): results of a randomised-controlled feasibility study. Aging-Clinical & Experimental Research. 2020; 32(12):2565-2585

26. Aunger JA, Greaves CJ, Davis ET, Greig CA. A novel behavioural INTErvention to REduce Sitting Time in older adults undergoing orthopaedic surgery (INTEREST): protocol for a randomised controlled feasibility study. Pilot & Feasibility Studies. 2019; 5:54

27. Axford J, Heron C, Ross F, Victor CR. Management of knee osteoarthritis in primary care: pain and depression are the major obstacles. Journal of Psychosomatic Research. 2008; 64(5):461-467

28. Azma K, RezaSoltani Z, Rezaeimoghaddam F, Dadarkhah A, Mohsenolhosseini S. Efficacy of tele-rehabilitation compared with office-based physical therapy in patients with knee osteoarthritis: A randomized clinical trial. Journal of Telemedicine and Telecare. 2018; 24(8):560-565

29. Bandak E, Christensen R, Overgaard A, Kristensen LE, Ellegaard K, Guldberg-Moller J et al. Exercise and education versus saline injections for knee osteoarthritis: a randomised controlled equivalence trial. Annals of the Rheumatic Diseases. 2021;

30. Barker KL, Room J, Knight R, Dutton S, Toye F, Leal J et al. Home-based rehabilitation programme compared with traditional physiotherapy for patients at risk of poor outcome after knee arthroplasty: the CORKA randomised controlled trial. BMJ Open. 2021; 11(8):e052598

31. Barker KL, Room J, Knight R, Dutton SJ, Toye F, Leal J et al. Outpatient physiotherapy versus home-based rehabilitation for patients at risk of poor outcomes after knee arthroplasty: CORKA RCT. Health technology assessment (Winchester, England). 2020; 24(65):1-116

32. Barlow JH, Turner AP, Wright CC. A randomized controlled study of the Arthritis Self-Management Programme in the UK. Health Education Research. 2000; 15(6):665-680

33. Bearne LM, Walsh NE, Jessep S, Hurley MV. Feasibility of an exercise-based rehabilitation programme for chronic hip pain. Musculoskeletal Care. 2011; 9(3):160-168

34. Beavers DP, Beavers KM, Loeser RF, Walton NR, Lyles MF, Nicklas BJ et al. The independent and combined effects of intensive weight loss and exercise training on bone mineral density in overweight and obese older adults with osteoarthritis. Osteoarthritis and Cartilage. 2014; 22(6):726-733

35. Bendrik R, Kallings LV, Broms K, Kunanusornchai W, Emtner M. Physical activity on prescription in patients with hip or knee osteoarthritis: A randomized controlled trial. Clinical Rehabilitation. 2021:2692155211008807

36. Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJ et al. Behavior Change Text Messages for Home Exercise Adherence in Knee Osteoarthritis: Randomized Trial. Journal of Medical Internet Research. 2020; 22(9):e21749

37. Bennell KL, Ahamed Y, Jull G, Bryant C, Hunt MA, Forbes AB et al. Physical therapist-delivered pain coping skills training and exercise for knee osteoarthritis: Randomized controlled trial. Arthritis Care and Research. 2016; 68(5):590-602

38. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: A randomized clinical trial. Arthritis Care and Research. 2017; 69(1):84-94

39. Bennell KL, Egerton T, Bills C, Gale J, Kolt GS, Bunker SJ et al. Addition of telephone coaching to a physiotherapist-delivered physical activity program in people with knee osteoarthritis: a randomised controlled trial protocol. BMC Musculoskeletal Disorders. 2012; 13:246

40. Bennell KL, Egerton T, Martin J, Abbott JH, Metcalf B, McManus F et al. Effect of physical therapy on pain and function in patients with hip osteoarthritis: a randomized clinical trial. JAMA. 2014; 311(19):1987-1997

41. Bennell KL, Egerton T, Pua YH, Abbott JH, Sims K, Metcalf B et al. Efficacy of a multimodal physiotherapy treatment program for hip osteoarthritis: a randomised placebo-controlled trial protocol. BMC Musculoskeletal Disorders. 2010; 11:238

42. Bennell KL, Hinman RS, Metcalf BR, Buchbinder R, McConnell J, McColl G et al. Efficacy of physiotherapy management of knee joint osteoarthritis: a randomised, double blind, placebo controlled trial. Annals of the Rheumatic Diseases. 2005; 64(6):906-912

43. Bennell KL, Keating C, Lawford BJ, Kimp AJ, Egerton T, Brown C et al. Better Knee, Better Meâ„¢: effectiveness of two scalable health care interventions supporting self-management for knee osteoarthritis-protocol for a randomized controlled trial. BMC Musculoskeletal Disorders. 2020; 21(1)

44. Bennell KL, Lawford BJ, Keating C, Brown C, Kasza J, Mackenzie D et al. Comparing Video-Based, Telehealth-Delivered Exercise and Weight Loss Programs With Online Education on Outcomes of Knee Osteoarthritis : A Randomized Trial. Annals of Internal Medicine. 2021;

45. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J et al. Effectiveness of an Internet-Delivered Exercise and Pain-Coping Skills Training Intervention for Persons With Chronic Knee Pain: A Randomized Trial. Annals of Internal Medicine. 2017; 166(7):453-462

46. Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S et al. Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial. Pain. 2018; 159(9):1833-1842

47. Bennell KL, Rini C, Keefe F, French S, Nelligan R, Kasza J et al. Effects of adding an internet-based pain coping skills training protocol to a standardized education and exercise program for people with persistent hip pain (hope trial): Randomized controlled trial protocol. Physical Therapy. 2015; 95(10):1408-1422

48. Bilgici A, Akdeniz O, Kuru O, Unlu S, Ulusoy H. The effect of aerobic exercise program versus a home based exercise therapy on pain and functional disability in patients with knee osteoarthritis. Journal of rheumatology and medical rehabilitation. 2005; 16(1):10-17

49. Bliddal H, Leeds AR, Stigsgaard L, Astrup A, Christensen R. Weight loss as treatment for knee osteoarthritis symptoms in obese patients: 1-year results from a randomised controlled trial. Annals of the Rheumatic Diseases. 2011; 70(10):1798-1803

50. Blixen CE, Bramstedt KA, Hammel JP, Tilley BC. A pilot study of health education via a nurse-run telephone self-management programme for elderly people with osteoarthritis. Journal of Telemedicine and Telecare. 2004; 10(1):44-49

51. Bobos P, Nazari G, Szekeres M, Lalone EA, Ferreira L, MacDermid JC. The effectiveness of joint-protection programs on pain, hand function, and grip strength levels in patients with hand arthritis: A systematic review and meta-analysis. Journal of Hand Therapy. 2018; 23:23

52. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. Journal of Medical Internet Research. 2013; 15(11):e257

53. Brand E, Nyland J, Henzman C, McGinnis M. Arthritis self-efficacy scale scores in knee osteoarthritis: a systematic review and meta-analysis comparing arthritis self-management education with or without exercise. Journal of Orthopaedic and Sports Physical Therapy. 2013; 43(12):895-910

54. Broderick JE, Keefe FJ, Bruckenthal P, Junghaenel DU, Schneider S, Schwartz JE et al. Nurse practitioners can effectively deliver pain coping skills training to osteoarthritis patients with chronic pain: A randomized, controlled trial. Pain. 2014; 155(9):1743-1754

55. Brosseau L, Thevenot O, MacKiddie O, Taki J, Wells GA, Guitard P et al. The Ottawa Panel guidelines on programmes involving therapeutic exercise for the management of hand osteoarthritis. Clinical Rehabilitation. 2018; 32(11):1449-1471

56. Brosseau L, Wells GA, Kenny GP, Reid R, Maetzel A, Tugwell P et al. The implementation of a community-based aerobic walking program for mild to moderate knee osteoarthritis: a knowledge translation randomized controlled trial: part II: clinical outcomes. BMC Public Health. 2012; 12:1073

57. Bryant C, Lewis P, Bennell KL, Ahamed Y, Crough D, Jull GA et al. Can physical therapists deliver a pain coping skills program? An examination of training processes and outcomes. Physical Therapy. 2014; 94(10):1443-1454

58. Buszewicz M, Rait G, Griffin M, Nazareth I, Patel A, Atkinson A et al. Self management of arthritis in primary care: randomised controlled trial. BMJ. 2006; 333(7574):879

59. Button K, Roos PE, Spasic I, Adamson P, van Deursen RW. The clinical effectiveness of self-care interventions with an exercise component to manage knee conditions: A systematic review. Knee. 2015; 22(5):360-371

60. Callaghan MJ, Oldham JA, Hunt J. An evaluation of exercise regimes for patients with osteoarthritis of the knee: a single-blind randomized controlled trial. Clinical Rehabilitation. 1995; 9(3):213-218

61. Cetin N, Aytar A, Atalay A, Akman MN. Comparing hot pack, short-wave diathermy, ultrasound, and TENS on isokinetic strength, pain, and functional status of women with osteoarthritic knees: a single-blind, randomized, controlled trial. American Journal of Physical Medicine and Rehabilitation. 2008; 87(6):443-451

62. Chang RW, Semanik PA, Lee J, Feinglass J, Ehrlich-Jones L, Dunlop DD. Improving physical activity in arthritis clinical trial (IMPAACT): study design, rationale, recruitment, and baseline data. Contemporary Clinical Trials. 2014; 39(2):224-235

63. Chang WJ, Bennell KL, Hodges PW, Hinman RS, Young CL, Buscemi V et al. Addition of transcranial direct current stimulation to quadriceps strengthening exercise in knee osteoarthritis: A pilot randomised controlled trial. PLoS ONE [Electronic Resource]. 2017; 12(6):e0180328

64. Cheing GL, Hui-Chan CW. Would the addition of TENS to exercise training produce better physical performance outcomes in people with knee osteoarthritis than either intervention alone? Clinical Rehabilitation. 2004; 18(5):487-497

65. Cheing GL, Hui-Chan CW, Chan KM. Does four weeks of TENS and/or isometric exercise produce cumulative reduction of osteoarthritic knee pain? Clinical Rehabilitation. 2002; 16(7):749-760

66. Chen H, Zheng X, Huang H, Liu C, Wan Q, Shang S. The effects of a home-based exercise intervention on elderly patients with knee osteoarthritis: a quasi-experimental study. BMC Musculoskeletal Disorders. 2019; 20(1):160

67. Chen LX, Mao JJ, Fernandes S, Galantino ML, Guo W, Lariccia P et al. Integrating acupuncture with exercise-based physical therapy for knee osteoarthritis: a randomized controlled trial. JCR: Journal of Clinical Rheumatology. 2013; 19(6):308-316

68. Chua SD, Jr., Messier SP, Legault C, Lenz ME, Thonar EJ, Loeser RF. Effect of an exercise and dietary intervention on serum biomarkers in overweight and obese adults with osteoarthritis of the knee. Osteoarthritis and Cartilage. 2008; 16(9):1047-1053

69. Clayton C, Feehan L, Goldsmith CH, Miller WC, Grewal N, Ye J et al. Feasibility and preliminary efficacy of a physical activity counseling intervention using Fitbit in people with knee osteoarthritis: the TRACK-OA study protocol. Pilot & Feasibility Studies. 2015; 1:30

70. Coelho Cde F, Leal-Junior EC, Biasotto-Gonzalez DA, Bley AS, de Carvalho Pde T, Politti F et al. Effectiveness of phototherapy incorporated into an exercise program for osteoarthritis of the knee: study protocol for a randomized controlled trial. Trials [Electronic Resource]. 2014; 15:221

71. Cohen J, Sauter S, DeVellis R, DeVellis B. Evaluation of arthritis self-management courses led by laypersons and by professionals. Arthritis and Rheumatism. 1986; 29(3):388-393

72. Coleman S, Briffa NK, Carroll G, Inderjeeth C, Cook N, McQuade J. Effects of self-management, education and specific exercises, delivered by health professionals, in patients with osteoarthritis of the knee. BMC Musculoskeletal Disorders. 2008; 9:133

73. Coleman S, Briffa NK, Carroll G, Inderjeeth C, Cook N, McQuade J. A randomised controlled trial of a self-management education program for osteoarthritis of the knee delivered by health care professionals. Arthritis Research & Therapy. 2012; 14(1):R21

74. Cortes Godoy V, Gallego Izquierdo T, Lazaro Navas I, Pecos Martin D. Effectiveness of massage therapy as co-adjuvant treatment to exercise in osteoarthritis of the knee: a randomized control trial. Journal of Back and Musculoskeletal Rehabilitation. 2014; 27(4):521-529

75. Crossley KM, Vicenzino B, Lentzos J, Schache AG, Pandy MG, Ozturk H et al. Exercise, education, manual-therapy and taping compared to education for patellofemoral osteoarthritis: a blinded, randomised clinical trial. Osteoarthritis and Cartilage. 2015; 23(9):1457-1464

76. Crossley KM, Vicenzino B, Pandy MG, Schache AG, Hinman RS. Targeted physiotherapy for patellofemoral joint osteoarthritis: a protocol for a randomised, single-blind controlled trial. BMC Musculoskeletal Disorders. 2008; 9:122

77. Crotty M, Prendergast J, Battersby MW, Rowett D, Graves SE, Leach G et al. Self-management and peer support among people with arthritis on a hospital joint replacement waiting list: a randomised controlled trial. Osteoarthritis and Cartilage. 2009; 17(11):1428-1433

78. Cuesta-Vargas AI, White M, Gonzalez-Sanchez M, Kuisma R. The optimal frequency of aquatic physiotherapy for individuals with chronic musculoskeletal pain: a randomised controlled trial. Disability and Rehabilitation. 2015; 37(4):311-318

79. Cuperus N, Hoogeboom TJ, Kersten CC, den Broeder AA, Vlieland TP, van den Ende CH. Randomized trial of the effectiveness of a non-pharmacological multidisciplinary face-to-face treatment program on daily function compared to a telephone-based treatment program in patients with generalized osteoarthritis. Osteoarthritis and Cartilage. 2015; 23(8):1267-1275

80. da Silva FS, de Melo FE, do Amaral MM, Caldas VV, Pinheiro IL, Abreu BJ et al. Efficacy of simple integrated group rehabilitation program for patients with knee

osteoarthritis: Single-blind randomized controlled trial. Journal of Rehabilitation Research and Development. 2015; 52(3):309-322

81. de Jong OR, Hopman-Rock M, Tak EC, Klazinga NS. An implementation study of two evidence-based exercise and health education programmes for older adults with osteoarthritis of the knee and hip. Health Education Research. 2004; 19(3):316-325

82. de Matos Brunelli Braghin R, Libardi EC, Junqueira C, Rodrigues NC, Nogueira-Barbosa MH, Renno ACM et al. The effect of low-level laser therapy and physical exercise on pain, stiffness, function, and spatiotemporal gait variables in subjects with bilateral knee osteoarthritis: a blind randomized clinical trial. Disability and Rehabilitation. 2019; 41(26):3165-3172

83. de Rezende MU, de Farias FES, da Silva CAC, Cernigoy CHA, de Camargo OP. Objective functional results in patients with knee osteoarthritis submitted to a 2-day educational programme: a prospective randomised clinical trial. BMJ Open Sport & Exercise Medicine. 2016; 2(1):e000200

84. de Rezende MU, Hissadomi MI, de Campos GC, Frucchi R, Pailo AF, Pasqualin T et al. One-Year Results of an Educational Program on Osteoarthritis: A Prospective Randomized Controlled Trial in Brazil. Geriatric Orthopaedic Surgery & Rehabilitation. 2016; 7(2):86-94

85. de Vos BC, Runhaar J, Bierma-Zeinstra SM. Effectiveness of a tailor-made weight loss intervention in primary care. European Journal of Nutrition. 2014; 53(1):95-104

86. Deveza LA, Hunter DJ, Wajon A, Bennell KL, Vicenzino B, Hodges P et al. Efficacy of combined conservative therapies on clinical outcomes in patients with thumb base osteoarthritis: protocol for a randomised, controlled trial (COMBO). BMJ Open. 2017; 7(1):e014498

87. Deveza LA, Robbins SR, Duong V, Bennell KL, Vicenzino B, Hodges PW et al. Efficacy of a Combination of Conservative Therapies vs an Education Comparator on Clinical Outcomes in Thumb Base Osteoarthritis: A Randomized Clinical Trial. JAMA Internal Medicine. 2021; 181(4):429-438

88. Devos-Comby L, Cronan T, Roesch SC. Do exercise and self-management interventions benefit patients with osteoarthritis of the knee? A metaanalytic review. Journal of Rheumatology. 2006; 33(4):744-756

89. Dias JM, Cisneros L, Dias R, Fritsch C, Gomes W, Pereira L et al. Hydrotherapy improves pain and function in older women with knee osteoarthritis: a randomized controlled trial. Brazilian Journal of Physical Therapy. 2017; 21(6):449-456

90. Dincer U, Cakar E, Ozdemir B, Kiralp MZ, Dursun H. Comparison of effects of combined physical therapy program and exercise on corrupted balance functions in patient with knee bilateral osteoarthritis. Romatizma. 2008; 23(1):9-13

91. Dobson F, Hinman RS, French S, Rini C, Keefe F, Nelligan R et al. Internet-mediated physiotherapy and pain coping skills training for people with persistent knee pain (IMPACT - knee pain): a randomised controlled trial protocol. BMC Musculoskeletal Disorders. 2014; 15:279

92. Dunning J, Butts R, Young I, Mourad F, Galante V, Bliton P et al. Periosteal electrical dry needling as an adjunct to exercise and manual therapy for knee osteoarthritis: A multicenter randomized clinical trial. Clinical Journal of Pain. 2018; 34(12):1149-1158

93. Dwyer L, Parkin-Smith GF, Brantingham JW, Korporaal C, Cassa TK, Globe G et al. Manual and manipulative therapy in addition to rehabilitation for osteoarthritis of the

knee: assessor-blind randomized pilot trial. Journal of Manipulative and Physiological Therapeutics. 2015; 38(1):1-21.e22

94. Dziedzic K, Nicholls E, Hill S, Hammond A, Handy J, Thomas E et al. Self-management approaches for osteoarthritis in the hand: a 2x2 factorial randomised trial. Annals of the Rheumatic Diseases. 2015; 74(1):108-118

95. Dziedzic KS, Healey EL, Porcheret M, Afolabi EK, Lewis M, Morden A et al. Implementing core NICE guidelines for osteoarthritis in primary care with a model consultation (MOSAICS): a cluster randomised controlled trial. Osteoarthritis and Cartilage. 2018; 26(1):43-53

96. Dziedzic KS, Hill S, Nicholls E, Hammond A, Myers H, Whitehurst T et al. Self management, joint protection and exercises in hand osteoarthritis: a randomised controlled trial with cost effectiveness analyses. BMC Musculoskeletal Disorders. 2011; 12:156

97. Ettinger WH, Jr., Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T et al. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). JAMA. 1997; 277(1):25-31

98. Farr JN, Going SB, McKnight PE, Kasle S, Cussler EC, Cornett M. Progressive resistance training improves overall physical activity levels in patients with early osteoarthritis of the knee: a randomized controlled trial. Physical Therapy. 2010; 90(3):356-366

99. Fernandes L, Storheim K, Sandvik L, Nordsletten L, Risberg MA. Efficacy of patient education and supervised exercise vs patient education alone in patients with hip osteoarthritis: a single blind randomized clinical trial. Osteoarthritis and Cartilage. 2010; 18(10):1237-1243

100. Fisher NM, Gresham G, Pendergast DR. Effects of a quantitative progressive rehabilitation program applied unilaterally to the osteoarthritic knee. Archives of Physical Medicine and Rehabilitation. 1993; 74(12):1319-1326

101. Fisken AL, Waters DL, Hing WA, Steele M, Keogh JW. Comparative effects of 2 aqua exercise programs on physical function, balance, and perceived quality of life in older adults with osteoarthritis. Journal of geriatric physical therapy (2001). 2015; 38(1):17-27

102. Fitzgibbon ML, Tussing-Humphreys L, Schiffer L, Smith-Ray R, Marquez DX, DeMott AD et al. Fit and Strong! Plus: Twelve and eighteen month follow-up results for a comparative effectiveness trial among overweight/obese older adults with osteoarthritis. Preventive Medicine. 2020; 141:106267

103. Focht BC, Garver MJ, Devor ST, Dials J, Lucas AR, Emery CF et al. Group-mediated physical activity promotion and mobility in sedentary patients with knee osteoarthritis: results from the IMPACT-pilot trial. Journal of Rheumatology. 2014; 41(10):2068-2077

104. Focht BC, Garver MJ, Lucas AR, Devor ST, Emery CF, Hackshaw KV et al. A group-mediated physical activity intervention in older knee osteoarthritis patients: effects on social cognitive outcomes. Journal of Behavioral Medicine. 2017; 40(3):530-537

105. Focht BC, Gauvin L, Rejeski WJ. The contribution of daily experiences and acute exercise to fluctuations in daily feeling states among older, obese adults with knee osteoarthritis. Journal of Behavioral Medicine. 2004; 27(2):101-121

106. Focht BC, Rejeski WJ, Ambrosius WT, Katula JA, Messier SP. Exercise, self-efficacy, and mobility performance in overweight and obese older adults with knee osteoarthritis. Arthritis and Rheumatism. 2005; 53(5):659-665

107. Foster NE, Healey EL, Holden MA, Nicholls E, Whitehurst DG, Jowett S et al. A multicentre, pragmatic, parallel group, randomised controlled trial to compare the clinical and cost-effectiveness of three physiotherapy-led exercise interventions for knee osteoarthritis in older adults: the BEEP trial protocol (ISRCTN: 93634563). BMC Musculoskeletal Disorders. 2014; 15:254

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

109. Gaines JM, Metter EJ, Talbot LA. The effect of neuromuscular electrical stimulation on arthritis knee pain in older adults with osteoarthritis of the knee. Applied Nursing Research. 2004; 17(3):201-206

110. Ganji R, Pakniat A, Armat MR, Tabatabaeichehr M, Mortazavi H. The Effect of Self-Management Educational Program on Pain Intensity in Elderly Patients with Knee Osteoarthritis: A Randomized Clinical Trial. Open Access Macedonian Journal of Medical Sciences. 2018; 6(6):1062-1066

111. Gay C, Guiguet-Auclair C, Pereira B, Goldstein A, Bareyre L, Coste N et al. Efficacy of self-management exercise program with spa therapy for behavioral management of knee osteoarthritis: research protocol for a quasi-randomized controlled trial (GEET one). BMC Complementary and Alternative Medicine. 2018; 18(1):279

112. Ghroubi S, Elleuch H, Kaffel N, Echikh T, Abid M, Elleuch MH. Contribution of exercise and diet in the management of knee osteoarthritis in the obese. Annales de readaptation et de medecine physique. 2008; 51(8):663-670

113. Goff AJ, De Oliveira Silva D, Merolli M, Bell EC, Crossley KM, Barton CJ. Patient education improves pain and function in people with knee osteoarthritis with better effects when combined with exercise therapy: a systematic review. Journal of Physiotherapy. 2021; 67(3):177-189

114. Gravas EMH, Osteras N, Nossum R, Eide REM, Klokkeide A, Matre KH et al. Does occupational therapy delay or reduce the proportion of patients that receives thumb carpometacarpal joint surgery? A multicentre randomised controlled trial. RMD Open. 2019; 5(2):e001046

115. Hall M, Castelein B, Wittoek R, Calders P, Van Ginckel A. Diet-induced weight loss alone or combined with exercise in overweight or obese people with knee osteoarthritis: A systematic review and meta-analysis. Seminars in Arthritis and Rheumatism. 2019; 48(5):765-777

116. Hansson EE, Jonsson-Lundgren M, Ronnheden AM, Sorensson E, Bjarnung A, Dahlberg LE. Effect of an education programme for patients with osteoarthritis in primary care--a randomized controlled trial. BMC Musculoskeletal Disorders. 2010; 11:244

117. Hay EM, Foster NE, Thomas E, Peat G, Phelan M, Yates HE et al. Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial. BMJ. 2006; 333(7576):995

118. Health Quality O. Structured Education and Neuromuscular Exercise Program for Hip and/or Knee Osteoarthritis: A Health Technology Assessment. Ontario Health Technology Assessment Series. 2018; 18(8):1-110

119. Helminen EE, Sinikallio SH, Valjakka AL, Vaisanen-Rouvali RH, Arokoski JP. Effectiveness of a cognitive-behavioural group intervention for knee osteoarthritis pain: a randomized controlled trial. Clinical Rehabilitation. 2015; 29(9):868-881

120. Heuts PH, de Bie R, Drietelaar M, Aretz K, Hopman-Rock M, Bastiaenen CH et al. Self-management in osteoarthritis of hip or knee: a randomized clinical trial in a primary healthcare setting. Journal of Rheumatology. 2005; 32(3):543-549

121. Higgins BT, Barlow DR, Heagerty NE, Lin TJ. Anterior vs. posterior approach for total hip arthroplasty, a systematic review and meta-analysis. Journal of Arthroplasty. 2015; 30(3):419-434

122. Hinman RS, Lawford BJ, Campbell PK, Briggs AM, Gale J, Bills C et al. Telephone-Delivered Exercise Advice and Behavior Change Support by Physical Therapists for People with Knee Osteoarthritis: Protocol for the Telecare Randomized Controlled Trial. Physical Therapy. 2017; 97(5):524-536

123. Holden MA, Burke DL, Runhaar J, van Der Windt D, Riley RD, Dziedzic K et al. Subgrouping and TargetEd Exercise pRogrammes for knee and hip OsteoArthritis (STEER OA): a systematic review update and individual participant data meta-analysis protocol. BMJ Open. 2017; 7(12):e018971

124. Hoogeboom TJ, Kwakkenbos L, Rietveld L, den Broeder AA, de Bie RA, van den Ende CH. Feasibility and potential effectiveness of a non-pharmacological multidisciplinary care programme for persons with generalised osteoarthritis: a randomised, multiple-baseline single-case study. BMJ Open. 2012; 2(4)

125. Hopman-Rock M, Westhoff MH. The effects of a health educational and exercise program for older adults with osteoarthritis for the hip or knee. Journal of Rheumatology. 2000; 27(8):1947-1954

126. Hsu YI, Chen YC, Lee CL, Chang NJ. Effects of Diet Control and Telemedicine-Based Resistance Exercise Intervention on Patients with Obesity and Knee Osteoarthritis: A Randomized Control Trial. International Journal of Environmental Research & Public Health [Electronic Resource]. 2021; 18(15):21

127. Huang MH, Chen CH, Chen TW, Weng MC, Wang WT, Wang YL. The effects of weight reduction on the rehabilitation of patients with knee osteoarthritis and obesity. Arthritis Care and Research. 2000; 13(6):398-405

128. Huang MH, Yang RC, Lee CL, Chen TW, Wang MC. Preliminary results of integrated therapy for patients with knee osteoarthritis. Arthritis and Rheumatism. 2005; 53(6):812-820

129. Huang TT, Sung CC, Wang WS, Wang BH. The effects of the empowerment education program in older adults with total hip replacement surgery. Journal of Advanced Nursing. 2017; 73(8):1848-1861

130. Hughes SL, Seymour RB, Campbell R, Pollak N, Huber G, Sharma L. Impact of the fit and strong intervention on older adults with osteoarthritis. Gerontologist. 2004; 44(2):217-228

131. Hughes SL, Seymour RB, Campbell RT, Desai P, Huber G, Chang HJ. Fit and Strong!: bolstering maintenance of physical activity among older adults with lower-extremity osteoarthritis. American Journal of Health Behavior. 2010; 34(6):750-763

132. Hughes SL, Seymour RB, Campbell RT, Huber G, Pollak N, Sharma L et al. Long-term impact of Fit and Strong! on older adults with osteoarthritis. Gerontologist. 2006; 46(6):801-814

133. Hughes SL, Tussing-Humphreys L, Schiffer L, Smith-Ray R, Marquez DX, DeMott AD et al. Fit & Strong! Plus Trial Outcomes for Obese Older Adults with Osteoarthritis. Gerontologist. 2018; 26:26

134. Hunt MA, Keefe FJ, Bryant C, Metcalf BR, Ahamed Y, Nicholas MK et al. A physiotherapist-delivered, combined exercise and pain coping skills training intervention for individuals with knee osteoarthritis: a pilot study. Knee. 2013; 20(2):106-112

135. Hunter DJ, Beavers DP, Eckstein F, Guermazi A, Loeser RF, Nicklas BJ et al. The Intensive Diet and Exercise for Arthritis (IDEA) trial: 18-month radiographic and MRI outcomes. Osteoarthritis and Cartilage. 2015; 23(7):1090-1098

136. Hurley M, Dickson K, Hallett R, Grant R, Hauari H, Walsh N et al. Exercise interventions and patient beliefs for people with hip, knee or hip and knee osteoarthritis: a mixed methods review. Cochrane Database of Systematic Reviews 2018, Issue 4. Art. No.: CD010842. DOI: 10.1002/14651858.CD010842.pub2.

137. Hurley MV, Walsh NE, Mitchell H, Nicholas J, Patel A. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis Care and Research. 2012; 64(2):238-247

138. Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Patel A, Williamson E et al. Clinical effectiveness of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain: a cluster randomized trial. Arthritis and Rheumatism. 2007; 57(7):1211-1219

139. Hurley MV, Walsh NE, Mitchell HL, Pimm TJ, Williamson E, Jones RH et al. Economic evaluation of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain. Arthritis and Rheumatism (Arthritis Care and Research). 2007; 57(7):1220-1229

140. Ikeda T, Jinno T, Masuda T, Aizawa J, Ninomiya K, Suzuki K et al. Effect of exercise therapy combined with branched-chain amino acid supplementation on muscle strengthening in persons with osteoarthritis. Hong kong physiotherapy journal. 2018; 38(1):23-31

141. Isaramalai SA, Hounsri K, Kongkamol C, Wattanapisitkul P, Tangadulrat N, Kaewmanee T et al. Integrating participatory ergonomic management in non-weight-bearing exercise and progressive resistance exercise on self-care and functional ability in aged farmers with knee osteoarthritis: a clustered randomized controlled trial. Clinical Interventions in Aging. 2018; 13:101-108

142. Ismail A, Moore C, Alshishani N, Yaseen K, Alshehri MA. Cognitive behavioural therapy and pain coping skills training for osteoarthritis knee pain management: a systematic review. Journal of Physical Therapy Science. 2017; 29(12):2228-2235

143. Jan MH, Lai JS. The effects of physiotherapy on osteoarthritic knees of females. Journal of the Formosan Medical Association. 1991; 90(10):1008-1013

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

145. Jordan KP, Edwards JJ, Porcheret M, Healey EL, Jinks C, Bedson J et al. Effect of a model consultation informed by guidelines on recorded quality of care of osteoarthritis

(MOSAICS): a cluster randomised controlled trial in primary care. Osteoarthritis and Cartilage. 2017; 25(10):1588-1597

146. Kao MJ, Wu MP, Tsai MW, Chang WW, Wu SF. The effectiveness of a self-management program on quality of life for knee osteoarthritis (OA) patients. Archives of Gerontology and Geriatrics. 2012; 54(2):317-324

147. Kars Fertelli T, Mollaoglu M, Sahin O. Aquatic exercise program for individuals with osteoarthritis: Pain, stiffness, physical function, self-efficacy. Rehabilitation Nursing Journal. 2019; 44(5):290-299

148. Kaufman BG, Allen KD, Coffman CJ, Woolson S, Caves K, Hall K et al. Cost and Quality of Life Outcomes of the STepped Exercise Program for Patients With Knee OsteoArthritis Trial. Value in Health. 2021;

149. Keays SL, Mason M, Newcombe PA. Individualized physiotherapy in the treatment of patellofemoral pain. Physiotherapy Research International. 2015; 20(1):22-36

150. Keefe FJ, Blumenthal J, Baucom D, Affleck G, Waugh R, Caldwell DS et al. Effects of spouse-assisted coping skills training and exercise training in patients with osteoarthritic knee pain: a randomized controlled study. Pain. 2004; 110(3):539-549

151. Kemp J, Moore K, Fransen M, Russell T, Freke M, Crossley KM. A pilot randomised clinical trial of physiotherapy (manual therapy, exercise, and education) for early-onset hip osteoarthritis post-hip arthroscopy. Pilot & Feasibility Studies. 2018; 4:16

152. Keogh A, Matthews J, Segurado R, Hurley DA. Feasibility of Training Physical Therapists to Deliver the Theory-Based Self-Management of Osteoarthritis and Low Back Pain Through Activity and Skills (SOLAS) Intervention Within a Trial. Physical Therapy. 2018; 98(2):95-107

153. Kigozi J, Nicholls E, Tooth S, Foster NE, Holden MA, Healey EL et al. Cost-utility analysis of interventions to improve effectiveness of exercise therapy for adults with knee osteoarthritis: The BEEP trial. Rheumatology Advances in Practice. 2018; 2(2):1-11

154. Kim IS, Chung SH, Park YJ, Kang HY. The effectiveness of an aquarobic exercise program for patients with osteoarthritis. Applied Nursing Research. 2012; 25(3):181-189

155. Klassbo M, Larsson G, Harms-Ringdahl K. Promising outcome of a hip school for patients with hip dysfunction. Arthritis and Rheumatism. 2003; 49(3):321-327

156. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: A cluster-randomized controlled trial. Physical Therapy. 2018; 98(7):560-570

157. Kloek Cjj PhD PT, Bossen D PhD PT, de Vries Hj Msc PT, de Bakker DHP, Veenhof C PhD PT, Dekker JP. Physiotherapists' experiences with a blended osteoarthritis intervention: a mixed methods study. Physiotherapy Theory & Practice. 2018:1-8

158. Kloek CJJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health. 2018; 18(1):1082

159. 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. Annals of Internal Medicine. 1992; 116(7):529-534

160. Kroon F, van dBL, Buchbinder R, Osborne R, Johnston R, Pitt V. Self-management education programmes for osteoarthritis. Cochrane Database of Systematic Reviews 2014, Issue 1. Art. No.: CD008963. DOI: 10.1002/14651858.CD008963.pub2.

161. Kumar S, Kumar A, Kumar R. Proprioceptive training as an adjunct in osteoarthritis of knee. Journal of musculoskeletal research. 2013; 16(1):1350002

162. Laufer Y, Shtraker H, Elboim Gabyzon M. The effects of exercise and neuromuscular electrical stimulation in subjects with knee osteoarthritis: a 3-month follow-up study. Clinical Interventions in Aging. 2014; 9:1153-1161

163. Lawford BJ, Hinman RS, Kasza J, Nelligan R, Keefe F, Rini C et al. Moderators of Effects of Internet-Delivered Exercise and Pain Coping Skills Training for People With Knee Osteoarthritis: Exploratory Analysis of the IMPACT Randomized Controlled Trial. Journal of Medical Internet Research. 2018; 20(5):e10021

164. Lee H, Wiggers J, Kamper SJ, Williams A, O'Brien KM, Hodder RK et al. Mechanism evaluation of a lifestyle intervention for patients with musculoskeletal pain who are overweight or obese: protocol for a causal mediation analysis. BMJ Open. 2017; 7(6):e014652

165. Lee HY. Comparison of effects among Tai-Chi exercise, aquatic exercise, and a self-help program for patients with knee osteoarthritis. Taehan kanho hakhoe chi. 2006; 36(3):571-580

166. Li LC, Lineker S, Cibere J, Crooks VA, Jones CA, Kopec JA et al. Capitalizing on the teachable moment: osteoarthritis physical activity and exercise net for improving physical activity in early knee osteoarthritis. JMIR Research Protocols. 2013; 2(1):e17

167. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: Feasibility and preliminary efficacy of the track-oa study. JMIR MHealth and UHealth. 2017; 5(6):e86

168. Lin EH, Katon W, Von Korff M, Tang L, Williams JW, Jr., Kroenke K et al. Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. JAMA. 2003; 290(18):2428-2429

169. Loeser RF, Beavers DP, Bay-Jensen AC, Karsdal MA, Nicklas BJ, Guermazi A et al. Effects of dietary weight loss with and without exercise on interstitial matrix turnover and tissue inflammation biomarkers in adults with knee osteoarthritis: the Intensive Diet and Exercise for Arthritis trial (IDEA). Osteoarthritis and Cartilage. 2017; 25(11):1822-1828

170. Loew L, Brosseau L, Kenny GP, Durand-Bush N, Poitras S, De Angelis G et al. An evidence-based walking program among older people with knee osteoarthritis: the PEP (participant exercise preference) pilot randomized controlled trial. Clinical Rheumatology. 2017; 36(7):1607-1616

171. Lord J, Victor C, Littlejohns P, Ross FM, Axford JS. Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee. Health Technology Assessment. 1999; 3(23):iii-55

172. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis and Rheumatism. 2008; 59(7):1009-1017

173. Magrans-Courtney T, Wilborn C, Rasmussen C, Ferreira M, Greenwood L, Campbell B et al. Effects of diet type and supplementation of glucosamine, chondroitin, and MSM on body composition, functional status, and markers of health in women with

knee osteoarthritis initiating a resistance-based exercise and weight loss program. Journal of the International Society of Sports Nutrition. 2011; 8(1):8

174. Maire J, Dugue B, Faillenet-Maire AF, Smolander J, Tordi N, Parratte B et al. Influence of a 6-week arm exercise program on walking ability and health status after hip arthroplasty: a 1-year follow-up pilot study. Journal of Rehabilitation Research and Development. 2006; 43(4):445-450

175. Marconcin P, Espanha M, Teles J, Bento P, Campos P, André R et al. A randomized controlled trial of a combined self-management and exercise intervention for elderly people with osteoarthritis of the knee: the PLE2NO program. Clinical Rehabilitation. 2018; 32(2):223-232

176. Marconcin P, Espanha M, Yazigi F, Campos P. The PLE(2)NO self-management and exercise program for knee osteoarthritis: Study Protocol for a Randomized Controlled Trial. BMC Musculoskeletal Disorders. 2016; 17:250

177. Marconcin P, Yazigi F, Teles J, Campos P, Espanha M. The effectiveness of a randomised clinical trial of PLE2 NO self-management and exercise programme for knee osteoarthritis to improve self-efficacy. Musculoskeletal Care. 2021; 02:02

178. Marra CA, Cibere J, Grubisic M, Grindrod KA, Gastonguay L, Thomas JM et al. Pharmacist-initiated intervention trial in osteoarthritis: a multidisciplinary intervention for knee osteoarthritis. Arthritis Care and Research. 2012; 64(12):1837-1845

179. Marra CA, Grubisic M, Cibere J, Grindrod KA, Woolcott JC, Gastonguay L et al. Cost-utility analysis of a multidisciplinary strategy to manage osteoarthritis of the knee: economic evaluation of the PhiTOA study. Arthritis Care and Research. 2013:epub

180. Mazzei DR, Ademola A, Abbott JH, Sajobi T, Hildebrand K, Marshall DA. Are education, exercise and diet interventions a cost-effective treatment to manage hip and knee osteoarthritis? A systematic review. Osteoarthritis and Cartilage. 2021; 29(4):456-470

181. Mazzuca SA, Brandt KD, Katz BP, Ragozzino LR, G'Sell P M. Can a nurse-directed intervention reduce the exposure of patients with knee osteoarthritis to nonsteroidal antiinflammatory drugs? JCR: Journal of Clinical Rheumatology. 2004; 10(6):315-322

182. McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR et al. Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis. Health Technology Assessment. 2004; 8(46):iii-iv, 1-61

183. McCarthy CJ, Mills PM, Pullen R, Roberts C, Silman A, Oldham JA. Supplementing a home exercise programme with a class-based exercise programme is more effective than home exercise alone in the treatment of knee osteoarthritis. Rheumatology. 2004; 43(7):880-886

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

185. McVeigh KH, Kannas SN, Ivy CC, Garner HW, Barnes CS, Heckman MG et al. Dynamic stabilization home exercise program for treatment of thumb carpometacarpal osteoarthritis: A prospective randomized control trial. Journal of Hand Therapy. 2021; epub

186. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-Week Digital Care Program for Chronic Knee Pain on Pain, Mobility, and Surgery

Risk: Randomized Controlled Trial. Journal of Medical Internet Research. 2018; 20(4):e156

187. Messier SP, Callahan LF, Beavers DP, Queen K, Mihalko SL, Miller GD et al. Weight-loss and exercise for communities with arthritis in North Carolina (we-can): design and rationale of a pragmatic, assessor-blinded, randomized controlled trial. BMC Musculoskeletal Disorders. 2017; 18(1):91

188. Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA et al. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the Arthritis, Diet, and Activity Promotion Trial. Arthritis and Rheumatism. 2004; 50(5):1501-1510

189. Messier SP, Mihalko S, Loeser RF, Legault C, Jolla J, Pfruender J et al. Glucosamine/chondroitin combined with exercise for the treatment of knee osteoarthritis: a preliminary study. Osteoarthritis and Cartilage. 2007; 15(11):1256-1266

190. Messier SP, Mihalko SL, Legault C, Miller GD, Nicklas BJ, DeVita P et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. JAMA. 2013; 310(12):1263-1273

191. Messier SP, Royer TD, Craven TE, O'Toole ML, Burns R, Ettinger WH, Jr. Long-term exercise and its effect on balance in older, osteoarthritic adults: results from the Fitness, Arthritis, and Seniors Trial (FAST). Journal of the American Geriatrics Society. 2000; 48(2):131-138

192. Mihalko SL, Cox P, Beavers DP, Miller GD, Nicklas BJ, Lyles M et al. Effect of intensive diet and exercise on self-efficacy in overweight and obese adults with knee osteoarthritis: The IDEA randomized clinical trial. Translational Behavioral Medicine. 2019; 9(2):227-235

193. Miller GD, Nicklas BJ, Davis C, Loeser RF, Lenchik L, Messier SP. Intensive weight loss program improves physical function in older obese adults with knee osteoarthritis. Obesity. 2006; 14(7):1219-1230

194. Miller GD, Rejeski WJ, Williamson JD, Morgan T, Sevick MA, Loeser RF et al. The Arthritis, Diet and Activity Promotion Trial (ADAPT): design, rationale, and baseline results. Controlled Clinical Trials. 2003; 24(4):462-480

195. Mizusaki Imoto A, Peccin S, Gomes da Silva KN, de Paiva Teixeira LE, Abrahao MI, Fernandes Moca Trevisani V. Effects of neuromuscular electrical stimulation combined with exercises versus an exercise program on the pain and the function in patients with knee osteoarthritis: a randomized controlled trial. BioMed Research International. 2013; 2013:272018

196. Moe RH, Grotle M, Kjeken I, Olsen IC, Mowinckel P, Haavardsholm EA et al. Effectiveness of an Integrated Multidisciplinary Osteoarthritis Outpatient Program versus Outpatient Clinic as Usual: A Randomized Controlled Trial. Journal of Rheumatology. 2016; 43(2):411-418

197. Moe RH, Uhlig T, Kjeken I, Hagen KB, Kvien TK, Grotle M. Multidisciplinary and multifaceted outpatient management of patients with osteoarthritis: protocol for a randomised, controlled trial. BMC Musculoskeletal Disorders. 2010; 11:253

198. Molgaard CM, Rathleff MS, Andreasen J, Christensen M, Lundbye-Christensen S, Simonsen O et al. Foot exercises and foot orthoses are more effective than knee

focused exercises in individuals with patellofemoral pain. Journal of Science and Medicine in Sport. 2018; 21(1):10-15

199. Murphy SL, Janevic MR, Lee P, Williams DA. Occupational Therapist-Delivered Cognitive-Behavioral Therapy for Knee Osteoarthritis: A Randomized Pilot Study. American Journal of Occupational Therapy. 2018; 72(5):7205205040p7205205041-7205205040p7205205049

200. Nahayatbin M, Ghasemi M, Rahimi A, Khademi-Kalantari K, Naimi SS, Tabatabaee SM et al. The effects of routine physiotherapy alone and in combination with either Tai Chi or closed kinetic chain exercises on knee osteoarthritis: a comparative clinical trial study. Iranian red crescent medical journal. 2018; 20(4)

201. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2020]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

202. Nejati P, Farzinmehr A, Moradi-Lakeh M. The effect of exercise therapy on knee osteoarthritis: a randomized clinical trial. Medical Journal of the Islamic Republic of Iran. 2015; 29:186

203. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a Self-directed Web-Based Strengthening Exercise and Physical Activity Program Supported by Automated Text Messages for People With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Internal Medicine. 2021; 181(6):776-785

204. Nelligan RK, Hinman RS, Kasza J, Schwartz S, Kimp A, Atkins L et al. Effect of a short message service (SMS) intervention on adherence to a physiotherapist-prescribed home exercise program for people with knee osteoarthritis and obesity: protocol for the ADHERE randomised controlled trial. BMC Musculoskeletal Disorders. 2019; 20(1):428

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

206. Nicklas BJ, Ambrosius W, Messier SP, Miller GD, Penninx BW, Loeser RF et al. Diet-induced weight loss, exercise, and chronic inflammation in older, obese adults: a randomized controlled clinical trial. American Journal of Clinical Nutrition. 2004; 79(4):544-551

207. Nour K, Laforest S, Gauvin L, Gignac M. Behavior change following a self-management intervention for housebound older adults with arthritis: an experimental study. International Journal of Behavioral Nutrition & Physical Activity. 2006; 3:12

208. Nunez M, Nunez E, Segur JM, Macule F, Quinto L, Hernandez MV et al. The effect of an educational program to improve health-related quality of life in patients with osteoarthritis on waiting list for total knee replacement: a randomized study. Osteoarthritis and Cartilage. 2006; 14(3):279-285

209. O'Brien KM, Wiggers J, Williams A, Campbell E, Hodder RK, Wolfenden L et al. Telephone-based weight loss support for patients with knee osteoarthritis: a pragmatic randomised controlled trial. Osteoarthritis and Cartilage. 2018; 26(4):485-494

210. Ogut H, Guler H, Yildizgoren MT, Velioglu O, Turhanoglu AD. Does kinesiology taping improve muscle strength and function in knee osteoarthritis? A single-blind, randomized and controlled study. Archives of Rheumatology. 2018; 33(3):335-343

211. Oh SL, Kim DY, Bae JH, Lim JY. Effects of rural community-based integrated exercise and health education programs on the mobility function of older adults with knee osteoarthritis. Aging clinical and experimental research. 2020; 04

212. Oppong R, Jowett S, Lewis M, Clarkson K, Paskins Z, Croft P et al. Cost-effectiveness of a model consultation to support self-management in patients with osteoarthritis. Rheumatology (Oxford, England). 57(6):1056-1063

213. Oppong R, Jowett S, Nicholls E, Whitehurst DG, Hill S, Hammond A et al. Joint protection and hand exercises for hand osteoarthritis: an economic evaluation comparing methods for the analysis of factorial trials. Rheumatology. 2014:epub

214. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2021. Available from: http://www.oecd.org/std/ppp Last accessed: 10/02/2022.

215. Osborne RH, Buchbinder R, Ackerman IN. Can a disease-specific education program augment self-management skills and improve Health-Related Quality of Life in people with hip or knee osteoarthritis? BMC Musculoskeletal Disorders. 2006; 7:90

216. Østerås N. Cost-utility analysis of a cluster randomized controlled trial implementing a structured osteoarthritis model in primary healthcare. Osteoarthritis and Cartilage. 2021; 29:S58-S59

217. Ozguclu E, Cetin A, Cetin M, Calp E. Additional effect of pulsed electromagnetic field therapy on knee osteoarthritis treatment: a randomized, placebo-controlled study. Clinical Rheumatology. 2010; 29(8):927-931

218. Palmer S, Domaille M, Cramp F, Walsh N, Pollock J, Kirwan J et al. Transcutaneous electrical nerve stimulation as an adjunct to education and exercise for knee osteoarthritis: a randomized controlled trial. Arthritis Care and Research. 2014; 66(3):387-394

219. Park J, McCaffrey R, Newman D, Cheung C, Hagen D. The effect of Sit 'n' Fit Chair Yoga among community-dwelling older adults with osteoarthritis. Holistic Nursing Practice. 2014; 28(4):247-257

220. Park YG, Kwon BS, Park JW, Cha DY, Nam KY, Sim KB et al. Therapeutic effect of whole body vibration on chronic knee osteoarthritis. Annals of Rehabilitation Medicine. 2013; 37(4):505-515

221. Patel A, Buszewicz M, Beecham J, Griffin M, Rait G, Nazareth I et al. Economic evaluation of arthritis self management in primary care. BMJ. 2009; 339:b3532

222. Paterson KL, Hinman RS, Metcalf BR, Campbell PK, Menz HB, Hunter DJ et al. Podiatry Intervention Versus Usual General Practitioner Care for Symptomatic Radiographic Osteoarthritis of the First Metatarsophalangeal Joint: A Randomized Clinical Feasibility Study. Arthritis Care and Research. 2021; 73(2):250-258

223. Perez-Marmol JM, Garcia-Rios MC, Ortega-Valdivieso MA, Cano-Deltell EE, Peralta-Ramirez MI, Ickmans K et al. Effectiveness of a fine motor skills rehabilitation program on upper limb disability, manual dexterity, pinch strength, range of fingers motion, performance in activities of daily living, functional independency, and general self-efficacy in hand osteoarthritis: A randomized clinical trial. Journal of Hand Therapy. 2017; 30(3):262-273

224. Peterson MG, 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 and Research. 1993; 6(1):11-16

225. Pitsillides A, Stasinopoulos D, Giannakou K. The effects of cognitive behavioural therapy delivered by physical therapists in knee osteoarthritis pain: A systematic review and meta-analysis of randomized controlled trials. Journal of Bodywork and Movement Therapies. 2021; 25:157-164

226. Piyakhachornrot N, Aree-Ue S, Putwatana P, Kawinwonggowit V. Impact of an integrated health education and exercise program in middle-aged Thai adults with osteoarthritis of the knee. Orthopaedic nursing / National Association of Orthopaedic Nurses. 2011; 30(2):134-142

227. Poulsen E, Christensen HW, Roos EM, Vach W, Overgaard S, Hartvigsen J. Non-surgical treatment of hip osteoarthritis. Hip school, with or without the addition of manual therapy, in comparison to a minimal control intervention: protocol for a three-armed randomized clinical trial. BMC Musculoskeletal Disorders. 2011; 12:88

228. Poulsen E, Hartvigsen J, Christensen HW, Roos EM, Vach W, Overgaard S. Patient education with or without manual therapy compared to a control group in patients with osteoarthritis of the hip. A proof-of-principle three-arm parallel group randomized clinical trial. Osteoarthritis and Cartilage. 2013; 21(10):1494-1503

229. Poulsen E, Hartvigsen J, Christensen HW, Roos EM, Vach W, Overgaard S. Patient education with or without manual therapy compared to a control group in patients with osteoarthritis ofthe hip A- a proof of principle threearm parallel group randomized clinical trial. Osteoarthritis and cartilage. 2013; 21:S145

230. Quilty B, Tucker M, Campbell R, Dieppe P. Physiotherapy, including quadriceps exercises and patellar taping, for knee osteoarthritis with predominant patello-femoral joint involvement: randomized controlled trial. Journal of Rheumatology. 2003; 30(6):1311-1317

231. Rafiq MT, Hamid MSA, Hafiz E, Chaudhary FA, Khan MI. Feasibility and acceptability of instructions of daily care in overweight and obese knee osteoarthritis participants. Current Rheumatology Reviews. 2021; 26:26

232. Rattanachaiyanont M, Kuptniratsaikul V. No additional benefit of shortwave diathermy over exercise program for knee osteoarthritis in peri-/post-menopausal women: an equivalence trial. Osteoarthritis and Cartilage. 2008; 16(7):823-828

233. Ravaud P, Flipo RM, Boutron I, Roy C, Mahmoudi A, Giraudeau B et al. ARTIST (osteoarthritis intervention standardized) study of standardised consultation versus usual care for patients with osteoarthritis of the knee in primary care in France: pragmatic randomised controlled trial. BMJ. 2009; 338:b421

234. Ravaud P, Giraudeau B, Logeart I, Larguier JS, Rolland D, Treves R et al. Management of osteoarthritis (OA) with an unsupervised home based exercise programme and/or patient administered assessment tools. A cluster randomised controlled trial with a 2x2 factorial design. Annals of the Rheumatic Diseases. 2004; 63(6):703-708

235. Rejeski WJ, Focht BC, Messier SP, Morgan T, Pahor M, Penninx B. Obese, older adults with knee osteoarthritis: weight loss, exercise, and quality of life. Health Psychology. 2002; 21(5):419-426

236. Rezende MU, Brito NLR, Farias FES, Silva CAC, Cernigoy CHA, Rodrigues da Silva JM et al. Improved function and strength in patients with knee osteoarthritis as a result of adding a two-day educational program to usual care. Prospective randomized trial. Osteoarthritis and Cartilage Open. 2021; 3 (1) (no pagination)(100137)

237. Robbins SR, Melo LRS, Urban H, Deveza LA, Asher R, Johnson VL et al. Stepped care approach for medial tibiofemoral osteoarthritis (STrEAMline): protocol for a randomised controlled trial. BMJ Open. 2017; 7(12):e018495

238. Rodrigues da Silva JM, de Rezende MU, Spada TC, da Silva Francisco L, Sabine de Farias FE, Clemente da Silva CA et al. Educational program promoting regular physical exercise improves functional capacity and daily living physical activity in subjects with knee osteoarthritis. BMC Musculoskeletal Disorders. 2017; 18(1):546

239. Rogind H, Bibow-Nielsen B, Jensen B, Moller HC, Frimodt-Moller H, Bliddal H. The effects of a physical training program on patients with osteoarthritis of the knees. Archives of Physical Medicine and Rehabilitation. 1998; 79(11):1421-1427

240. Rosemann T, Joos S, Laux G, Gensichen J, Szecsenyi J. Case management of arthritis patients in primary care: a cluster-randomized controlled trial. Arthritis and Rheumatism. 2007; 57(8):1390-1397

241. Runhaar J, Deroisy R, van Middelkoop M, Barretta F, Barbetta B, Oei EH et al. The role of diet and exercise and of glucosamine sulfate in the prevention of knee osteoarthritis: Further results from the PRevention of knee Osteoarthritis in Overweight Females (PROOF) study. Seminars in Arthritis and Rheumatism. 2016; 45(Suppl 4):S42-48

242. Saccomanno MF, Donati F, Careri S, Bartoli M, Severini G, Milano G. Efficacy of intra-articular hyaluronic acid injections and exercise-based rehabilitation programme, administered as isolated or integrated therapeutic regimens for the treatment of knee osteoarthritis. Knee Surgery, Sports Traumatology, Arthroscopy. 2016; 24(5):1686-1694

243. Sanchez Romero EA, Fernandez-Carnero J, Calvo-Lobo C, Ochoa Saez V, Burgos Caballero V, Pecos-Martin D. Is a combination of exercise and dry needling effective for knee OA? Pain Medicine. 2020; 21(2):349-363

244. Saw MM, Kruger-Jakins T, Edries N, Parker R. Significant improvements in pain after a six-week physiotherapist-led exercise and education intervention, in patients with osteoarthritis awaiting arthroplasty, in South Africa: a randomised controlled trial. BMC Musculoskeletal Disorders. 2016; 17:236

245. Schafer AGM, Zalpour C, von Piekartz H, Hall TM, Paelke V. The Efficacy of Electronic Health-Supported Home Exercise Interventions for Patients With Osteoarthritis of the Knee: Systematic Review. Journal of Medical Internet Research. 2018; 20(4):e152

246. Schlenk EA, Fitzgerald GK, Rogers JC, Kwoh CK, Sereika SM. Promoting physical activity in older adults with knee osteoarthritis and hypertension: A randomized controlled trial. Journal of Aging & Physical Activity. 2020; 29(2):207-218

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

248. Schrubbe LA, Ravyts SG, Benas BC, Campbell LC, Cene CW, Coffman CJ et al. Pain coping skills training for African Americans with osteoarthritis (STAART): study protocol of a randomized controlled trial. BMC Musculoskeletal Disorders. 2016; 17(1):359

249. Sevick MA, Miller GD, Loeser RF, Williamson JD, Messier SP. Cost-effectiveness of exercise and diet in overweight and obese adults with knee osteoarthritis. Medicine and Science in Sports and Exercise. 2009; 41(6):1167-1174

250. Sharma M, Singh A, Dhillon MS, Kaur S. Comparative impact of nonpharmacological interventions on pain of knee osteoarthritis patients reporting at a tertiary care institution: A randomized controlled trial. Indian Journal of Palliative Care. 2018; 24(4):478-485

251. Shavianidze GO. A comparative evaluation of the effect of balneopsychotherapy, balneotherapy and psychotherapy on osteoarthrosis patients. Voprosy kurortologii, fizioterapii, i lechebnoi fizicheskoi kultury. 1991; (5):27-28

252. Shea MK, Houston DK, Nicklas BJ, Messier SP, Davis CC, Miller ME et al. The effect of randomization to weight loss on total mortality in older overweight and obese adults: the ADAPT Study. Journals of Gerontology Series A-Biological Sciences & Medical Sciences. 2010; 65(5):519-525

253. Skou ST, Rasmussen S, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O et al. The efficacy of 12 weeks non-surgical treatment for patients not eligible for total knee replacement: a randomized controlled trial with 1-year follow-up. Osteoarthritis and Cartilage. 2015; 23(9):1465-1475

254. Skou ST, Roos EM, Laursen M, Arendt-Nielsen L, Rasmussen S, Simonsen O et al. Cost-effectiveness of 12 weeks of supervised treatment compared to written advice in patients with knee osteoarthritis: a secondary analysis of the 2-year outcome from a randomized trial. Osteoarthritis and Cartilage. 2020; 28(7):907-916

255. Smith-Ray RL, Fitzgibbon ML, Tussing-Humphreys L, Schiffer L, Shah A, Huber GM et al. Fit and Strong! Plus: design of a comparative effectiveness evaluation of a weight management program for older adults with osteoarthritis. Contemporary Clinical Trials. 2014; 37(2):178-188

256. Somers TJ, Blumenthal JA, Guilak F, Kraus VB, Schmitt DO, Babyak MA et al. Pain coping skills training and lifestyle behavioral weight management in patients with knee osteoarthritis: a randomized controlled study. Pain. 2012; 153(6):1199-1209

257. Soni A, Joshi A, Mudge N, Wyatt M, Williamson L. Supervised exercise plus acupuncture for moderate to severe knee osteoarthritis: a small randomised controlled trial. Acupuncture in Medicine. 2012; 30(3):176-181

258. Stamm TA, Machold KP, Smolen JS, Fischer S, Redlich K, Graninger W et al. Joint protection and home hand exercises improve hand function in patients with hand osteoarthritis: a randomized controlled trial. Arthritis and Rheumatism. 2002; 47(1):44-49

259. Steinhilber B, Haupt G, Miller R, Boeer J, Grau S, Janssen P et al. Feasibility and efficacy of an 8-week progressive home-based strengthening exercise program in patients with osteoarthritis of the hip and/or total hip joint replacement: a preliminary trial. Clinical Rheumatology. 2012; 31(3):511-519

260. Steinhilber B, Haupt G, Miller R, Janssen P, Krauss I. Exercise therapy in patients with hip osteoarthritis: Effect on hip muscle strength and safety aspects of exercise-results of a randomized controlled trial. Modern Rheumatology. 2017; 27(3):493-502

261. Stener-Victorin E, Kruse-Smidje C, Jung K. Comparison between electro-acupuncture and hydrotherapy, both in combination with patient education and patient education alone, on the symptomatic treatment of osteoarthritis of the hip. Clinical Journal of Pain. 2004; 20(3):179-185

262. Stoffer-Marx MA, Klinger M, Luschin S, Meriaux-Kratochvila S, Zettel-Tomenendal M, Nell-Duxneuner V et al. Functional consultation and exercises improve grip strength

in osteoarthritis of the hand - a randomised controlled trial. Arthritis Research & Therapy. 2018; 20(1):253

263. 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 and Research. 1998; 11(4):228-233

264. Svege I, Fernandes L, Nordsletten L, Holm I, Risberg MA. Long-term effect of exercise therapy and patient education on impairments and activity limitations in people with hip osteoarthritis: Secondary outcome analysis of a randomized clinical trial. Physical Therapy. 2016; 96(6):818-827

265. Svege I, Nordsletten L, Fernandes L, Risberg MA. Exercise therapy may postpone total hip replacement surgery in patients with hip osteoarthritis: a long-term follow-up of a randomised trial. Annals of the Rheumatic Diseases. 2015; 74(1):164-169

266. Tak E, Staats P, Van Hespen A, Hopman-Rock M. The effects of an exercise program for older adults with osteoarthritis of the hip. Journal of Rheumatology. 2005; 32(6):1106-1113

267. Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. Journal of the American Geriatrics Society. 2003; 51(3):387-392

268. Talbot LA, Gaines JM, Ling SM, Metter EJ. A home-based protocol of electrical muscle stimulation for quadriceps muscle strength in older adults with osteoarthritis of the knee. Journal of Rheumatology. 2003; 30(7):1571-1578

269. Taylor SS, Oddone EZ, Coffman CJ, Jeffreys AS, Bosworth HB, Allen KD. Cognitive Mediators of Change in Physical Functioning in Response to a Multifaceted Intervention for Managing Osteoarthritis. International Journal of Behavioral Medicine. 2018; 25(2):162-170

270. Tegiacchi T. Interaction of energetic points, tendinomuscular meridian and 5 elements in the treatment of osteoarthritis of the knee in patients over 45 years old: a randomized controlled trial. Jams Journal of Acupuncture & Meridian Studies. 2018; 18:18

271. Teirlinck CH, Luijsterburg PA, Dekker J, Bohnen AM, Verhaar JA, Koopmanschap MA et al. Effectiveness of exercise therapy added to general practitioner care in patients with hip osteoarthritis: a pragmatic randomized controlled trial. Osteoarthritis and Cartilage. 2016; 24(1):82-90

272. Thomas KS, Miller P, Doherty M, Muir KR, Jones AC, O'Reilly SC. Cost effectiveness of a two-year home exercise program for the treatment of knee pain. Arthritis and Rheumatism (Arthritis Care and Research). 2005; 53(3):388-394

273. Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. BMJ. 2002; 325(7367):752

274. Umapathy H, Bennell K, Dickson C, Dobson F, Fransen M, Jones G et al. My joint pain: web-based osteoarthritis management resource improves quality of care. Osteoarthritis and cartilage. 2015; 23:A201

275. van Gool CH, Penninx BW, Kempen GI, Rejeski WJ, Miller GD, van Eijk JT et al. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. Arthritis and Rheumatism. 2005; 53(1):24-32

276. Vas J, Mendez C, Perea-Milla E, Vega E, Panadero MD, Leon JM et al. Acupuncture as a complementary therapy to the pharmacological treatment of osteoarthritis of the knee: randomised controlled trial. BMJ. 2004; 329(7476):1216

277. Victor CR, Triggs E, Ross F, Lord J, Axford JS. Lack of benefit of a primary care-based nurse-led education programme for people with osteoarthritis of the knee. Clinical Rheumatology. 2005; 24(4):358-364

278. Villadsen A, Overgaard S, Holsgaard-Larsen A, Christensen R, Roos EM. Immediate efficacy of neuromuscular exercise in patients with severe osteoarthritis of the hip or knee: a secondary analysis from a randomized controlled trial. Journal of Rheumatology. 2014; 41(7):1385-1394

279. Wallis JA, Webster KE, Levinger P, Singh PJ, Fong C, Taylor NF. A walking program for people with severe knee osteoarthritis did not reduce pain but may have benefits for cardiovascular health: a phase II randomised controlled trial. Osteoarthritis and Cartilage. 2017; 25(12):1969-1979

280. Walsh N, Jones L, Phillips S, Thomas R, Odondi L, Palmer S et al. Facilitating Activity and Self-management for people with Arthritic knee, hip or lower back pain (FASA): A cluster randomised controlled trial. Musculoskeletal Science & Practice. 2020; 50:102271

281. Wang V, Allen K, Van Houtven CH, Coffman C, Sperber N, Mahanna EP et al. Supporting teams to optimize function and independence in Veterans: a multi-study program and mixed methods protocol. Implementation Science. 2018; 13(1):58

282. Wang Y, Lombard C, Hussain SM, Harrison C, Kozica S, Brady SRE et al. Effect of a low-intensity, self-management lifestyle intervention on knee pain in community-based young to middle-aged rural women: a cluster randomised controlled trial. Arthritis Research & Therapy. 2018; 20(1):74

283. Warsi A, LaValley MP, Wang PS, Avorn J, Solomon DH. Arthritis self-management education programs: A meta-analysis of the effect on pain and disability. Arthritis and Rheumatism. 2003; 48(8):2207-2213

284. Woods B, Manca A, Weatherly H, Saramago P, Sideris E, Giannopoulou C et al. Cost-effectiveness of adjunct non-pharmacological interventions for osteoarthritis of the knee. PLoS ONE [Electronic Resource]. 2017; 12(3):e0172749

285. Yan H, Su Y, Chen L, Zheng G, Lin X, Chen B et al. Rehabilitation for the management of knee osteoarthritis using comprehensive traditional Chinese medicine in community health centers: study protocol for a randomized controlled trial. Trials [Electronic Resource]. 2013; 14:367

286. Yilmaz M, Sahin M, Algun ZC. Comparison of effectiveness of the home exercise program and the home exercise program taught by physiotherapist in knee osteoarthritis. Journal of Back and Musculoskeletal Rehabilitation. 2019; 32(1):161-169

287. Yip YB, Sit JW, Fung KK, Wong DY, Chong SY, Chung LH et al. Effects of a self-management arthritis programme with an added exercise component for osteoarthritic knee: randomized controlled trial. Journal of Advanced Nursing. 2007; 59(1):20-28

288. Yip YB, Sit JW, Fung KK, Wong DY, Chong SY, Chung LH et al. Impact of an Arthritis Self-Management Programme with an added exercise component for osteoarthritic knee sufferers on improving pain, functional outcomes, and use of health care

services: An experimental study. Patient Education and Counseling. 2007; 65(1):113-121

289. Yip YB, Sit JW, Wong DY, Chong SY, Chung LH. A 1-year follow-up of an experimental study of a self-management arthritis programme with an added exercise component of clients with osteoarthritis of the knee. Psychology Health & Medicine. 2008; 13(4):402-414

290. Yurtkuran M, Kocagil T. TENS, electroacupuncture and ice massage: comparison of treatment for osteoarthritis of the knee. American Journal of Acupuncture. 1999; 27(3-4):133-140

291. Zacharias A, Green RA, Semciw AI, Kingsley MI, Pizzari T. Efficacy of rehabilitation programs for improving muscle strength in people with hip or knee osteoarthritis: a systematic review with meta-analysis. Osteoarthritis and Cartilage. 2014; 22(11):1752-1773

292. Zammit G, Menz H, Munteanu S, Landorf K, Gilheany M. Interventions for treating osteoarthritis of the big toe joint. Cochrane Database of Systematic Reviews 2010, Issue 9. Art. No.: CD007809. DOI: 10.1002/14651858.CD007809.pub2.

293. Zgibor JC, Ye L, Boudreau RM, Conroy MB, Vander Bilt J, Rodgers EA et al. Community-based healthy aging interventions for older adults with arthritis and multimorbidity. Journal of Community Health. 2017; 42(2):390-399

294. Zhou SF, Xue CC. Acupuncture as an adjunct to exercise-based physiotherapy does not improve the pain of knee osteoarthritis. Australian journal of acupuncture and chinese medicine. 2008; 3(1):53-55

295. Zhou YF, Wang JY, Yan JL. Clincal research on the treatment of degenerative knee arthritis by massage combined with electric acupuncture. China journal of chinese medicine [zhong yi xue bao]. 2015; 30(4):603-605

Appendices

Appendix A – Review protocols

Review protocol for the clinical and cost-effectiveness of treatment packages for the management of osteoarthritis

ID	Field	Content
0.	PROSPERO registration number	CRD42020221541
1.	Review title	What is the clinical and cost-effectiveness of treatment packages (that include combinations of interventions) for the management of osteoarthritis?
2.	Review question	5.1 What is the clinical and cost-effectiveness of treatment packages (that include combinations of interventions) for the management of osteoarthritis?
3.	Objective	To evaluate the clinical and cost-effectiveness of treatment packages, where combinations of interventions are used together, for the management of osteoarthritis.
4.	Searches	The following databases will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE Searches will be restricted by: • English language • Human studies • Letters and comments are excluded

		Other searches: • Inclusion lists of relevant systematic reviews will be checked by the reviewer. The searches may be re-run 6 weeks before final submission of the review and further studies retrieved for inclusion if relevant. The full search strategies for MEDLINE database will be published in the final review.
5.	Condition or domain being studied	Osteoarthritis (of any joint) in adults (defined as a clinical diagnosis of osteoarthritis with or without imaging)
6.	Population	Inclusion: • Adults (age ≥16 years) with osteoarthritis affecting any joint Exclusion: • Children (age <16 years) • People with conditions that may make them susceptible to osteoarthritis or often occur alongside osteoarthritis (including: crystal arthritis, inflammatory arthritis, septic arthritis, hemochromatosis, haemophilic arthropathy, diseases of childhood that may predispose to osteoarthritis, and malignancy). • Studies with an unclear population (e,g, proportion of participants with osteoarthritis unclear) • Spinal osteoarthritis
7.	Intervention/Exposure/Test	Treatment packages (minimum intervention duration 1 week). A treatment package is defined as any intervention for osteoarthritis (including: exercise, manual therapy, electrotherapy, acupuncture, devices, pharmacological management [including oral, topical, transdermal and intra-articular formulations], arthroscopic procedures) combined with one of the following:

		3. Behaviour change interventions (for example: joint protection principles, cognitive-behavioural therapy) 4. An education programme, including those based on behavioural theory (defined as education sessions provided by one or more healthcare professionals over multiple sessions where the study provides clear information about the content included in the education sessions)
8.	Comparator/Reference standard/Confounding factors	• Non-combined active treatment for osteoarthritis, started at the time of trial initiation ○ Exercise ○ Manual therapy ○ Electrotherapy ○ Acupuncture ○ Devices ○ Pharmacological management (oral, topical, transdermal or intra-articular therapy) ○ Arthroscopic procedures ○ Other (education programmes, behaviour change interventions) • Standard care (non-organised) or no treatment **No treatment defined as either (1) doing nothing or (2) very low intensity intervention such as advice*
9.	Types of study to be included	• Systematic reviews of RCTs • Parallel RCTs
10.	Other exclusion criteria	• Non-English language studies • Non-randomised/observational studies • Crossover RCTs • Abstracts will be excluded as it is expected there will be sufficient full text published studies available.
11.	Context	N/A
12.	Primary outcomes (critical outcomes)	Stratify by ≤/>3 months (longest time-point in each): • Health-related quality of life [validated patient-reported outcomes, continuous data prioritised]

		• Pain [validated patient-reported outcomes, continuous data prioritised] • Physical function [validated patient-reported outcomes, continuous data prioritised] *The COMET database was searched and several core outcome sets were identified for specific sites of osteoarthritis (including hand, knee and hip). The committee took these into account when defining outcomes:* https://onlinelibrary.wiley.com/doi/full/10.1002/acr.22868 https://www.ncbi.nlm.nih.gov/pubmed/26136489 https://www.ncbi.nlm.nih.gov/pubmed/30647185 The committee did not include stiffness or global scores as Delphi discussions by the OMERACT group have found these to not be as important to people with osteoarthritis or clinicians. The outcomes included were universal for all groups allowing for broader comparisons.
13.	Secondary outcomes (important outcomes)	• Psychological distress [validated patient-reported outcomes, continuous data prioritised] • Osteoarthritis flares [dichotomous data prioritised] • Discontinuation [dichotomous data]
14.	Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. All references identified by the searches and from other sources will be screened for inclusion. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. The full text of potentially eligible studies will be retrieved and will be assessed in line with the criteria outlined above. EviBASE will be used for data extraction. Study investigators may be contacted for missing data where time and resources allow.
15.	Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual For intervention reviews the following checklists will be used according to the study design being assessed: • Systematic reviews: Risk of Bias in Systematic Reviews (ROBIS)

		• Randomised Controlled Trial: Cochrane RoB (2.0) 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately • a sample of the data extractions • correct methods are used to synthesise data • a sample of the risk of bias assessments Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.
16.	Strategy for data synthesis	• Pairwise meta-analyses will be performed using Cochrane Review Manager (RevMan5). • GRADEpro will be used to assess the quality of evidence for each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) will be appraised for each outcome. Publication bias is tested for when there are more than 5 studies for an outcome. The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/ • Where meta-analysis is not possible, data will be presented and quality assessed individually per outcome. • WinBUGS will be used for network meta-analysis, if possible given the data identified. Heterogeneity between studies in the effect measures will be assessed using the I^2 statistic and visual inspection. We will consider an I^2 value great than 50% as indicative of substantial heterogeneity. If significant heterogeneity is identified during meta-analysis then subgroup analysis, using subgroups predefined by the GC, will take place. If this does not explain the heterogeneity, the results will be presented using a random-effects model.
17.	Analysis of sub-groups	Subgroup analysis to be conducted if heterogeneity in the meta-analysis is present: • Site of osteoarthritis

		• Diagnosis with or without imaging (indicative of severity) • Multimorbidity (high versus low morbidity score; as defined by study, measured by validated instruments e.g. Charlson Comorbidity Index) • Age (≤/> 75 years) • Length of package (≤/>6 weeks) • Behaviour change interventions or education program		
18.	Type and method of review	☒	Intervention	
		☐	Diagnostic	
		☐	Prognostic	
		☐	Qualitative	
		☐	Epidemiologic	
		☐	Service Delivery	
		☐	Other (please specify)	
19.	Language	English		
20.	Country	England		
21.	Anticipated or actual start date	23/08/2019		
22.	Anticipated completion date	25/08/2021		
23.	Stage of review at time of this submission	Review stage	Started	Completed
		Preliminary searches	☑	☐

		Piloting of the study selection process	☐	☐
		Formal screening of search results against eligibility criteria	☐	☐
		Data extraction	☐	☐
		Risk of bias (quality) assessment	☐	☐
		Data analysis	☐	☐
24.	Named contact	5a. Named contact National Guideline Centre 5b Named contact e-mail [Guideline email]@nice.org.uk [Developer to check with Guideline Coordinator for email address] 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE) and the National Guideline Centre		
25.	Review team members	From the National Guideline Centre: Carlos Sharpin [Guideline lead] Julie Nielson [Senior systematic reviewer]		

		George Wood [Systematic reviewer] David Wonderling [Senior health economist] Joseph Runicles [Information specialist] Amber Hernaman [Project manager]
26.	Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.
27.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
28.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10127
29.	Other registration details	
30.	Reference/URL for published protocol	
31.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.

32.	Keywords	Adults; Advice; Combinations; Education; Non-pharmacological; Osteoarthritis; Pharmacological; Physiotherapy; Treatment packages	
33.	Details of existing review of same topic by same authors		
34.	Current review status	☒	Ongoing
		☐	Completed but not published
		☐	Completed and published
		☐	Completed, published and being updated
		☐	Discontinued
35..	Additional information	N/A	
36.	Details of final publication	www.nice.org.uk	

Table 16. Health economic review protocol

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken for all years using population-specific terms and a health economic study filter – see appendix B below.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2005, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Studies published in 2005 or later, that were included in the previous guidelines, will be reassessed for inclusion and may be included or selectively excluded based on their relevance to the questions covered in this update and whether more applicable evidence is also identified. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual (2014).[201] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations' then it will be included in the guideline. A health economic evidence table will be completed and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations' then it will usually be excluded from the guideline. If it is excluded then a health economic evidence table will not be completed and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included.

Where there is discretion

The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below.

The health economist will be guided by the following hierarchies.

Setting:

- UK NHS (most applicable).
- OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden).
- OECD countries with predominantly private health insurance systems (for example, Switzerland).
- Studies set in non-OECD countries or in the USA will be excluded before being assessed for applicability and methodological limitations.

Health economic study type:

- Cost–utility analysis (most applicable).
- Other type of full economic evaluation (cost–benefit analysis, cost-effectiveness analysis, cost–consequences analysis).
- Comparative cost analysis.
- Non-comparative cost analyses including cost-of-illness studies will be excluded before being assessed for applicability and methodological limitations.

Year of analysis:

- The more recent the study, the more applicable it will be.
- Studies published in 2005 or later (including any such studies included in the previous guidelines) but that depend on unit costs and resource data entirely or predominantly from before 2005 will be rated as ‘Not applicable’.
- Studies published before 2005 (including any such studies included in the previous guidelines) will be excluded before being assessed for applicability and methodological limitations.

Quality and relevance of effectiveness data used in the health economic analysis:

- The more closely the clinical effectiveness data used in the health economic analysis match with the outcomes of the studies included in the clinical review the more useful the analysis will be for decision-making in the guideline.

Appendix B – Literature search strategies

- What is the clinical and cost-effectiveness of treatment packages (that include combinations of interventions) for the management of osteoarthritis?

The literature searches for this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual.[201]

For more information, please see the Methodology review published as part of the accompanying documents for this guideline.

B.1 Clinical search literature search strategy

Searches were constructed using an Osteoarthritis population. All results were then sifted for each question. Search filters were applied to the search where appropriate.

Table 17: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 17 November 2021	Randomised controlled trials Systematic review studies Exclusions (animals studies, letters, comments)
Embase (OVID)	1974 – 17 November 2021	Randomised controlled trials Systematic review studies Exclusions (animals studies, letters, comments)
The Cochrane Library (Wiley)	Cochrane Reviews to 2021 Issue 11 of 12 CENTRAL to 2021 Issue 11 of 12	None

Medline (Ovid) search terms

1.	exp osteoarthritis/
2.	(osteoarthriti* or osteo-arthriti* or osteoarthrotic or osteoarthros*).ti,ab.
3.	(degenerative adj2 arthritis).ti,ab.
4.	coxarthrosis.ti,ab.
5.	gonarthrosis.ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14

16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice or rodent*).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	randomized controlled trial.pt.
28.	controlled clinical trial.pt.
29.	randomi#ed.ti,ab.
30.	placebo.ab.
31.	randomly.ti,ab.
32.	Clinical Trials as topic.sh.
33.	trial.ti.
34.	or/27-33
35.	Meta-Analysis/
36.	exp Meta-Analysis as Topic/
37.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
38.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
39.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
40.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
41.	(search* adj4 literature).ab.
42.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
43.	cochrane.jw.
44.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
45.	or/35-44
46.	26 and (34 or 45)

Embase (Ovid) search terms

1.	exp osteoarthritis/
2.	(osteoarthriti* or osteo-arthriti* or osteoarthrotic or osteoarthros*).ti,ab.
3.	(degenerative adj2 arthritis).ti,ab.
4.	coxarthrosis.ti,ab.
5.	gonarthrosis.ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.

12.	or/7-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice or rodent*).ti.
22.	or/14-21
23.	6 not 22
24.	Limit 23 not English language
25.	random*.ti,ab.
26.	factorial*.ti,ab.
27.	(crossover* or cross over*).ti,ab.
28.	((doubl* or singl*) adj blind*).ti,ab.
29.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
30.	crossover procedure/
31.	single blind procedure/
32.	randomized controlled trial/
33.	double blind procedure/
34.	or/25-33
35.	systematic review/
36.	meta-analysis/
37.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
38.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
39.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
40.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
41.	(search* adj4 literature).ab.
42.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
43.	cochrane.jw.
44.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
45.	or/35-44
46.	24 and (34 or 45)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Osteoarthritis] explode all trees
#2.	(osteoarthriti* or osteo-arthriti* or osteoarthrotic or osteoarthros*):ti,ab
#3.	(degenerative near/2 arthritis):ti,ab
#4.	coxarthrosis:ti,ab
#5.	gonarthrosis:ti,ab

#6.	(or #1-#5)

B.2 Health Economics literature search strategy

Health economic evidence was identified by conducting a broad search relating to a Gout population in NHS Economic Evaluation Database (NHS EED – this ceased to be updated after March 2015) and the Health Technology Assessment database (HTA – this ceased to be updates after March 2018). NHS EED and HTA databases are hosted by the Centre for Research and Dissemination (CRD). Additional searches were run on Medline and Embase for health economics studies and quality of life studies. Searches for quality of life studies were run for general information.

Table 18: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	1 January 2014 – 17 November 2021	Health economics studies Quality of life studies Exclusions (animals studies, letters, comments)
Embase	1 January 2014 – 17 November 2021	Health economics studies Quality of life studies Exclusions (animals studies, letters, comments)
Centre for Research and Dissemination (CRD)	HTA - Inception – 31 March 2018 NHSEED - Inception to 31 March 2015	None

Medline (Ovid) search terms

1.	exp osteoarthritis/
2.	(osteoarthriti* or osteo-arthriti* or osteoarthrotic or osteoarthros*).ti,ab.
3.	(degenerative adj2 arthritis).ti,ab.
4.	coxarthrosis.ti,ab.
5.	gonarthrosis.ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16

18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice or rodent*).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	Economics/
28.	Value of life/
29.	exp "Costs and Cost Analysis"/
30.	exp Economics, Hospital/
31.	exp Economics, Medical/
32.	Economics, Nursing/
33.	Economics, Pharmaceutical/
34.	exp "Fees and Charges"/
35.	exp Budgets/
36.	budget*.ti,ab.
37.	cost*.ti.
38.	(economic* or pharmaco?economic*).ti.
39.	(price* or pricing*).ti,ab.
40.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
41.	(financ* or fee or fees).ti,ab.
42.	(value adj2 (money or monetary)).ti,ab.
43.	or/27-42
44.	quality-adjusted life years/
45.	sickness impact profile/
46.	(quality adj2 (wellbeing or well being)).ti,ab.
47.	sickness impact profile.ti,ab.
48.	disability adjusted life.ti,ab.
49.	(qal* or qtime* or qwb* or daly*).ti,ab.
50.	(euroqol* or eq5d* or eq 5*).ti,ab.
51.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
52.	(hui or hui1 or hui2 or hui3).ti,ab.
53.	(health* year* equivalent* or hye or hyes).ti,ab.
54.	discrete choice*.ti,ab.
55.	rosser.ti,ab.
56.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.

57.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
58.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
59.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
60.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
61.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
62.	or/44-61
63.	26 and (43 or 62)

Embase (Ovid) search terms

1.	exp osteoarthritis/
2.	(osteoarthriti* or osteo-arthriti* or osteoarthrotic or osteoarthros*).ti,ab.
3.	(degenerative adj2 arthritis).ti,ab.
4.	coxarthrosis.ti,ab.
5.	gonarthrosis.ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice or rodent*).ti.
22.	or/14-21
23.	6 not 22
24.	Limit 23 to English language
25.	health economics/
26.	exp economic evaluation/
27.	exp health care cost/
28.	exp fee/
29.	budget/
30.	funding/
31.	budget*.ti,ab.

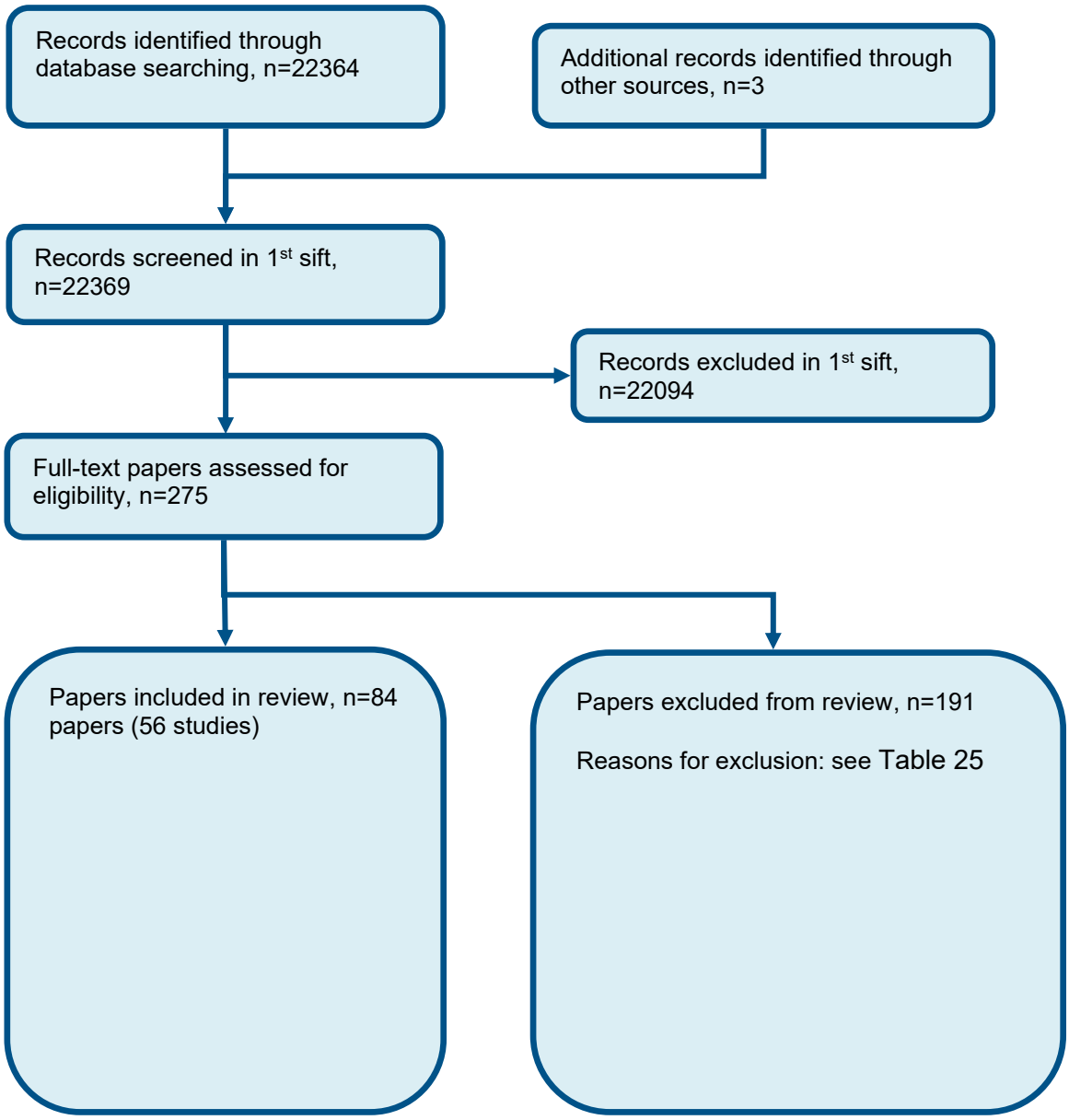
32.	cost*.ti.
33.	(economic* or pharmaco?economic*).ti.
34.	(price* or pricing*).ti,ab.
35.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
36.	(financ* or fee or fees).ti,ab.
37.	(value adj2 (money or monetary)).ti,ab.
38.	or/25-37
39.	quality adjusted life year/
40.	"quality of life index"/
41.	short form 12/ or short form 20/ or short form 36/ or short form 8/
42.	sickness impact profile/
43.	(quality adj2 (wellbeing or well being)).ti,ab.
44.	sickness impact profile.ti,ab.
45.	disability adjusted life.ti,ab.
46.	(qal* or qtime* or qwb* or daly*).ti,ab.
47.	(euroqol* or eq5d* or eq 5*).ti,ab.
48.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
49.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
50.	(hui or hui1 or hui2 or hui3).ti,ab.
51.	(health* year* equivalent* or hye or hyes).ti,ab.
52.	discrete choice*.ti,ab.
53.	rosser.ti,ab.
54.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
55.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
56.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
57.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
58.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
59.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
60.	or/39-59
61.	24 and (38 or 60)

NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Osteoarthritis EXPLODE ALL TREES
#2.	((osteoarthriti* or osteo-arthriti* or osteoarthrotic or osteoarthros*))
#3.	((degenerative adj2 arthritis))
#4.	(coxarthrosis)
#5.	(gonarthrosis)
#6.	#1 OR #2 OR #3 OR #4 OR #5
#7.	(#6) IN NHSEED
#8.	(#6) IN HTA

Appendix C – Effectiveness evidence study selection

Figure 1: Flow chart of clinical study selection for the review of treatment packages for the management of osteoarthritis



Appendix D – Effectiveness evidence

Study (subsidiary papers)	Adams 2021[4] (Adams 2019[3])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=349)
Countries and setting	Conducted in United Kingdom; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks of therapy, 12 weeks overall
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Base of thumb osteoarthritis reporting at least moderate hand pain (>5) and dysfunction (>9) on the Australian Canadian outcome measure.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Adults aged >30 years with symptomatic base of thumb osteoarthritis reporting at least moderate hand pain (>5) and dysfunction (>9) on the Australian Canadian outcome measure; show signs and symptoms of thumb base osteoarthritis on clinical enquiry and examination; no other household member participating in the trial; able to give written informed consent; available to attend Occupational Therapy/Physiotherapy/Hand Therapy sessions.
Exclusion criteria	Consultations with therapy department or treatment for this thumb problem (excluding pain killers and anti-inflammatories) in the previous six months; intra-articular joint injection to wrist, fingers or thumb in the previous two months; fractures or significant injury or surgery to the wrist or hand within the previous six months; red flags (i.e. history of serious illness or disease, such as gout, psoriatic arthritis, ankylosing spondylitis, connective tissue disorders, resulting in inflammatory arthritis in the hand/s or progressive neurological signs, or acute swollen hand joint); diagnosis of dementia or significant disorder likely to affect communication; already received thumb splints for thumb base osteoarthritis; skin disease that may interfere or contraindicate splint wear; participant of a drug or medical device trial in the last 12 weeks.
Recruitment/selection of patients	Conducted across 17 National Health service hospitals in England. Conducted between March 2017 and December 2018 by the Oxford Clinical Trials Research Unit, UK.
Age, gender and ethnicity	Age - Mean (SD): 62.6 (9.6). Gender (M:F): 75:274. Ethnicity: White British = 338
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Thumb
Extra comments	Severity: Not stated/unclear Duration of symptoms (median [IQR]): Between 2 (0,4) and 1 (0,3).

Study (subsidiary papers)	Adams 2021[4] (Adams 2019[3])
Indirectness of population	No indirectness
Interventions	(n=116) Intervention 1: Treatment package - Devices and education programme. A supported self-management programme (including education) with one of two verum thumb splints, either a Procool thumb carpometacarpal restriction black splint or a beige Orfilight 2.5 mm 3/32" microperforated trouser leg splint custom made using a standard template. The self-management programme consisted of 90 minute 1:1 therapist intervention over two hospital visits, including: 1) information on thumb base pain and instructions on how to carry out a hand exercise programme for thumb base pain. The exercise programme was supported by a trial specific colour hand exercise booklet. This booklet contained four main sections: 1. Causes of Thumb Osteoarthritis, 2. Symptoms of Thumb Osteoarthritis, 3. Treatment of Thumb Osteoarthritis, 4. Hand Exercises. The hand exercise programme involved a warm up exercise, Level 1, Level 2 and Level 3 hand exercises. Participants were requested to repeat the hand exercises at least 3 times a week for at least 20 minutes each time. They were advised to always become aware of their hand and thumb position and to avoid positions of thumb deformity. Participants were advised to start the hand exercises with the Warm-Up Exercise. Participants were asked to warm up their hand by placing their hand in a bowl of warm water and gently move their thumb in a circular direction. After one minute changing direction and carrying out these gentle moves for at least 2 minutes. They gradually increased the level of exercise with level 1 exercises including active range of motion exercises for thumb abduction, extension and thumb opposition, level 2 exercises including resistive range of motion exercises for thumb abduction and extension using latex free rubber bands; level 3 exercises including functional pinch tasks using 2 point pinch, 3 point pinch and lateral pinch activities using daily objects such as plates, pens, paper and clothes pegs and grip and turn tasks using daily objects such as a key and bottle tops. Other elements provided included: the Arthritis Research UK Osteoarthritis booklet, a discussion with the therapist about the potential facilitators and barriers to engaging with self-management and a self-management contract sheet; a hand exercise diary.. Duration 8 weeks. Concurrent medication/care: No additional information.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (8 weeks). (n=116) Intervention 2: Non-combined active treatment - Education programme. Self-management intervention only. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (8 weeks). (n=117) Intervention 3: Treatment package - Devices and education programme. Self management program and placebo splint. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (8 weeks). Comments: This group was excluded from the analysis as the comparison including it was not included in the protocol

Study (subsidiary papers)	Adams 2021[4] (Adams 2019[3])
Funding	Academic or government funding (This work was funded by UK Versus Arthritis (Grant Project Number 21019). Versus Arthritis approved the appointment of a Trial Steering Committee and Data Management Committee to scrutinize and oversee the running of this trial. All splints for the trial were purchased.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: DEVICES AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: EQ-5D-5L Index at 12 weeks; Group 1: mean 0.63 (SD 0.22); n=84, Group 2: mean 0.61 (SD 0.24); n=83; EQ-5D-5L Index -0.11-1 Top=High is good outcome; Comments: Baseline devices and education programme: 0.59 (0.21). Baseline education programme: 0.58 (0.23).
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25, Reason: 16 withdrew before 12 weeks, 9 lost to follow-up/questionnaire not completed; Group 2 Number missing: 26, Reason: 16 withdrew before 12 weeks, 10 lost to follow-up/questionnaire not completed

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: AUSCAN hand pain index at 12 weeks; Group 1: mean 9.7 (SD 3.9); n=91, Group 2: mean 9.7 (SD 4); n=90; AUSCAN hand pain index 0-20 Top=High is poor outcome; Comments: Baseline devices and education programme: 11.9 (3.2). Baseline education programme: 12.0 (2.9).
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25, Reason: 16 withdrew before 12 weeks, 9 lost to follow-up/questionnaire not completed; Group 2 Number missing: 26, Reason: 16 withdrew before 12 weeks, 10 lost to follow-up/questionnaire not completed

Protocol outcome 3: Physical function at ≤3 months
- Actual outcome: AUSCAN hand function index at 12 weeks; Group 1: mean 17.3 (SD 7.9); n=85, Group 2: mean 18.2 (SD 7.5); n=84; AUSCAN hand function index 0-36 Top=High is good outcome; Comments: Baseline devices and education programme: 21.5 (6.6). Baseline education programme: 21.3 (5.7).
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25, Reason: 16 withdrew before 12 weeks, 9 lost to follow-up/questionnaire not completed; Group 2 Number missing: 26, Reason: 16 withdrew before 12 weeks, 10 lost to follow-up/questionnaire not completed

Protocol outcome 4: Discontinuation at ≤3 months
- Actual outcome: Discontinuation at 12 weeks; Group 1: 25/116, Group 2: 26/116; Comments: Devices and education programme: 16 withdrew before 12 weeks, 9 lost to follow up/questionnaire not completed. Education program: 16 withdrew before 12 weeks, 10 lost to follow-up/questionnaire not completed.
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover -

Study (subsidiary papers)	Adams 2021[4] (Adams 2019[3])
Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 25, Reason: 16 withdrew before 12 weeks, 9 lost to follow-up/questionnaire not completed; Group 2 Number missing: 26, Reason: 16 withdrew before 12 weeks, 10 lost to follow-up/questionnaire not completed	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 mo

Study	ADAPT trial: Rejeski 2002[235]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=278)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 18 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee pain on most days of the month, limitations in activity and radiographic tibiofemoral osteoarthritis on weight-bearing anteroposterior x-rays
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age at least 60 years; calculated BMI at least 28; knee pain on most days of the month; sedentary activity pattern with less than 20 minutes of formal exercise once per week for the past 6 months; self-reported difficulty in at least one of the following activities ascribed to knee pain: walking 0.25 miles (3-4 city blocks), climbing stairs, bending, stooping, kneeling, shopping, housecleaning; or other self-care activities, such as getting in and out of bed, standing up from a chair, lifting and carrying groceries, or getting in and out of the bathtub; radiographic evidence of tibio-femoral osteoarthritis as determined by a single observer and on the basis of weight-bearing anteroposterior x-rays; willingness to undergo testing and intervention procedures.
Exclusion criteria	A serious medical condition that prevented safe participation in an exercise program, such as coronary artery disease, severe hypertension, peripheral vascular disease, stroke, congestive heart failure, chronic obstructive pulmonary disease, insulin-dependent diabetes, psychiatric disease, renal disease, liver disease, active cancer other than skin cancer and anaemia; a Mini-Mental score <24; an inability to walk without a cane or other assistive device; participation in another research study; excessive alcohol use with a cutoff of at least 14 drinks per week; an inability to complete the protocol, in the opinion of the clinical staff, because of frailty, illness or other reasons.
Recruitment/selection of patients	People were recruited through mass mailing and other more focused strategies (e.g. letters to minority churches)
Age, gender and ethnicity	Age - Mean (SD): 68.52 (6.30). Gender (M:F): 78:200. Ethnicity: Caucasian = 211, Other = 67

Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Low morbidity score (48.73% had hypertension, 15.51% had cardiovascular disease, 9.49% had diabetes). 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=68) Intervention 1: Treatment package - Exercise and behaviour change intervention. Exercise and dietary weight loss. Exercise was a 3-day-per-week program including an aerobic phase (15 minutes), a resistance-training phase (15 minutes), a second aerobic phase (15 minutes) and a cool-down phase (15 minutes). The first 4 months were facility based and then people could opt in for a home based program after completing a 2 month transition phase. The aerobic exercise included walking within a heart rate range of 50-75% of the heart-rate reserve, whereas the resistance-training portion of the program consisted of two sets of 12 repetitions of the following exercises: leg extension, leg curl, heel raise and step up. Cuff weights and weighted vests were used to provide resistance, and a 1-1.5 min rest interval separated by each exercise. Weight was increased after the person performed two sets of 12 repetitions for 2 consecutive days. Dietary weight loss advise was provided to aim for a 5% weight loss that would be maintained throughout the 18 month intervention period. People were give three phases of support: intensive (months 1-4), transition (months 5-6) and maintenance (months 7-18). The major emphasis of the intensive phase was to heighten awareness of the importance of and the need to change eating habits to lower caloric intake. Behaviour change was facilitated through the use of self-regulatory skills. These skills included self-monitoring, goal setting, cognitive restructuring, problem solving, and environmental management. One introductory individual session was followed by 16 weekly se sessions and 1 individual session each month. Each group session included problem solving, reviewing a specific topic, and food tasting of several well-balanced, low-fat, nutritious meals prepared with widely available food products. The transition phase included 8 weeks of biweekly contacts (three group sessions and one individual session). The goals included: assisting people who had not reached their weight goals to reestablish new goals and maintaining and preventing relapse in those participants who had reached their weight-loss goals. The maintenance phase included monthly meetings and phone contacts alternated every 2 weeks. Additionally, newsletters were mailed at regular intervals that provided nutritional information and notice of upcoming meetings. The maintenance phase included: assisting people who had reached their weight loss goals to maintain this weight loss and providing counsel for people who had had a

	difficult time losing weight and adhering to the intervention. . Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Weight loss advice). 2. Length of package: > 6 weeks (18 months). (n=69) Intervention 2: Non-combined active treatment - Exercise. Exercise only. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (18 months). (n=73) Intervention 3: Non-combined active treatment - Behaviour change intervention. Weight loss advice only. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Weight loss advice). 2. Length of package: > 6 weeks (18 months). (n=68) Intervention 4: Non-combined active treatment - Education programme. Healthy lifestyle advice included a discussion group and education sessions. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (18 months). Comments: This group was not included in the analysis as it was not an individual component of the treatment package, yet was organised and too intensive to count as no treatment/standard care, and so was not a valid comparison
Funding	Academic or government funding (Support for this study was provided by National Institute on Aging Grants AG14131 and 5P60 AG10484 and General Clinical Research Center Grant M01-RR00211)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: SF-36 composite physical health at 18 months; Group 1: mean 40.31 (SD 7.09); n=68, Group 2: mean 37.61 (SD 7.06); n=69; SF-36 composite physical health 0-100 Top=High is good outcome; Comments: Reported final values (adjusted means) and standard errors. Reported treatment package: 40.31 (0.86). Reported exercise: 37.61 (0.85). Baseline treatment package: 35.39 (1.28). Baseline exercise: 34.50 (1.14).

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, BMI, gender, obesity, high blood pressure, cardiovascular disease, diabetes, race, income and education; Group 1 Number missing: 16, Reason: Unclear reason. 76% stayed in the trial.; Group 2 Number missing: 12, Reason: Unclear reason. 82% stayed in the trial. - Actual outcome: SF-36 composite mental health at 18 months; Group 1: mean 53.84 (SD 6.76); n=68, Group 2: mean 54.06 (SD 6.73); n=69; SF-36 composite mental health 0-100 Top=High is good outcome; Comments: Reported final values (adjusted means) and standard errors. Reported treatment package: 53.84 (0.82). Reported exercise: 54.06 (0.81). Baseline treatment package: 52.85 (1.31). Baseline exercise: 54.28 (1.00). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, BMI, gender, obesity, high blood pressure, cardiovascular disease, diabetes, race, income and education; Group 1 Number missing: 16, Reason: Unclear reason. 76% stayed in the trial.; Group 2 Number missing: 12, Reason: Unclear reason. 82% stayed in the trial. RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION Protocol outcome 1: Quality of life at >3 months - Actual outcome: SF-36 composite physical health at 18 months; Group 1: mean 40.31 (SD 7.09); n=68, Group 2: mean 38.15 (SD 6.92); n=73; SF-36 composite physical health 0-100 Top=High is good outcome; Comments: Reported final values (adjusted means) and standard errors. Reported treatment package: 40.31 (0.86). Reported diet: 38.15 (0.81). Baseline treatment package: 35.39 (1.28). Baseline diet: 35.17 (1.05). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, BMI, gender, obesity, high blood pressure, cardiovascular disease, diabetes, race, income and education; Group 1 Number missing: 16, Reason: Unclear reason. 76% stayed in the trial.; Group 2 Number missing: 15, Reason: Unclear reason. 80% stayed in the trial. - Actual outcome: SF-36 composite mental health at 18 months; Group 1: mean 53.84 (SD 6.76); n=68, Group 2: mean 54.39 (SD 6.66); n=73; SF-36 composite mental health 0-100 Top=High is good outcome; Comments: Reported final values (adjusted means) and standard errors. Reported treatment package: 53.84 (0.82). Reported diet: 54.39 (0.78). Baseline treatment package: 52.85 (1.31). Baseline diet: 52.69 (1.04). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, BMI, gender, obesity, high blood pressure, cardiovascular disease, diabetes, race, income and education; Group 1 Number missing: 16, Reason: Unclear reason. 76% stayed in the trial.; Group 2 Number missing: 15, Reason: Unclear reason. 80% stayed in the trial.	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study	Alasfour 2020[7]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=40)
Countries and setting	Conducted in Saudi Arabia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Diagnosed by the physician with unilateral or bilateral chronic knee osteoarthritis (diagnosis at least 6 months) with mild to moderate pain intensity (score no more than 7 on the Arabic Numeric Pain Rating Scale)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Saudi women aged at least 50 years; diagnosed by the physician with unilateral or bilateral chronic knee osteoarthritis (diagnosis at least 6 months) with mild to moderate pain intensity (score no more than 7 on the Arabic Numeric Pain Rating Scale) and who were able to ambulate independently. The participants had to be literate and familiar with using a smartphone or tablet.
Exclusion criteria	People with comorbidities that affected their health and wellness (e.g. neurological conditions, unstable cardiopulmonary conditions, mental disorders with a score <24 on the Mini Mental State Examination); those who were waiting for surgical interventions; those who had a recent history of trauma (within less than 3 months) (e.g., fall or accident); those who have engaged in lower-limb strengthening exercises within the previous 6 months.
Recruitment/selection of patients	Recruited from various physical therapy clinics in Riyadh City.
Age, gender and ethnicity	Age - Mean (SD): 54.4 (4.4). Gender (M:F): 0:40. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated/unclear Duration of symptoms: Not stated/unclear
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Treatment package - Exercise and behaviour change intervention. The "My Dear Knee" application. An app designed for Android and iPhone Operative System devices. Provides a guide for exercise performance in the Arabic language. Includes alerts and a monitoring system controlled by the physical therapist. The app provides automatic recording of exercise adherence, including the time and completed sessions.. Duration 6 weeks. Concurrent medication/care: All

	participants from both groups had the same exercise program. This was a simple strengthening exercise program for lower-limb muscles (mainly for knee extensor and hip abductor muscles). It was modified from previous programs proven to significantly reduce knee pain and improve function. This program included: 1) isometric quadriceps contraction; 2) isotonic quadriceps contraction; 3) isotonic hamstring contraction; 4) isotonic quadriceps contraction with resistance band; 5) straight leg raising, 6) side-lying hip abduction; 7) partial squats; 8) dynamic stepping exercise; 9) sidestepping with a resistance band around the thighs or ankles. All participants from both groups stated with two exercises for the first week. After that two new exercises were added on a weekly basis till week 4, at week 5 only one last exercises was added. The exercise program consisted of one set with 10 repetitions per exercise and a 10 second rest between each exercise.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: ≤ 6 weeks (6 weeks). (n=20) Intervention 2: Non-combined active treatment - Exercise. Exercise program only.. Duration 6 weeks. Concurrent medication/care: All participants from both groups had the same exercise program. This was a simple strengthening exercise program for lower-limb muscles (mainly for knee extensor and hip abductor muscles). It was modified from previous programs proven to significantly reduce knee pain and improve function. This program included: 1) isometric quadriceps contraction; 2) isotonic quadriceps contraction; 3) isotonic hamstring contraction; 4) isotonic quadriceps contraction with resistance band; 5) straight leg raising, 6) side-lying hip abduction; 7) partial squats; 8) dynamic stepping exercise; 9) sidestepping with a resistance band around the thighs or ankles. All participants from both groups stated with two exercises for the first week. After that two new exercises were added on a weekly basis till week 4, at week 5 only one last exercises was added. The exercise program consisted of one set with 10 repetitions per exercise and a 10 second rest between each exercise.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Academic or government funding (The authors would like to thank Deanship of scientific research for funding and supporting this research through the initiative of DSR Graduate Students Research Support (GSR).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: Arabic Numeric pain rating scale at 6 weeks; Group 1: mean 3.56 (SD 2.1); n=20, Group 2: mean 5.18 (SD 2.43); n=20; Numeric pain rating scale 0-10 Top=High is poor outcome; Comments: Baseline app: 5.78 (1.21). Baseline paper: 6.00 (0.86).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2, Reason: 1 illness (cold/flu), 1 out of city (traveled).; Group 2 Number missing: 3, Reason: 2 no response to contact, 1 illness (fall accident).

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: Arabic reduced WOMAC physical function subscale at 6 weeks; Group 1: mean 3 (SD 1.91); n=20, Group 2: mean 5.18 (SD 3.24); n=20; Reduced WOMAC phyiscal function subscale 0-28 Top=High is poor outcome; Comments: Baseline app: 8.11 (3.62). Baseline paper: 6.47 (2.93).

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2, Reason: 1 illness (cold/flu), 1 out of city (traveled).; Group 2 Number missing: 3, Reason: 2 no response to contact, 1 illness (fall accident). Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 6 weeks; Group 1: 2/20, Group 2: 3/20; Comments: App: 1 illness (cold/flu), 1 out of city (travelled). Exercise only: 2 no response to contact, 1 illness (fall accident). Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2, Reason: 1 illness (cold/flu), 1 out of city (traveled).; Group 2 Number missing: 3, Reason: 2 no response to contact, 1 illness (fall accident).	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	**Alfieri 2020[8]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=83)
Countries and setting	Conducted in Brazil; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Clinical and radiographic diagnosis of unilateral or bilateral knee osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Older than 50 years of age; presented clinical and radiographic diagnosis of unilateral or bilateral knee osteoarthritis (Evaluated by an x-ray images); Kellgren Lawrence grading scale 1 to 4; pain perception equal to or above 4cm in visual analogue scale.
Exclusion criteria	Patients with any other chronic diseases such as fibromyalgia, rheumatic arthritis, neurologic or cardiac diseases and uncontrolled hypertension; as well as the ones with total or partial prosthesis in one or both knees or hips; people who missed four or more consecutive treatment sessions and the ones who started in any other type of physical exercise during the course of the study.

Recruitment/selection of patients	People referred to exercise treatment or physical therapy by the public primary health attention.
Age, gender and ethnicity	Age - Mean (SD): 64.0 (7.8). Gender (M:F): 8:31. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren Lawrence grade 1-4 Duration of symptoms: Not stated/unclear
Indirectness of population	No indirectness
Interventions	(n=29) Intervention 1: Treatment package - Exercise and education programme. Exercise plus lifestyle group. In addition to exercise, 8 sessions of lectures and group discussions on the following topics: nutritional counselling (three 1-hour meetings with a nutritionist) discussion and follow-up on the importance of maintaining a healthy weight, and healthy eating based on the consumption of fresh and minimally processed foods instead of processed nutrition; self-management of the disease (three 1-hour meetings with a psychologist): educational interventions aiming at developing self-care strategies, discussions about the beneficial effects of relationships with family, friends and other social support providers, coexistence with pain and pain management, disease and pain coping and improvement of living conditions and social relations; and health education (two 1 hour meetings with a physical therapist and physical educator: guidance on the disease and its symptoms, performing daily activities without unnecessary physical efforts, lifestyle guidance (the importance of rest, of being physically active, healthy eating, sun exposure, breathing fresh air, drinking plenty of water, having good relationships, cultivating spirituality and avoiding harmful products such as tobacco and alcohol). Treatment sessions were two times/week during 8 weeks.. Duration 8 weeks. Concurrent medication/care: A therapeutic exercise program including warm-up, flexibility, active muscle strengthening exercises, balance and proprioception exercises. In warm-up, participants were oriented to perform brisk walking and play ball games with feet and hands. Stretching exercises targeted the following muscle groups: hip flexors, extensors and adductors, knee flexors and extensors and plantar flexors. Strengthening exercises were performed using the volunteer's own body resistance against gravity. Exercises for feet plantar flexors, dorsiflexors, knee and hip extensors and flexors, and abdominal muscles were performed. Exercises combining sensory stimulation of feet plantar surface and dynamic and static balance were also proposed. Volunteers were instructed to walk forward, backward and sideways on different surfaces, with and without visual information.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (8 weeks). (n=32) Intervention 2: Non-combined active treatment - Exercise. Exercise program only. Duration 8 weeks. Concurrent medication/care: A therapeutic exercise program including warm-up, flexibility, active muscle strengthening exercises, balance and proprioception exercises. In warm-up, participants were oriented to perform brisk walking and play ball games with feet and hands. Stretching exercises targeted the following muscle groups: hip flexors, extensors and adductors, knee flexors and extensors and plantar flexors. Strengthening exercises were performed using the volunteer's own body resistance against gravity. Exercises for feet plantar flexors, dorsiflexors, knee and hip extensors and flexors, and abdominal muscles were performed. Exercises combining sensory stimulation of feet plantar surface and dynamic and static balance were also

	proposed. Volunteers were instructed to walk forward, backward and sideways on different surfaces, with and without visual information.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (8 weeks).
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus EXERCISE Protocol outcome 1: Pain at ≤3 months - Actual outcome: WOMAC pain at 8 weeks; Group 1: mean 41.8 (SD 28); n=22, Group 2: mean 43.5 (SD 21.1); n=17; WOMAC pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 55.2 (26.1). Baseline exercise alone: 48.1 (18.6). Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, height, BMI, gender and baseline values of symptoms; Group 1 Number missing: 7, Reason: 4 schedule mismatch, 3 did not justify; Group 2 Number missing: 15, Reason: 5 schedule mismatch, 5 did not justify, 1 surgical intervention not related to osteoarthritis, 1 went on a trip, 3 started hydrotherapy Protocol outcome 2: Physical function at ≤3 months - Actual outcome: WOMAC functionality at 8 weeks; Group 1: mean 38.4 (SD 30.9); n=22, Group 2: mean 35.2 (SD 18.6); n=17; WOMAC function 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 42.1 (28.0). Baseline exercise alone: 38.7 (19.2). Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, height, BMI, gender and baseline values of symptoms; Group 1 Number missing: 7, Reason: 4 schedule mismatch, 3 did not justify; Group 2 Number missing: 15, Reason: 5 schedule mismatch, 5 did not justify, 1 surgical intervention not related to osteoarthritis, 1 went on a trip, 3 started hydrotherapy Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 8 weeks; Group 1: 7/29, Group 2: 15/32; Comments: Treatment package: 4 schedule mismatch, 3 did not justify. Exercise alone: 5 schedule mismatch, 5 did not justify, 1 surgical intervention not related to osteoarthritis, 1 went on a trip, 3 started hydrotherapy. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, weight, height, BMI, gender and baseline values of symptoms; Group 1 Number missing: 7, Reason: 4 schedule mismatch, 3 did not justify; Group 2 Number missing: 15, Reason: 5 schedule mismatch, 5 did not justify, 1 surgical intervention not related to osteoarthritis, 1 went on a trip, 3 started hydrotherapy	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study (subsidiary papers)	Allen 2021[18] (Kaufman 2021[148])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=345)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 9 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Physician diagnosis of knee osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Veteran enrolled at the Durham VA Medical Center (VAMC, physician diagnosis of knee osteoarthritis, current knee joint symptoms
Exclusion criteria	Currently meeting physical activity guidelines, currently completing Physical Therapy (PT) visits for knee OA Gout (in knee) Rheumatoid arthritis, fibromyalgia, or other systemic rheumatic disease Dementia Psychosis Active substance abuse disorder Meniscus or anterior cruciate ligament (ACL) tear in the past 6 months Total joint replacement, other major lower extremity surgery in the past 6 months or planned in the next 9 months Severe hearing impairment Serious/terminal illness Other health problem that would prohibit participation in the study and/or warrant immediate PT Current participation in another OA intervention study Unstable angina History of ventricular tachycardia Unstable chronic obstructive pulmonary disease (two hospitalizations within the previous 12 months and/or on oxygen) Uncontrolled hypertension (diastolic blood pressure >110 mm/Hg or systolic > 200mm/Hg) Stroke with moderate to severe aphasia

Recruitment/selection of patients	Conducted at 2 veterans affairs sites: Durham and Greenville, North Carolina.
Age, gender and ethnicity	Age - Mean (SD): 60.0 (10.3). Gender (M:F): 292M/53F. Ethnicity: Person of colour = 229, Hispanic ethnicity = 8
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated/unclear Duration of symptoms (mean [SD]): 16.4 (11.2) years
Indirectness of population	No indirectness
Interventions	(n=230) Intervention 1: Treatment package - Exercise and behaviour change intervention. STEP-KOA programme. Began with access to an internet-based exercise program for knee osteoarthritis (step 1). After 3 months, people not meeting OMERACT-OARSI response criteria progressed to step 2: biweekly telephone coaching to address barriers to physical activity. After 3 months of step 2, participants still not meeting response criteria went on to step 3: in-person physiotherapy visits. Some people initially met response criteria at 3 months but had regression by 6 months and no longer met response criteria compared to baseline; these participants were advanced to step 2 at 6 months. Those who missed their 3 or 6 month assessment, remained in their assigned step at that time point. The internet-based training programme provided personalised exercise recommendations, including progression of activities, with 7 exercise levels. Each level included stretching and strengthening exercises, along with aerobic exercise recommendations. Static pictures and videos of assigned stretching and strengthening exercises were provided. People were instructed to complete the exercises at least 3 times per week. At any time, they could ask to move to a harder or an easier exercise level but could move to a harder level only if their score on the modified WOMAC was better than or equal to their previous score. People were given ankle weights and elastic resistance bands, and those without internet access were given an iPad and data plan during the intervention period. The six biweekly telephone-based physical activity coaching sessions were delivered to address osteoarthritis-related and other barriers to exercise, provide social support for physical activity, reinforce the benefits of physical activity and use motivational interviewing strategies to address any ambivalence about physical activity. During each call, the coach led participants in goal setting for their weekly physical activity by using SMART principles. They were encouraged to perform strengthening exercises 2 to 3 times per week and to aim for a long-term goal of 150 minutes of physical activity per week (based on guidelines), but goals were tailored to participants' functional abilities. PHysiotherapy visits were based on usual care for knee osteoarthritis and included a personalised exercise program; instruction in activity pacing and joint protection; and evaluation of mobility, stability, function, knee alignment, limb length inequalities, muscle weakness, inflexibility and need for mobility aids, knee braces and shoe orthotics. Veterans Affairs physical therapists delivered the intervention. The first session lasted 1 hour, and the remaining visits 30 minutes.. Duration 9 months. Concurrent medication/care: No additional information.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (9 months). (n=115) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). People received educational materials via mail every 2 weeks for 9 months. The intervention included a comprehensive set of topics related to osteoarthritis and its management, described previously and based on established treatment guidelines.. Duration 9 months.

	Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (9 months).
Funding	Academic or government funding (Funding from the Department of Veterans Affairs, Health Services Research and Development Service)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: EQ-5D-5L at 3 months; Group 1: mean 0.72 (SD 0.17); n=162, Group 2: mean 0.7 (SD 0.17); n=96; EQ-5D-5L -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.67 (0.15). Baseline usual care: 0.68 (0.15).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 68, Reason: 162 followed up; Group 2 Number missing: 19, Reason: 96 followed up

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: EQ-5D-5L at 9 months; Group 1: mean 0.06 (SD 0.17); n=163, Group 2: mean 0.02 (SD 0.15); n=90; EQ-5D-5L -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.67 (0.15). Baseline usual care: 0.68 (0.15).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 67, Reason: 163 followed up; Group 2 Number missing: 25, Reason: 90 followed up

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: WOMAC pain at 3 months; Group 1: mean -1 (SD 3.9); n=230, Group 2: mean -0.1 (SD 3.3); n=115; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported mean change scores and 95% confidence intervals. Treatment package = -1.0 (-1.5 to -0.5). Standard care = -0.1 (-0.7 to 0.5). Reported baseline means, but not SD. Baseline treatment package: 9.9. Baseline standard care: 9.9.
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 66, Reason: 164 followed up; Group 2 Number missing: 15, Reason: 100 followed up

Protocol outcome 4: Pain at >3 months
- Actual outcome: WOMAC pain at 9 months; Group 1: mean -1 (SD 3.9); n=230, Group 2: mean 0.4 (SD 3.8); n=115; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported mean change scores and 95% confidence intervals. Treatment package = -1.0 (-1.5 to -0.5). Standard care = 0.4 (-0.3 to 1.1). Reported baseline means, but not SD. Baseline treatment package: 9.9. Baseline standard care: 9.9.
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 67, Reason: 163 followed up; Group 2 Number missing: 20, Reason: 95 followed up

	Protocol outcome 5: Physical function at ≤3 months - Actual outcome: WOMAC function at 3 months; Group 1: mean -3.2 (SD 11.6); n=230, Group 2: mean 0.4 (SD 10.7); n=115; WOMAC function 0-68 Top=High is poor outcome; Comments: Reported mean change scores and 95% confidence intervals. Treatment package = -3.2 (-4.7 to -1.7). Standard care = 0.4 (-1.5 to 2.4). Reported baseline means, but not SD. Baseline treatment package: 33.3. Baseline standard care: 33.3. Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 66, Reason: 164 followed up; Group 2 Number missing: 15, Reason: 100 followed up Protocol outcome 6: Physical function at >3 months - Actual outcome: WOMAC function at 9 months; Group 1: mean -3.7 (SD 13.2); n=230, Group 2: mean 1 (SD 12.3); n=115; WOMAC function 0-68 Top=High is poor outcome; Comments: Reported mean change scores and 95% confidence intervals. Treatment package = -3.7 (-5.4 to -2.0). Standard care = 1.0 (-1.3 to 3.2). Reported baseline means, but not SD. Baseline treatment package: 33.3. Baseline standard care: 33.3. Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 67, Reason: 163 followed up; Group 2 Number missing: 20, Reason: 95 followed up Protocol outcome 7: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 3 months; Group 1: 66/230, Group 2: 15/115; Comments: Treatment package: 2 excluded, 19 withdrew, 45 unable to contact. Standard care: 0 excluded, 2 withdrew, 13 unable to contact. Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 66, Reason: 164 followed up; Group 2 Number missing: 15, Reason: 100 followed up Protocol outcome 8: Discontinuation at >3 months - Actual outcome: Discontinuation at 9 months; Group 1: 67/230, Group 2: 20/115; Comments: Treatment package: Reasons unclear (excluded, withdrew or unable to contact) but stated that 163 people were followed up. Standard care: Reasons unclear (excluded, withdrew or unable to contact) but stated that 95 people were followed up. Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 67, Reason: 163 followed up; Group 2 Number missing: 20, Reason: 95 followed up
Protocol outcomes not reported by the study	Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Arnold 2010[24]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=83)
Countries and setting	Conducted in Canada; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 11 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People with hip pain for at least 6 months who were diagnosed with hip osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age 65 years or older; presence of hip pain 6 months or longer; diagnosed with hip osteoarthritis and presenting with 1 fall risk factor, a timed up-and-go test score of 10s or more; a history of at least one fall in the past 12 months
Exclusion criteria	Joint surgery within the last 6 months; current participation in a group exercise program incorporating balance training or aquatics twice a week or more; the presence of any medical or neurological condition that significantly affected independence in mobility
Recruitment/selection of patients	No additional information
Age, gender and ethnicity	Age - Mean (SD): 74.4 (6.3). Gender (M:F): 23:56. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: High morbidity score (Number of comorbidities (mean [SD]): 2.1 (1.3)). 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=28) Intervention 1: Treatment package - Exercise and education programme. Aquatic exercise sessions lasting 45 minutes (delivered twice a week for 11 weeks). The goals were to improve mobility, strength and balance. The exercise protocol consisted of warm-up exercises (variations of walking in the water and stretching the upper and lower body); lower and upper extremity strengthening exercises (using floats, noodles, sponges, and paddles for added resistance); trunk-control exercises (abdominal strengthening in floating positions, trunk control in standing positions); posture practice and balance exercises (mobility games, variation in walking and

	standing balance activities), and cooldown (gentle stretch and breathing). Combined with an education program with 30 minute sessions delivered once a week for 11 weeks. 4 sessions were in a multipurpose room with mats, mirrors and space to walk, while the other 7 were in a common meeting space with tables and chairs. The education sessions were conducted by a physical therapist with 20 years of experience working with older adults. The goals were to increase the transfer of exercises learned in the pool to the ability to successfully perform activities of daily living, increase knowledge of individual fall risk factors and fall-prevention strategies, and improve confidence in the ability to avoid a fall and recover from a fall at home and in the community. People in this group also received a booklet with information for each session and had the opportunity to set individual goals regarding exercise and fall-prevention strategies. In 4 sessions, people practiced functional tasks such as sit-to-stand, walking, dual-task walking, and getting up and down from the floor. The purpose of this practice was to reinforce the transition of exercises in water to improving functional tasks on land and also to increase confidence related to fall risk. This additional practice added approximately 1.5 hours of "physical" practice of balance-related activities to this group's experience. The rest of the educational content focused on knowledge building, group discussion, sharing goals and solutions, and positive reinforcement from the group leader. This cognitive-behavioural approach was designed to help persuade individuals to change behaviors and adopt positive fall-prevention strategies, to motivate them to participate in exercise, and to increase their understanding that physiological changes associated with exercise such as fatigue or muscle soreness are not signs of failure or dysfunction.. Duration 11 weeks. Concurrent medication/care: People were allowed to start new therapies if necessary.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (11 weeks). (n=27) Intervention 2: Non-combined active treatment - Exercise. Exercise component only. Duration 11 weeks. Concurrent medication/care: People were allowed to start new therapies if necessary.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (11 weeks). (n=27) Intervention 3: Standard care (non-organised) or no treatment - No treatment. People were instructed to not begin an exercise program during the control period and were told they would be offered a treatment after 11 weeks. Duration 11 weeks. Concurrent medication/care: People were allowed to start new therapies if necessary..

	Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (11 weeks).
Funding	Academic or government funding (The Saskatchewan-Canadian Institutes of Health Research Regional Partnerships Program (Sask-CIHR RPP) provided a 2-year fellowship grant for the primary author, and the Physiotherapy Foundation of Canada provided operational funding)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus EXERCISE Protocol outcome 1: Discontinuation at ≤3 months - Actual outcome: Dropped out at 11 weeks; Group 1: 5/28, Group 2: 8/27; Comments: Treatment package: 1 mobility, 1 medical, 1 personal, 1 transportation, 1 surgery. Exercise: 1 personal, 2 medical, 2 surgery, 2 pain, 1 allergy. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, comrobidities, prescription medications, length of hip pain, BMI, fall history, use a walking aid, previous hip joint replacement, unilateral hip involvement and baseline values of outcomes; Group 1 Number missing: 5, Reason: Mobility = 1, medical = 1, personal = 1, transportation = 1, surgery = 1; Group 2 Number missing: 8, Reason: Medical = 1, surgery = 2, pain = 2, allergy = 1, personal = 1 RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus NO TREATMENT Protocol outcome 1: Discontinuation at ≤3 months - Actual outcome: Dropped out at 11 weeks; Group 1: 5/28, Group 2: 8/27; Comments: Treatment package: 1 mobility, 1 medical, 1 personal, 1 transportation, 1 surgery. No treatment: 4 medical, 1 personal, 1 surgery, 1 transportation, 1 deceased. Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, comrobidities, prescription medications, length of hip pain, BMI, fall history, use a walking aid, previous hip joint replacement, unilateral hip involvement and baseline values of outcomes; Group 1 Number missing: 5, Reason: Mobility = 1, medical = 1, personal = 1, transportation = 1, surgery = 1; Group 2 Number missing: 6, Reason: Medical = 2, surgery = 1, deceased = 1, transportation = 1, personal = 1	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Bearne 2011[33]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=48)
Countries and setting	Conducted in United Kingdom; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 5 weeks of intervention, 6 weeks follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Chronic hip pain for more than 6 months duration who were diagnosed with a clinical diagnosis of hip osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with chronic hip pain of more than 6 months durations who were at least 50 years or older with a clinical diagnosis of hip osteoarthritis
Exclusion criteria	Had received physiotherapy for hip pain within the past 6 months; had primary pain from other joints (e.g. back, knees or ankles) which interfered with assessment; had unstable co-existing medical problems (e.g. cardiovascular, respiratory or neurological disorders); had received an intra-articular injection to the hip within 6 months of study commencement; were currently taking systemic steroids; were unable or unwilling to exercise or unable or unwilling to give informed consent
Recruitment/selection of patients	People were recruited from two general practitioner practices in the south of England over an 11-month period
Age, gender and ethnicity	Age - Mean (range): 66 (52-78). Gender (M:F): 14:34. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms (mean [range]): 5.0 (1-40) years
Indirectness of population	No indirectness
Interventions	(n=24) Intervention 1: Treatment package - Exercise and education programme. In addition to usual management by their GP, people received ten 75 minute group exercise and self-management sessions (up to eight participants per group, twice a week for five weeks), supervised by an experiences, qualified clinical physiotherapist (band 6) in a physiotherapy outpatient department. Each session comprised of two parts: supervised exercise (for 45 minutes, people completed an exercise circuit consisting of: strengthening and stretching exercises for the hip abductors, flexors and

	gluteal musculature; cycling on a static exercise bike; therapeutic resistance bands to increase hip muscle strength and dynamic control (maintaining joint stability and motor control during movement); functional and balance/coordination exercises. As the quantity and quality of these exercises improved, they were progressed and more challenging exercises were introduced. The physiotherapist prescribed exercises for each participant according to their abilities, and monitored and revised the performance of these exercises) and education, coping and self-management (at the end of each exercise session, people took part in a 30-minute 'interactive discussion' emphasizing simple coping strategies, self-care, pain control, joint protection and problem-solving to enable lifestyle changes to promote joint health and self-management. The sessions emphasized the importance of attaining and maintaining correct bodyweight and incorporating regular exercise and physical activity into the daily routine. All interactive discussions were facilitated by the physiotherapist who supervised the exercise classes. A handbook containing information that reinforced the discussion topics and exercises completed in the sessions was provided.. Duration 5 weeks. Concurrent medication/care: All people were allowed to continue routine management prescribed by their GPs, including referral to secondary care. Medication for co-existent conditions continued as needed.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (5 weeks). (n=24) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care only. Duration 5 weeks. Concurrent medication/care: All people were allowed to continue routine management prescribed by their GPs, including referral to secondary care. Medication for co-existent conditions continued as needed.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (5 weeks).
Funding	Academic or government funding (The project was funded by the Physiotherapy Research Foundation, administered by the Chartered Society of Physiotherapy, MH. and N.W. are funded by the Arthritis Research UK)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain at 6 weeks; Group 1: mean 3.7 (SD 2); n=24, Group 2: mean 4.7 (SD 3.2); n=24; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.0 (2.65). Baseline usual care: 5.2 (4.22).

Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 2, Reason: 2 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments; Group 2 Number missing: 6, Reason: 6 withdrew: 1 moved, 5 lost to follow up

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 6 months; Group 1: mean 4.4 (SD 3.1); n=24, Group 2: mean 3.8 (SD 3.4); n=24; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.0 (2.65). Baseline usual care: 5.2 (4.22).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 5, Reason: 5 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments, 1 surgery, 2 lost to follow up; Group 2 Number missing: 7, Reason: 7 withdrew: 1 moved, 5 lost to follow up, 1 other committments

Protocol outcome 3: Physical function at ≤3 months
- Actual outcome: WOMAC function at 6 weeks; Group 1: mean 11.1 (SD 7.9); n=24, Group 2: mean 13.8 (SD 10.6); n=24; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 14.3 (9.0). Baseline usual care: 17.3 (12.5).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 2, Reason: 2 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments; Group 2 Number missing: 6, Reason: 6 withdrew: 1 moved, 5 lost to follow up

Protocol outcome 4: Physical function at >3 months
- Actual outcome: WOMAC function at 6 months; Group 1: mean 13.5 (SD 10.1); n=24, Group 2: mean 13.5 (SD 12.1); n=24; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 14.3 (9.0). Baseline usual care: 17.3 (12.5).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 5, Reason: 5 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments, 1 surgery, 2 lost to follow up; Group 2 Number missing: 7, Reason: 7 withdrew: 1 moved, 5 lost to follow up, 1 other commitments

Protocol outcome 5: Psychological distress at ≤3 months
- Actual outcome: HADS anxiety at 6 weeks; Group 1: mean 4.6 (SD 2.6); n=24, Group 2: mean 4.1 (SD 3); n=24; HADS anxiety 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 5.0 (2.6). Baseline usual care: 4.1 (2.6).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 2, Reason: 2 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments; Group 2 Number missing: 6, Reason: 6 withdrew: 1 moved, 5 lost to follow up
- Actual outcome: HADS depression at 6 weeks; Group 1: mean 2.4 (SD 1.8); n=24, Group 2: mean 2.9 (SD 2.1); n=24; HADS depression 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 3.04 (2.3). Baseline usual care: 2.88 (2.8).

Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 2, Reason: 2 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments; Group 2 Number missing: 6, Reason: 6 withdrew: 1 moved, 5 lost to follow up Protocol outcome 6: Psychological distress at >3 months - Actual outcome: HADS anxiety at 6 months; Group 1: mean 4 (SD 3); n=24, Group 2: mean 4.5 (SD 3); n=24; HADS anxiety 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 5.0 (2.6). Baseline usual care: 4.1 (2.6). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 5, Reason: 5 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments, 1 surgery, 2 lost to follow up; Group 2 Number missing: 7, Reason: 7 withdrew: 1 moved, 5 lost to follow up, 1 other committments - Actual outcome: HADS depression at 6 months; Group 1: mean 2.4 (SD 2.2); n=24, Group 2: mean 2.5 (SD 1.2); n=24; HADS depression 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 3.04 (2.3). Baseline usual care: 2.88 (2.8). Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 5, Reason: 5 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments, 1 surgery, 2 lost to follow up; Group 2 Number missing: 7, Reason: 7 withdrew: 1 moved, 5 lost to follow up, 1 other committments Protocol outcome 7: Discontinuation at ≤3 months - Actual outcome: Withdrawing from the study at 6 weeks; Group 1: 2/24, Group 2: 6/24; Comments: Treatment package: 1 failed to begin rehabilitation, 1 withdrew because of other commitments. Usual care: 1 moved away from the area, 5 were lost to follow up. Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 2, Reason: 2 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments; Group 2 Number missing: 6, Reason: 6 withdrew: 1 moved, 5 lost to follow up Protocol outcome 8: Discontinuation at >3 months - Actual outcome: Withdrawing from the study at 6 months; Group 1: 5/24, Group 2: 7/24; Comments: Treatment package: 1 failed to begin rehabilitation, 1 withdrew because of other commitments, 1 underwent surgery, 2 lost to follow up. Usual care: 1 moved away from the area, 5 were lost to follow up, 1 withdrew due to other commitments. Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in WOMAC function and HADS anxiety; Group 1 Number missing: 5, Reason: 5 withdrew: 1 failed to begin rehabilitation, 1 withdrew due to other commitments, 1 surgery, 2 lost to follow up; Group 2 Number missing: 7, Reason: 7 withdrew: 1 moved, 5 lost to follow up, 1 other committments	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study (subsidiary papers)	Bennell 2016[37] (Bennell 2015[47])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=222)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of treatment, 52 weeks follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee osteoarthritis fulfilling the American College of Rheumatology criteria (pain on most days in the past month and radiographic changes) with knee pain for at least 3 months
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People of ages at least 50 years; knee osteoarthritis fulfilling the American College of rheumatology criteria (pain on most days in the past month and radiographic changes); knee pain for at least 3 months; average pain during the previous week of at least 40 on a 100mm VAS; at least moderate difficulty with daily activities.
Exclusion criteria	Systemic arthritic conditions such as rheumatoid arthritis; medical condition precluding safe exercise such as uncontrolled hypertension or heart condition; self-reported history of serious mental illness such as schizophrenia, or self-reported diagnosis of current clinical depression; neurological condition such as Parkinson's disease, multiple sclerosis or stroke; knee surgery including arthroscopy within the past 6 months or total joint replacement; awaiting or planning any back or lower limb surgery within the next 12 months; current or past (within 3 months) oral or intra-articular corticosteroid use; physiotherapy, chiropractic or acupuncture treatment or exercises specifically for the knee withint he past 6 months; walking exercise for >30 minutes continuously daily; participating in a regular (more than twice a week) structured and/or supervised exercise program such as attending exercise classes in a gym or use of a personal trainer; participating in or previous participation in a formal PCST program; inability to walk unaided; inadequate written and spoken English; inability to comply with the study protocol such as inability to attend physical therapy sessions or attend assessment appointments at the University
Recruitment/selection of patients	People were recruited from multiple sites in Melbourne and Brisbane, Australia
Age, gender and ethnicity	Age - Mean (SD): 63.4 (8.1). Gender (M:F): 89:133. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee

Extra comments	Severity: Radiographic grade 2-4, median grade 3 Duration of symptoms (median [IQR]): Exercise = 6 (3-10), PCST = 5.5 (4-10), treatment package = 5.5 (2-10).
Indirectness of population	No indirectness
Interventions	(n=73) Intervention 1: Treatment package - Exercise and behaviour change intervention. Pain coping skills training and exercise intervention. Pain coping skills training involved 10 weekly sessions. The first session educated about the pain gate control theory, sessions 1-4 focused on employing behavioural pain coping strategies, including progressive muscle relaxation. Sessions 5-9 focused on cognitive pain coping strategies and taught cognitive restructuring techniques to identify maladaptive thoughts and how to replace them with helpful coping thoughts and identifying and challenging negative thoughts and replacing them with calming self-statements. These sessions utilized pleasant imagery, attention diversion, distraction and problem-solving techniques to aid in coping with pain. The final session provided a review of the entire treatment program and dealt with relapse prevention, developing a pain coping plan for the future and identification of coping strategies no longer being used. Each session lasted 45 minutes. The exercise treatment was a standardised home-based exercise program designed to strengthen the lower limb muscles. People were taught 6 exercises targeting the quadriceps, hamstrings and hip abductor muscles. Resistance was applied via the use of ankle cuffs with optional weight poles (0.5kg each), resistance elastic bands or body weight. Intensity was determined by the participant's ability to complete 10 repetitions for a given exercise and by perceived difficulty using the modified Borg rating of perceived exertion scale for resistance training. therapist monitored progression and ensured correct technique over time. Home exercises were prescribed 4 times per week, aiming for a dosage of 3 sets of 10 repetitions, during the 12 week treatment phase, reducing to 3 times/week during the 9 month follow up. Handouts with descriptions of the prescribed exercises were provided, as well as a study log book. Exercise sessions with the physical therapist lasted 25 minutes. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Pain coping skills training). 2. Length of package: > 6 weeks (12 weeks). (n=75) Intervention 2: Non-combined active treatment - Exercise. Exercise therapy only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness

	Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks). (n=74) Intervention 3: Non-combined active treatment - Behaviour change intervention. Pain coping skills training only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Pain coping skills training). 2. Length of package: > 6 weeks (12 weeks).
Funding	Principal author funded by industry (Supported by Australian Health management, National Health and Medical Research Council (631717). Dr Bennell has received grants from the National Health and Medical Research Council (Fellowship 1058440), the Australian Research Council, and Medibank Private. Dr Ahamed has received an Australian postgraduate award to conduct this study. Dr. Jull has received grants from the National Health and Medical Research Council and the Australian Research Council. Dr. Bryant has received funding from Beyond Blue and the Collier Charitable Trust. Dr Hunt has received grants from the Arthritis Society (Canada) and the Natural Sciences and Engineering Research Council of Canada. Dr Forebes has received grants from the National Health and Medical Research Council and the Department of Veterans Affairs. Mr Harris has received grants from the National Health and Medical Research Council, the Australian Research Council, and the Medibank Health Research Fund. Dr Kenardy has received grants from the National Health and Medical Research Council (1035261), the Australian Research Council, the NIH, the Patient-Centered Outcomes Research Institute, the Motor Accident Insurance Commission of Queensland, the Commonwealth of Australia-Department of Families, Housing, Community Services, and Indigenous Affairs, the Motor Accident Authority of New South Wales, and Medibank Private. Dr Nicholas has received grants from the national Health and Medical Research Council, the Australian Research Council, the Australian Health Ministers Advisory Council, the Motor Accidents Authority of New South Wales, Beyond Blue, Self-Insurance Corporation of New South Wales, Beyond Blue, Self-Insurance Corporation of New South Wales, the New South Wales Ministry of Health, and EML Insurance. Dr. Keefe has received grant funding from the NIH and the American Cancer Society. Dr Bennell has received honoraria and/or consultation fees from Physitrack and ASICS Oceania (less than $10000 each). Dr Jull has received honoraria from Elsevier journal editorship (more than $10000). Dr Keefe has received travel support and honoraria from the North American Spine Society.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AQOL II at 12 weeks; Group 1: mean 0.1 (SD 0.1); n=73, Group 2: mean 0.1 (SD 0.2); n=75; AQOL II -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.74 (0.12). Baseline exercise: 0.71 (0.14).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: AQOL II at 52 weeks; Group 1: mean 0.1 (SD 0.1); n=74, Group 2: mean 0.1 (SD 0.1); n=75; AQOL II -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.74 (0.12). Baseline exercise: 0.71 (0.14).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: WOMAC pain at 12 weeks; Group 1: mean -4.4 (SD 3); n=73, Group 2: mean -3.3 (SD 3.1); n=75; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 9.0 (2.8). Baseline exercise: 8.6 (2.7).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain

Protocol outcome 4: Pain at >3 months
- Actual outcome: WOMAC pain at 52 weeks; Group 1: mean -3.8 (SD 3.4); n=73, Group 2: mean -3.2 (SD 3.7); n=75; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 9.0 (2.8). Baseline exercise: 8.6 (2.7).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: WOMAC function at 12 weeks; Group 1: mean -19.9 (SD 9.1); n=73, Group 2: mean -15.1 (SD 10.9); n=75; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 35.6 (7.3). Baseline exercise: 34.3 (7.2).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain

Protocol outcome 6: Physical function at >3 months
- Actual outcome: WOMAC function at 52 weeks; Group 1: mean -19.1 (SD 10.1); n=73, Group 2: mean -15.9 (SD 12.5); n=75; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 35.6 (7.3). Baseline exercise: 34.3 (7.2).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).

Protocol outcome 7: Psychological distress at ≤3 months
- Actual outcome: DASS21 Depression at 12 weeks; Group 1: mean -0.9 (SD 4.8); n=73, Group 2: mean -0.7 (SD 5.8); n=75; DASS21 Depression 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.4 (7.2). Baseline exercise: 5.7 (7.1).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain
- Actual outcome: DASS21 Anxiety at 12 weeks; Group 1: mean -0.9 (SD 4.1); n=73, Group 2: mean -1.1 (SD 3.9); n=75; DASS21 Anxiety 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.2 (5.2). Baseline exercise: 5.4 (6.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain
- Actual outcome: DASS21 Stress at 12 weeks; Group 1: mean -0.5 (SD 5.6); n=73, Group 2: mean -1.5 (SD 7.6); n=75; DASS21 Stress 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 8.5 (7.5). Baseline exercise: 9.4 (9.3).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain

Protocol outcome 8: Psychological distress at >3 months
- Actual outcome: DASS21 Depression at 52 weeks; Group 1: mean -1.4 (SD 6); n=73, Group 2: mean -0.5 (SD 5.8); n=75; DASS21 Depression 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.4 (7.2). Baseline exercise: 5.7 (7.1).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).
- Actual outcome: DASS21 Anxiety at 52 weeks; Group 1: mean -2 (SD 4.9); n=73, Group 2: mean -0.7 (SD 6.2); n=75; DASS21 Anxiety 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.2 (5.2). Baseline exercise: 5.4 (6.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).
- Actual outcome: DASS21 Stress at 52 weeks; Group 1: mean -2.1 (SD 6.3); n=73, Group 2: mean -0.8 (SD 8.6); n=75; DASS21 Stress 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 8.5 (7.5). Baseline exercise: 9.4 (9.3).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).

Protocol outcome 9: Discontinuation at ≤3 months
- Actual outcome: Participants lost at 12 weeks; Group 1: 5/73, Group 2: 8/75; Comments: Treatment package: 2 no longer interested, 1 other illness, 1 family illness, 1 no time. Exercise: 1 unable to contact, 4 other illness, 2 no time, 1 increased pain.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 1

unable to contact, 4 other illness, 2 no time, 1 increased pain

Protocol outcome 10: Discontinuation at >3 months
- Actual outcome: Participants lost at 52 weeks; Group 1: 13/73, Group 2: 14/75; Comments: Treatment package: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact. Exercise: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 14, Reason: 2 unable to contact, 7 other illness, 2 no time, 1 increased pain, 1 did not return mail, 1 knee pain, 1 family issue, 1 no longer interested (2 rejoined).

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AQOL II at 12 weeks; Group 1: mean 0.1 (SD 0.1); n=73, Group 2: mean 0.1 (SD 0.1); n=74; AQOL II -0.04-1 Top=High is good outcome; Comments: Baseline treatment package: 0.74 (0.12). Baseline behaviour change: 0.71 (0.16).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: AQOL II at 52 weeks; Group 1: mean 0.1 (SD 0.1); n=73, Group 2: mean 0.1 (SD 0.1); n=74; AQOL II -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.74 (0.12). Baseline behaviour change: 0.71 (0.16).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined).

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: WOMAC pain at 12 weeks; Group 1: mean -4.4 (SD 3); n=73, Group 2: mean -2.6 (SD 3.6); n=74; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 9.0 (2.8). Baseline behaviour change: 8.7 (2.8).

Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness

Protocol outcome 4: Pain at >3 months
- Actual outcome: WOMAC pain at 52 weeks; Group 1: mean -3.8 (SD 3.4); n=73, Group 2: mean -2.6 (SD 3.3); n=74; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 9.0 (2.8). Baseline behaviour change: 8.7 (2.8).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined).

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: WOMAC function at 12 weeks; Group 1: mean -19.9 (SD 9.1); n=73, Group 2: mean -11.2 (SD 10.3); n=74; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 35.6 (7.3). Baseline behaviour change: 35.0 (7.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness

Protocol outcome 6: Physical function at >3 months
- Actual outcome: WOMAC function at 52 weeks; Group 1: mean -19.1 (SD 10.1); n=73, Group 2: mean -12.3 (SD 10.7); n=74; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 35.6 (7.3). Baseline behaviour change: 35.0 (7.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined).

Protocol outcome 7: Psychological distress at ≤3 months
- Actual outcome: DASS21 Depression at 12 weeks; Group 1: mean -0.9 (SD 4.8); n=73, Group 2: mean -0.6 (SD 6.3); n=74; DASS21 Depression 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.4 (7.2). Baseline behaviour change: 6.4 (8.5).

Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness
- Actual outcome: DASS21 Anxiety at 12 weeks; Group 1: mean -0.9 (SD 4.1); n=73, Group 2: mean -1.9 (SD 4.1); n=74; DASS21 Anxiety 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.2 (5.2). Baseline behaviour change: 6.5 (6.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness
- Actual outcome: DASS21 Stress at 12 weeks; Group 1: mean -0.5 (SD 5.6); n=73, Group 2: mean -0.3 (SD 6.1); n=74; DASS21 Stress 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 8.5 (7.5). Baseline behaviour change: 9.4 (9.3).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness

Protocol outcome 8: Psychological distress at >3 months
- Actual outcome: DASS21 Depression at 52 weeks; Group 1: mean -1.4 (SD 6); n=73, Group 2: mean -0.9 (SD 4.2); n=74; DASS21 Depression 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.4 (7.2). Baseline behaviour change: 6.4 (8.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined).
- Actual outcome: DASS21 Anxiety at 52 weeks; Group 1: mean -2 (SD 4.9); n=73, Group 2: mean -2.1 (SD 4); n=74; DASS21 Anxiety 0-42 Top=High is poor outcome; Comments: Baseline treatment package: 5.2 (5.2). Baseline behaviour change: 6.5 (6.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined).
- Actual outcome: DASS21 Stress at 52 weeks; Group 1: mean -2.1 (SD 6.3); n=73, Group 2: mean -1.7 (SD 6.7); n=74; DASS21 Stress 0-42 Top=High is

poor outcome; Comments: Baseline treatment package: 8.5 (7.5). Baseline behaviour change: 9.4 (9.3). Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined). Protocol outcome 9: Discontinuation at ≤3 months - Actual outcome: Participants lost at 12 weeks; Group 1: 5/73, Group 2: 8/74; Comments: Treatment package: 2 no longer interested, 1 other illness, 1 family illness, 1 no time. Behaviour change: 4 no longer interested, 2 no time, 1 other illness, 1 family illness. Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 5, Reason: 2 no longer interested, 1 other illness, 1 family illness, 1 no time; Group 2 Number missing: 8, Reason: 4 no longer interested, 2 no time, 1 other illness, 1 family illness Protocol outcome 10: Discontinuation at >3 months - Actual outcome: Participants lost at 52 weeks; Group 1: 13/73, Group 2: 13/74; Comments: Treatment package: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact. Behaviour change: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined). Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, body mass, BMI, symptom duration, unilateral symptoms, level of education, employment status, comorbidities, radiographic disease severity, medication use and baseline values of outcomes; Group 1 Number missing: 13, Reason: 2 no longer interested, 4 other illness, 1 family illness, 1 no time, 3 did not return mail, 2 unable to contact.; Group 2 Number missing: 13, Reason: 6 no longer interested, 2 no time, 2 other illness, 1 family illness, 2 did not return mail, 1 unable to contact, 1 knee replacement (2 rejoined).	
Protocol outcomes not reported by the study	Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study (subsidiary papers)	Bennell 2017[38] (Bennell 2012[39])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=168)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: Intervention delivered over 6 months, 18 months total follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: American College of Rheumatology clinical criteria for knee osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age at least 50 years, average knee pain at least 4 on an 11-point NRS, American College of Rheumatology clinical criteria for knee osteoarthritis and a classification as sedentary or insufficiently physically active according to the Active Australia Survey
Exclusion criteria	An inability to to safely participate in moderate-intensity exercise; undertaking regular lower-extremity strengthening exercises or receiving nondrug management for knee pain from a health professional more than once within the past 6 months; knee surgery or intraarticular corticosteroid injection within the past 6 months; history of joint replacement on study knee or on waiting list; systemic arthritic conditions or current or past (within 4 weeks) oral corticosteroid use; other condition affecting lower-extremity function more than knee pain; unable to use/access a telephone; and a score of at least 21 on the depression subscale of the Depression, Anxiety and Stress Scale
Recruitment/selection of patients	People from metropolitan and regional communities in Victoria, Australia were recruited between July 2012 and August 2013 via advertisements in print, on the radio and in social media and via their research volunteer database
Age, gender and ethnicity	Age - Mean (SD): 62.3 (7.5). Gender (M:F): 62:106. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: <2->10 years, median time 2-10 years
Indirectness of population	No indirectness
Interventions	(n=84) Intervention 1: Treatment package - Exercise and behaviour change intervention. Exercise program (with some education) and coaching sessions.

	Exercise included verbal and written education/information about OA, the benefits of physical activity/exercise and strategies to enhance adherence (but not formally organised) with a progressive individualized home exercise program comprising 4-6 lower extremity exercises (at least 3 knee extensor strengthening exercises and at least 1 hip abductor strengthening exercise from a predetermined list with 1-2 optional exercises based on assessment) performed 3 times per week and promoted increased general physical activity, including provision of a pedometer for optional self-monitoring/motivation and assistance with formulating short-term goals. Coaching included 6 additional sessions where the coach discussed the person's preference, confidence and success in the exercise to help reinforce desired behavioural changes. The coaching used HealthChange methodology, using features from motivational interviewing, solution-focused counseling and cognitive behavioural therapy. Duration 6 months (5 exercise sessions, 6 coaching sessions). Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (6 months (delivered slowly)). (n=84) Intervention 2: Non-combined active treatment - Exercise. Exercise only. Duration 6 months (5 exercise sessions). Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks
Funding	Principal author funded by industry (Supported by the National Health and Medical Research Council (grant 631717). Drs Bennell and Harris are supported by fellowships from the National Health and Medical Research Council (1058440 and 1079777, respectively), and by the Australian Research Council and Medibank Health Research Fund. Dr Forbes' work was supported by grants from the National Health and Medical Research Council and the Department of Veterans Affairs. Dr Kolt's work was supported by grants from the National Health and Medical Research Council, the Australian Research Council, the Health Research Council of New Zealand and the New Zealand Ministry of Health. Dr Hunter's work was supported by the National Health and Medical Research Council, the Australian Research Council and the NIH. Dr Hinman is supported by an Australian Research Council Future Fellowship (FT130100175) and by the National Health and Medical Research Council and Medibank Health Research Fund. Dr Bennell has received royalities from Physitrack and Asics Oceania. Dr Hinman has received royalties from Asics Oceania and has

	received fees from the journal Physical Therapy for her role as editorial board member.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: AQOL II at 18 months; Group 1: mean 0 (SD 0.2); n=66, Group 2: mean 0 (SD 0.2); n=62; AQOL II -0.11-1 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: 0.00 (0.1, 0.00). Reported exercise: 0.00 (0.1, 0.00). Baseline treatment package: 0.7 (0.1). Baseline exercise: 0.7 (0.1).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, symptom duration, height, body mass, BmI, gender, arthritis self-efficacy, self-efficacy for physical activity, level of education, employment status, current medication/supplement use, medication/supplement type and baseline values of outcomes; Group 1 Number missing: 18, Reason: Treatment package: 7 withdrew, 7 unable to contact, 2 other illness, 1 increased knee pain, 1 moved interstate, 1 lack of time (1 rejoined).; Group 2 Number missing: 22, Reason: Exercise: 13 unable to contact, 5 withdrew, 3 family illness, 1 other illness, 1 passed away, 1 moved overseas (2 rejoined)

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 18 months; Group 1: mean -3.5 (SD 3.9); n=66, Group 2: mean -3.7 (SD 5.4); n=62; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -3.5 (-2.6, -4.5). Reported exercise: -3.7 (-2.3, -5.0). Baseline treatment package: 8.1 (2.7). Baseline exercise: 8.5 (2.9).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, symptom duration, height, body mass, BmI, gender, arthritis self-efficacy, self-efficacy for physical activity, level of education, employment status, current medication/supplement use, medication/supplement type and baseline values of outcomes; Group 1 Number missing: 18, Reason: Treatment package: 7 withdrew, 7 unable to contact, 2 other illness, 1 increased knee pain, 1 moved interstate, 1 lack of time (1 rejoined).; Group 2 Number missing: 22, Reason: Exercise: 13 unable to contact, 5 withdrew, 3 family illness, 1 other illness, 1 passed away, 1 moved overseas (2 rejoined)

Protocol outcome 3: Physical function at >3 months
- Actual outcome: WOMAC function at 18 months; Group 1: mean -14.5 (SD 12.9); n=66, Group 2: mean -12.6 (SD 15.1); n=62; WOMAC function 0-68 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -14.6 (-11.5, -17.7). Reported exercise: -12.6 (-8.8, -16.3). Baseline treatment package: 27.3 (11.1). Baseline exercise: 30.3 (10.1).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, symptom duration, height, body mass, BmI, gender, arthritis self-efficacy, self-efficacy for physical activity, level of education, employment status, current medication/supplement use, medication/supplement type and baseline values of outcomes; Group 1 Number missing: 18, Reason: Treatment package: 7 withdrew, 7 unable to contact, 2 other illness, 1 increased knee pain, 1 moved interstate, 1 lack of time (1 rejoined).; Group 2 Number missing: 22, Reason: Exercise: 13 unable to contact, 5 withdrew, 3 family illness, 1 other illness, 1 passed away, 1 moved overseas (2 rejoined)

Protocol outcome 4: Discontinuation at >3 months - Actual outcome: Lost to month 18 assessment at 18 months; Group 1: 18/84, Group 2: 22/84; Comments: Treatment package: 7 withdrew, 7 unable to contact, 2 other illness, 1 increased knee pain, 1 moved interstate, 1 lack of time (1 rejoined). Exercise: 13 unable to contact, 5 withdrew, 3 family illness, 1 other illness, 1 passed away, 1 moved overseas (2 rejoined) Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, symptom duration, height, body mass, BmI, gender, arthritis self-efficacy, self-efficacy for physical activity, level of education, employment status, current medication/supplement use, medication/supplement type and baseline values of outcomes; Group 1 Number missing: 18, Reason: Treatment package: 7 withdrew, 7 unable to contact, 2 other illness, 1 increased knee pain, 1 moved interstate, 1 lack of time (1 rejoined).; Group 2 Number missing: 22, Reason: Exercise: 13 unable to contact, 5 withdrew, 3 family illness, 1 other illness, 1 passed away, 1 moved overseas (2 rejoined)	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study (subsidiary papers)	**Bennell 2020[36] (Bennell 2020[43])**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=110)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee pain on most days of the last month with knee pain for at least 3 months, average overall pain severity of at least 4 on an 11-point numeric rating scale and tibiofemoral osteophytes on x-ray
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age at least 50 years; knee pain on most days of the past month; knee pain for at least 3 months; average overall pain severity at least 4 on an 11-point numeric rating scale; tibiofemoral osteophytes on x-ray; obesity (BMI at least 30kg/m²); own a mobile phone with text messaging
Exclusion criteria	Lateral more than medial joint space narrowing on x-ray; knee surgery/joint injection in the past 6 months or planned surgery in the next 9 months; current or past (4 weeks) oral corticosteroid use; systemic arthritic conditions; past knee fracture or malignancy; past hip/knee joint replacement/tibial osteotomy; other condition affecting lower limb function; participation in knee

	strengthening or neuromuscular/functional exercise in the past 6 months or planning to start exercise in the next 9 months; unable to walk unaided; unable to commit to study requirements.
Recruitment/selection of patients	People who completed the TARGET trial were recruited (a trial where people visited a physiotherapist five times over 12 weeks for prescription of either a weight-bearing functional exercise program or a non-weight-bearing quadriceps strengthening exercise program). Target trial participants were recruited from the community in Melbourne, Australia between September 2017 and May 2019 via advertisements through consumer organisations, social media, community locations, media and our volunteer database.
Age, gender and ethnicity	Age - Mean (SD): 62.3 (6.8). Gender (M:F): 36:74. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee (People also had problems in other joints, including hand, neck, back, hip, foot and shoulder, but all had knee osteoarthritis).
Extra comments	Severity: Kellgren Lawrence grade 2-4, median grade 3 Duration of symptoms (mean [SD]): 8.2 (7.5) years
Indirectness of population	No indirectness
Interventions	(n=56) Intervention 1: Treatment package - Exercise and behaviour change intervention. An SMS intervention. People received a 24 week automated, semi-interactive SMS intervention delivered via mobile phone to support adherence to the home exercise program. The development of the SMS intervention was based on the Behaviour Change Wheel framework. Behaviour change techniques linked to each barrier or facilitator were used to construct the content of the SMS messages. People received up to five text messages weekly, with message frequency reducing over 24 weeks. Each week to fortnight people received a message asking them to self-report the number of home exercise sessions completed in the past week. People who completed no more than 2 sessions then received a message prompting them to select their main reason for not performing exercise sessions as prescribed from a predetermined list (forgot, too tired, knee hurts so cannot exercise, worried exercise is causing pain, exercise is not helping, boring, lack of time, life stress and none of the above apply to me). Barrier selection then triggered a message providing a suggestion tailored to address the selected barrier. Those who chose the barrier option of non of the above apply to me received a message encouraging them to continue exercise, but the message was not linked to a specific behaviour change technique. People who reported being adherent to exercise received a positive reinforcement message. Program automation ensured that different messages were received each time. All people, irrespective of adherence, also received regular motivational messages (twice weekly initially then once fortnightly by 24 weeks) containing suggestions linked to exercise facilitators. To enhance engagement, participants received special occasion messages (e.g. birthday). Message lengths ranged from 105 to 420 characters, with literacy demands assessed as grade 5.4, well below the maximum eight-grade reading level recommended for consumer health care information.. Duration 24 weeks. Concurrent medication/care: People continued their previously allocated home exercise program as an unsupervised program for 24 weeks but to reduce the frequency from four times per week to three times per week.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (24 weeks).

	(n=54) Intervention 2: Non-combined active treatment - Exercise. No SMS text messaging intervention.. Duration 24 weeks. Concurrent medication/care: People continued their previously allocated home exercise program as an unsupervised program for 24 weeks but to reduce the frequency from four times per week to three times per week.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (24 weeks).
Funding	Academic or government funding (This study was funded by the National Health and Medical Research Council Program Grant (1091302). KLB is supported by a National Health and Medical Research Council Investigator Fellowship (1174431). RKN is supported by an Australian Government Research Training Program Scholarship. RSH is supported by a National Health and Medical Research Council Senior Research Fellowship (1154217).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: KOOS quality of life at 24 weeks; Group 1: mean -2.2 (SD 23); n=56, Group 2: mean -2.3 (SD 16.2); n=54; KOOS quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 44.4 (19.9). Baseline exercise: 47.9 (21.7).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8, Reason: 7 did not return messages, 1 chose to withdraw; Group 2 Number missing: 3, Reason: 2 did not return messages, 1 chose to withdraw

Protocol outcome 2: Pain at >3 months
- Actual outcome: KOOS pain at 24 weeks; Group 1: mean -0.8 (SD 14.9); n=56, Group 2: mean -2.6 (SD 14.1); n=54; KOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 64.3 (14.9). Baseline exercise: 63.2 (19.8).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8, Reason: 7 did not return messages, 1 chose to withdraw; Group 2 Number missing: 3, Reason: 2 did not return messages, 1 chose to withdraw

Protocol outcome 3: Physical function at >3 months
- Actual outcome: KOOS function at 24 weeks; Group 1: mean 0 (SD 18.5); n=56, Group 2: mean -0.5 (SD 14); n=54; KOOS function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 72.2 (15.6). Baseline exercise: 70.6 (20.7).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8, Reason: 7 did not return messages, 1 chose to withdraw; Group 2 Number missing: 3, Reason: 2 did not return messages, 1 chose to withdraw

Protocol outcome 4: Discontinuation at >3 months
- Actual outcome: Discontinuation at 24 weeks; Group 1: 8/56, Group 2: 3/54; Comments: Treatment package: 7 did not return messages, 1 chose to withdraw. Control: 2 did not return messages, 1 chose to withdraw.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8, Reason: 7 did not return messages, 1 chose to withdraw; Group 2 Number missing: 3, Reason: 2 did not return messages, 1 chose to withdraw	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Brosseau 2012[56]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=222)
Countries and setting	Conducted in Canada; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 months of intervention, with an additional 6 months of follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Mild to moderate unilateral or bilateral osteoarthritis of the knee according to the American College of Rheumatology clinical and radiographic/magnetic resonance imagery criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Mild to moderate unilateral or bilateral osteoarthritis of the knee according to the American College of Rheumatology clinical and radiographic/magnetic resonance imagery criteria; reported pain for at least 3 months; expected their medication to change during the study period; demonstrated an ability to ambulated for a minimum of 20 minutes at their own pace with minimal reports of pain (at least 3 out of 10 on a visual analogue pain rating scale); were able to be treated as an outpatient; were available three times a week over a period of 12 months
Exclusion criteria	Participated in regular physical or aerobic sports at least 2 times per week for more than 20 minutes per session during the previous 6 months; severe osteoarthritis of the knee or other weight bearing joints of the lower extremity; no written consent from their physician to participate in the study; pain at rest or at night; received rehabilitation treatment, corticosteroids injection, or any other pain-related treatment besides medication for arthritis within the last 12 months; uncontrolled hypertension (systolic blood pressure >160mmHg confirmed by the screening initial VO2 max test at the Ottawa Heart institute); other illnesses, such as rheumatoid arthritis (judged by the patient or study physician to make participation in this study inadvisable); significant cognitive deficit resulting in an ability to understand or comply with instructions; surgery planned in the next year; intention to move away from Ottawa region in the next year; an inability to communicate in English or French; an unwillingness to sign informed consent
Recruitment/selection of patients	No additional information

Age, gender and ethnicity	Age - Mean (SD): 63.4 (8.6). Gender (M:F): 74:166. Ethnicity: White = 197, Black = 5, Hispanic = 8, Asian or Pacific Islander = 10, "American Indian" or Alaskan native = 1, Other = 1
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms (mean [SD]): 10.3 (9.26)
Indirectness of population	No indirectness
Interventions	(n=69) Intervention 1: Treatment package - Exercise and behaviour change intervention. Walking and behavioural intervention, including a supervised walking program delivered over a 12 month period three times a week with 45 minute aerobic walking phases achieving approximately 50 to 70% of the subjects' pre-determined maximum heart rate, a behavioural intervention using the adapted Program for Arthritis Control through Education and Exercise program using short- and long-term goal setting during classes, an educational component delivered by the instructor on the benefits of physical activity, monthly face-to-face counselling where people received moral support/encouragement and exploring potential barriers and phone counseling to achieve goal setting.. Duration 12 months. Concurrent medication/care: Everyone was given educational pamphlets and a pedometer as a measurement tool for exercise. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (12 months). (n=79) Intervention 2: Non-combined active treatment - Exercise. Walking program only. Duration 12 months. Concurrent medication/care: Everyone was given educational pamphlets and a pedometer as a measurement tool for exercise.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 months). (n=74) Intervention 3: Standard care (non-organised) or no treatment - Standard care (non-organised). Non-organised care (self-directed). Duration 12 months. Concurrent medication/care: Everyone was given educational pamphlets and a pedometer as a measurement tool for exercise.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 months).

Funding	Academic or government funding (This study was completed with the support of a research grant obtained from the Canadian Institutes of Health Research (CIHR) (Grant #MCT82367); University Research Chair (salary support for research staff) and the Ministry of Human Resources (summer student program) (Canada). This RCT won a prize for the best community-based project from the City of Gatineau (Canada) in 2009.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: SF-36 physical component at 18 months; Group 1: mean 40.909 (SD 11.038); n=42, Group 2: mean 42.82 (SD 9.24); n=44; SF-36 physical component 0-100 Top=High is good outcome; Comments: Baseline treatment package: 43.645 (8.656). Baseline exercise: 40.516 (8.598).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 35, Reason: 35 discontinued (reasons not given)
- Actual outcome: SF-36 mental component at 18 months; Group 1: mean 53.922 (SD 9.023); n=42, Group 2: mean 51.993 (SD 11); n=44; SF-36 mental component 0-100 Top=High is good outcome; Comments: Baseline treatment package: 53.812 (8.639). Baseline exercise: 52.914 (10.835).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 34, Reason: 34 discontinued (reasons not given); Group 2 Number missing: 35, Reason: 35 discontinued (reasons not given)

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 18 months; Group 1: mean 26.16 (SD 17.97); n=42, Group 2: mean 23.6 (SD 15.09); n=43; WOMAC pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 26.81 (14.92). Baseline exercise: 31.15 (14.29).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 36, Reason: 36 discontinued (reasons not given)

Protocol outcome 3: Physical function at >3 months
- Actual outcome: WOMAC function at 18 months; Group 1: mean 24.15 (SD 17.24); n=42, Group 2: mean 18.2 (SD 14.63); n=43; WOMAC function 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 27.65 (18.22). Baseline exercise: 28.16 (15.41).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1

Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 36, Reason: 36 discontinued (reasons not given)

Protocol outcome 4: Discontinuation at ≤3 months
- Actual outcome: Dropped out at 3 months; Group 1: 10/69, Group 2: 10/79; Comments: Taken from the part 1 article for this study.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 10, Reason: 10 discontinued (reasons not given); Group 2 Number missing: 10, Reason: 10 discontinued (reasons not given)

Protocol outcome 5: Discontinuation at >3 months
- Actual outcome: Dropped out at 18 months; Group 1: 27/69, Group 2: 35/79; Comments: Reasons not given
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 35, Reason: 35 discontinued (reasons not given)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: SF-36 physical component at 18 months; Group 1: mean 40.909 (SD 11.038); n=42, Group 2: mean 45.149 (SD 8.93); n=36; SF-36 physical component 0-100 Top=High is good outcome; Comments: Baseline treatment package: 43.645 (8.656). Baseline no treatment: 41.996 (9.100).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 38, Reason: 38 discontinued (reasons not given)
- Actual outcome: SF-36 mental component at 18 months; Group 1: mean 53.922 (SD 9.023); n=42, Group 2: mean 53.101 (SD 9.914); n=36; SF-36 mental component 0-100 Top=High is good outcome; Comments: Baseline treatment package: 53.812 (8.639). =Baseline no treatment: 53.556 (8.995).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 34, Reason: 34 discontinued (reasons not given); Group 2 Number missing: 38, Reason: 38 discontinued (reasons not given)

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 18 months; Group 1: mean 26.16 (SD 17.97); n=42, Group 2: mean 23.5 (SD 17.78); n=35; WOMAC pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 26.81 (14.92). Baseline no treatment: 30.30 (16.47).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration

of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 39, Reason: 39 discontinued (reasons not given) Protocol outcome 3: Physical function at >3 months - Actual outcome: WOMAC function at 18 months; Group 1: mean 24.15 (SD 17.24); n=42, Group 2: mean 19.4 (SD 17.08); n=35; WOMAC function 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 27.65 (18.22). Baseline no treatment: 26.89 (16.34). Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 39, Reason: 39 discontinued (reasons not given) Protocol outcome 4: Discontinuation at ≤3 months - Actual outcome: Dropped out at 3 months; Group 1: 10/69, Group 2: 17/74; Comments: Taken from the part 1 article. Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 10, Reason: 10 discontinued (reasons not given); Group 2 Number missing: 17, Reason: 17 discontinued (reasons not given) Protocol outcome 5: Discontinuation at >3 months - Actual outcome: Dropped out at 18 months; Group 1: 27/69, Group 2: 38/74; Comments: Reasons not given Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, affected knee, duration of symptoms, weight, BMI, walking aid use, racial background, marital status, level of education, medication use and baseline values of outcomes; Group 1 Number missing: 27, Reason: 27 discontinued (reasons not given); Group 2 Number missing: 38, Reason: 38 discontinued (reasons not given)	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study (subsidiary papers)	Crossley 2015[75] (Crossley 2008[76])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=92)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of intervention, 9 months follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Anterior or retro-patellar pain with lateral patellofemoral osteophytes on weight-bearing skyline radiographs
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged at least 40 years; have anterior or petro-patellar pain that was aggravated by two or more PFJ-loaded activities (e.g. stair ambulation, rising from sitting or squatting; have an average pain score of at least 3 on an 11-point scale during aggravating activities and on most days during the past month; and have evidence of lateral PFJ osteophytes on weight-bearing skyline radiographs.
Exclusion criteria	Pain from other lower-limb sites; predominantly tibiofemoral joint symptoms on clinical examination; current or previous (prior 12 months) physiotherapy for knee pain; recent knee injections (prior 3 months); previous or planned (following 6 months) knee surgery; physical inability to undertake testing; other medical conditions; inability to understand written and spoken English; and a body mass index greater than 34 kg/m²; additionally people with median > lateral patellofemoral osteophytes or moderate-to-severe concomitant tibiofemoral joint osteoarthritis (Kellgren and Lawrence grade >2) were excluded).
Recruitment/selection of patients	People were recruited by advertisements in print and radio median, posters in sporting clubs, health and medical practices and referrals from practitioners.
Age, gender and ethnicity	Age - Mean (SD): 54.4 (9.9). Gender (M:F): 39:53. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren Lawrence grade 0-2, median grade 0 (this study looks at people with patellofemoral osteoarthritis) Duration of symptoms: Not stated
Indirectness of population	No indirectness

Interventions	(n=44) Intervention 1: Treatment package - Combination and education programme. Exercise, education, manual therapy and taping. Eight treatments (approximately 60 minutes duration) were provided once a week for 4 weeks and then once every 2 weeks for 8 weeks for each group. The package was standardised to consist of: functional retraining exercises for the quadriceps and hip muscles; quadriceps and hip muscle strengthening; patellar taping; manual-therapy (PFJ, TFJ and soft tissue mobilisation); osteoarthritis education (eight sessions, 1a. what is arthritis?, 1b. osteoarthritis, 1c. tips for osteoarthritis of the hip or knee, 2. healthy eating and arthritis, 3. physical activity, 4. dealing with pain, 5. medicines and arthritis, 6a. complementary therapies, 6b. glucosamine and chondroitin, 6c. fish oils, 7a. arthritis and emotions, 7b. saving energy, 8. taking control of your osteoarthritis (booklet). The standard elements were tailored to each person's clinical presentation as well as the presence of co-morbidities (e.g. back and hip pain or pathology). The load was adjusted over time. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks). (n=48) Intervention 2: Non-combined active treatment - Education programme. Education sessions only delivered over the same time period. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (This trial was funded by the National Health and Medical Research Council (NHMRC, Project #508966). RSH (FT#130100175) is funded in part by Australian Research Council Future Fellowship.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: KOOS QoL at 12 weeks; Group 1: mean 54.7 (SD 20); n=39, Group 2: mean 49.8 (SD 13.8); n=42; KOOS QoL 0-100 Top=High is good outcome; Comments: Baseline treatment package: 44.3 (14.2). Baseline education only: 39.5 (15.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 5, Reason: 4 did not receive allocated intervention (but were

still potentially included in the analysis). 1 moved interstate/overseas, 2 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment; Group 2 Number missing: 6, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 3 lost contact.

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: KOOS QoL at 9 months; Group 1: mean 56 (SD 19.6); n=35, Group 2: mean 52 (SD 15.2); n=34; KOOS QoL 0-100 Top=High is good outcome; Comments: Baseline treatment package: 44.3 (14.2). Baseline education only: 39.5 (15.5).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 9, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 4 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment, 2 not interested; Group 2 Number missing: 14, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 6 lost contact, 5 not interested

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: KOOS pain at 12 weeks; Group 1: mean 76.3 (SD 13.4); n=39, Group 2: mean 69.4 (SD 14.2); n=42; KOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 64.0 (14.7). Baseline education only: 63.4 (14.3).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 5, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 2 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment; Group 2 Number missing: 6, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 3 lost contact.

Protocol outcome 4: Pain at >3 months
- Actual outcome: KOOS pain at 9 months; Group 1: mean 75.5 (SD 16.5); n=35, Group 2: mean 73.5 (SD 14.4); n=34; KOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 64.0 (14.7). Baseline education only: 63.4 (14.3).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 9, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 4 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment, 2 not interested; Group 2 Number missing: 14, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 6 lost contact, 5 not interested

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: KOOS activities of daily living at 12 weeks; Group 1: mean 83.8 (SD 12.8); n=39, Group 2: mean 76.6 (SD 14.6); n=42; KOOS activites of daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 72.2 (14.9). Baseline education only: 70.8 (16.9).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 5, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 2 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment; Group 2 Number missing: 6, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 3 lost contact.

Protocol outcome 6: Physical function at >3 months
- Actual outcome: KOOS activities of daily living at 9 months; Group 1: mean 82.1 (SD 14.8); n=35, Group 2: mean 77.7 (SD 16); n=34; KOOS activities of daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 72.2 (14.9). Baseline education only: 70.8 (16.9).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 9, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 4 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment, 2 not interested; Group 2 Number missing: 14, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 6 lost contact, 5 not interested

Protocol outcome 7: Discontinuation at ≤3 months
- Actual outcome: Lost to follow-up at 12 weeks; Group 1: 5/44, Group 2: 6/48; Comments: Treatment package: 1 moved interstate/overseas, 2 lost contact, 2 unwilling to commit/attend appointment. Education only: 1 mother ill, 2 could not attend appointment, 3 lost contact.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 5, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 2 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment; Group 2 Number missing: 6, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 3 lost contact.

Protocol outcome 8: Discontinuation at >3 months
- Actual outcome: Lost to follow-up at 9 months; Group 1: 9/44, Group 2: 14/48; Comments: Treatment package: 1 moved interstate/overseas, 4 lost contact, 2 unwilling to commit/attend appointment, 2 not interested. Education only: 1 mother ill, 2 could not attend appointment, 6 lost contact, 5 not interested.
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, mass, BMI, gender, Kellgren Lawrence grade, osteophyte severity and baseline values of outcomes; Group 1 Number missing: 9, Reason: 4 did not receive allocated intervention (but were still potentially included in the analysis). 1 moved interstate/overseas, 4 lost contact, 1 moved interstate/overseas, 2 unwilling to commit/attend appointment, 2 not interested; Group 2 Number missing: 14, Reason: 2 did not receive allocated intervention (but were still potentially included in the analysis). 1 mother ill, 2 could not attend appointment, 6 lost contact, 5 not interested

Protocol outcomes not reported by the study	Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Da silva 2015[80]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=41)
Countries and setting	Conducted in Brazil; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: A clinical diagnosis of chronic knee osteoarthritis (based on the criteria of the American College of Rheumatology)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with symptomatic clinical diagnosis of chronic knee osteoarthritis and moderate to very severe knee pain according to the Lequesne algofunctional index. People were referred from rheumatologists and were using stable doses of anti-inflammatory drugs. People had experienced pain within the last year in or around the knee occurring on most days for at least 3 months.
Exclusion criteria	Cognitive dysfunction; previous participation in a similar rehabilitation program; medical contraindication to mild to moderate physical activity; other causes of pain in the lower limb; refusal to continue the study; two consecutive or three non-consecutive absences
Recruitment/selection of patients	People were referred from rheumatologists
Age, gender and ethnicity	Age - Mean (SD): 58.5 (7.1). Gender (M:F): 4:26. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Low morbidity score (Diabetes Mellitus: 3, Hypertension: 18, Hypercholesterolemia: 2). 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Symptoms in the last year on most days for at least 3 months
Indirectness of population	No indirectness
Interventions	(n=19) Intervention 1: Treatment package - Exercise and education programme. A group rehabilitation program that consisted of 60 minute sessions performed twice a week for 8 weeks including educational aspects about knee osteoarthritis (15 minutes) followed by several physical activities (45 minutes). The educational programs included the following themes: identification of personal objectives and recognition of individual functional capabilities; weight control and constituents of a healthy diet,

	including possible benefits of omega-3; explanation of pain perceptions and biopsychosocial model of pain; nonpharmacological procedures of pain management and use of ice and heat when appropriate; home exercise and home relaxation techniques. Physical activities included the following: warm-up for 10 min with a stationary bike and stretching; exercises for the strength of the lower and upper limbs; body mobility, functional and balance exercises; relaxation. Fifty to sixth percent of the estimated maximum load was used. . Duration 8 weeks. Concurrent medication/care: Everyone had one self-management class session with a general orientation about osteoarthritis delivered in a 90 minute lecture. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (8 weeks). (n=22) Intervention 2: Standard care (non-organised) or no treatment - No treatment. No additional treatment. Duration 8 weeks. Concurrent medication/care: Everyone had one self-management class session with a general orientation about osteoarthritis delivered in a 90 minute lecture. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (8 weeks).
Funding	No funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus NO TREATMENT

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: SF-36 physical function at 8 weeks; Group 1: mean 65.33 (SD 11.57); n=15, Group 2: mean 51.33 (SD 21.25); n=15; SF-36 physical function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 39.67 (15.86). Baseline control: 47.67 (29.99).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 role physical at 8 weeks; Group 1: mean 88.33 (SD 20.85); n=15, Group 2: mean 35 (SD 39.87); n=15; SF-36 role physical 0-100 Top=High is good outcome; Comments: Baseline treatment package: 30.00 (35.61). Baseline control: 28.33 (31.15).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 bodily pain at 8 weeks; Group 1: mean 57.6 (SD 12.48); n=15, Group 2: mean 42.8 (SD 21.52); n=15; SF-36 bodily pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 44.47 (11.78). Baseline control: 41.27 (17.88).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 general health at 8 weeks; Group 1: mean 69 (SD 18.59); n=15, Group 2: mean 55.27 (SD 17.86); n=15; SF-36 general health 0-100 Top=High is good outcome; Comments: Baseline treatment package: 52.40 (24.50). Baseline control: 52.07 (20.78).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 vitality at 8 weeks; Group 1: mean 72 (SD 15.56); n=15, Group 2: mean 58.33 (SD 16.22); n=15; SF-36 vitality 0-100 Top=High is good outcome; Comments: Baseline treatment package: 56.00 (19.20). Baseline control: 60.00 (12.54).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 social function at 8 weeks; Group 1: mean 91.67 (SD 12.2); n=15, Group 2: mean 90.83 (SD 13.75); n=15; SF-36 social function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 86.67 (13.75). Baseline control: 87.50 (16.37).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 role emotional at 8 weeks; Group 1: mean 86.67 (SD 30.37); n=15, Group 2: mean 53.2 (SD 32.99); n=15; SF-36 role emotional 0-100 Top=High is good outcome; Comments: Baseline treatment package: 53.27 (43.33). Baseline control: 51.00 (39.65).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons
- Actual outcome: SF-36 mental health at 8 weeks; Group 1: mean 75.2 (SD 18.77); n=15, Group 2: mean 61.07 (SD 20.92); n=15; SF-36 mental health 0-100 Top=High is good outcome; Comments: Baseline treatment package: 71.20 (21.97). Baseline control: 57.87 (15.03).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: Lequesne index pain subscale at 8 weeks; Group 1: mean 2.6 (SD 1.55); n=15, Group 2: mean 4 (SD 1.56); n=15; Lequesne index pain subscale 0-8 Top=High is poor outcome; Comments: Baseline treatment package: 4.93 (1.33). Baseline control: 4.47 (1.46).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons Protocol outcome 3: Physical function at ≤3 months - Actual outcome: Lequesne index function subscale at 8 weeks; Group 1: mean 2.3 (SD 1.36); n=15, Group 2: mean 3.13 (SD 1.45); n=15; Lequesne index function subscale 0-8 Top=High is poor outcome; Comments: Baseline treatment package: 3.57 (1.08). Baseline control: 3.23 (1.53). Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons Protocol outcome 4: Discontinuation at ≤3 months - Actual outcome: Dropouts at 8 weeks; Group 1: 4/19, Group 2: 7/22; Comments: Treatment package: 1 absence, 3 personal reasons. No treatment: 2 health problems, 5 personal reasons. Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Different baseline values of SF-36 physical function, bodily pain, vitality, role emotional and mental health; Group 1 Number missing: 4, Reason: 1 absence, 3 personal reasons; Group 2 Number missing: 7, Reason: 2 health problems, 5 personal reasons	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	**Deveza 2021[87]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=204)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up

Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks intervention, 12 weeks overall
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Thumb base pain at least half of the days in the past month, average pain rated at 40 or greater on a 0 to 100mm visual analog scale over the 30 days and in the 48 hours prior to screening, score of 6 or higher on the Functional Index of Hand Osteoarthritis and radiographic evidence of osteoarthritis at the first metacarpal joint, read by a trained rheumatologist.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Define
Exclusion criteria	Define
Recruitment/selection of patients	Conducted at the Royal North Shore hospital, a tertiary-care academic hospital in Australia.
Age, gender and ethnicity	Age - Mean (SD): 65.6 (8.1). Gender (M:F): Define. Ethnicity: Australian = 97, British = 37, Irish = 14, Other = 56.
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Thumb
Extra comments	Severity: Kellgren Lawrence grade 2-3, median grade 3. Duration of symptoms: <1 to >5 years. Median 1-5 years.
Indirectness of population	No indirectness
Interventions	(n=102) Intervention 1: Treatment package - Combination and education programme. Education, splint, hand exercises and diclofenac sodium 1% gel. The splint was a prefabricated neoprene splint (Comfort Cool Thumb CMC Restriction Splint) that incorporated the thumb base and wrist and was recommended for use during daily activities for a minimum of 4 hours per day (removing the splint during rest, sleep, exercising and bathing). The hand exercises consisted of 5 exercises to optimize range of motion and improve neuromuscular control of thumb alignment, muscular endurance and proprioception. These were thumb opposition, paper tearing, line tracing on a ball, using chopsticks to pick up objects and squeezing a ball. Participants were instructed to perform the exercises at home 3 times per week. The program was adjusted as necessary at week 2. The topical NSAID diclofenac diethylammonium gel (11.6 mg/g) (diclofenac sodium 1% gel) to apply daily over the thumb base 3 times per day. They received a spatula with a permanent pen mark to standardize the amount to be used (corresponding to approximately 200mg in an area of 40 cm2).. Duration 6 weeks. Concurrent medication/care: Both groups were provided with education about osteoarthritis and ergonomic principles (formerly known as "joint protection") using a 9-page educational booklet and 2 individual, face-to-face sessions with the study physiotherapist. The educational booklet did not provide information about exercises or splints.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks

	(n=102) Intervention 2: Non-combined active treatment - Education programme. Education only. Duration 6 weeks. Concurrent medication/care: Both groups were provided with education about osteoarthritis and ergonomic principles (formerly known as "joint protection") using a 9-page educational booklet and 2 individual, face-to-face sessions with the study physiotherapist. The educational booklet did not provide information about exercises or splints.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks
Funding	Academic or government funding (This work was supported by an NHMRC Program Grant (APP1091302) and the Lincoln Centre for Research Into Bone and Joint Diseases. Dr Hunter is supported by an NHMRC Practitioner Fellowship. Dr Hodges is supported by an NHMRC Senior Principal Research Fellowship (APP1102905).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: VAS pain at 12 weeks; Group 1: mean 35.5 (SD 22.1); n=96, Group 2: mean 43.9 (SD 23.5); n=98; VAS 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 57.3 (13.1). Baseline control: 58.4 (14.1).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Treatment package: 6 discontinued intervention, 6 lost to follow up (2 unable to contact, 3 personal circumstances, 1 worsening of symptoms).; Group 2 Number missing: 7, Reason: Education: 1 incomplete assessment, 3 discontinued intervention, 3 lost to follow up (2 no perceived benefit, 1 unable to contact).

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: Functional Index for Hand Osteoarthritis at 12 weeks; Group 1: mean 7.6 (SD 4.4); n=96, Group 2: mean 9.5 (SD 4.4); n=98; Functional Index of Hand Osteoarthritis 0-30 Top=High is poor outcome; Comments: Baseline treatment package: 10.5 (4.1). Baseline control: 10.8 (4.0).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Treatment package: 6 discontinued intervention, 6 lost to follow up (2 unable to contact, 3 personal circumstances, 1 worsening of symptoms).; Group 2 Number missing: 7, Reason: Education: 1 incomplete assessment, 3 discontinued intervention, 3 lost to follow up (2 no perceived benefit, 1 unable to contact).

Protocol outcome 3: Discontinuation at ≤3 months
- Actual outcome: Discontinuation at 12 weeks; Group 1: 12/102, Group 2: 7/102; Comments: Treatment package: 6 discontinued intervention, 6 lost to follow up (2 unable to contact, 3 personal circumstances, 1 worsening of symptoms). Education: 1 incomplete assessment, 3 discontinued intervention, 3 lost to follow up (2 no perceived benefit, 1 unable to contact).
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Treatment package: 6 discontinued intervention, 6 lost to follow up (2 unable to contact, 3 personal circumstances, 1 worsening of symptoms).; Group 2 Number missing: 7, Reason: Education: 1 incomplete assessment, 3 discontinued intervention, 3 lost to follow up (2 no perceived benefit, 1 unable to contact).

Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Dias 2017[89]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=73)
Countries and setting	Conducted in Brazil; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People diagnosed with osteoarthritis in at least one knee based on the clinical and radiographic criteria of the American College of Rheumatology
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Older women with clinical and radiological diagnosis of knee osteoarthritis by the American College of Rheumatology criteria; aged 65 years or older; demonstrating no cognitive limitations to do aquatic activities assessed by the mini-mental state test
Exclusion criteria	Lower limb joint replacement surgery; history of recent trauma in lower-limbs; using any walking support (such as walking stick or crutches); have received physiotherapy or any other rehabilitation treatment in the past 3 months; present with open wounds or skin disease and urinary or faecal incontinence; severe radiological diagnosis of knee osteoarthritis (level IV according to the criteria of Kellgren and Lawrence) and unable to safely enter or exit the pool
Recruitment/selection of patients	Recruited from community centers in the city of Belo Horizonte, MG, Brazil
Age, gender and ethnicity	Age - Mean (SD): 70.9 (5.1). Gender (M:F): 0:65. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=37) Intervention 1: Treatment package - Exercise and education programme. Hydrotherapy and educational protocol. A standardized hydrotherapy protocol including progressive exercises, which were implemented twice a week for 6 weeks. The program included three stages: war-up (5 min), strengthening exercises (30 min), and a cool down session (5 min). Exercises included lower limb strengthening exercising including closed kinetic chain exercises using floats as well as

	multidirectional walking tasks. People were instructed to perform the exercises on the maximal possible intensity. The educational protocol was desired to provide educational information about the diagnosis, symptoms, prognosis and basic care of knee osteoarthritis during daily activities. This consisted of one lecture delivered in groups of six participants in a classroom. They also received weekly advice through telephone discussions about controlling knee loading during daily activities. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (6 weeks). (n=36) Intervention 2: Non-combined active treatment - Education programme. Education programme only. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Academic or government funding (We acknowledge the financial support of the Brazilian funding agencies Conselho Nacional de Desenvolvimento Cientifico e Tacnologico (CNPq) and Coordenacae de Aperfeicoamento de Pessoal de Nivel Superior)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain at 6 weeks; Group 1: mean -13.33 (SD 16.23); n=37, Group 2: mean -2.3 (SD 15.1); n=36; WOMAC pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 51.1 (20.4). Baseline education: 50.9 (19.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, dominant side, joint involvement, knee with complaint and baseline values of outcomes; Group 1 Number missing: 4, Reason: 2 discontinued intervention, 2 clinical conditions; Group 2 Number missing: 4, Reason: 1 spontaneous desistance, 1 accessing difficulty, 2 clinical conditions

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: WOMAC function at 6 weeks; Group 1: mean -16.4 (SD 17.5); n=37, Group 2: mean -5.1 (SD 9.6); n=36; WOMAC function 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 52.7 (20.6). Baseline education: 55.3 (21.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, dominant side, joint involvement, knee with complaint and baseline values of outcomes; Group 1 Number missing: 4, Reason: 2 discontinued intervention, 2 clinical conditions; Group 2 Number

missing: 4, Reason: 1 spontaneous desistance, 1 accessing difficulty, 2 clinical conditions Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Lost to follow up at 6 weeks; Group 1: 4/37, Group 2: 4/36; Comments: Treatment package: 2 discontinued intervention, 2 clinical conditions. Education: 1 spontaneous desistance, 1 accessing difficulty, 2 clinical conditions Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, dominant side, joint involvement, knee with complaint and baseline values of outcomes; Group 1 Number missing: 4, Reason: 2 discontinued intervention, 2 clinical conditions; Group 2 Number missing: 4, Reason: 1 spontaneous desistance, 1 accessing difficulty, 2 clinical conditions	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Dwyer 2015[93]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=83)
Countries and setting	Conducted in South Africa, USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 4 weeks of treatment, 5 weeks of follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Mild-moderate knee osteoarthritis based on the American College of Rheumatology and the Kellgren Lawrence grade (suitable grades being grades 0 to 3)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age at least 38 and no more than 80 years; knee pain for at least 1 year and able to stand and walk without severe varus/valgus deformity and/or severe instability (instability being defined as a significant increase in the anterior drawer or varus/valgus movement when compared to the opposite knee); a minimum of 1 of the 3 clinical criteria below for a diagnosis of knee osteoarthritis (sensitivity 89% and specificity 88%) a) knee pain and crepitus with active motion and morning stiffness of less than or equal to 30 minutes or; b) knee pain and crepitus with active motion and morning stiffness >30 minutes and bony enlargement or; c) knee pain and no crepitus and bony enlargement (bony enlargement being determined on palpation and supplemented by observations on radiographs); no history of knee surgery in the past 6 months; Kellgren and Lawrence grade of 0-3 on plain-film radiographs; ability to stand and walk without assistance for most of the day, as keeping active and performing exercises would otherwise be difficult; a participant was required to have a score of at least 720/2400 on the WOMAC; no previous manual and/or manipulative therapy for their knee pain
Exclusion criteria	Kellgren-Lawrence grade 4 knee degenerative changes on plain-film radiographs, indicating severe knee osteoarthritis; possibility of serious pathological or psychiatric disorders; possibility of a disorder that would prevent the person from performing exercises or contraindications to manual and manipulative therapy
Recruitment/selection of patients	People were recruited from the areas of 2 chiropractic university-based outpatient teaching clinics, 1 in the city of Durban, South Africa, at Durban University of Technology and the other in Los Angeles, California, at Cleveland Chiropractic

	College, Los Angeles, in the United States. People were recruited by advertisements on campus, local radio and local newspapers.
Age, gender and ethnicity	Age - Mean (SD): 62.2 (11.1). Gender (M:F): 29:49. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Grade 1-2, median grade 1 Duration of symptoms (mean [SD]): 83.9 (96.1) months
Indirectness of population	No indirectness
Interventions	(n=28) Intervention 1: Treatment package - Combination and education programme. Manual therapy and rehabilitation (including exercise and education). 6 treatment sessions of manual therapy over a 4 week treatment period for around 120 minutes in total. The treatment comprised joint mobilization (grades 1-4) and joint manipulation (grade 5; high-velocity, low-amplitude, thrust-type manipulation) of the affected kinematic chain (kne, hip, foot and spine). Manipulation, mobilization and soft tissue treatment were based on techniques previously described. Manipulation was applied to joints with restricted range of motion, identified by joint motion palpation by the treating clinician, using the high-velocity and low-amplitude manipulations noted above or, as described in textbooks and other peer-reviewed papers. Forced end-ROM grade 4 mobilisations or grade 5 thrust manipulations were avoided, particularly in flexion and extension, where it was likely to worsen symptoms or could not be tolerated by the participant. The rehabilitation program included patient education, exercise prescription, soft tissue treatment and passive stretches to the knee and elsewhere along the full kinetic chain. Education consisted of information about the diagnosis and prognosis, and advice on health promotion and lifestyle. The content and timing of treatment were import in that advice, education and training were provided to participants at the onset of their treatment program and reinforced at 2 other points during the treatment period. This was to reinforce the need for rehabilitation and to encourage compliance. Each treatment session was approximately 20 minutes with 12 sessions over the 4 weeks. Duration 4 weeks. Concurrent medication/care: Leaflet advice about the diagnosis, prognosis, and lifestyle advice was provided to all participants. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (4 weeks). (n=27) Intervention 2: Non-combined active treatment - Manual therapy. Manual therapy only. Duration 4 weeks. Concurrent medication/care: Leaflet advice about the diagnosis, prognosis, and lifestyle advice was provided to all participants.

	Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (4 weeks). (n=28) Intervention 3: Treatment package - Exercise and education programme. Exercise and education only. Duration 4 weeks. Concurrent medication/care: Leaflet advice about the diagnosis, prognosis, and lifestyle advice was provided to all participants. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (4 weeks). Comments: This group was not included in the analysis as there was not a valid comparator to compare it to
Funding	Academic or government funding (The NCMIC Foundation supported the development of the manuscript)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus MANUAL THERAPY

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain at 5 weeks; Group 1: mean 97.7 (SD 86.8); n=28, Group 2: mean 102.3 (SD 88.9); n=27; WOMAC pain 0-500 Top=High is poor outcome; Comments: Reported mean (standard error). Reported treatment package: 97.7 (16.4). Reported manual therapy: 102.3 (17.1). Baseline treatment package: 216.8 (17.0). Baseline manual therapy: 227.3 (17.2).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 2, Reason: Overall reasons given only (for all three groups): 3 no shows, 1 moved to another town, 1 underwent unrelated surgery; Group 2 Number missing: 1, Reason: Overall reasons given only (for all three groups): 3 no shows, 1 moved to another town, 1 underwent unrelated surgery

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: WOMAC function at 5 weeks; Group 1: mean 378.9 (SD 261.4); n=28, Group 2: mean 389.7 (SD 293.1); n=27; WOMAC function 0-1800 Top=High is poor outcome; Comments: Reported mean (standard error). Reported treatment package: 378.9 (62.0). Reported manual therapy: 389.7 (49.4). Baseline treatment package: 411.7 (52.0) - this appears to be a typo and a copy of the group 2 1 week follow up result. Baseline manual therapy: 759.0 (47.6).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 2, Reason: Overall reasons given only (for all three groups): 3 no shows, 1 moved to another town, 1 underwent unrelated surgery; Group 2 Number missing: 1, Reason: Overall reasons given only (for all three groups): 3 no

shows, 1 moved to another town, 1 underwent unrelated surgery Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Drop outs at 5 weeks; Group 1: 2/28, Group 2: 1/27; Comments: Overall reasons given only (for all three groups): 3 no shows, 1 moved to another town, 1 underwent unrelated surgery Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 2, Reason: Overall reasons given only (for all three groups): 3 no shows, 1 moved to another town, 1 underwent unrelated surgery; Group 2 Number missing: 1, Reason: Overall reasons given only (for all three groups): 3 no shows, 1 moved to another town, 1 underwent unrelated surgery	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	**Dziedzic 2018[95] (Jordan 2017[145], Oppong 2018[212])**
Study type	Cluster RCT (Practices randomised; Parallel)
Number of studies (number of participants)	1 (n=8)

Countries and setting	Conducted in United Kingdom; Setting: General practice (primary care)
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Joint pain in the past year (in the hand, hip, knee, foot) self-reported in the Health Survey in people 45 years and over
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Males and females; 45 years and over; registered with a MOSAICS study practice; joint pain in the past year (in the hand, hip, knee, foot) self-reported in the Health Survery; consent to further content from the study team (consent sought as part of the Health Survey); consent to medical record review (consent sought as part of the Health Survey); consulting with joint pain in the cluster trial recruitment period; triggered the e-template in EMIS system in the consultation; returned post-consultation baseline Questionnaire.
Exclusion criteria	Excluded via GP screen of practice list; unable to give fully informed consent e.g., learning difficulties or dementia; resident in a care or nursing home; history of serious disease e.g., malignancy, terminal illness; unable to consult in the general practice surgery; flagged as excluded from research in that practice; declined to take part in the post-consultation baseline Questionnaire.
Recruitment/selection of patients	People who responded to surveys sent out by 10 general practices.
Age, gender and ethnicity	Age - Mean (SD): 67.3 (10.5). Gender (M:F): 212:313. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Mixed (hand, hip, knee, foot – proportions not stated)
Extra comments	Severity: Not stated/unclear Duration of symptoms: Not stated/unclear
Indirectness of population	No indirectness
Interventions	(n=4) Intervention 1: Treatment package - Combination and behaviour change intervention. An enhanced GP consultation to make, give and explain the diagnosis, and provide initial care for older adults presenting with peripheral joint pain; an osteoarthritis guidebook offered by the GP to patients to support osteoarthritis self-management; advice on analgesia; up to four follow-up practice nurse consultations to guide people in self management for osteoarthritis with advice on weight management if required, general exercise and physical activity with goal-setting as appropriate.

	Duration 12 months (difficult to define time for the intervention, but will potentially lead to follow up for an extended period of time so the maximal time of follow up was used). Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: ≤ 6 weeks (the package was likely delivered for less than 6 weeks, but follow up may continue for a long time after this). (n=4) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Control practices received no training, guidebook or dedicated nurse osteoarthritis clinic and continued usual care as in the pre-randomisation period. Duration 12 months (difficult to define time for the intervention, but will potentially lead to follow up for an extended period of time so the maximal time of follow up was used). Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: ≤ 6 weeks (the package was likely delivered for less than 6 weeks, but follow up may continue for a long time after this).
Funding	Academic or government funding (Independent research funded by the National Institute for Health Research (NIHR) Programme Grant (RP-PG-0407-10386).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED) OR NO TREATMENT

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: EQ-5D at 3 months; Group 1: mean 0.615 (SD 0.280); n=4, Group 2: mean 0.631 (SD 0.264); n=4; EQ-5D -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.573 (0.298). Baseline manual therapy: 0.588 (0.272).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: EQ-5D at 12 months; Group 1: mean 0.651 (SD 0.262); n=4, Group 2: mean 0.674 (SD 0.224); n=4; EQ-5D -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.573 (0.298). Baseline manual therapy: 0.588 (0.272).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the

baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: Arthritis self-efficacy pain subscale at 3 months; Group 1: mean 5.82 (SD 2.18); n=4, Group 2: mean 5.82 (SD 2.13); n=4; Arthritis self-efficacy pain subscale 0-10 Top=High is good outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 4: Pain at >3 months
- Actual outcome: Arthritis self-efficacy pain subscale at 12 months; Group 1: mean 5.83 (SD 2.24); n=4, Group 2: mean 6.04 (SD 2.17); n=4; Arthritis self-efficacy pain subscale 0-10 Top=High is good outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: Physical Activity Scale for the Elderly at 3 months; Group 1: mean 123.6 (SD 72.0); n=4, Group 2: mean 149.1 (SD 90.6); n=4; Physical Activity Scale for the Elderly 0-793 Top=High is good outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 6: Physical function at >3 months
- Actual outcome: Physical Activity Scale for the Elderly at 12 months; Group 1: mean 134.2 (SD 69.6); n=4, Group 2: mean 148.2 (SD 77.9); n=4; Physical Activity Scale for the Elderly 0-793 Top=High is good outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 7: Psychological distress at ≤3 months
- Actual outcome: PHQ-8 at 3 months; Group 1: mean 4.36 (SD 4.50); n=4, Group 2: mean 3.85 (SD 4.64); n=4; PHQ-8 0-24 Top=High is poor outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 8: Psychological distress at >3 months
- Actual outcome: PHQ-8 at 12 months; Group 1: mean 4.06 (SD 4.74); n=4, Group 2: mean 3.96 (SD 4.81); n=4; PHQ-8 0-24 Top=High is poor outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 9: Psychological distress at ≤3 months
- Actual outcome: GAD-7 at 3 months; Group 1: mean 3.16 (SD 4.32); n=4, Group 2: mean 2.90 (SD 4.60); n=4; GAD-7 0-21 Top=High is poor outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 10: Psychological distress at >3 months
- Actual outcome: GAD-7 at 12 months; Group 1: mean 2.90 (SD 4.31); n=4, Group 2: mean 2.75 (SD 3.84); n=4; GAD-7 0-21 Top=High is poor outcome; Comments: Baseline values not reported.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 11: Discontinuation at >3 months
- Actual outcome: Discontinuation at 12 months; Group 1: 21/237, Group 2: 27/288; Comments: Intervention: 17 did not wish to take further part in the study, 1 due to ill health, 2 non-responders, 1 no longer at the address. Control: 20 did not wish to take further part in the study again, 2 withdrawal due to ill health, 3 non responders, 2 no longer at the address.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, height, weight, BMI, onset, duration of symptoms, side affected, occupation, regular exercise, grade of knee osteoarthritis and baseline values of outcomes (note, there is a likely typo for the baseline value for physical function which makes this unclear); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcomes not reported by the study	Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Farr 2010[98]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=293)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 9 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Pain on 4 or more days of the week in one or both knees for at least 4 months during the previous year with radiographic status of grade 2 osteoarthritis in at least one knee. All people met the American College of Rheumatology classification criteria for early osteoarthritis of the knee
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age between 35 and 68 years to ensure an early-onset knee osteoarthritis sample; pain on 4 or more days of the week in one or both knees for at least 4 months during the previous year; less than 5 years' symptom duration; radiographic status of grade 2 osteoarthritis (and no higher) in at least one knee, as defined by the Kellgren and Lawrence classification; disability due to knee osteoarthritis as assessed with the WOmac index
Exclusion criteria	No additional information
Recruitment/selection of patients	People were recruited from the Tucson, Arizona, general community and surrounding areas using mass mailings, media advertisements, periodic media coverage, and requests to local physicians for patient referrals
Age, gender and ethnicity	Age - Mean (SD): 55.1 (7.0). Gender (M:F): 43:128. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren Lawrence grade 2 Duration of symptoms: Less than 5 years
Indirectness of population	No indirectness
Interventions	(n=100) Intervention 1: Treatment package - Exercise and behaviour change intervention. Combined resistance training and self-management. Exercise focused on 4 core areas: stretching and balance; range of motion and flexibility; isotonic muscle strengthening; aerobics. People met with certified physical trainers 3 times a week for 9 months, with a minimum of 1 days of rest between training sessions, to complete 1

	hour exercise regimens. Each session consisted of: 10 minute warm up on either a bicycle ergometer or treadmill at 50% maximum heart rate; 5 to 10 minutes of stretching and balance exercises; 10 minutes of range of motion exercises; 30 minutes of RT exercises; 5 minutes of coll-down. Specific exercises included leg press, leg curl, hip abduction and adduction, straight leg lift, incline dumbbell press, seated row, and calf raise. Self-management training was designed to target coping skills, promoting the use of more adaptive strategies and fewer avoidance or passive strategies based on existing self-help programs. The 9 month program began with 12 weekly 90 minute classroom sessions in which participants completed education modules addressing an overview of osteoarthritis, general exercise principles and physical activity recommendations, stress management, foot care, pain management, analgesic and anti-inflammatory medications, nutrition for health, coping mechanisms, communication with health care providers, and healthy lifestyle practices. Classroom sessions were followed by 24 weeks of a structured telephone intervention program that reinforced self-management skills.. Duration 9 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Self management training). 2. Length of package: > 6 weeks (9 months). (n=95) Intervention 2: Non-combined active treatment - Exercise. Exercise component only. Duration 9 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (9 months). (n=98) Intervention 3: Non-combined active treatment - Behaviour change intervention. Self-management training only. Duration 9 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (9 months).
Funding	Academic or government funding (The project was supported by National Institutes of Health/National Institute of Arthritis and musculoskeletal and Skin Diseases grant R01-AR-047595. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Arthritis and Musculoskeletal and Skin Diseases or the National Institutes of Health.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain at 3 months; Group 1: mean 67.1 (SD 68.8); n=100, Group 2: mean 47.6 (SD 50.9); n=95; WOMAC pain 0-500 Top=High is poor outcome; Comments: Baseline treatment package: 81.9 (67.3). Baseline exercise: 84.3 (70.1). Baseline behaviour intervention: 82.2 (68.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, height, weight, BMI, knee osteoarthritis severity and baseline values of outcome; Group 1 Number missing: 24, Reason: Treatment package: 9 did not receive allocated intervention (not interested, time commitment, leaving local area, personal). 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given).; Group 2 Number missing: 23, Reason: Exercise: 11 did not receive allocated intervention (not interested, lost to follow-up, concomitant health problems, time commitment, personal, leaving local area). 12 discontinued due to not interested, time commitment, concomitant health problems, leaving local area, personal, lost to follow-up, inflammatory arthritis.

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 9 months; Group 1: mean 56.2 (SD 75.3); n=100, Group 2: mean 48.6 (SD 61.3); n=95; WOMAC pain 0-500 Top=High is poor outcome; Comments: Baseline treatment package: 81.9 (67.3). Baseline exercise: 84.3 (70.1). Baseline behaviour intervention: 82.2 (68.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, height, weight, BMI, knee osteoarthritis severity and baseline values of outcome; Group 1 Number missing: 24, Reason: Treatment package: 9 did not receive allocated intervention (not interested, time commitment, leaving local area, personal). 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given).; Group 2 Number missing: 23, Reason: Exercise: 11 did not receive allocated intervention (not interested, lost to follow-up, concomitant health problems, time commitment, personal, leaving local area). 12 discontinued due to not interested, time commitment, concomitant health problems, leaving local area, personal, lost to follow-up, inflammatory arthritis.

Protocol outcome 3: Discontinuation at >3 months
- Actual outcome: Discontinued intervention at 9 months; Group 1: 15/100, Group 2: 12/95; Comments: Treatment package: 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given). Exercise: 12 discontinued due to not interested, time commitment, concomitant health problems, leaving local area, personal, lost to follow-up, inflammatory arthritis.
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, height, weight, BMI, knee osteoarthritis severity and baseline values of outcome; Group 1 Number missing: 24, Reason: Treatment package: 9 did not receive allocated intervention (not interested, time commitment, leaving local area, personal). 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given).; Group 2 Number missing: 23, Reason: Exercise: 11 did not receive allocated intervention (not interested, lost to follow-up, concomitant health problems, time commitment, personal, leaving local area). 12 discontinued due to not interested, time commitment, concomitant health problems, leaving local area, personal, lost to follow-up, inflammatory arthritis.

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Pain at ≤3 months - Actual outcome: WOMAC pain at 3 months; Group 1: mean 67.1 (SD 68.8); n=100, Group 2: mean 72 (SD 66.3); n=98; WOMAC pain 0-500 Top=High is poor outcome; Comments: Baseline treatment package: 81.9 (67.3). Baseline exercise: 84.3 (70.1). Baseline behaviour intervention: 82.2 (68.3) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, height, weight, BMI, knee osteoarthritis severity and baseline values of outcome; Group 1 Number missing: 24, Reason: Treatment package: 9 did not receive allocated intervention (not interested, time commitment, leaving local area, personal). 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given).; Group 2 Number missing: 25, Reason: Behaviour intervention: 19 did not receive allocated intervention (not interested, not adherent, time commitment, concomitant health problems, personal, other). 6 discontinued due to lost to follow-up, not adherent, time commitment. Protocol outcome 2: Pain at >3 months - Actual outcome: WOMAC pain at 9 months; Group 1: mean 56.2 (SD 75.3); n=100, Group 2: mean 62.9 (SD 81); n=98; WOMAC pain 0-500 Top=High is poor outcome; Comments: Baseline treatment package: 81.9 (67.3). Baseline exercise: 84.3 (70.1). Baseline behaviour intervention: 82.2 (68.3) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, height, weight, BMI, knee osteoarthritis severity and baseline values of outcome; Group 1 Number missing: 24, Reason: Treatment package: 9 did not receive allocated intervention (not interested, time commitment, leaving local area, personal). 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given).; Group 2 Number missing: 25, Reason: Behaviour intervention: 19 did not receive allocated intervention (not interested, not adherent, time commitment, concomitant health problems, personal, other). 6 discontinued due to lost to follow-up, not adherent, time commitment. Protocol outcome 3: Discontinuation at >3 months - Actual outcome: Discontinued intervention at 9 months; Group 1: 15/100, Group 2: 6/98; Comments: Treatment package: 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given). Behaviour intervention: 6 discontinued due to lost to follow-up, not adherent, time commitment. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, height, weight, BMI, knee osteoarthritis severity and baseline values of outcome; Group 1 Number missing: 24, Reason: Treatment package: 9 did not receive allocated intervention (not interested, time commitment, leaving local area, personal). 15 discontinued due to lost to follow-up, concomitant health problems, not interested, time commitment, not adherent, other (the breakdown of what contributed is not given).; Group 2 Number missing: 25, Reason: Behaviour intervention: 19 did not receive allocated intervention (not interested, not adherent, time commitment, concomitant health problems, personal, other). 6 discontinued due to lost to follow-up, not adherent, time commitment.	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study (subsidiary papers)	Fernandes 2010[99] (Svege 2016[264], Svege 2015[265])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=109)
Countries and setting	Conducted in Norway; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of intervention, 16 months of follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Radiographical and symptomatic hip osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People aged 40 and 80 years who had experienced hip pain for the past 3 months or longer; radiographically certified minimum joint space <4mm for patients <70 years old and <3mm for people at least 70 years old and a Harris Hip Score of between 60 and 95 points. The included people had to have both radiographic and symptomatic hip osteoarthritis.
Exclusion criteria	Total hip replacement in the index joint; had been diagnosed with knee osteoarthritis; had knee pain; low back pain; rheumatoid arthritis; osteoporosis; cancer; cardiovascular disease; did not tolerate exercise; dysfunction in lower extremities due to accident or disease; were pregnant; did not understand Norwegian
Recruitment/selection of patients	People were recruited by one university hospital, one local hospital, one rehabilitation center, general medical practitioners and by advertisement in a local newspaper in Oslo, Norway.
Age, gender and ethnicity	Age - Mean (SD): 57.8 (9.9). Gender (M:F): 50:59. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms (mean [SD]): 48.4 (52.1) months
Indirectness of population	No indirectness
Interventions	(n=55) Intervention 1: Treatment package - Exercise and education programme. "Hip School" - patient education and supervised exercise. Education consisted of three group-based sessions and one individual physical therapy visit, 2 months after completing the group sessions. The exercise was a therapeutic exercise program specifically designed for people with hip osteoarthritis. The group started exercises

	within a week of completing the patient education sessions. The exercise program included 26 exercises, including warm-up, strengthening exercises, functional exercises and flexibility exercises. This program was supervised twice a week, but access to the gym was provided throughout weekdays for a period of 12 weeks. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks). (n=54) Intervention 2: Non-combined active treatment - Education programme. Patient education only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: SF-36 physical function at 16 months; Group 1: mean 75.5 (SD 20.5); n=40, Group 2: mean 71.3 (SD 20.8); n=35; SF-36 physical function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 70.9 (18.5). Baseline education: 71.6 (17.1).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 role physical at 16 months; Group 1: mean 82.3 (SD 25.5); n=41, Group 2: mean 75.7 (SD 29); n=37; SF-36 role physical 0-100 Top=High is good outcome; Comments: Baseline treatment package: 80.4 (23.2). Baseline education: 74.3 (26.3).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 bodily pain at 16 months; Group 1: mean 70.5 (SD 18.6); n=41, Group 2: mean 61.4 (SD 24.3); n=37; SF-36 bodily pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 62.8 (16.0). Baseline education: 57.4 (19.1).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain

duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 general health at 16 months; Group 1: mean 71.3 (SD 20.7); n=38, Group 2: mean 67.6 (SD 22.1); n=36; SF-36 general health 0-100 Top=High is good outcome; Comments: Baseline treatment package: 69.5 (21.8). Baseline education: 68.5 (17.7).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 vitality at 16 months; Group 1: mean 59 (SD 21); n=41, Group 2: mean 61.7 (SD 20.6); n=37; SF-36 vitality 0-100 Top=High is good outcome; Comments: Baseline treatment package: 58.0 (20.3). Baseline education: 58.3 (20.4).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 social function at 16 months; Group 1: mean 91.2 (SD 15.9); n=41, Group 2: mean 84.1 (SD 26.9); n=37; SF-36 social function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 80.4 (18.6). Baseline education: 85.9 (23.4).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 role emotional at 16 months; Group 1: mean 90.7 (SD 15.5); n=41, Group 2: mean 90.5 (SD 21.7); n=37; SF-36 role emotional 0-100 Top=High is good outcome; Comments: Baseline treatment package: 94.3 (13.1). Baseline education: 91.5 (19.3).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.
- Actual outcome: SF-36 mental health at 16 months; Group 1: mean 81.8 (SD 14.9); n=40, Group 2: mean 82.8 (SD 15.4); n=37; SF-36 mental health 0-100 Top=High is good outcome; Comments: Baseline treatment package: 81.8 (15.4). Baseline education: 82.2 (13.1).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.

Protocol outcome 2: Pain at >3 months - Actual outcome: WOMAC pain at 16 months; Group 1: mean 17.3 (SD 14.5); n=42, Group 2: mean 22.3 (SD 18.4); n=36; WOMAC pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 26.0 (16.1). Baseline education: 27.3 (17.9). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond. Protocol outcome 3: Physical function at >3 months - Actual outcome: WOMAC function at 16 months; Group 1: mean 15.1 (SD 13.7); n=41, Group 2: mean 22.8 (SD 18.6); n=36; WOMAC function 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 21.1 (15.3). Baseline education: 23.6 (15.7). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond. Protocol outcome 4: Discontinuation at >3 months - Actual outcome: Reasons for not attending at 16 months; Group 1: 13/55, Group 2: 18/54; Comments: Treatment packages: 6 total hip replacement, 7 did not respond. Education programme: 11 total hip replacement, 7 did not respond. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, MJS, HHS, pain duration, uni/bilateral joint space narrowing and baseline values of outcomes. Different values for SF-36 role physical, bodily pain, general health and role emotional.; Group 1 Number missing: 13, Reason: Treatment packages: 6 total hip replacement, 7 did not respond.; Group 2 Number missing: 18, Reason: Education programme: 11 total hip replacement, 7 did not respond.	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study (subsidiary papers)	Focht 2005[106] (Focht 2004[105], Messier 2004[188], Miller 2003[194], Van gool 2005[275], Shea 2010[252])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=316)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 18 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee pain on most days with radiographic evidence of grade 1-3 tibiofemoral or patellofemoral osteoarthritis based on weight-bearing anteroposterior and sunrise view radiographs
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age at least 60 years; calculated body mass index of at least 28kg/m²; knee pain on most days of the month; sedentary activity pattern with <20 minutes of formal exercise once weekly for the past 6 months; self-reported difficulty in at least one of the following activities ascribed to knee pain: walking one quarter of a mile, climbing stairs, bending, stopping, kneeling, shopping, house cleaning or other self-care activities, getting in and out of bed, standing up from a chair, lifting and carrying groceries, o r getting in and out of the bathtub; radiographic evidence of grade 1-3 tibiofemoral or patellofemoral osteoarthritis based on weight-bearing anteroposterior and sunrise view radiographs; willingness to undergo testing and intervention procedures
Exclusion criteria	Serious medical condition that prevented safe participation in an exercise program, including symptomatic heart or vascular disease (angina, peripheral vascular disease, congestive heart failure), severe hypertension, recent stroke, chronic obstructive pulmonary disease, severe insulin-dependent diabetes mellitus, psychiatric disease, renal disease, liver disease, active cancer other than skin cancer, and anaemia; a mini-mental state examination score of <24; inability to finish the 18-month study or unlikely to be compliant; inability to walk without a cane or other assistive device; participation in another research study; reported alcohol consumption of >14 drinks per week; ST segment depression of at least 2mm at an exercise level of 4 METS or less, hypotesnion, or complex arrhythmias during a graded exercise test; inability to complete the protocol, in the opinion of the clinical staff, because of frailty , illness or other reasons

Recruitment/selection of patients	People were recruited from mass mailings within the target area, targeted mailing to employees of the university and medical center; presentations to various groups of older adults, mass media advertisement, and placement of posters (with pull-off reply cards) in strategic locations. They tried to enhance recruitment of racial minorities, including ads and interviews on minority-run radio stations, newspaper ads in predominantly African American publications, letters to churches attended mainly by minotiries, and inserts in these church bulletins
Age, gender and ethnicity	Age - Mean (SD): 68.7 (6.3). Gender (M:F): 89:227. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: High morbidity score (70-84% were obese, 53-58% had arthritis in other joints, 44-54% had hypertension, 23-34% had coronary heart disease, 6-12% had diabetes). 4. Site of osteoarthritis: Knee
Extra comments	Severity: Mean Kellgren Lawrence score: 2.3 (0.7) Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=76) Intervention 1: Treatment package - Exercise and behaviour change intervention. Diet and exercise. The dietary intervention strategy was conducted by trained registered dieticians. They worked with the health psychologist in the development and delivery of the behavioural aspects of the intervention. The weight-loss goal for these two groups was a mean loss of at least 5% of initial body weight. This was achieved through weekly meetings with a registered dietitian discussing healthful food selection with portion and dietary fat control to decrease energy intake, emphasizing an increased awareness in the consequences of, and the need to change, dietary habits. People were counselled to reduce energy intake by around 500 calories per day in order to achieve the desired weight loss. Group and individual sessions took place. Examples of group program topics including health eating, reading labels, shopping, food preparation, meal ideas, restaurants, ethnic eating, special occasions, and old and new routines. Meetings were weekly for 4 months, then biweekly for months 5-6, then monthly for months 7-18 (but with biweekly phone calls). The exercise therapy included a program 3 times a week for 60 minutes per session including a warm-up phase (5 minutes), an aerobic phase (15 minutes), a strength phase (20 minutes), a second aerobic phase (15 minutes) and a cool down phase (5 minutes). The exercise intensity for the aerobic exercise was 50-85% of the heart rate reserve. Strength training included: leg extension, leg curl, heel raise, and step-ups using ankle cuff weights and a weighted vest. 2 sets of 12 repetitions were performed for each.

	. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Weight loss). 2. Length of package: > 6 weeks (18 months). (n=80) Intervention 2: Non-combined active treatment - Exercise. Exercise component only. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (18 months). (n=82) Intervention 3: Non-combined active treatment - Behaviour change intervention. Weight loss intervention only. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Weight loss). 2. Length of package: > 6 weeks (18 months). (n=78) Intervention 4: Standard care (non-organised) or no treatment - Standard care (non-organised). No intervention. Participants had regular meetings for 1 hour monthly over the first 3 months to provide attention, social interaction, and some health education 9discussing osteoarthritis, obesity and exercise, and the healthy lifestyle program).. Duration 18 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (18 months).
Funding	Study funded by industry (Supported by the National Institute of Aging (grants AG14131 and 5P60 AG10484) and the General Clinical Research Center (grant M01-RR07122).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Pain at >3 months
- Actual outcome: WOMAC pain at 18 months; Group 1: mean -2.2 (SD 4.1); n=76, Group 2: mean -0.4 (SD 4.3); n=80; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -2.20 (-3.12, -1.28). Reported exercise: -0.40 (-1.32, 0.52). Baseline treatment package (mean [SE]): 7.27 (0.41). Baseline exercise: 6.64 (0.39).

Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnicity, weight, height, BMI, annual household income, education, comorbid illnesses, Kellgren Lawrence score, WOMAC function and baseline values of outcomes; Group 1 Number missing: 18, Reason: Reasons not given; Group 2 Number missing: 16, Reason: Reasons not given

Protocol outcome 2: Discontinuation at >3 months
- Actual outcome: Did not complete the study at 18 months; Group 1: 18/76, Group 2: 16/80; Comments: Reasons not given
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnicity, weight, height, BMI, annual household income, education, comorbid illnesses, Kellgren Lawrence score, WOMAC function and baseline values of outcomes; Group 1 Number missing: 18, Reason: Reasons not given; Group 2 Number missing: 16, Reason: Reasons not given

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Pain at >3 months
- Actual outcome: WOMAC pain at 18 months; Group 1: mean -2.2 (SD 4.1); n=76, Group 2: mean -1.07 (SD 4.1); n=82; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -2.20 (-3.12, -1.28). Reported diet: -1.07 (-1.95, -0.19). Baseline treatment package (mean [SE]): 7.27 (0.41). Baseline diet: 6.58 (0.40).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnicity, weight, height, BMI, annual household income, education, comorbid illnesses, Kellgren Lawrence score, WOMAC function and baseline values of outcomes; Group 1 Number missing: 18, Reason: Reasons not given; Group 2 Number missing: 19, Reason: Reasons not given

Protocol outcome 2: Discontinuation at >3 months
- Actual outcome: Did not complete the study at 18 months; Group 1: 18/76, Group 2: 19/82; Comments: Reason not given
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnicity, weight, height, BMI, annual household income, education, comorbid illnesses, Kellgren Lawrence score, WOMAC function and baseline values of outcomes; Group 1 Number missing: 18, Reason: Reasons not given; Group 2 Number missing: 19, Reason: Reasons not given

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Pain at >3 months
- Actual outcome: WOMAC pain at 18 months; Group 1: mean -2.2 (SD 4.1); n=76, Group 2: mean -1.23 (SD 4); n=78; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -2.20 (-3.12, -1.28). Reported healthy lifestyle: -1.23 (-2.11, -0.35). Baseline treatment package (mean [SE]): 7.27 (0.41). Baseline healthy lifestyle: 7.25 (0.39).

Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnicity, weight, height, BMI, annual household income, education, comorbid illnesses, Kellgren Lawrence score, WOMAC function and baseline values of outcomes; Group 1 Number missing: 18, Reason: Reasons not given; Group 2 Number missing: 11, Reason: Reasons not given Protocol outcome 2: Discontinuation at >3 months - Actual outcome: Did not complete the study at 18 months; Group 1: 18/76, Group 2: 11/78; Comments: Reasons not given Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnicity, weight, height, BMI, annual household income, education, comorbid illnesses, Kellgren Lawrence score, WOMAC function and baseline values of outcomes; Group 1 Number missing: 18, Reason: Reasons not given; Group 2 Number missing: 11, Reason: Reasons not given	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study (subsidiary papers)	Focht 2014[103] (Focht 2017[104])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=80)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Radiographically confirmed, symptomatic knee osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age >55 years; knee pain on most days of the month; less than 20 minutes/week of structured exercise during the prior 6 months; self-reported difficulty with at least 1 of the following activities because of knee pain: walking 0.25 miles, climbing stairs, bending, stooping, kneeling, shopping, housecleaning, or self-care activities such as getting in or out of bed, standing up from a chair, lifting and carrying groceries, or getting in or out of a bathtub; radiographic evidence of Kellgren-Lawrence scale stage 2-3 (mild to moderate) tibiofemoral osteoarthritis; willingness to participate in the study protocol
Exclusion criteria	Serious medical conditions such as active cardiovascular disease, cancer, or pulmonary disease; inability to walk without a cane or other assistive device; physician-documented radiographic evidence of knee joint varus or valgus malalignment; participation in another research study; any more than 21 alcoholic drinks per week; osteoarthritis severity >3 on the Kellgren-Lawrence scale; inability to complete out 12-month study or unlikely to be compliant because of conflicts; other safety/adherence concerns noted by the clinical staff
Recruitment/selection of patients	No additional information
Age, gender and ethnicity	Age - Mean (SD): 63.5 (6.9). Gender (M:F): 13:67. Ethnicity: White = 55, African American = 20, Asian = 2, Latino = 3
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren Lawrence grade 2-3 Duration of symptoms: Not stated
Indirectness of population	No indirectness

Interventions	(n=40) Intervention 1: Treatment package - Exercise and behaviour change intervention. Exercise and CBT intervention. 27, 80 minute, center-based sessions for a total of 36 total contact hours. Included 60 minutes of exercise (the same as the exercise only group - 30-40 minutes of moderate intensity aerobic exercise and 20 minutes of lower body strength training, performed for 1-3 sets of 8-12 repetitions, with 3 exercise sessions per week for 3 months) and 20 minutes of group-based cognitive behavioural activity counseling that focused on the use of key self-regulatory skills (self-monitoring, group and individual goal setting, barrier problem solving, action planning, relaxation/pain management strategies) to promote independent self-regulation of physical activity and prevent knee osteoarthritis disability.. Duration 3 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (CBT). 2. Length of package: > 6 weeks (3 months). (n=40) Intervention 2: Non-combined active treatment - Exercise. Exercise only. Duration 3 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (3 months).
Funding	Academic or government funding (Supported by the National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant #R21 AR054595)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Discontinuation at ≤3 months
- Actual outcome: Did not complete follow up at 12 weeks; Group 1: 7/40, Group 2: 9/40; Comments: Reasons only given for all participants. 9 missed/lost contact, 7 dropped out.
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, ethnicity, education, income, BMI and baseline values of outcomes; Group 1 Number missing: 7, Reason: Reasons only given for all participants. 9 missed/lost contact, 7 dropped out.; Group 2 Number missing: 9, Reason: Reasons only given for all participants. 9 missed/lost contact, 7 dropped out.

Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Gaines 2004[109]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=38)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of intervention, 16 weeks of follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Radiographic and clinical evidence of knee osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	60 years of age or older with radiographic and clinical evidence of knee osteoarthritis
Exclusion criteria	Presence of a cardiac pacemaker; cognitive impairment (a score less than 24 on the Mini-mental State Examination); uncontrolled conditions of diabetes, hyper- or hypotension; cardiac disease
Recruitment/selection of patients	No additional information
Age, gender and ethnicity	Age - Other: Mean: 70.8. Gender (M:F): 8:30. Ethnicity: White = 33, 'Non-white' = 5
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Grades 1-4, median grades 1-2 Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Treatment package - Electrotherapy and education programme. Neuromuscular electrical stimulation with the Arthritis Self-management program. The Arthritis Self-Help course was the standard of care for all and was taught as 12 hour community-based courses. It was designed to provide accurate information about arthritis, instill positive attitudes towards self-management (including pain management) and assist in developing personalized action plans (including exercise) for the management of arthritis. NMES was delivered by a portal electrical home stimulator. The parameters were: a rectangular waveform; pulsed, symmetric, biphasic current; 50 bursts per seccond; a ramp up time of 3 seconds each with an "on" time of 10 seconds followed by a 50-second "off" time. High impedance, reusable, self-adhesive electrodes were positioned over the vastus medialis oblique and proximal vastus lateralis of the index leg. People were asked to use the NMES device for 15

	minutes per day 3 days a week on the index leg for a total of 36 sessions. During the first 4 weeks, the intensity of electrical stimulation was set to induce a muscle contraction that was 10-20% of the isometric maximum voluntary contraction. Over the 12 weeks of the protocol, the electrical current intensity levels incrementally increased to achieve higher percentages: 20-30% during weeks 5 to 8, and 30-40% during weeks 9 to 12. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme (Arthritis self-management program). 2. Length of package: > 6 weeks (12 weeks). (n=18) Intervention 2: Non-combined active treatment - Education programme. Arthritis self-management program only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme (Arthritis self-management program). 2. Length of package: > 6 weeks (12 weeks).
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: ELECTROTHERAPY AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AIMS-2 pain subscale at 12 weeks; Group 1: mean 5.18 (SD 2.11); n=20, Group 2: mean 5.99 (SD 2.4); n=18; AIMS-2 pain subscale 0-10 Top=High is good outcome; Comments: Baseline treatment package: 4.85 (2.20). Baseline education: 3.61 (2.26).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in outcome at baseline (underestimating the benefit from the education programme alone); Group 1 Number missing: -, Reason: Reports that 4 people withdrew from the study, but doesn't state if this was before randomisation and which groups they were assigned to if it was after randomisation; Group 2 Number missing: -

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: Pain Rating Index-Total of the McGill Pain questionnaire at 12 weeks; Group 1: mean 14.95 (SD 13.07); n=20, Group 2: mean 10.63 (SD 4.84); n=18; McGill Pain Questionnaire 0-78 Top=High is poor outcome; Comments: Baseline treatment package: 19.68 (11.03). Baseline education: 14.00 (10.32).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in outcome at baseline (overestimating the benefit from the education programme alone); Group 1 Number missing: -, Reason: Reports that 4 people withdrew from the study, but doesn't state if this was before randomisation and which groups they were assigned to if it was after randomisation; Group 2 Number missing: -

Protocol outcome 3: Pain at >3 months - Actual outcome: Pain Rating Index-Total of the McGill Pain questionnaire at 16 weeks; Group 1: mean 19.38 (SD 13.66); n=20, Group 2: mean 10.44 (SD 5.25); n=18; McGill Pain questionnaire 0-78 Top=High is poor outcome; Comments: Baseline treatment package: 19.68 (11.03). Baseline education: 14.00 (10.32) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in outcome at baseline (overestimating the benefit from the education programme alone); Group 1 Number missing: -, Reason: Reports that 4 people withdrew from the study, but doesn't state if this was before randomisation and which groups they were assigned to if it was after randomisation; Group 2 Number missing: -	
Protocol outcomes not reported by the study	Quality of life at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study	HOPE trial: Bennell 2018[46]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=144)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks of intervention, 52 weeks of follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Hip osteoarthritis with hip pain for at least 3 months on most days of the past month
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age at least 50 years; hip pain for >3 months on most days of the past month; average hip pain during walking at least 4 on a 11-point NRS in the previous week; able to attend a trial physiotherapy clinic; computer/internet access; can commit to be involved in the study for 12 months; could read/understand English
Exclusion criteria	Hip joint replacement on symptomatic side; awaiting joint replacement surgery within 12 months or any knee surgery in the previous 12 months; use of oral or intraarticular corticosteroids in past 3 months; use of oral or intraarticular corticosteroids in past 3 months; systemic arthritic condition; cognitive behavioral treatment for pain in the past 12 months; physiotherapy treatment or exercises for the back, hip or knee in past 6 months; any other muscular, joint or neurological condition affecting lower limb function; a score >21 on depression subscale of the Depression, Anxiety and Stress Scale
Recruitment/selection of patients	Community-dwelling people from Victoria and Queensland, Australia, between March 2014 and April 2015 through print, radio, social media, medical practitioners and their volunteer database
Age, gender and ethnicity	Age - Mean (SD): 61.3 (7.2). Gender (M:F): 62:82. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms: <2 years to >10 years, median 2-10 years.
Indirectness of population	No indirectness
Interventions	(n=73) Intervention 1: Treatment package - Exercise and behaviour change intervention. Pain coping skills training in the form of eight 35- to 45-minute modules at

	a rate of 1 per week and to practice skills daily. Modules included progressive muscle relaxation, brief relaxation practices, activity-rest cycling, pleasant activity scheduling, cognitive restructuring, pleasant imagery, distraction techniques and problem solving. Between weeks 8 and 24 people undertook a home-based exercise program 3 times per week. People attended 5 face-to-face 30-minute individual sessions with a physiotherapist. A physiotherapist prescribed an individualized exercise program designed to strengthen lower limb muscles and increase hip joint range or motion. Programs contained a quadriceps strengthening, a hip abductor strengthening and a hip stretch/flexibility exercise as well as 2 to 3 other exercises chosen at the physiotherapists' discretion. Duration 24 weeks. Concurrent medication/care: All people received 8 information sheets (covering arthritis, osteoarthritis, managing pain, physical activity, saving energy, health eating, emotions and tips for hip osteoarthritis) produced by Arthritis Australia. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Pain coping skills training). 2. Length of package: > 6 weeks (24 weeks). (n=71) Intervention 2: Non-combined active treatment - Exercise. Exercise training only. Duration 24 weeks. Concurrent medication/care: All people received 8 information sheets (covering arthritis, osteoarthritis, managing pain, physical activity, saving energy, health eating, emotions and tips for hip osteoarthritis) produced by Arthritis Australia. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (24 weeks).
Funding	Academic or government funding (The trial was funded by the National Health and Medical Research Council Program grant (#631717). The funders had no role int he study other than to provide funded. The PCST program was co-developed by 2 of the investigators with funding from the National Institute of Arthritis and musculoskeletal and Skin Diseases, part of the United States National Institutes of Health (award no. R01 AR057346).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: AQoL II at 52 weeks; Group 1: mean 0.02 (SD 0.13); n=73, Group 2: mean 0.02 (SD 0.13); n=71; AQoL II -0.04-1 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: 0.02 (0.05 to -0.01). Reported exercise: 0.02 (0.05 to -0.01). Baseline treatment package: 0.8 (0.1). Baseline exercise: 0.8 (0.1).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 52 weeks; Group 1: mean -2.9 (SD 4.6); n=73, Group 2: mean -3.3 (SD 5.4); n=71; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -2.9 (-1.9 to -4.0). Reported exercise: -3.3 (-2.0 to -4.5). Baseline treatment package: 8.7 (2.9). Baseline exercise: 8.3 (2.8).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined

Protocol outcome 3: Physical function at >3 months
- Actual outcome: WOMAC function at 52 weeks; Group 1: mean -8.8 (SD 16.4); n=73, Group 2: mean -10.7 (SD 17); n=71; WOMAC function 0-68 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -8.8 (-5.1 to -12.6). Reported exercise: -10.7 (-6.8 to -14.7). Baseline treatment package: 27.9 (10.5). Baseline exercise: 26.4 (11.2).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined

Protocol outcome 4: Psychological distress at >3 months
- Actual outcome: DASS anxiety at 52 weeks; Group 1: mean -0.2 (SD 1.1); n=73, Group 2: mean -0.1 (SD 1.3); n=71; DASS anxiety 0-42 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -0.2 (0 to -0.5). Reported exercise: -0.1 (0.2 to -0.4). Baseline treatment package: 2.9 (3.8). Baseline exercise: 2.8 (3.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined
- Actual outcome: DASS depression at 52 weeks; Group 1: mean -0.2 (SD 1.5); n=73, Group 2: mean -0.1 (SD 1.5); n=71; DASS depression 0-42 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -0.2 (0.2 to -0.5). Reported exercise: -0.1 (0.2 to -0.5). Baseline treatment package: 2.9 (3.8). Baseline exercise: 2.8 (3.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined - Actual outcome: DASS stress at 52 weeks; Group 1: mean -0.4 (SD 1.5); n=73, Group 2: mean -0.2 (SD 1.5); n=71; DASS stress 0-42 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: -0.4 (0 to -0.7). Reported exercise: -0.2 (0.2 to -0.5). Baseline treatment package: 2.9 (3.8). Baseline exercise: 2.8 (3.4). Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined Protocol outcome 5: Discontinuation at >3 months - Actual outcome: Lost to assessment at 52 weeks; Group 1: 8/73, Group 2: 10/71; Comments: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined. Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, BMI, bilateral hip involvement, symptom duration, level of education, employment status, current drug/supplement use, treatment expectation and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 6 unable to contact, 1 health issue, 2 changed circumstances, 1 increased hip pain, 2 rejoined.; Group 2 Number missing: 12, Reason: Exercise: 1 health issue, 8 unable to contact, 1 not suitable for research, 1 increased pain, 1 rejoined	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Hopman-rock 2000[125]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=105)
Countries and setting	Conducted in Netherlands; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks with a total of 6 months follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Radiographs of the hips and knees confirming osteoarthritis of Kellgren Grade at least 2. Following the classification criteria of the American College of Rheumatology.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Self-reported osteoarthritis of the knee or hip (to be confirmed later by radiographic and/or clinical criteria of the American College of Rheumatology) and age 55 to 75 years
Exclusion criteria	People who were on the waiting list for knee or hip replacement
Recruitment/selection of patients	No additional information
Age, gender and ethnicity	Age - Mean (SD): 65.3 (5.5). Gender (M:F): 18:87. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: High morbidity score (Number of other chronic conditions (mean [SD]): 2.5 (1.6)). 4. Site of osteoarthritis: Mixed (Hip and/or knee).
Extra comments	Severity: Kellgren Lawrence score of at least 2 in 795 of people Duration of symptoms: <1 year to >20 years, median 3-10 years
Indirectness of population	No indirectness
Interventions	(n=56) Intervention 1: Treatment package - Exercise and education programme. 6 weekly sessions of an education program with an exercise component. Each session was 2 hours in duration. The first hour was guided by a peer educator and the following topics were discussed: pathophysiology of osteoarthritis, lifestyle and physical activity, pain management, the importance of weight reduction and diet, ergonomic aspects, and medical aspects of osteoarthritis (treatments, radiographs). Additionally, questions were answered by an invited occupational therapist and general practitioner. The second hour was an exercise program directed by a physical therapist. Fifteen minutes was spent on education about the balance between rest and activity, preferable types of activity and how to incorporate them in a daily lifestyle, and

	practical advice on physical activity, such as the benefits of walking. People learned the exercises of the program, which consisted of warming up exercises, exercises for the knee and hip (independently of the site of major pain), and cooling down including relaxation exercises. All exercises were performed with the help of a chair, and alternatives were offered to participants who preferred to remain seated. Dynamic exercises were alternated with static exercises and a standard resistance protocol. All educational information, addresses of relevant organisations and the whole exercise program were written up in a course book for participants was provided. People were advised to exercise at home at least 3 times a week. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (6 weeks). (n=49) Intervention 2: Standard care (non-organised) or no treatment - No treatment. No treatment. Education booklets (and a gift voucher) were given after the follow up ended. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Academic or government funding (Supported by a grant from The Netherlands Health Research and Development Council)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus NO TREATMENT

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: VAS quality of life at 6 weeks; Group 1: mean 60.6 (SD 19.6); n=56, Group 2: mean 53.9 (SD 18); n=49; VAS quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 60.3 (19.2). Baseline no treatment: 59.6 (15.3).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, marital status, education, BMI, Kellgren score in hip or knee of at least 2, number of other chronic conditions, duration of complaints and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: VAS quality of life at 6 months; Group 1: mean 54.8 (SD 20.2); n=56, Group 2: mean 55.7 (SD 16.5); n=49; VAS quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 60.3 (19.2). Baseline no treatment: 59.6 (15.3).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, marital status, education, BMI, Kellgren score in hip or knee of at least 2, number of other chronic conditions, duration of complaints and baseline values of outcomes; Group 1 Number

missing: 0; Group 2 Number missing: 0 Protocol outcome 3: Pain at ≤3 months - Actual outcome: VAS pain at 6 weeks; Group 1: mean 27.2 (SD 21.4); n=56, Group 2: mean 25.2 (SD 23.5); n=49; VAS pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 33.0 (22.4). Baseline no treatment: 29.4 (20.0). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, marital status, education, BMI, Kellgren score in hip or knee of at least 2, number of other chronic conditions, duration of complaints and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 4: Pain at >3 months - Actual outcome: VAS pain at 6 months; Group 1: mean 34.7 (SD 20.8); n=56, Group 2: mean 37.9 (SD 20.3); n=49; VAS pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 33.0 (22.4). Baseline no treatment: 29.4 (20.0). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, marital status, education, BMI, Kellgren score in hip or knee of at least 2, number of other chronic conditions, duration of complaints and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study	**Hsu 2021[126]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=63)
Countries and setting	Conducted in Taiwan; Setting: Outpatient follow up

Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee osteoarthritis diagnosed when x-ray findings indicated a Kellgren and Lawrence grade of no more than 3 and visual analog scale at least 4 out of 10.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Older than 55 years and a body mass index of 27-35 kg/m². Obesity per the definition established by the National Health Agency. Knee osteoarthritis diagnosed when x-ray findings indicated a Kellgren and Lawrence grade of no more than 3 and visual analog scale at least 4 out of 10.
Exclusion criteria	Inability to live independently; Kellgren and Lawrence grade >3; history of hip or knee replacement surgery; history of myocardial infarction; Kellgren and Lawrence grade >3; history of hip or knee replacement surgery; history of myocardial infarction; pregnancy or lactation; physical function testing due to conditions such as unstable angina, myocardial infarction, heart failure, severe heart rhythm disorder or second- or third-degree heart conduction block, cardiac aneurysm or aortic aneurysm, or myocarditis or pericarditis, chronic obstructive pulmonary disease accompanied by pulmonary heart disease, untreated or unstable asthma, severe pulmonary hypertension or pulmonary embolism; malignant hypertension.
Recruitment/selection of patients	The study was conducted at Kaohsiung Chang Gung Memorial Hospital.
Age, gender and ethnicity	Age - Mean (SD): 65.3 (4.0). Gender (M:F): 23:40. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren Lawrence grade (mean [SD]): 1.73 (0.78) (grades I-III) Duration of symptoms: Not stated/unclear
Indirectness of population	No indirectness
Interventions	(n=22) Intervention 1: Treatment package - Exercise and behaviour change intervention. Both the diet control and the elastic band resistance program interventions. The diet control consisted of dietary advice (from the clinical dietitian), health education and manuals and handouts during their first visit to the medical center. Each participant was asked to follow a balanced low-energy diet of 1200 kcal/day and update their record sheet at least three times a week. The clinical dietitian followed up with and advised the participants through active phone calls or a communication application once a week for 12 weeks. While performing active calls or mobile application, patinet's interventions were actively instructed by the clinical dietitian based on the individual's nutritional needs and preferences of each participant. The exercise involved an elastic band resistance exercise. This incorporated seated, open-chain exercises to strengthen the major muscle groups of the lower extremities. The exercise regime included hip joint extension/flexion, abduction/adduction, external/internal rotation, knee joint extension/flexion and ankle joint plantarflexion/dorsiflexion movements. Each participant

	performed 10 repetitions/set of five sets/day of the aforementioned exercise movements 3 days a week for 12 weeks. Exercise intensity was increased by applying more force to the band to provide greater resistance or by switching to a thicker resistance band that created more resistance. A repetition maximum of 10 was used. Each movement was taught by clinical staff with further instruction using telemedicine. Thereafter, compliance was tracked and instruction provided by clinical staff once every week through active phone calls or a communication application for 12 weeks. In addition they were provided brochures with highlighted notes that served as reminders.. Duration 12 weeks. Concurrent medication/care: All people continued their previous therapies.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (12 weeks). (n=22) Intervention 2: Non-combined active treatment - Exercise. Exercise only. Duration 12 weeks. Concurrent medication/care: All people continued their previous therapies.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (12 weeks). (n=22) Intervention 3: Non-combined active treatment - Behaviour change intervention. Dietary advice intervention only. Duration 12 weeks. Concurrent medication/care: All people continued their previous therapies.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (This research was supported by the Department of Medical Research, Kaohsiung Chang Gung Memorial Hospital (Project Number: BMRPG9H0581). This work was in part supported by the NSYSU-KMU joint research project (NSYSUKMU110-P004).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain at 12 weeks; Group 1: mean -2.95 (SD 1.12); n=21, Group 2: mean -1.9 (SD 1.48); n=21; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 7.43 (2.01). Baseline exercise: 6.05 (1.99).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, body height, weight and body mass index, body fat percentage, waistline size, Kellgren-Lawrence grade, drugs, baseline VAS and baseline outcome values; Group 1 Number missing: 1, Reason: Treatment package: 1 family refused.; Group 2 Number missing: 1, Reason: Exercise: 1 loss of contact.

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: WOMAC function at 12 weeks; Group 1: mean -8.62 (SD 3.58); n=21, Group 2: mean -5.1 (SD 1.7); n=21; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 28.57 (5.76). Baseline exercise: 22.86 (4.30).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, body height, weight and

body mass index, body fat percentage, waistline size, Kellgren-Lawrence grade, drugs, baseline VAS and baseline outcome values; Group 1 Number missing: 1, Reason: Treatment package: 1 family refused.; Group 2 Number missing: 1, Reason: Exercise: 1 loss of contact.

Protocol outcome 3: Discontinuation at ≤3 months
- Actual outcome: Discontinuation at 12 weeks; Group 1: 1/22, Group 2: 1/22; Comments: Treatment package: 1 family refused. Exercise: 1 loss of contact.
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, body height, weight and body mass index, body fat percentage, waistline size, Kellgren-Lawrence grade, drugs, baseline VAS and baseline outcome values; Group 1 Number missing: 1, Reason: Treatment package: 1 family refused.; Group 2 Number missing: 1, Reason: Exercise: 1 loss of contact.

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain at 12 weeks; Group 1: mean -2.95 (SD 1.12); n=21, Group 2: mean -2.14 (SD 1.28); n=21; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 7.43 (2.01). Baseline diet control: 6.48 (2.21).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, body height, weight and body mass index, body fat percentage, waistline size, Kellgren-Lawrence grade, drugs, baseline VAS and baseline outcome values; Group 1 Number missing: 1, Reason: Treatment package: 1 family refused.; Group 2 Number missing: 1, Reason: Diet control: 1 go abroad for half a month.

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: WOMAC function at 12 weeks; Group 1: mean -8.62 (SD 3.58); n=21, Group 2: mean -5.76 (SD 2.84); n=21; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 28.57 (5.76). Baseline diet control: 25.19 (5.62).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, body height, weight and body mass index, body fat percentage, waistline size, Kellgren-Lawrence grade, drugs, baseline VAS and baseline outcome values; Group 1 Number missing: 1, Reason: Treatment package: 1 family refused.; Group 2 Number missing: 1, Reason: Diet control: 1 go abroad for half a month.

Protocol outcome 3: Discontinuation at ≤3 months
- Actual outcome: Discontinuation at 12 weeks; Group 1: 1/22, Group 2: 1/22; Comments: Treatment package: 1 family refused. Diet control: 1 go abroad for half a month.
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, body height, weight and body mass index, body fat percentage, waistline size, Kellgren-Lawrence grade, drugs, baseline VAS and baseline outcome values; Group 1 Number missing: 1, Reason: Treatment package: 1 family refused.; Group 2 Number missing: 1, Reason: Diet control: 1 go abroad for half a month.

Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Huang 2000[127]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=126)
Countries and setting	Conducted in China; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People with osteoarthritis stage 2-4 according to the Altman criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with osteoarthritis of the knees and BMI >25 (in males) or >30 (in females). Altman grade 2-4.
Exclusion criteria	People who had different degrees of severity of osteoarthritis in bilateral knees.
Recruitment/selection of patients	Outpatient follow up
Age, gender and ethnicity	Age - Other: Mean: 54.8 years. Gender (M:F): 14:112. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Altman grade 2-4 Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=42) Intervention 1: Treatment package - Combination and behaviour change intervention. Weight reduction therapy, with auricular acupuncture and exercise, and electrotherapy. Auricular acupuncture was performed using specific auricle points, inserting 2mm stainless press needles alternating between the two auricles. Acupuncture was performed once a week with 8 implantations in each course. Diet control was supported through a counseling session regarding their body parameters. People were asked to keep detailed records of their food intake. In the second session they analysed these results and provided advise to reduce the number of calories required (500kcal/day less than the amount required by calculation). The calories were split into 3 components: 15-20% protein, 25% fat, and 55-60% carbohydrates. Aerobic exercise was achieved through an ergonomic bicycle. The pedal rate was typically 60 revolutions per minute for untrained cyclists. The aim was to achieve a heart rate under 60% of the maximal oxygen consumption level for the home program. 3

	sessions weekly was suggested. Electrotherapy included ultrasound and TENS treatment for pain relief as the modalities used at the rehabilitation department. Ultrasound was performed at a frequency of 1 MHz, and a spatial and temporal peak intensity of 2.5 W/cm². ultrasound was pulsed at a duty cycle of 20% for 3 minutes at each position. The TENS was applied with dense-disperse wave to relieve pain, and the TENS pads were applied over the local tender points for 15 minutes in each treatment. Each course of treatment consisted of 3 treatments per week for 12 weeks. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Weight loss). 2. Length of package: > 6 weeks (12 weeks). (n=42) Intervention 2: Non-combined active treatment - Electrotherapy. Electrotherapy only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks). (n=42) Intervention 3: Treatment package - Combination and behaviour change intervention. Only diet, exercise and acupuncture. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: 2. Length of package: Comments: This group was not included as there were no valid comparisons that this fell into in the protocol
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND BEHAVIOUR CHANGE INTERVENTION versus ELECTROTHERAPY

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: VAS at 12 weeks; Group 1: mean -4 (SD 2); n=42, Group 2: mean -1.9 (SD 1.7); n=42; VAS 0-10 Top=High is poor outcome; Comments: Reports results subgrouped by radiographic severity. Values combined to calculate the overall value for each group. Reported treatment package (II): -2.9 (1.7). Reported treatment package (III): -4.5 (1.7). Reported treatment package (IV): -5.1 (2.0). Reported electrotherapy (II): -1.6 (1.4). Reported electrotherapy (III): -1.8 (2.1). Reported electrotherapy (IV): -2.7 (1.1). Baseline treatment package (II): 4.6 (1.3). Baseline treatment package (III): 7.0 (1.4). Baseline treatment package (IV): 8.7 (2.0). Baseline electrotherapy (II): 4.5 (1.0). Baseline electrotherapy (III): 6.7 (1.2). Baseline electrotherapy (IV): 8.3 (1.8).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported baseline values of outcomes. Generally very limited reporting.; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study (subsidiary papers)	Hughes 2004[130] (Hughes 2010[131])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=150)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks of intervention, 6 months follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee osteoarthritis with at least 3 of the following 6: age >60 years, morning stiffness with a duration <30 minutes, crepitus on active motion, tenderness of the bony margins of the joint, bony enlargement on examination, a lack of palpable warmth of the synovium. Hip osteoarthritis if pain is present in combination with either: hip internal rotation at least 15 degrees, pain present on internal rotation of the hip, morning stiffness of the hip for a time no more than 60 minutes, and age <60 years or; hip internal rotation <15 degrees, and hip flexion at least 115 degrees.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Older people with mild to moderate lower extremity osteoarthritis
Exclusion criteria	People with severe, limiting cardiovascular disease, active thrombophlebitis, recent pulmonary embolus, an acute systemic illness, poorly controlled diabetes and other health conditions that might preclude exercise training
Recruitment/selection of patients	Conducted at several different senior centers and senior housing residences with volunteers being recruited by newsletter, announcements in the local media and presentations to local senior groups
Age, gender and ethnicity	Age - Mean (SD): 73.6 (6.6). Gender (M:F): 24:126. Ethnicity: White-Caucasian = 123, African American = 19, Hispanic = 4, Other = 1
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Low morbidity score (Unclear. Around 60% had a cardiovascular disease, 5.5% had an asthma, 4% had emphysema, 11% had diabetes, 5% had cancer). 4. Site of osteoarthritis: Mixed (Hip or knee).
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness

Interventions	(n=80) Intervention 1: Treatment package - Exercise and behaviour change intervention. Fit & Strong intervention. 90 minute sessions held three times per week for 8 weeks. The first 60 minutes included both resistance training and fitness walking. The last 30 minutes included an adapted version of a group discussion educational component. All exercises were accompanied by music. Strengthening exercises for the lower extremities and trunk utilized a graded task-specific approach (sit to stand and postural stabilisation) using weights to add progression. Fitness walking progressed for a maximum duration at baseline to 40 minutes over time with an exercise intensity of 40-60% of maximum heart rate. The education/behaviour change component used social cognitive theory to increase individuals' confidence in their ability to achieve a desired outcome. The health education discussed the efficacy of exercise, but also the ability of the person to manage their own pain and other arthritis-related symptoms. People were asked to identify specific tasks they could do and how they were going to achieve them (and how exercise would help). Staff helped to reinforce these ideas as people were encouraged to complete a home based exercise program of their design to help with their symptoms. Duration 8 weeks. Concurrent medication/care: All people were given a copy of 'The Arthritis Helpbook' includingself-care materials and hand-outs.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Fit & Strong intervention). 2. Length of package: > 6 weeks (8 weeks). (n=70) Intervention 2: Standard care (non-organised) or no treatment - No treatment. No additional treatment. Duration 8 weeks. Concurrent medication/care: All people were given a copy of 'The Arthritis Helpbook' including self-care materials and hand-outs.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (8 weeks).
Funding	Academic or government funding (The program was developed using a grant from the Chicago Chapter of the Arthritis Foundation. The research was also supported by funding from the National Institute on Arthritis and Musculoskeletal Disease (Grant AR30692) and by the National Institute on Ageing and the Roybal Center for Research on Applied Gerontology (Grant AG 15890).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus NO TREATMENT

Protocol outcome 1: Pain at ≤3 months

- Actual outcome: WOMAC pain at 2 months; Group 1: mean 4.9 (SD 3.4); n=68, Group 2: mean 6.2 (SD 3.4); n=43; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.9 (3.9). Baseline control: 6.5 (3.9).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, educaion, income, race, ACR class, comorbid conditions and baseline values of outcomes; Group 1 Number missing: 12, Reason: No reasons given; Group 2 Number missing: 27, Reason: No reasons given

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 6 months; Group 1: mean 5.1 (SD 3.7); n=60, Group 2: mean 6.7 (SD 3.9); n=36; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.9 (3.9). Baseline control: 6.5 (3.9).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, educaion, income, race, ACR class, comorbid conditions and baseline values of outcomes; Group 1 Number missing: 20, Reason: No reasons given; Group 2 Number missing: 34, Reason: No reasons given

Protocol outcome 3: Physical function at ≤3 months
- Actual outcome: WOMAC physical function at 2 months; Group 1: mean 17.3 (SD 12.6); n=68, Group 2: mean 22.3 (SD 12.8); n=43; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 21.1 (11.9). Baseline control: 25.0 (13.9).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, educaion, income, race, ACR class, comorbid conditions and baseline values of outcomes; Group 1 Number missing: 12, Reason: No reasons given; Group 2 Number missing: 27, Reason: No reasons given

Protocol outcome 4: Physical function at >3 months
- Actual outcome: WOMAC physical function at 6 months; Group 1: mean 18.3 (SD 12.6); n=60, Group 2: mean 24.1 (SD 14.6); n=36; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 21.1 (11.9). Baseline control: 25.0 (13.9).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, educaion, income, race, ACR class, comorbid conditions and baseline values of outcomes; Group 1 Number missing: 20, Reason: No reasons given; Group 2 Number missing: 34, Reason: No reasons given

Protocol outcome 5: Discontinuation at ≤3 months
- Actual outcome: Discontinuation at 8 weeks; Group 1: 12/80, Group 2: 27/70; Comments: No reasons given
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, educaion, income, race, ACR class, comorbid conditions and baseline values of outcomes; Group 1 Number missing: 12, Reason: No reasons given; Group 2 Number missing: 27, Reason: No reasons given

Protocol outcome 6: Discontinuation at >3 months - Actual outcome: Discontinuation at 6 months; Group 1: 20/80, Group 2: 34/70; Comments: No reasons given Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, educaion, income, race, ACR class, comorbid conditions and baseline values of outcomes; Group 1 Number missing: 20, Reason: No reasons given; Group 2 Number missing: 34, Reason: No reasons given	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Hughes 2006[132]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=215)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: Treatment for 8 weeks, total follow up 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Osteoarthritis of the hip or knee as per a modified version of the American College of Rheumatology functional classes
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Knee osteoarthritis with at least 3 of the following 6: age >60 years, morning stiffness with a duration <30 minutes, crepitus on active motion, tenderness of the bony margins of the joint, bony enlargement on examination, a lack of palpable warmth of the synovium. Hip osteoarthritis if pain is present in combination with either: hip internal rotation at least 15 degrees, pain present on internal rotation of the hip, morning stiffness of the hip for a time no more than 60 minutes, and age <60 years or; hip internal rotation <15 degrees, and hip flexion at least 115 degrees.
Exclusion criteria	People with severe, limiting cardiovascular disease, active thrombophlebitis, recent pulmonary embolus, an acute systemic illness, poorly controlled diabetes, and other health conditions that might preclude exercise training
Recruitment/selection of patients	People were community dwelling older adults who were recruited by newsletter, announcements in the local media, and presentations to local senior groups
Age, gender and ethnicity	Age - Other: Mean: 73.3. Gender (M:F): 36:179. Ethnicity: White-Caucasian = 155, African American = 48, Hispanic = 5, Asian-Pacific Islander = 4, Other = 2
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Low morbidity score (Around 60% had hypertension, around 44% had cardiovascular disease, around 6.5% had asthma, around 4% had emphysema, around 13% had diabetes, around 4% had cancer). 4. Site of osteoarthritis: Mixed (Hip or knee).
Extra comments	Severity: American Rheumatism Association classes I-III, median class II Duration of symptoms: Not stated
Indirectness of population	No indirectness

Interventions	(n=115) Intervention 1: Treatment package - Exercise and behaviour change intervention. Fit & Strong intervention. 90 minute sessions held three times per week for 8 weeks. The first 60 minutes included both resistance training and fitness walking. The last 30 minutes included an adapted version of a group discussion educational component. All exercises were accompanied by music. Strengthening exercises for the lower extremities and trunk utilized a graded task-specific approach (sit to stand and postural stabilisation) using weights to add progression. Fitness walking progressed for a maximum duration at baseline to 40 minutes over time with an exercise intensity of 40-60% of maximum heart rate. The education/behaviour change component used social cognitive theory to increase individuals' confidence in their ability to achieve a desired outcome. The health education discussed the efficacy of exercise, but also the ability of the person to manage their own pain and other arthritis-related symptoms. People were asked to identify specific tasks they could do and how they were going to achieve them (and how exercise would help). Staff helped to reinforce these ideas as people were encouraged to complete a home based exercise program of their design to help with their symptoms. Duration 8 weeks. Concurrent medication/care: All people were given a copy of 'The Arthritis Helpbook' including self-care materials and hand-outs. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Fit & Strong intervention). 2. Length of package: > 6 weeks (8 weeks). (n=100) Intervention 2: Standard care (non-organised) or no treatment - No treatment. No additional treatment. Duration 8 weeks. Concurrent medication/care: All people were given a copy of 'The Arthritis Helpbook' including self-care materials and hand-outs. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (8 weeks).
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus NO TREATMENT

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: Geri-AIMS pain subscale at 2 months; Group 1: mean 4.67 (SD 0.85); n=115, Group 2: mean 4.65 (SD 0.86); n=100; Comments: Baseline treatment package: 4.58 (0.94). Baseline control: 4.64 (0.95).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race,

ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 32, Reason: Reasons not given; Group 2 Number missing: 45, Reason: Reasons not given

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: Geri-AIMS pain subscale at 12 months; Group 1: mean 4.77 (SD 0.82); n=115, Group 2: mean 4.61 (SD 0.91); n=100; Geri-AIMS pain subscale 0-10 Top=High is good outcome; Comments: Baseline treatment package: 4.58 (0.94). Baseline control: 4.64 (0.95).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 57, Reason: Reasons not given; Group 2 Number missing: 68, Reason: Reasons not given

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: WOMAC pain at 2 months; Group 1: mean 4.89 (SD 3.53); n=115, Group 2: mean 6.45 (SD 4.01); n=100; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 6.32 (3.84). Baseline control: 7.04 (3.84).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 32, Reason: Reasons not given; Group 2 Number missing: 45, Reason: Reasons not given

Protocol outcome 4: Pain at >3 months
- Actual outcome: WOMAC pain at 12 months; Group 1: mean 5.39 (SD 3.72); n=115, Group 2: mean 5.31 (SD 4.42); n=100; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 6.32 (3.84). Baseline control: 7.04 (3.84).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 57, Reason: Reasons not given; Group 2 Number missing: 68, Reason: Reasons not given

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: WOMAC physical function at 2 months; Group 1: mean 17.45 (SD 12.25); n=115, Group 2: mean 22.57 (SD 12.21); n=100; WOMAC physical function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 22.68 (11.75). Baseline control: 27.11 (14.24).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 32, Reason: Reasons not given; Group 2 Number missing: 45, Reason: Reasons not given

Protocol outcome 6: Physical function at >3 months
- Actual outcome: WOMAC physical function at 12 months; Group 1: mean 17.81 (SD 11.15); n=115, Group 2: mean 20.15 (SD 14.71); n=100; WOMAC physical function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 22.68 (11.75). Baseline control: 27.11 (14.24).

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 57, Reason: Reasons not given; Group 2 Number missing: 68, Reason: Reasons not given Protocol outcome 7: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 2 months; Group 1: 32/115, Group 2: 45/100; Comments: Reasons not given Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 32, Reason: Reasons not given; Group 2 Number missing: 45, Reason: Reasons not given Protocol outcome 8: Discontinuation at >3 months - Actual outcome: Discontinuation at 12 months; Group 1: 57/115, Group 2: 68/100; Comments: Reasons not given Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, education, income, race, ARA class, comorbid conditions, and baseline values of outcomes; Group 1 Number missing: 57, Reason: Reasons not given; Group 2 Number missing: 68, Reason: Reasons not given	
Protocol outcomes not reported by the study	Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study (subsidiary papers)	Hurley 2007[138] (Hurley 2012[137], Hurley 2007[139])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=418)
Countries and setting	Conducted in United Kingdom; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks of treatment, 30 months of follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People with chronic knee pain of mild, moderate or severe magnitude for more than 6 months
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People age 50 years or older who had consulted a primary care physician for mild, moderate or severe knee pain of >6 months' duration (many people were labeled as osteoarthritis based on their clinical presentation and history). People were not excluded if they used assistive walking devices; had stable comorbidities common in this age group (e.g. type II diabetes, cardiovascular or respiratory disorders); or had back, lower or upper limb pain
Exclusion criteria	Lower limb arthroplasty; physiotherapy for knee pain in the preceding 6 months; unstable medical conditions; inability/unwillingness to exercise; wheelchair dependence; inability to understand English
Recruitment/selection of patients	People were recruited form their primary care phyisican. Primary care practices were the unit of randomisation.
Age, gender and ethnicity	Age - Mean (range): 67 (50-91). Gender (M:F): 124:294. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms (median [IQR]): Usual care: 6 (3-15), Individual rehab: 7 (3-15), Group rehab: 5 (2.5-11).
Indirectness of population	Serious indirectness: Chronic knee pain (no clear statement about the presence of osteoarthritis)
Interventions	(n=278) Intervention 1: Treatment package - Exercise and behaviour change intervention. ESCAPE-knee pain program. Comprised of integrated patient education, with simple self-management and pain coping strategies, delivered in the first 15-20 minutes of each rehabilitation session followed by 35-45 minutes of individualized

	progressive exercise programs. The content of the self-management, coping and education settings included goal setting, pacing and activity-rest cycling, drug management and action plan review, diet and healthy eating, intermediate home exercise regimen and program review, pain gate and review of action plans, managing flares in pain, advanced home exercise regimen and reviewing action plans, mini-relaxation and deep breath techniques and information regarding pursuing activity and exercise in the community. Exercises focused on strength, balance, coordination, control, endurance and function. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Education/behaviour change intervention, but the majority of the education was behaviour change themed). 2. Length of package: ≤ 6 weeks (6 weeks). Comments: This group included two groups that were combined (one looking at individual rehabilitation and one looking at group rehabilitation). These were combined due to class effect. (n=140) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual primary care. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Other author(s) funded by industry (Dr Hurley's work was supported by an Arthritis Research Campaign Research Fellowship. Dr Jones has received consultancies (less than $10,000) from AstraZeneca)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: EQ-5D at 6 months; Group 1: mean 0.64 (SD 0.27); n=229, Group 2: mean 0.66 (SD 0.3); n=140; EQ-5D -0.11-1 Top=High is good outcome; Comments: Reported adjusted (by baseline values) mean (final values) and 95% confidence intervals. Reported treatment package: 0.64 (0.61, 0.68). Reported usual care: 0.66 (0.60, 0.71). Baseline group rehab: 0.60 (0.30). Baseline individual rehab: 0.59 (0.28). Baseline usual care: 0.60 (0.32). Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 51, Reason: Reasons not given; Group 2 Number missing: 30, Reason: Reasons not given

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: WOMAC pain at 6 weeks; Group 1: mean 5.2 (SD 1.7); n=237, Group 2: mean 7.1 (SD 1.8); n=128; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 7.5 (1.7). Baseline usual care: 7.7 (1.7).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 40, Reason: Reasons not given; Group 2 Number missing: 13, Reason: Reasons not given

Protocol outcome 3: Pain at >3 months
- Actual outcome: WOMAC pain at 30 months; Group 1: mean 5.9 (SD 2.6); n=189, Group 2: mean 6.4 (SD 2); n=94; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 7.5 (1.7). Baseline usual care: 7.7 (1.7).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 89, Reason: Reasons not given; Group 2 Number missing: 46, Reason: Reasons not given

Protocol outcome 4: Physical function at ≤3 months
- Actual outcome: WOMAC physical function at 6 weeks; Group 1: mean 20 (SD 5.9); n=237, Group 2: mean 25.9 (SD 6.3); n=140; WOMAC physical function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 27.1 (6.7). Baseline usual care: 27.2 (7.0).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 40, Reason: Reasons not given; Group 2 Number missing: 13, Reason: Reasons not given

Protocol outcome 5: Physical function at >3 months
- Actual outcome: WOMAC physical function at 30 months; Group 1: mean 22.3 (SD 8.7); n=189, Group 2: mean 23.8 (SD 6.3); n=94; WOMAC physical function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 27.1 (6.7). Baseline usual care: 27.2 (7.0).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 89, Reason: Reasons not given; Group 2 Number missing: 46, Reason: Reasons not given

Protocol outcome 6: Psychological distress at >3 months
- Actual outcome: HADS anxiety at 6 months; Group 1: mean 5.97 (SD 3.98); n=229, Group 2: mean 5.32 (SD 1.95); n=113; HADS anxiety 0-21 Top=High is poor outcome; Comments: Reported adjusted (by baseline values) mean (final values) and 95% confidence intervals. Reported treatment package: 5.97 (5.46, 6.49). Reported usual care: 5.32 (4.96, 5.68). Baseline group rehab: 6.6 (4.5). Baseline individual rehab: 6.3 (3.9). Baseline usual care: 6.7 (4.6).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms,

height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 51, Reason: Reasons not given; Group 2 Number missing: 30, Reason: Reasons not given - Actual outcome: HADS depression at 6 months; Group 1: mean 4.28 (SD 3.17); n=229, Group 2: mean 3.93 (SD 1.57); n=113; HADS depression 0-21 Top=High is poor outcome; Comments: Reported adjusted (by baseline values) mean (final values) and 95% confidence intervals. Reported treatment package: 4.28 (3.87, 4.69). Reported usual care: 3.93 (3.64, 4.22). Baseline group rehab: 5.0 (3.4). Baseline individual rehab: 4.5 (3.2). Baseline usual care: 5.1 (3.7). Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 51, Reason: Reasons not given; Group 2 Number missing: 30, Reason: Reasons not given Protocol outcome 7: Discontinuation at ≤3 months - Actual outcome: Lost to follow up at 6 weeks; Group 1: 41/278, Group 2: 12/140; Comments: Reasons not given Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 40, Reason: Reasons not given; Group 2 Number missing: 13, Reason: Reasons not given Protocol outcome 8: Discontinuation at >3 months - Actual outcome: Lost to follow up at 30 months; Group 1: 89/278, Group 2: 46/140; Comments: Reasons not given. Original study stated that at 6 months, only 5 withdrew because of exercise-related adverse events, 3 had exacerbation of pain (2 knee, 1 hip) and 2 with cardiac pacemakers had concerns about exercising, despite reassurance Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, duration of symptoms, height, body mass, body mass index, and baseline values of outcomes; Group 1 Number missing: 89, Reason: Reasons not given; Group 2 Number missing: 46, Reason: Reasons not given	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Psychological distress at ≤3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study (subsidiary papers)	IMPACT trial: Bennell 2017[45] (Lawford 2018[163], Lin 2003[168])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=148)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: Intervention delivered over 3 months, additional follow up for 9 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Chronic knee pain and reduced physical function
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People age 50 years or older; knee pain for more than 3 months and on most days of the previous month; knee pain during walking (score of at least 4 on a 11-point numerical rating scale) in the previous week; mild to moderate physical dysfunction (score >20 out of 68 on the physical function subscale of WOMAC); an active e-mail account and a computer with Internet access
Exclusion criteria	Joint replacement in the symptomatic knee; awaiting joint replacement surgery; intra-articular corticosteroid injection or knee surgery in the previous 6 months or planned joint surgery in the subsequent 9 months; treatment for knee pain or participation in a strengthening exercise of PCST program in the previous 6 months; systemic arthritic condition; neurological condition affecting the lower limb or limiting exercise; pain at another site that was worse than knee pain or limited exercise; high-level depression (score >21 on the depression subscale of the DASS-21).
Recruitment/selection of patients	People from the community in Australia were recruited via print, radio and social media advertisements and their database
Age, gender and ethnicity	Age - Mean (SD): 61.2 (7.1). Gender (M:F): 65:83. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: <2 years - >10 years, median 2-10 years
Indirectness of population	Serious indirectness: Not explicitly stated to be osteoarthritic pain, but fulfills all of the criteria otherwise

Interventions	(n=74) Intervention 1: Treatment package - Exercise and behaviour change intervention. Videoconferencing sessions with a physiotherapist for home exercise and a pain coping skills training program. The pain coping skills training program (PainCOACH) included eight 35- to 45- minute modules that were interactive and automated, and advised to practice pain-coping skills daily (including progressive relaxation, activity-rest cycling, scheduling pleasant activities, changing negative thoughts, pleasant imagery and distraction techniques, and problem solving). The physiotherapy sessions were completed over 12 weeks in 7 sessions (delivered weeks 2, 3, 4, 6, 8, 10 and 12). Sessions lasted 45 minutes in weeks 2 and 12 and 30 minutes in other weeks. The physiotherapist performed a brief assessment and prescribed a lower-limb strengthening home exercise program to be performed 3 times per week. Exercise progression was provided by varying the exercises, repetitions, load or difficulty. People were provided with instructions, video demonstrations and equipment. Duration 12 weeks. Concurrent medication/care: All people had access to internet educational material about exercise and physical activity, pain management, emotions, healthy eating, complementary therapies, and medications (www.arthritisaustralia.com.au) that they were encouraged to access at their leisure. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Pain coping skills training). 2. Length of package: > 6 weeks (12 weeks). (n=74) Intervention 2: Standard care (non-organised) or no treatment - No treatment. No additional treatment (just internet educational material). Duration 12 weeks. Concurrent medication/care: All people had access to internet educational material about exercise and physical activity, pain management, emotions, healthy eating, complementary therapies, and medications (www.arthritisaustralia.com.au) that they were encouraged to access at their leisure. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (This study was funded by the Australian National Health and Medical Research Council (program grant 1091302). Prof. Bennell is supported by a National Health and medical Research Council fellowship (1058440). Dr. Rini received funding from a Multidisciplinary Clinical Research Center funded by the U.S. National Institute of Arthritis and Musculoskeletal and Skin Diseases through the Thurston Arthritis Research Center at the University North Carolina (P60AR064166). Dr Hinman is supported by an Australian Research Council Future

	Fellowship (FT130100175). Dr Abbott was funded by a Sir Charles Hercus Health Research Fellowship from the Health Research Council of New Zealand.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus NO TREATMENT

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AQoL-2 at 3 months; Group 1: mean 0.1 (SD 0.2); n=70, Group 2: mean 0 (SD 0.2); n=69; AQoL-2 -0.04-1 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported intervention: 0.1 (0.1 to 0). Reported control: 0 (0.1 to 0). Baseline intervention: 0.7 (0.2). Baseline control: 0.7 (0.1).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 4, Reason: 3 unable to contact, 1 family issue; Group 2 Number missing: 5, Reason: 4 unable to contact, 1 family illness

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: AQoL-2 at 9 months; Group 1: mean 0.1 (SD 0.2); n=66, Group 2: mean 0 (SD 0); n=67; AQoL-2 -0.04-1 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported intervention: 0.1 (0.1 to 0). Reported control: 0 (0 to 0). Baseline intervention: 0.7 (0.2). Baseline control: 0.7 (0.1).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 8, Reason: 5 unable to contact, 1 family issue, 1 health issue, 1 deceased; Group 2 Number missing: 7, Reason: 4 unable to contact, 1 family illness, 1 deceased, 1 declined

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: WOMAC pain at 3 months; Group 1: mean -3.9 (SD 3.4); n=70, Group 2: mean -1.5 (SD 3.4); n=69; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported intervention: -3.9 (-3.1 to -4.7). Reported control: -1.5 (-0.7 to -2.3). Baseline intervention: 9.0 (2.4). Baseline control: 9.2 (2.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 4, Reason: 3 unable to contact, 1 family issue; Group 2 Number missing: 5, Reason: 4 unable to contact, 1 family illness

Protocol outcome 4: Pain at >3 months
- Actual outcome: WOMAC pain at 9 months; Group 1: mean -3.7 (SD 3.3); n=66, Group 2: mean -2.3 (SD 3.6); n=67; WOMAC pain 0-20 Top=High is poor

outcome; Comments: Reported change scores and 95% confidence intervals. Reported intervention: -3.7 (-2.9 to -4.5). Reported control: -2.3 (-1.4 to -3.1). Baseline intervention: 9.0 (2.4). Baseline control: 9.2 (2.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 8, Reason: 5 unable to contact, 1 family issue, 1 health issue, 1 deceased; Group 2 Number missing: 7, Reason: 4 unable to contact, 1 family illness, 1 deceased, 1 declined

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: WOMAC physical function at 3 months; Group 1: mean -14.4 (SD 11.1); n=70, Group 2: mean -4.9 (SD 11); n=69; WOMAC physical function 0-68 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported intervention: -14.4 (-11.8 to -17.0). Reported control: -4.9 (-2.3 to -7.5). Baseline intervention: 33.1 (8.0). Baseline control: 32.5 (8.3).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 4, Reason: 3 unable to contact, 1 family issue; Group 2 Number missing: 5, Reason: 4 unable to contact, 1 family illness

Protocol outcome 6: Physical function at >3 months
- Actual outcome: WOMAC physical function at 9 months; Group 1: mean -13.9 (SD 11.4); n=69, Group 2: mean -6.6 (SD 11.1); n=67; WOMAC physical function 0-68 Top=High is poor outcome; Comments: Reported change scores and 95% confidence intervals. Reported intervention: -13.9 (-11.2 to -16.6). Reported control: -6.6 (-4.0 to -9.3). Baseline intervention: 33.1 (8.0). Baseline control: 32.5 (8.3).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 8, Reason: 5 unable to contact, 1 family issue, 1 health issue, 1 deceased; Group 2 Number missing: 7, Reason: 4 unable to contact, 1 family illness, 1 deceased, 1 declined

Protocol outcome 7: Discontinuation at ≤3 months
- Actual outcome: Lost to follow-up at 3 months; Group 1: 4/74, Group 2: 5/74; Comments: Treatment package: 3 unable to contact, 1 family issue. Control: 4 unable to contact, 1 family illness.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 4, Reason: 3 unable to contact, 1 family issue; Group 2 Number missing: 5, Reason: 4 unable to contact, 1 family illness

Protocol outcome 8: Discontinuation at >3 months

- Actual outcome: Lost to follow-up at 9 months; Group 1: 8/74, Group 2: 7/74; Comments: Treatment package: 5 unable to contact, 1 family issue, 1 health issue, 1 deceased. Control: 4 unable to contact, 1 family illness, 1 deceased, 1 declined Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, geographic location, height, weight, BmI, symptom duration, level of education, employment status, current drug/supplement use, expectation of treatment outcomes, years using the internet, internet use for social media, including Skype, self-rated ability to use the internet; Group 1 Number missing: 8, Reason: 5 unable to contact, 1 family issue, 1 health issue, 1 deceased; Group 2 Number missing: 7, Reason: 4 unable to contact, 1 family illness, 1 deceased, 1 declined	
Protocol outcomes not reported by the study	Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Isaramalai 2018[141]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=108)
Countries and setting	Conducted in Thailand; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 9 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Symptomatic knee osteoarthritis, as determined by the clinical and radiographic criteria of the American College of Rheumatology and the Kellgren-Lawrence radiographic grading scale (<4)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Para rubber farmers aged at least 60 years who currently had symptomatic knee osteoarthritis, as determined by the clinical and radiographic criteria of the American College of Rheumatology and the Kellgren-Lawrence radiographic grading scale (<4)
Exclusion criteria	People with a history of major knee injury, knee surgery or steroid injection; those with a contraindication to strengthening exercise, such as uncontrolled hypertension, inflamed knee during exercise, cognitive dysfunction, or planning for knee surgery
Recruitment/selection of patients	No additional information
Age, gender and ethnicity	Age - Mean (SD): 66.2 (5.2). Gender (M:F): 17:58. Ethnicity: No additional information
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren and Lawrence grade of knee osteoarthritis 1-3, median grade 2 Duration of symptoms (mean [IQR]): PEM-NEW = 3 (2,5), PEM-PRE = 2 (2,3.3), ST = 3 (2,5).
Indirectness of population	No indirectness
Interventions	(n=63) Intervention 1: Treatment package - Exercise and behaviour change intervention. Exercise by one of two forms: progressive strengthening exercise or non-weight bearing exercise, and a community based education session and behaviour change intervention and follow up home visits to provide extra support. The behaviour change intervention included: a twenty-minute job hazard analysis (discussing ergonomic risk factors in the work process that increase the severity of knee osteoarthritis), a one-hour health education session (20 minutes teaching and 40 minutes exercise demonstration on ergonomic management), and thirty minute mutual

	goal setting (where identified risk factors were then used to make goals and action plans). Home-based interventions were conducted every other week. Self-directed exercise was performed at least 3 days per week for 8 weeks. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (8 weeks). Comments: The two groups were combined due to class effect (n=45) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care only. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (8 weeks).
Funding	Academic or government funding (This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: WOMAC pain subscale at 9 weeks; Group 1: mean 5.28 (SD 5.49); n=50, Group 2: mean 11.72 (SD 6.61); n=25; WOMAC pain 0-20 Top=High is poor outcome; Comments: Reported PEM-NWE = 4.80 (5.71). Reported PEM-PRE = 6.60 (4.41). Baseline PEM-NEW = 13.04 (7.27). Baseline PEM-PRE = 14.48 (12.25). Baseline standard therapy: 15.28 (8.63).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, tapping size, working years, years of pain onset, sex, body mass index, waist circumference, Kellgren and Lawrence grade and baseline values of outcomes; Group 1 Number missing: 13, Reason: Did not receive allocated intervention; Group 2 Number missing: 20, Reason: Did not receive allocated intervention

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: WOMAC function subscale at 9 weeks; Group 1: mean 11.68 (SD 15.36); n=50, Group 2: mean 32.24 (SD 23.16); n=25; WOMAC function 0-68 Top=High is poor outcome; Comments: Reported PEM-NWE = 12.84 (16.01). Reported PEM-PRE = 10.52 (14.58). Baseline PEM-NEW = 33.32 (16.4). Baseline PEM-PRE = 30.12 (23.84). Baseline standard therapy: 37.16 (27.71).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, tapping size, working years, years of pain onset, sex, body mass index, waist circumference, Kellgren and Lawrence grade and baseline values of outcomes; Group 1 Number missing: 13, Reason: Did not receive allocated intervention; Group 2 Number missing: 20, Reason: Did not receive allocated intervention

Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Did not receive allocated intervention at 9 weeks; Group 1: 13/63, Group 2: 20/45; Comments: No reasons given Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, tapping size, working years, years of pain onset, sex, body mass index, waist circumference, Kellgren and Lawrence grade and baseline values of outcomes; Group 1 Number missing: 13, Reason: Did not receive allocated intervention; Group 2 Number missing: 20, Reason: Did not receive allocated intervention	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Jessep 2009[144]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=67)
Countries and setting	Conducted in United Kingdom; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 5 weeks of intervention, 12 months follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Mild, moderate or severe non-specific knee pain lasting more than 6 months with no identifiable recent cause; these people would be diagnosed as having clinical osteoarthritis based on their clinical presentation and history
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Over 50 years of age; had consulted a primary care physician for mild, moderate or severe non-specific knee pain lasting for more than 6 months with no identifiable recent cause (these people would be diagnosed as having clinical osteoarthritis based on their clinical presentation and history)
Exclusion criteria	Knee pain emanating from knee trauma within the past year; lower limb arthroplasty; physiotherapy for knee pain in the preceding 12 months; intra-articular injections in the preceding 6 months; unstable medical or psychological conditions; unable or unwilling to exercise; unable to walk 100 metres; insufficient command of English to complete the assessment and undertake the intervention
Recruitment/selection of patients	People were recruited from two local primary care practices
Age, gender and ethnicity	Age - Mean (range): 67 (51 to 81). Gender (M:F): 20:44. Ethnicity: No additional information
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms (mean [range]): 13 (0.5 to 55) years
Indirectness of population	No indirectness
Interventions	(n=29) Intervention 1: Treatment package - Exercise and behaviour change intervention. ESCAPE-Knee pain: 10 sessions held twice a week for 5 weeks, with a review session 4 months after completion of the program. Each session began with an informal themed group discussion led by a supervising physiotherapist for 15-20

	minutes, followed by a 40-minute self-paced, progressive exercise circuit to improve quadriceps strength, dynamic control, balance, coordination and function. After completion, people received a written, tailored home exercise regimen. At 4 months messages were reinforced. Duration 5 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: ≤ 6 weeks (5 weeks). (n=35) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Outpatient physiotherapy (no additional information). Duration 5 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (5 weeks).
Funding	Academic or government funding (Physiotherapy Research Foundation Project Number PRF/03/3)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: EQ-5D at 5 weeks; Group 1: mean 0.81 (SD 0.11); n=29, Group 2: mean 0.77 (SD 0.2); n=35; EQ-5D -0.11-0.1 Top=High is good outcome; Comments: Baseline treatment package: 0.73 (0.14). Baseline standard therapy: 0.76 (0.09).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 3, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica; Group 2 Number missing: 4, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: EQ-5D at 12 months; Group 1: mean 0.78 (SD 0.174); n=29, Group 2: mean 0.73 (SD 0.23); n=35; EQ-5D -0.11-1.0 Top=High is good outcome; Comments: Baseline treatment package: 0.73 (0.14). Baseline standard therapy: 0.76 (0.09).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 8, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 had related knee surgery, 1 moved away, 1 stopped attending; Group 2 Number missing: 8, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending

Protocol outcome 3: Pain at ≤3 months

- Actual outcome: WOMAC pain at 5 weeks; Group 1: mean 3.2 (SD 2.7); n=29, Group 2: mean 4 (SD 3.6); n=35; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.6 (3.4). Baseline standard therapy: 5.7 (3.2).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 3, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica; Group 2 Number missing: 4, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems

Protocol outcome 4: Pain at >3 months
- Actual outcome: WOMAC pain at 12 months; Group 1: mean 3.2 (SD 3.3); n=29, Group 2: mean 4.2 (SD 4); n=35; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.6 (3.4). Baseline standard therapy: 5.7 (3.2).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 8, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 had related knee surgery, 1 moved away, 1 stopped attending; Group 2 Number missing: 8, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: WOMAC function at 5 weeks; Group 1: mean 11.6 (SD 9.5); n=29, Group 2: mean 11.4 (SD 12.2); n=35; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 16.1 (11.8). Baseline standard therapy: 15.9 (10.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 3, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica; Group 2 Number missing: 4, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems

Protocol outcome 6: Physical function at >3 months
- Actual outcome: WOMAC function at 12 months; Group 1: mean 11.5 (SD 12.1); n=29, Group 2: mean 12.2 (SD 13.7); n=35; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 16.1 (11.8). Baseline standard therapy: 15.9 (10.4).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 8, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 had related knee surgery, 1 moved away, 1 stopped attending; Group 2 Number missing: 8, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending

Protocol outcome 7: Psychological distress at ≤3 months
- Actual outcome: HADS anxiety at 5 weeks; Group 1: mean 4.4 (SD 3.5); n=29, Group 2: mean 4.2 (SD 3.3); n=35; HADS anxiety 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 4.2 (2.9). Baseline standard therapy: 3.6 (2.4).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain

and baseline values of outcomes; Group 1 Number missing: 3, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica; Group 2 Number missing: 4, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems
- Actual outcome: HADS depression at 5 weeks; Group 1: mean 2.4 (SD 1.3); n=29, Group 2: mean 3 (SD 2.4); n=35; HADS depression 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 2.7 (1.7). Baseline standard therapy: 2.7 (1.7).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 3, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica; Group 2 Number missing: 4, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems

Protocol outcome 8: Psychological distress at >3 months
- Actual outcome: HADS anxiety at 12 months; Group 1: mean 4.9 (SD 3.9); n=29, Group 2: mean 4.5 (SD 2.9); n=35; HADS anxiety 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 4.2 (2.9). Baseline standard therapy: 3.6 (2.4).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 8, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 had related knee surgery, 1 moved away, 1 stopped attending; Group 2 Number missing: 8, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending
- Actual outcome: HADS depression at 12 months; Group 1: mean 2.7 (SD 1.9); n=29, Group 2: mean 3.2 (SD 2.4); n=35; HADS depression 0-21 Top=High is poor outcome; Comments: Baseline treatment package: 2.7 (1.7). Baseline standard therapy: 2.7 (1.7).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 8, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 had related knee surgery, 1 moved away, 1 stopped attending; Group 2 Number missing: 8, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending

Protocol outcome 9: Discontinuation at ≤3 months
- Actual outcome: Lost to follow up at 5 weeks; Group 1: 3/29, Group 2: 4/35; Comments: Treatment package: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica. Standard care: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 3, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica; Group 2 Number missing: 4, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems

Protocol outcome 10: Discontinuation at >3 months
- Actual outcome: Lost to follow up at 12 months; Group 1: 7/29, Group 2: 8/35; Comments: Treatment package: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 related knee surgery, 1 moved away, 1 stopped attending. Standard care: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending.
Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, duration of knee pain and baseline values of outcomes; Group 1 Number missing: 8, Reason: 1 withdrew, 1 developed heart problems, 1 diagnosed with polymyalgia rheumatica, 1 developed hip complications, 1 had related knee surgery, 1 moved away, 1 stopped attending; Group 2 Number missing: 8, Reason: 1 withdrew, 1 diagnosed with spinal stenosis, 2 developed unrelated health problems, 1 had heart surgery, 1 moved away, 2 stopped attending	
Protocol outcomes not reported by the study	Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Kao 2012[146]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=259)
Countries and setting	Conducted in Taiwan; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 4 weeks of intervention, total follow up of 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Diagnosis by medical history and a physical examination (including an x-ray showing osteophytes)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	An age greater than 50 years old; having morning stiffness lasting less than 30 minutes, or existing crepitus when moving the legs; an X-ray showing osteophytes
Exclusion criteria	Having ever had knee replacement surgery; unable to maintain balance while standing independently; comorbidities with any medical condition that could be exacerbated by the protocol, such as unstable heart disease
Recruitment/selection of patients	People were recruited from four districts of Taipei City
Age, gender and ethnicity	Age - Mean (SD): 67.7 (10.6). Gender (M:F): 48:147. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Low morbidity score (No comorbidities: 76, High blood pressure: 94, Diabetes mellitus: 27, Hyperlipidaemia: 25, Heart disease: 29). 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=134) Intervention 1: Treatment package - Exercise and behaviour change intervention. A treatment package containing a behaviour change component, an education component and an exercise component. This consisted of four 80 minute classes held once a week with 10-15 participants. These were led by a physical therapist. The program included receiving patient education, viewing a DVD, doing exercise, and participating in four weekly discussion sessions. Education aimed to assist people to maintain a healthy lifestyle, seek support, solve problems and make an action plan. The topics of the four classes involved: anatomy, pathology and common treatment; protection and pain reducing techniques; exercise and relieving

	pressure caused by the osteoarthritis induced disability. After teaching, there was a 20 minute exercise program aiming at stretching and strengthening the whole body's muscles, especially in the lower extremities. The final part of the class was a 40 minute discussion. This used a self-efficacy promoting strategy. People discussed their experiences, set their own goals and practiced these ideas at home, allowing them to share the outcomes to the rest of the group at the next meeting. Duration 4 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Mixture of both, but more of a behaviour change component). 2. Length of package: ≤ 6 weeks (4 weeks). (n=125) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Standard care available to all participants. Duration 4 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (4 weeks).
Funding	Academic or government funding (Received grant support from the Department of Health, Taipei City Government (96001-62-001))

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: SF-36 physical component scale at 8 weeks; Group 1: mean 0.19 (SD 10.7); n=134, Group 2: mean -0.76 (SD 6.2); n=125; SF-36 physical component 0-100 Top=High is good outcome; Comments: Baseline treatment package: 40.9 (12.2). Baseline control: 42.8 (10.5).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low, Comments - States it is a 'quasi-experimental' study. States that there was cluster randomisation but not information given.; Indirectness of outcome: No indirectness ; Baseline details: Reported age, number taking treatment, gender, marital status, education level, having health education for knee pain, medication, other chronic disease and baseline values of outcomes; Group 1 Number missing: 20, Reason: 8 drop out, 12 lost to follow up; Group 2 Number missing: 34, Reason: 15 drop out, 19 lost to follow up
- Actual outcome: SF-36 mental component scale at 8 weeks; Group 1: mean 0.86 (SD 8.5); n=134, Group 2: mean -1.7 (SD 6); n=125; SF-36 mental component 0-100 Top=High is good outcome; Comments: Baseline treatment package: 47.9 (10.6). Baseline control: 49.2 (9.7).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low, Comments - States it is a 'quasi-experimental' study. States that there was cluster randomisation but not information given.; Indirectness of outcome: No indirectness ; Baseline details: Reported age, number taking treatment, gender, marital status, education level, having health education for knee pain, medication, other chronic disease and baseline values of outcomes; Group 1 Number missing: 20, Reason: 8 drop out, 12 lost to follow up; Group 2 Number missing: 34, Reason: 15 drop out, 19 lost to follow up

Protocol outcome 2: Discontinuation at ≤3 months - Actual outcome: Drop out and lost to follow up at 8 weeks; Group 1: 20/134, Group 2: 34/125; Comments: Treatment package: 8 dropped out (couldn't contact, busy, withdraw, not in city), 12 lost to follow up (couldn't contact, busy, much better, not in city). Standard care: 15 dropped out (couldn't contact, busy, withdraw), 19 lost to follow up (couldn't contact, busy, withdraw) Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low, Comments - States it is a 'quasi-experimental' study. States that there was cluster randomisation but not information given.; Indirectness of outcome: No indirectness ; Baseline details: Reported age, number taking treatment, gender, marital status, education level, having health education for knee pain, medication, other chronic disease and baseline values of outcomes; Group 1 Number missing: 20, Reason: 8 drop out, 12 lost to follow up; Group 2 Number missing: 34, Reason: 15 drop out, 19 lost to follow up	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Keefe 2004[150]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=72)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Persistent knee pain due to osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Persistent knee pain due to osteoarthritis diagnosed by a board-certified rheumatologist
Exclusion criteria	Comorbid medical conditions that could affect their health status over the course of the trial (e.g. a recent myocardial infarction); an abnormal cardiac response to exercise (e.g. exercise-induced ventricular tachycardia, abnormal blood pressure response); other known organic disease that would contraindicate safe participation in the study (e.g. chronic obstructive pulmonary disease, congestive heart failure, or cancer)
Recruitment/selection of patients	People and their spouses were recruited from rheumatology clinics and advertisements placed in newspapers
Age, gender and ethnicity	Age - Mean (SD): 59.5 (11.4). Gender (M:F): 33:39. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Treatment package - Exercise and behaviour change intervention. Spouse assisted coping skills training and exercise. Spouse assisted coping skills training consisted of 12 weekly, 2 hour sessions. The training discussed: pain being a complex experience, which as the gate control theory suggests, can be influenced by thoughts, feelings and behaviours; that people and their spouses can acquire and maintain skills for managing pain through frequent practice; that osteoarthritis is a couples issue that affects each partner and their relationship, so involving he spouse can be quite helpful. The sessions discussed coping skills and

	encouraged couples to practice in the group and at home. methods taught included attention diversion skills (relaxation, imagery and distraction), activity-based skills (activity-rest cycling, pleasant activity scheduling) and cognitive coping strategies (cognitive restructuring and self-instructional methods for dealing with severe pain). Training was provided in communication skills, behavioural rehearsal, mutual goal setting, joint home practice and in vivo practice. Exercise training included 3 supervised group exercise sessions per week for 12 consecutive weeks. The spouses did not attend the exercise sessions. The program included: cardiopulmonary endurance training; strength training; flexibility/range of motion training. People participated in 30 minutes of aerobic training three days per week at an intensity of 50-70% of heart rate reserve gradually increasing to 70-85% over the 12 weeks (this was achieved through biking, walking or water aerobics). People also participated in 30 minutes of strength training two days per week. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Spouse-assisted coping skills training). 2. Length of package: > 6 weeks (12 weeks). (n=16) Intervention 2: Non-combined active treatment - Exercise. Exercise therapy only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks). (n=18) Intervention 3: Non-combined active treatment - Behaviour change intervention. Spouse assisted coping skills training. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (12 weeks). (n=18) Intervention 4: Standard care (non-organised) or no treatment - Standard care (non-organised). Standard care. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (This research was supported by National Institute of Arthritis and Musculoskeletal Diseases Grant No. AR-35270)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AIMS pain at 12 weeks; Group 1: mean 4.26 (SD 1.45); n=20, Group 2: mean 3.19 (SD 1.85); n=16; AIMS pain 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 5.20 (1.20). Baseline exercise: 3.91 (1.64).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex and baseline values of outcomes. Difference in outcomes at baseline for exercise and standard care in AIMS pain, when compared to the others.; Group 1 Number missing: -; Group 2 Number missing: -
- Actual outcome: AIMS psychological distress at 12 weeks; Group 1: mean 2.21 (SD 1.21); n=20, Group 2: mean 1.88 (SD 0.87); n=16; AIMS psychological disability 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 2.57 (1.14). Baseline exercise: 2.36 (1.22).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex and baseline values of outcomes. Difference in outcomes at baseline for exercise and standard care in AIMS pain, when compared to the others.; Group 1 Number missing: -; Group 2 Number missing: -

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AIMS pain at 12 weeks; Group 1: mean 4.26 (SD 1.45); n=20, Group 2: mean 4 (SD 1.56); n=18; AIMS pain 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 5.20 (1.20). Baseline behaviour change: 5.44 (1.88).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex and baseline values of outcomes. Difference in outcomes at baseline for exercise and standard care in AIMS pain, when compared to the others.; Group 1 Number missing: -; Group 2 Number missing: -
- Actual outcome: AIMS psychological distress at 12 weeks; Group 1: mean 2.21 (SD 1.21); n=20, Group 2: mean 2.38 (SD 1.38); n=18; AIMS psychological disability 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 2.57 (1.14). Baseline behaviour change: 2.83 (1.64).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex and baseline values of outcomes. Difference in outcomes at baseline for exercise and standard care in AIMS pain, when compared to the others.; Group 1 Number missing: -; Group 2 Number missing: -

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months

- Actual outcome: AIMS pain at 12 weeks; Group 1: mean 4.26 (SD 1.45); n=20, Group 2: mean 4.03 (SD 2.08); n=18; AIMS pain 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 5.20 (1.20). Baseline standard care: 3.91 (1.73). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex and baseline values of outcomes. Difference in outcomes at baseline for exercise and standard care in AIMS pain, when compared to the others.; Group 1 Number missing: -; Group 2 Number missing: - - Actual outcome: AIMS psychological distress at 12 weeks; Group 1: mean 2.21 (SD 1.21); n=20, Group 2: mean 1.8 (SD 1.04); n=18; AIMS psychological disability 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 2.57 (1.14). Baseline standard care: 1.85 (0.33). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex and baseline values of outcomes. Difference in outcomes at baseline for exercise and standard care in AIMS pain, when compared to the others.; Group 1 Number missing: -; Group 2 Number missing: -	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study	Kemp 2018[151]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=17)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Early-onset hip osteoarthritis (defined as chondropathy Outerbridge grade at least 1).
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 18 to 50 years; arthroscopy for intra-articular hip pathology during the past 4 to 14 months; evidence of early-onset at the time of hip arthroscopy which is equivalent to OARSI grade 2 = surface discontinuity and usually not visible on radiographs; pain in the hip of at least 30mm on a visual analogue scale on aggravating activities
Exclusion criteria	Pain not confirmed by physical examination of the hip; concurrent symptoms of hip bursitis or tendinitis; surgical complications including infection; planned lower limb surgery in the following 12 months; physical inability to weight-bear fully or undertake testing procedures; inability to understand written and spoken English
Recruitment/selection of patients	The study was undertaken in a private physiotherapy clinic in Hobart, Tasmania, Australia
Age, gender and ethnicity	Age - Mean (SD): 35.7 (9.9). Gender (M:F): 8:9. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years (Early-onset). 2. Diagnosis with or without imaging: Diagnosed without imaging (Diagnosed with arthroscopy). 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	Serious indirectness: People were post-hip arthroscopy (on average 8-9 months afterwards)
Interventions	(n=10) Intervention 1: Treatment package - Combination and education programme. A treatment package that was semi-standardised including: manual hip joint and soft tissue mobilisation and stretching; hip muscle retraining; trunk muscle retraining; functional, proprioceptive and sports- or activity- specific retraining; enhancing physical activity; education. The physiotherapy intervention was progressed based on

	response to exercise load, thus maximising the training effects, and included supervised exercises during each visit. In addition, a home exercise program was encouraged to be performed independently four times per week, using a structured exercise manual. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks). (n=7) Intervention 2: Non-combined active treatment - Education programme. Education only delivered at the same frequency and duration as the treatment package. Encompassed individualised health education sessions covering topics such as exercise, diet, weight loss, and appropriate stretching. People were also provided with a treatment manual containing specific education information sheets. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (This study was funded by the Australian Physiotherapy Research Foundation Beryl Haynes Memorial Tagged Grant (T13-BH007). Funding was used as part-payment for the physiotherapy sessions provided to both groups in the private physiotherapy clinic.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: HOOS quality of life at 12 weeks; Group 1: mean 3 (SD 16); n=10, Group 2: mean -5 (SD 18); n=7; HOOS quality of life 0-100 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: 3 (-7 to 13). Reported education: -5 (-18 to 9). Baseline treatment package: 49.0 (25.0). Baseline control: 51.0 (15.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, time since surgery, height, weight, bMI and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: HOOS pain at 12 weeks; Group 1: mean 10 (SD 19); n=10, Group 2: mean -2 (SD 21); n=7; HOOS pain 0-100 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: 10 (-2 to 22). Reported education: -2 (-18 to 13). Baseline treatment package: 76.8 (17.4). Baseline control: 69.6 (22.5).
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, time since surgery, height, weight, bMI and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 3: Physical function at ≤3 months - Actual outcome: HOOS activities of daily living at 12 weeks; Group 1: mean 8 (SD 13); n=10, Group 2: mean -7 (SD 14); n=7; HOOS activities of daily living 0-100 Top=High is good outcome; Comments: Reported change scores and 95% confidence intervals. Reported treatment package: 8 (0 to 16). Reported education: -7 (-17 to 4). Baseline treatment package: 80.3 (15.9). Baseline control: 86.9 (10.9). Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, time since surgery, height, weight, bMI and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 4: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 12 weeks; Group 1: 0/10, Group 2: 0/7 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, time since surgery, height, weight, bMI and baseline values of outcomes; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Klassbo 2003[155]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=145)
Countries and setting	Conducted in Sweden; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: Treatment for 6 months, total follow up 1 year
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: All people had to have fulfilled diagnostic tests (radiography) and clinical criteria, defined as pain in the hip region lasting more than 3 months and manifestations of impaired hip joint range of motion and/or muscle function
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Hip dysfunction, which lacking established diagnostic tests or clinical criteria
Exclusion criteria	Trauma; fractures; congenital malalignments; other hip joint diseases; inflammatory joint or neuromuscular diseases; low back, sacroiliac, or knee problems overshadowing the hip problems; inclusion criteria for total hip replacement (severe pain and persisting resting pain despite pharmacologic treatment; tried all other kinds of pain treatments; disturbed night sleep; walking ability not exceeding 200-300 meters, even with walking aid)
Recruitment/selection of patients	People were consecutively recruited by physicians in primary care and orthopedic units
Age, gender and ethnicity	Age - Mean (SD): 61.8 (10.4). Gender (M:F): 59:86. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Between <6 months and 10+ years, median time >2 years <5 years
Indirectness of population	No indirectness
Interventions	(n=77) Intervention 1: Treatment package - Exercise and education programme. Hip school. Instructions on home based exercises and an education program, consisting of three education sessions and 1 individual follow up session at 2 months after the other sessions. The content of the meetings were: first group (where is the hip?; tissues belonging to a joint; diagnosing hip osteoarthritis; who gets hip osteoarthritis?;

	hip osteoarthritis and pain; natural course at group level), second group (muscles involved; diagnosing decreased range of motion; proposed physical activity; not too much and not too little; to preserve/enhanced range of motion; therapeutic exercise sheet), third group (pain; self management of pain; pros and cons of pain treatments; physical therapy; pharmacology; surgery). Duration 6 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (6 months). (n=68) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual treatment. Duration 6 months. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (6 months).
Funding	Academic or government funding (Supported by grants from the Research Center of Primary Care, Varmland County Council, the Varmland social insurance office, the Swedish Association of Registered Physical Therapists Memorial Fund, the Swedish Federation of County Councils, the Karolinska Institutet, and the Swedish Rheumatism Association)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED) Protocol outcome 1: Discontinuation at >3 months - Actual outcome: Dropouts at 6 months; Group 1: 17/77, Group 2: 9/68; Comments: Reasons not given Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, time since onset, first visit to doctor, medicine intake per month, walking distance, physical activity inde, satisfaction with activity and baseline values of outcomes; Group 1 Number missing: 17, Reason: No additional information; Group 2 Number missing: 8, Reason: No additional information	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Kloek 2018[156]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=218)
Countries and setting	Conducted in Netherlands; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of treatment, 12 months total follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People hip/knee osteoarthritis according to the clinical criteria of the American College of Rheumatology
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	An age of 40 to 80 years and hip/knee osteoarthritis according to the clinical criteria of the American College of Rheumatology.
Exclusion criteria	Being on a waiting list for a hip or knee replacement surgery; contraindications for physical activity without supervision according to the Physical Activity Readiness Questionnaire; sufficiently physically active according to the physical therapist; participation in a physical therapy and/or physical activity program in the past 6 months; no access to internet; inability to understand the Dutch language
Recruitment/selection of patients	People who visited a participating physical therapist were invited. Also, recruitment advertisements were placed in local newspapers, and information brochures were sent to general practitioners.
Age, gender and ethnicity	Age - Mean (SD): 63.1 (8.7). Gender (M:F): 67:141. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Low morbidity score (0 comorbidities: 124, 1 comorbidity: 40, at least 2 comorbidities: 44). 4. Site of osteoarthritis: Mixed (Hip and/or knee).
Extra comments	Severity: Not stated Duration of Symptoms: <1 to at least 5 years, median time 1-5 years
Indirectness of population	No indirectness
Interventions	(n=109) Intervention 1: Treatment package - Exercise and education programme. E-exercise. An intervention over 12 weeks with a combination of about 5 face-to-face sessions with a physical therapist and an online application focusing on behavioural graded activity, exercises and information. The sessions discussed exercises, provided support and was used to formulate goals. The online part consisted of 3 modules: grade activity (the duration was gradually increased until the individual short-

	term goal was met); strength and stability (each week the participant was asked to perform 2 video-supported exercises on 3 different days, and the number of repetitions was increased gradually every 4 weeks); information (each week a new video was generated about osteoarthritis etiology, pain management, weight management, motivation, medication, and social influences on pain). Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks). (n=99) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual physical therapy according to a Dutch Osteoarthritis guideline. This recommends the same 3 elements as e-exercise: information, physical exercise and strength and stability exercises. No restrictions were given with regard to the number of face-to-face sessions. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (The study was funded by ZonMw (ZonMw Research Program Sport, Ref. no. 525001007), the Dutch Rheumatoid Arthritis Foundation, and the Royal Dutch Society for Physiotherapy)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: KOOS/HOOS quality of life at 3 months; Group 1: mean 49.1 (SD 30.2); n=87, Group 2: mean 53 (SD 31.9); n=87; KOOS/HOOS quality of life 0-100 Top=High is good outcome; Comments: Reported final values and 95% confidence intervals. Reported treatment package: 49.1 (42.7 to 55.4). Reported usual care: 53.0 (46.3 to 59.7). Baseline treatment package: 45.0 (39.2 to 50.8). Baseline usual care: 44.2 (38.1 to 50.4).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, BMI, location of osteoarthritis, duration of symptoms, education, number of comorbidities and baseline values of outcomes; Group 1 Number missing: 22, Reason: Reasons unclear; Group 2 Number missing: 12, Reason: Reasons unclear

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: KOOS/HOOS quality of life at 12 months; Group 1: mean 52.5 (SD 36.6); n=65, Group 2: mean 56.1 (SD 38.4); n=69; KOOS/HOOS quality of life 0-100 Top=High is good outcome; Comments: Reported final values and 95% confidence intervals. Reported treatment package: 52.5 (43.6 to 61.4). Reported usual care: 56.1 (47.0 to 65.1). Baseline treatment package: 45.0 (39.2 to 50.8). Baseline usual care: 44.2 (38.1 to 50.4).

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, BMI, location of osteoarthritis, duration of symptoms, education, number of comorbidities and baseline values of outcomes; Group 1 Number missing: 44, Reason: Reasons unclear; Group 2 Number missing: 30, Reason: Reasons unclear

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: KOOS/HOOS pain at 3 months; Group 1: mean 55.8 (SD 40.5); n=87, Group 2: mean 48.8 (SD 42.4); n=87; KOOS/HOOS pain 0-100 Top=High is good outcome; Comments: Reported final values and 95% confidence intervals. Reported treatment package: 55.8 (47.3 to 64.3). Reported usual care: 48.8 (39.9 to 57.7). Baseline treatment package: 50.4 (42.1 to 58.8). Baseline usual care: 43.9 (35.2 to 52.7).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, BMI, location of osteoarthritis, duration of symptoms, education, number of comorbidities and baseline values of outcomes; Group 1 Number missing: 22, Reason: Reasons unclear; Group 2 Number missing: 12, Reason: Reasons unclear

Protocol outcome 4: Pain at >3 months
- Actual outcome: KOOS/HOOS pain at 12 months; Group 1: mean 65.9 (SD 47.7); n=65, Group 2: mean 61.6 (SD 49.8); n=69; KOOS/HOOS pain 0-100 Top=High is good outcome; Comments: Reported final values and 95% confidence intervals. Reported treatment package: 65.9 (54.3 to 77.5). Reported usual care: 61.6 (49.9 to 73.4). Baseline treatment package: 50.4 (42.1 to 58.8). Baseline usual care: 43.9 (35.2 to 52.7).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, BMI, location of osteoarthritis, duration of symptoms, education, number of comorbidities and baseline values of outcomes; Group 1 Number missing: 44, Reason: Reasons unclear; Group 2 Number missing: 30, Reason: Reasons unclear

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: KOOS/HOOS physical function at 3 months; Group 1: mean 56.8 (SD 27.8); n=87, Group 2: mean 56.3 (SD 29); n=87; KOOS/HOOS physical function 0-100 Top=High is good outcome; Comments: Reported final values and 95% confidence intervals. Reported treatment package: 56.8 (51.0 to 62.7). Reported usual care: 56.3 (50.2 to 62.4). Baseline treatment package: 52.7 (47.3 to 58.0). Baseline usual care: 50.7 (45.1 to 56.4).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, BMI, location of osteoarthritis, duration of symptoms, education, number of comorbidities and baseline values of outcomes; Group 1 Number missing: 22, Reason: Reasons unclear; Group 2 Number missing: 12, Reason: Reasons unclear

Protocol outcome 6: Physical function at >3 months
- Actual outcome: KOOS/HOOS physical function at 12 months; Group 1: mean 59.8 (SD 34.3); n=65, Group 2: mean 58 (SD 35.8); n=69; KOOS/HOOS physical function 0-100 Top=High is good outcome; Comments: Reported final values and 95% confidence intervals. Reported treatment package: 59.8 (51.4 to 68.1). Reported usual care: 58.0 (49.6 to 66.5). Baseline treatment package: 52.7 (47.3 to 58.0). Baseline usual care: 50.7 (45.1 to 56.4).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, BMI, location of

osteoarthritis, duration of symptoms, education, number of comorbidities and baseline values of outcomes; Group 1 Number missing: 44, Reason: Reasons unclear; Group 2 Number missing: 30, Reason: Reasons unclear	
Protocol outcomes not reported by the study	Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study (subsidiary papers)	Kovar 1992[159] (Sullivan 1998[263])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=102)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Clinical and radiographic osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable:
Inclusion criteria	Age 40 years or more; a documented diagnosis of chronic, stable, primary osteoarthritis of one or both knee joints in association with at least a 4-month history of symptomatic knee pain occurring during weight-bearing activities (patients with multiple joint involvement, those who had undergone major joint surgery, or had a lower joint prosthesis were also eligible); radiographic evidence of primary osteoarthritis of one or both knee joints as demonstrated by: joint-space narrowing, marginal spur formation, or subchondral cyst formation; the use of any of the various common, over-the-counter non-steroidal anti-inflammatory drugs 2 or more days per week; nonparticipation in a regular program of physical activity at the time of enrollment
Exclusion criteria	Serious medical conditions for which exercise would be contraindicated, such as unstable angina, significant aortic stenosis, myocardial infarction within the last 3 months, or advanced chronic obstructive pulmonary disease; asymptomatic primary osteoarthritis of one or both knees; dementia or the inability to give informed consent; nonambulation due to amputation, stroke, or incapacitating arthritis; involvement in another treatment program or study protocol
Recruitment/selection of patients	People were recruited from cooperating physicians at the Hospital for Special Surgery, a major referral center for patients with musculoskeletal and rheumatic diseases located at the New York Hospital Cornell Medical Center; people seen in the outpatient rheumatology and orthopaedic clinics of the hospital; people identified through the New York Chapter of the Arthritis Foundation and various community-based sites in the vicinity of the hospital
Age, gender and ethnicity	Age - Mean (SD): 69.5 (10.3). Gender (M:F): 17:85. Ethnicity: Black = 8, Hispanic = 3, White = 91

Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms (mean [SD]): 11.5 (11.5) years
Indirectness of population	No indirectness
Interventions	(n=52) Intervention 1: Treatment package - Exercise and education programme. Indoor supervised fitness walking and patient education, the goal of which was to increase the functional capacity of patients by encouraging the adoption and maintenance of regular fitness walking. The program comprised 24 90-minute walking and education sessions that were designed and led by a registered physical therapist. Sessions occurred thrice weekly and included light stretching and strengthening exercises; guest speakers on the medical aspects of osteoarthritis and exercise; group discussion about barriers and benefits of walking; instruction in proper walking techniques and the maintenance of a walking program; supportive encouragement and up to 30 minutes of walking. The walking portion was conducted in a hospital corridor where people walked on a tiled floor surface that was hard and smooth. The people wore supportive athletic shoes or shoes designed specifically for walking, cushioned athletic socks, and loose-fitting clothing. Each person received an instructional guidebook with educational materials printed in large, bold-face type. The program and instructional materials were designed after conducting a patient-needs assessment and a review of the literature on walking programs; concepts from self-efficacy theory and educational strategies from behavioural psychology were incorporated into the program to help patients adhere to the walking regimen. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (8 weeks). (n=50) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Standard care only. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (8 weeks).
Funding	Academic or government funding (Grant support in part by a dissertation research grant to Dr Kovar from the Arthritis Foundation and by National Institutes of Health Multipurpose Arthritis Center Program Grant No. 1 P60 AR38520-01A1 from the National Institute for Arthritis and Musculoskeletal and Skin Diseases)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: AIMS physical activity at 8 weeks; Group 1: mean 3.74 (SD 2.69); n=47, Group 2: mean 5.96 (SD 2.32); n=45; AIMS physical activity 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 6.15 (2.27). Baseline usual care: 5.72 (2.49).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 5, Reason: Treatment package: 1 total knee replacement, 1 died due to a cause unrelated to the intervention, 1 hip fracture due to fall during a program session, 1 feared that walking might exacerbate a heart condition, 1 family problems.; Group 2 Number missing: 5, Reason: Standard care: 1 fractured hip after a fall, 1 sprained ankle, 1 dental problem, 2 family problems.
- Actual outcome: AIMS arthritis impact at 8 weeks; Group 1: mean 2.86 (SD 1.88); n=47, Group 2: mean 3.06 (SD 1.91); n=45; AIMS arthritis impact 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 4.56 (2.14). Baseline usual care: 3.85 (2.38).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 5, Reason: Treatment package: 1 total knee replacement, 1 died due to a cause unrelated to the intervention, 1 hip fracture due to fall during a program session, 1 feared that walking might exacerbate a heart condition, 1 family problems.; Group 2 Number missing: 5, Reason: Standard care: 1 fractured hip after a fall, 1 sprained ankle, 1 dental problem, 2 family problems.
- Actual outcome: AIMS arthritis pain at 8 weeks; Group 1: mean 3.77 (SD 1.73); n=47, Group 2: mean 4.77 (SD 2.12); n=45; AIMS arthritis pain 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 5.15 (1.99). Baseline usual care: 4.87 (2.31).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 5, Reason: Treatment package: 1 total knee replacement, 1 died due to a cause unrelated to the intervention, 1 hip fracture due to fall during a program session, 1 feared that walking might exacerbate a heart condition, 1 family problems.; Group 2 Number missing: 5, Reason: Standard care: 1 fractured hip after a fall, 1 sprained ankle, 1 dental problem, 2 family problems.
- Actual outcome: AIMS medications use at 8 weeks; Group 1: mean 3.64 (SD 1.92); n=47, Group 2: mean 2.9 (SD 2.02); n=45; AIMS medications use 0-6 Top=High is good outcome; Comments: Baseline treatment package: 2.80 (1.65). Baseline usual care: 2.64 (1.68).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 5, Reason: Treatment package: 1 total knee replacement, 1 died due to a cause unrelated to the intervention, 1 hip fracture due to fall during a program session, 1 feared that walking might exacerbate a heart condition, 1 family problems.; Group 2 Number missing: 5, Reason: Standard care: 1 fractured hip after a fall, 1 sprained ankle, 1 dental problem, 2 family problems.

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: AIMS physical activity at 1 year; Group 1: mean 6.07 (SD 2.95); n=29, Group 2: mean 6.18 (SD 2.75); n=23; AIMS physical activity 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 6.15 (2.27). Baseline usual care: 5.72 (2.49).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 23, Reason: Reports only 29 people followed up; Group 2 Number missing: 27, Reason: Reports only 23 people followed up
- Actual outcome: AIMS arthritis impact at 1 year; Group 1: mean 3.25 (SD 2.6); n=29, Group 2: mean 3.8 (SD 2.06); n=23; AIMS arthritis impact 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 4.56 (2.14). Baseline usual care: 3.85 (2.38).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 23, Reason: Reports only 29 people followed up; Group 2 Number missing: 27, Reason: Reports only 23 people followed up
- Actual outcome: AIMS arthritis pain at 1 year; Group 1: mean 4.59 (SD 2.4); n=29, Group 2: mean 5.5 (SD 2.07); n=23; AIMS arthritis pain 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 5.15 (1.99). Baseline usual care: 4.87 (2.31).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 23, Reason: Reports only 29 people followed up; Group 2 Number missing: 27, Reason: Reports only 23 people followed up
- Actual outcome: AIMS medications use at 1 year; Group 1: mean 3.34 (SD 2.16); n=29, Group 2: mean 3.6 (SD 2.25); n=23; AIMS medication use 1-6 Top=High is good outcome; Comments: Baseline treatment package: 2.80 (1.65). Baseline usual care: 2.64 (1.68).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 23, Reason: Reports only 29 people followed up; Group 2 Number missing: 27, Reason: Reports only 23 people followed up
- Actual outcome: AIMS general health perception at 1 year; Group 1: mean 3.71 (SD 2.8); n=29, Group 2: mean 3.26 (SD 1.87); n=23; AIMS general health perception 0-10 Top=High is poor outcome; Comments: No baseline values reported
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 23, Reason: Reports only 29 people followed up; Group 2 Number missing: 27, Reason: Reports only 23 people followed up

Protocol outcome 3: Pain at >3 months
- Actual outcome: VAS pain at 1 year; Group 1: mean 4.96 (SD 2.82); n=29, Group 2: mean 5.43 (SD 3.14); n=23; VAS 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 4.13 (2.55). Baseline control: 6.26 (3.15).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 23, Reason: Reports only 29 people followed up; Group 2 Number missing: 27, Reason: Reports only 23 people followed up

Protocol outcome 4: Discontinuation at ≤3 months
- Actual outcome: Attrition at 8 weeks; Group 1: 5/52, Group 2: 5/50; Comments: Treatment package: 1 total knee replacement, 1 died due to a cause

unrelated to the intervention, 1 hip fracture due to fall during a program session, 1 feared that walking might exacerbate a heart condition, 1 family problems. Standard care: 1 fractured hip after a fall, 1 sprained ankle, 1 dental problem, 2 family problems. Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, sex, race, marital status, education and baseline values of outcomes; Group 1 Number missing: 5, Reason: Treatment package: 1 total knee replacement, 1 died due to a cause unrelated to the intervention, 1 hip fracture due to fall during a program session, 1 feared that walking might exacerbate a heart condition, 1 family problems.; Group 2 Number missing: 5, Reason: Standard care: 1 fractured hip after a fall, 1 sprained ankle, 1 dental problem, 2 family problems.	
Protocol outcomes not reported by the study	Pain at ≤3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study (subsidiary papers)	Li 2017[167] (Clayton 2015[69])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=34)
Countries and setting	Conducted in Canada; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 4 weeks of intervention, 8 weeks of total follow up (though at 4 to 8 weeks the control group had a delayed start of the intervention, therefore data from only 4 weeks will be used)
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Physician-confirmed diagnosis of knee osteoarthritis, or pass 2 criteria for early osteoarthritis (being age 50 years or older and having experience pain or discomfort in or around the knee during the previous year lasting 28 or more separate or consecutive days). 98% also met the American College of Rheumatology clinical criteria for knee osteoarthritis.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with physician-confirmed diagnosis of knee osteoarthritis
Exclusion criteria	A diagnosis of inflammatory arthritis, connective tissue diseases, fibromyalgia, or gout; had used disease-modifying antirheumatic drugs or gout medications; had knee arthroplasty; were on the waitlist to receive total knee arthroplasty; had acute knee injury in the past 6 months; did not have an e-mail address or daily access to a personal computer with Internet access; had a body mass index of 40kg/m² or more; had received a steroid injection in the last 6 months; had received hyaluronate injection in a knee in the last 6 months; were using medications that impaired activity tolerance (such as beta-blockers); had an inappropriate level of risk for increasing their unsupervised physical activity
Recruitment/selection of patients	No additional information
Age, gender and ethnicity	Age - Mean (SD): 55.5 (8.6). Gender (M:F): 8:28. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness

Interventions	(n=17) Intervention 1: Treatment package - Exercise and behaviour change intervention. A 1.5 education session about physical activity, a FitbitFlex to encourage aerobic exercise, and individual weekly activity counseling with a physiotherapist by telephone. The education session discussed the benefits of physical activity, the detrimental effects of sedentary behaviour, and ways to be active without aggravating osteoarthritis symptoms. People were advised to wear the fitness band 24 hours a day except during water-based activity or when charging. The activity goals were progressively modified during the 4 weekly 20 minute phone calls. The counseling component followed the brief action planning approach, whereby the participants identified their activity goals, developed an action plan, identified barriers and solutions and then rated their confidence in executing the plan (until their confidence was at least 7 out of 10). Duration 4 weeks. Concurrent medication/care: No additional information Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Activity counselling (and one education session)). 2. Length of package: ≤ 6 weeks (4 weeks). (n=17) Intervention 2: Standard care (non-organised) or no treatment - No treatment. Delayed intervention. Duration 4 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (4 weeks).
Funding	Funding not stated (No additional information)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus NO TREATMENT

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: KOOS knee-related quality of life at 4 weeks; Group 1: mean 51.8 (SD 19.5); n=17, Group 2: mean 48.9 (SD 19.3); n=17; KOOS quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 53.3 (18.4). Baseline no treatment: 47.4 (16.1).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, marital status, gross annual household income, osteoarthritis diagnosis, health status, comorbidities, body mass index and baseline values of outcomes. Outcomes are different at baseline for all KOOS subscales; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: KOOS pain at 4 weeks; Group 1: mean 71.4 (SD 17.5); n=17, Group 2: mean 71.6 (SD 15.2); n=17; KOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 74.5 (16.2). Baseline no treatment: 68.6 (16.1).

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, marital status, gross annual household income, osteoarthritis diagnosis, health status, comorbidities, body mass index and baseline values of outcomes. Outcomes are different at baseline for all KOOS subscales; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 3: Physical function at ≤3 months - Actual outcome: KOOS activities of daily living at 4 weeks; Group 1: mean 75.1 (SD 19.7); n=17, Group 2: mean 79.1 (SD 18.9); n=17; KOOS activities of daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 81.8 (17.1). Baseline no treatment: 78.3 (15.9). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, marital status, gross annual household income, osteoarthritis diagnosis, health status, comorbidities, body mass index and baseline values of outcomes. Outcomes are different at baseline for all KOOS subscales; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 4: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 4 weeks; Group 1: 0/17, Group 2: 0/17 Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, marital status, gross annual household income, osteoarthritis diagnosis, health status, comorbidities, body mass index and baseline values of outcomes. Outcomes are different at baseline for all KOOS subscales; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Mcknight 2010[184]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=273)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 2 years, phase 1 lasting 9 months and phase 2 lasting 15 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Pain on most days in 1 or both knees for less than 4 years with a Kellgren Lawrence score of 2 in one or both knees
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Between the age of 35 and 64 years; reported pain on most days in 1 or both knees; duration of symptoms of less than 5 years; had Kellgren and Lawrence classification grade 2 radiographic evidence of knee osteoarthritis in one or both knees; self-reported disability due to knee pain for at least 3 of the following: descending or ascending stairs, walking, kneeling, or performing daily activities
Exclusion criteria	An uncontrolled medical condition that precluded safe participation or prevented completion of the study (e.g. heart disease, blood pressure or respiratory conditions); any neurological condition that could affect coordination; inflammatory arthritis (e.g. rheumatoid or psoriatic arthritis); previous knee surgery; Kellgren Lawrence grades III or IV radiographic evidence of osteoarthritis in one or both knees; a BMI >37.5 - individuals over the limit were advised to follow a weight loss program and achieve stable weight for 6 months prior to participation; a knee corticosteroid injection in the previous 3 months; plans to move from the local area; plans to become pregnant during the study period; more than 120 minutes per week of any vigorous (e.g. exercise, walking, household chores, etc.) physical activity; participated in any form of resistance training
Recruitment/selection of patients	People were recruited from the local community by mass mailings, television/newspaper advertisements, and flyers.
Age, gender and ethnicity	Age - Mean (SD): 52.6 (7.2). Gender (M:F): 63:210. Ethnicity: White = 86.3%-96.3%. No information about other participants.
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee

Extra comments	Severity: Kellgren Lawrence grade of 2 Duration of symptoms: Less than 5 years
Indirectness of population	No indirectness
Interventions	(n=95) Intervention 1: Treatment package - Exercise and behaviour change intervention. Strength training and self-management sessions. Strength training involved three core areas: stretching and balance, range of motion and flexibility, and isotonic muscle strengthening. Sessions were completed three times per week and each session consisted of the following: a 10 minute walking warm-up at 50% maximum heart rate; 5-10 minutes of stretching and balance exercises; 10 minutes of range of motion/flexibility exercises; 30 minutes of strength-training exercises; 5 minutes of cool-down which includes walking and/or static stretching of the muscles. This was conducted over 9 months, with the remaining 15 months being support from trainers to develop self-directed long-term exercising habits (through two weekly phone contact and quarterly "booster" sessions). Self-management through a two-phase self-management intervention targeted at coping and self-efficacy skills. The first 9 months included 12 weekly 90 minute sessions facilitated by a program manager and local health professionals. These were followed by weekly phone calls to boost knowledge and behaviours from classroom sessions, as well as providing practical, one-on-one problem solving discussions to tailor the treatment to each participant's needs. Coping skills focused on promoting more adaptive strategies and reducing avoidant or passive strategies. Over time weekly phone calls became biweekly, monthly and then bimonthly. Duration 2 years. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (2 years). (n=91) Intervention 2: Non-combined active treatment - Exercise. Strength training only. Duration 2 years. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (2 years). (n=87) Intervention 3: Non-combined active treatment - Behaviour change intervention. Behaviour change intervention only. Duration 2 years. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (2 years).

Funding	Other author(s) funded by industry (This work was supported by the National Institute of Arthritis, Musculoskeletal, and Skin Diseases R01-AR-047595 (PI-Yocum/Going). An author was employed by Bristol Meyers-Squibb.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Discontinuation at >3 months
- Actual outcome: Lost to follow up at 24 months; Group 1: 25/95, Group 2: 27/91; Comments: Treatment package: 5 lost to follow up, 20 discontinued due to health problems, not interested, time commitment, non compliant, adverse events and other. Exercise: 6 lost to follow up. 21 discontinued due to not interested, personal, other, knee replacement, time commitment.
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, female, ethnicity, college education, BMI, VAS, SF-36, Depression, compliance and baseline values of outcomes; Group 1 Number missing: 25, Reason: Treatment package: 5 lost to follow up, 20 discontinued due to health problems, not interested, time commitment, non compliant, adverse events and other. ; Group 2 Number missing: 27, Reason: Exercise: 6 lost to follow up. 21 discontinued due to not interested, personal, other, knee replacement, time commitment.

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Discontinuation at >3 months
- Actual outcome: Lost to follow up at 24 months; Group 1: 25/95, Group 2: 20/87; Comments: Treatment package: 5 lost to follow up, 20 discontinued due to health problems, not interested, time commitment, non compliant, adverse events and other. Behaviour change intervention: 10 lost to follow up, 10 discontinued due to non compliance, time commitment, and inflammatory arthritis.
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, female, ethnicity, college education, BMI, VAS, SF-36, Depression, compliance and baseline values of outcomes; Group 1 Number missing: 25, Reason: Treatment package: 5 lost to follow up, 20 discontinued due to health problems, not interested, time commitment, non compliant, adverse events and other. ; Group 2 Number missing: 20, Reason: Behaviour change intervention: 10 lost to follow up, 10 discontinued due to non compliance, time commitment, and inflammatory arthritis.

Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Mecklenburg 2018[186]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=162)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee pain for at least 1 month in the past 12 months
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age over 18; knee pain for at least 1 month in the last 12 months; participating in the collaborating employers' health plans; provision of informed consent. They did not include knee osteoarthritis as an inclusion criterion, though did assess the presence of osteoarthritis through 6 self-reported clinical criteria, whereby 3 or more positive criteria suggested osteoarthritis: age over 50 years, stiffness for ,30 minutes in the morning, crepitus, bony tenderness, bony enlargement and no palpable warmth.
Exclusion criteria	A prior diagnosis of rheumatoid arthritis; surgery on the knee less than 3 months ago; an injury to the knee less than 3 months ago.
Recruitment/selection of patients	People were recruited through emails and posters distributed through the participants' employers
Age, gender and ethnicity	Age - Mean (SD): 46 (12). Gender (M:F): 98:57. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: not stated Duration of symptoms: At least 1 month in the past 12 weeks
Indirectness of population	Serious indirectness: Chronic knee pain, not explicitly stated as osteoarthritis
Interventions	(n=101) Intervention 1: Treatment package - Exercise and behaviour change intervention. Hinge Health 12 weeks digital care package for chronic knee pain. Participants received a tablet computer with the Hinge Health application installed, and two custom Bluetooth sensors with straps to be used on the upper and lower leg during the in-app exercise therapy. People were assigned a personal coach that provided support and accountability throughout the program and were placed in a team to provide peer support through a discussion feed within the app. On a weekly

	basis, people were set the goal of completing 3 sessions of sensor-guided exercise therapy, reading one to two education articles, logging their symptoms at least twice, performing cognitive behavioural therapy (subset of weeks only), working at weight loss (if overweight), and tracking at least three 30-minute sessions of aerobic activities. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (CBT and weight loss advice (also educational articles)). 2. Length of package: > 6 weeks (12 weeks). (n=61) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care and access to three education articles (talking about the importance of self-care, how to deal with setbacks in knee pain and how to manage communication and relationships when living with chronic knee pain). Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks).
Funding	Principal author funded by industry (All authors except JH work at Hinge Health, Inc. Author JH is a paid domain expert consultant.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: KOOS pain at 12 weeks; Group 1: mean 30.3 (SD 17.1); n=101, Group 2: mean 38.4 (SD 17.2); n=61; KOOS pain 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 41.0 (14.1). Baseline control: 41.4 (16.5).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, gender, factors about surgery and background, taking antidepressants/opioids, self-efficacy, surgery on the knee in the past, knee osteoarthritis and baseline values of outcomes; Group 1 Number missing: 43, Reason: 14 people did not respond to invitation, 22 people did not complete week 12 survey, 1 skiing accident, 6 personal reasons; Group 2 Number missing: 25, Reason: 7 entered into treatment due to administrative error, 18 did not complete week 12 survey

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: KOOS physical function short-form at 12 weeks; Group 1: mean 44.6 (SD 16.7); n=101, Group 2: mean 52.5 (SD 16.2); n=61; KOOS physical function short-form 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 53.8 (12.3). Baseline control: 54.5 (15.7).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - High, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, gender, factors about

surgery and background, taking antidepressants/opioids, self-efficacy, surgery on the knee in the past, knee osteoarthritis and baseline values of outcomes; Group 1 Number missing: 43, Reason: 14 people did not respond to invitation, 22 people did not complete week 12 survey, 1 skiing accident, 6 personal reasons; Group 2 Number missing: 25, Reason: 7 entered into treatment due to administrative error, 18 did not complete week 12 survey Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Discontinued at 12 weeks; Group 1: 43/101, Group 2: 25/61; Comments: Treatment package: 14 did not response to invitation, 1 skiing accident, 6 personal reasons, 22 did not complete week 12 survey. Standard care: 7 entered into treatment due to administrative error, 18 did not complete week 12 survey Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, BMI, gender, factors about surgery and background, taking antidepressants/opioids, self-efficacy, surgery on the knee in the past, knee osteoarthritis and baseline values of outcomes; Group 1 Number missing: 43, Reason: 14 people did not respond to invitation, 22 people did not complete week 12 survey, 1 skiing accident, 6 personal reasons; Group 2 Number missing: 25, Reason: 7 entered into treatment due to administrative error, 18 did not complete week 12 survey	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Nunez 2006[208]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=100)
Countries and setting	Conducted in Spain; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 9 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People with knee osteoarthritis according to the Kellgren and Lawrence criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People referred by the Orthopaedic Surgery Department to their therapeutic education and functional readaptation unit, with knee osteoarthritis according to the Kellgren and Lawrence criteria, who had been on a waiting list for total knee replacement for less than 6 months and who consented to participate in the study
Exclusion criteria	Functional illiteracy; inflammatory musculoskeletal disease; metabolic or neoplastic disease and severe psychopathology or comorbidity; defined as a diagnosis in the medical record severe enough that the patient could not complete the TEFR program.
Recruitment/selection of patients	People referred by the Orthopedic Surgery Department to a tertiary care center
Age, gender and ethnicity	Age - Mean (SD): 71.1 (6.7). Gender (M:F): 29:71. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms (mean [SD]): 11.8 (10.6) months
Indirectness of population	No indirectness
Interventions	(n=51) Intervention 1: Treatment package - Exercise and behaviour change intervention. Exercise and self-management training. The self-management component included two 30 minute visits at the first week and 3 months, and two group sessions of about 90 minutes at weeks 3 and 4, with a maximum of 10-12 people. The contents were centered on the consequences of the disease on daily life and included principles of economy/energy conservation and joint protection; evaluation and control of pain (rest and positioning, ice and heat, necessary length of application) and treatment recommended for the management of knee osteoarthritis; demonstration and use of assistive devices and tables of physical exercises with no

	burden on the lower limbs, with specific knee exercises to maintain and improve the strength of muscles acting around the knee, the range of motion at the knee joint and locomotor function; and general exercises to mobilize the joints and strengthen the musculature of the rest of the body. People were instructed to increase the number of repetitions up to a maximum of 30 times, twice a day, for the knee exercises and 10 times, once a day, for the general exercises according to individual tolerance to pain. The exercises were taught in group sessions. People were instructed to practice the exercises at home in the week previous to the second group session, in which all people carried out the complete table of exercises, supervised by the educator. Duration 9 months. Concurrent medication/care: Both groups received 3-4g/day of paracetamol alone or no more than 2g/day of paracetamol combined with 2400mg/day of ibuprofen or other NSAIDs (the dose of NSAIDs varying according to individual patient needs). Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (9 months). (n=49) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Standard care only. Duration 9 months. Concurrent medication/care: Both groups received 3-4g/day of paracetamol alone or no more than 2g/day of paracetamol combined with 2400mg/day of ibuprofen or other NSAIDs (the dose of NSAIDs varying according to individual patient needs). Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (9 months).
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: SF-36 physical function at 9 months; Group 1: mean 27.2 (SD 15.49); n=43, Group 2: mean 23.47 (SD 18.97); n=37; SF-36 physical function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 21.34 (13.51). Baseline usual care: 27.50 (19.07).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 physical role at 9 months; Group 1: mean 34.76 (SD 44.68); n=43, Group 2: mean 49.31 (SD 42.04); n=37; SF-36 physical role 0-100 Top=High is good outcome; Comments: Baseline treatment package: 26.83 (38.07). Baseline usual care: 40.97 (42.74).

Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 bodily pain at 9 months; Group 1: mean 38.61 (SD 21.93); n=43, Group 2: mean 30.33 (SD 24.62); n=37; SF-36 bodily pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 32.20 (22.62). Baseline usual care: 37.97 (25.76).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 general health at 9 months; Group 1: mean 50.12 (SD 22.52); n=43, Group 2: mean 56.42 (SD 20.88); n=37; SF-36 general health 0-100 Top=High is good outcome; Comments: Baseline treatment package: 44.56 (20.23). Baseline usual care: 55.75 (22.35).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 vitality at 9 months; Group 1: mean 51.34 (SD 23.8); n=43, Group 2: mean 54.58 (SD 25.11); n=37; SF-36 vitality 0-100 Top=High is good outcome; Comments: Baseline treatment package: 43.29 (24.15). Baseline usual care: 49.44 (25.54).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 social function at 9 months; Group 1: mean 61.24 (SD 30.81); n=51, Group 2: mean 62.53 (SD 32.15); n=37; SF-36 social function 0-100 Top=High is good outcome; Comments: Baseline treatment package: 58.12 (27.84). Baseline usual care: 62.08 (35.94).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 emotional role at 9 months; Group 1: mean 57.71 (SD 47.16); n=43, Group 2: mean 62.03 (SD 47.26); n=37; SF-36 emotional role 0-100 Top=High is good outcome; Comments: Baseline treatment package: 56.10 (47.99). Baseline usual care: 67.58 (42.56).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out

in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
- Actual outcome: SF-36 mental health at 9 months; Group 1: mean 57.27 (SD 23.82); n=43, Group 2: mean 63.81 (SD 25.94); n=37; SF-36 mental health 0-100 Top=High is good outcome; Comments: Baseline treatment package: 51.68 (25.19). Baseline usual care: 65.78 (22.62).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five

Protocol outcome 2: Pain at >3 months
- Actual outcome: WOMAC pain at 9 months; Group 1: mean 10.07 (SD 3.33); n=43, Group 2: mean 10.89 (SD 3.73); n=37; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 12.51 (8.55). Baseline usual care: 9.92 (3.69).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five

Protocol outcome 3: Physical function at >3 months
- Actual outcome: WOMAC function at 9 months; Group 1: mean 35.26 (SD 10.48); n=43, Group 2: mean 40.89 (SD 12.64); n=37; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 39.81 (13.75). Baseline usual care: 36.89 (11.49).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five

Protocol outcome 4: Discontinuation at >3 months
- Actual outcome: Dropout at 9 months; Group 1: 8/51, Group 2: 12/49; Comments: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three. Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in all subscales of SF-36 and WOMAC at baseline; Group 1 Number missing: 8, Reason: Treatment package: Death in two people, severe pathology in two, loss of contact in one, drop out in three.; Group 2 Number missing: 12, Reason: Standard care: death in two, severe pathology in one, loss of contact in two, transfer out to other autonomous communities in two, drop out in five

Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	**Oh 2021[211]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=60)
Countries and setting	Conducted in South Korea; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 5 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Clinically and radiologically defined degenerative osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Clinically defined degenerative arthritis; radiologically diagnosed degenerative arthritis (Kellgren-Lawrence grade 2 or less); the ability to walk independently.
Exclusion criteria	Severe arthritis with joint stiffness; neurological comorbidity (stroke, spinal cord disease, etc.); uncontrolled cardiovascular or metabolic diseases; lower limb injuries in the last 6 months.
Recruitment/selection of patients	Residents of Sunchang County who participated in the health education program "Osteoarthritis intervention project for Sunchang County: degenerative arthritis management and prevention" co-hosted by the Health & Longevity Research Institute in Sunchang County and the Public Health Service Project at Seoul National University Bundang Hospital (SNUHB). We collaborated with the Sunchang Health and Medical Center and local public health centers affiliated with Sunchang Health and Medical Center in Ingye, Yulbuk, Geumpyeong, Dongsan, Osan and Mokdong to contact older adults who could participate in the health education program.
Age, gender and ethnicity	Age - Mean (SD): 71.5 (5.8). Gender (M:F): Not stated/unclear. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee

Extra comments	Severity: Not stated/unclear Duration of symptoms: Not stated/unclear
Indirectness of population	No indirectness
Interventions	(n=40) Intervention 1: Treatment package - Exercise and education programme. Education and a self-directed home-based resistance training program. This was performed using a loop band (TheraBand, Akron, OH, USA) in a sitting position on a chair, a standing position with a chair, and a lying position on a mat. The exercise program consisted of a warm-up, the main exercises, and a cool-down, 2 or 3 days a week for 5 months. The warm-up and cool-down consisted of 14 movements; the main exercises included 12 movements with resistance using low-resistance yellow loop bands. The exercise intensity was increased gradually by increasing the number of repetitions every 4 weeks.. Duration 5 months. Concurrent medication/care: Both groups participated in the health education program, which consisted of 50 minutes, once a month for 5 months and was conducted by a multidisciplinary team consisting of doctors, physical education professionals, nurses, nutritionists and exercise experts. The health education program covered 1) the prevention and management of osteoarthritis; 2) lifestyle modification for pain management; 3) self-care strategies for pain relief; 4) nutrition for weight management; 5) ways to improve health-related quality of life.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (5 months). (n=20) Intervention 2: Non-combined active treatment - Education programme. Education programme only. Duration 5 months. Concurrent medication/care: Both groups participated in the health education program, which consisted of 50 minutes, once a month for 5 months and was conducted by a multidisciplinary team consisting of doctors, physical education professionals, nurses, nutritionists and exercise experts. The health education program covered 1) the prevention and management of osteoarthritis; 2) lifestyle modification for pain management; 3) self-care strategies for pain relief; 4) nutrition for weight management; 5) ways to improve health-related quality of life.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (5 months).
Funding	Academic or government funding (This research was supported by Basic Science Research Program through National Research Foundation of Korea funded by the Ministry of Education (NRF-2016R1D1A1B03935518) and by the grant funded by Sunchange County through the Institute on Aging Seoul National University (0564-2016006).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Pain at >3 months
- Actual outcome: WOMAC pain at 5 months; Group 1: mean 5.06 (SD 4.39); n=21, Group 2: mean 10.33 (SD 5.22); n=11; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 5.00 (4.50). Baseline control: 6.67 (5.15).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, blood

	pressure, medical history, current treatment and baseline values of outcomes; Group 1 Number missing: 19, Reason: Treatment package: 5 dropped out before the end of the intervention, 14 could not be evaluated afterwards because they were busy farming (10) or could not travel (4).; Group 2 Number missing: 10, Reason: Education: 5 dropped out before the end of the intervention, 5 could not be evaluated as they were busy farming. Protocol outcome 2: Physical function at >3 months - Actual outcome: WOMAC function at 5 months; Group 1: mean 16.22 (SD 10.87); n=21, Group 2: mean 30.89 (SD 14.09); n=11; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 15.06 (11.21). Baseline control: 22.11 (19.32). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, blood pressure, medical history, current treatment and baseline values of outcomes; Group 1 Number missing: 19, Reason: Treatment package: 5 dropped out before the end of the intervention, 14 could not be evaluated afterwards because they were busy farming (10) or could not travel (4).; Group 2 Number missing: 10, Reason: Education: 5 dropped out before the end of the intervention, 5 could not be evaluated as they were busy farming. Protocol outcome 3: Discontinuation at >3 months - Actual outcome: Dropouts at 5 months; Group 1: 14/40, Group 2: 10/20; Comments: Treatment package: 5 dropped out before the end of the intervention, 14 could not be evaluated afterwards because they were busy farming (10) or could not travel (4). Education: 5 dropped out before the end of the intervention, 5 could not be evaluated as they were busy farming. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI, blood pressure, medical history, current treatment and baseline values of outcomes; Group 1 Number missing: 19, Reason: Treatment package: 5 dropped out before the end of the intervention, 14 could not be evaluated afterwards because they were busy farming (10) or could not travel (4).; Group 2 Number missing: 10, Reason: Education: 5 dropped out before the end of the intervention, 5 could not be evaluated as they were busy farming.
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	**Paterson 2021[222]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=30)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up.

Line of therapy	Unclear
Duration of study	Intervention + follow up: 3 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: First metatarsophalangeal osteoarthritis defined as radiographic osteoarthritis (a score of at least 2 for osteophytes or joint space narrowing on either the anteroposterior and lateral views, according to a radiographic atlas), self-reported pain at least 4 for an 11-point numerical rating scale in the corresponding first MTP joint region on most days of the previous month.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age > 40 years with symptomatic radiographic first MTP joint OA.
Exclusion criteria	Inflammatory or systemic arthritis; current or past 6 months use of existing foot orthoses or intra-articular foot injections; a history of musculoskeletal foot surgery; other muscular, joint or neurologic condition affecting lower extremity function; inability to walk unaided; pain in any other location that is worse than their first MTP joint pain; significant hallux valgus deformity (grade 3 or 4 on the Manchester Scale); or current treatment by a podiatrist or physical therapist for foot pain.
Recruitment/selection of patients	People were recruited from the community using advertisements, their network of medical and allied health practitioners and their volunteer database.
Age, gender and ethnicity	Age - Mean (SD): 59.0 (7.83). Gender (M:F): 19M/ 11F. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Toe
Extra comments	Severity: Osteophyte grade 2-3, joint space narrowing grade 1-3 (median grade for both = 2). Duration of symptoms (mean [SD]): 8.5 (6.5) years
Indirectness of population	No indirectness
Interventions	(n=15) Intervention 1: Treatment package - Combination and behaviour change intervention. Foot orthoses and a self management program. The foot orthoses (a wedged insole) were fitted to participants' footwear at an initial clinic visit, and they were advised to wear the device for >6 hours/day. Orthoses were reviewed at follow up visits 1 and 5 weeks later, and additional modifications were made to address comfort or adverse events. The home management program was performed twice daily. This included a exercises including: isometric flexor hallucis longus strength exercises, 3 sets of 10-20 repetitions, first MTP joint distraction (1 minute) and distal glides (1 minute), and soft tissue massage to the plantar foot (5 minutes) using a massage ball. The self management advice and plan included: wearing shoes with adequate depth and width; advice on analgesia (a maximum of 4 grams/day of paracetamol if needed), weight management (general advice and dietitian referral if needed) and physical activity (30 minutes on most days).. Duration 3 months. Concurrent medication/care: Usual care was provided to all. People in this treatment group attended one 15-minute visit with a GP at which they received advice and/or prescription of analgesics and antiinflammatory medication at the discretion of the GP. In addition, the GP was also provided advice on weight management (general advice and dietitian referral if needed) and physical activity (30 minutes on most days).

	Participants were permitted additional visits if they experienced an ongoing problem related to the treatment, and this addition was documented by the GP.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (3 months). (n=15) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care only. Duration 3 months. Concurrent medication/care: Usual care was provided to all. People in this treatment group attended one 15-minute visit with a GP at which they received advice and/or prescription of analgesics and antiinflammatory medication at the discretion of the GP. In addition, the GP was also provided advice on weight management (general advice and dietitian referral if needed) and physical activity (30 minutes on most days). Participants were permitted additional visits if they experienced an ongoing problem related to the treatment, and this addition was documented by the GP.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (3 months).
Funding	Equipment / drugs provided by industry (Foot Science International supplied the orthoses. Supported by a National Health and Medical Research Council Program Grant (grant 1091302) and grants from Arthritis Australia and the Australian Podiatry Education and Research Foundation. Dr. Hinman's work was supported by a National Health and Medical Research Council Senior Research Fellowship (grant 1154217). Dr. Menz's work was supported by a National Health and Medical Research Council Senior Research Fellowship (grant APP1135995). Dr. Bennell's work was supported by a National Health and Medical Research Council Principal Research Fellowship (grant APP1058440).)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: Foot Health Status Questionnaire Pain domain at 3 months; Group 1: mean 22.5 (SD 17.3); n=14, Group 2: mean 24.3 (SD 24.3); n=12; Foot Health Status Questionnaire Pain domain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 52.3 (19.0). Baseline usual care: 50.1 (17.6).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, body mass index, symptom duration, side of symptoms, radiographic grades, FPI scores and hallux valgus stage.; Group 1 Number missing: 1, Reason: Reasons not provided; Group 2 Number missing: 3, Reason: Reasons not provided

Protocol outcome 2: Physical function at ≤3 months
- Actual outcome: Foot Health Status Questionnaire Function domain at 3 months; Group 1: mean 18.3 (SD 15.8); n=14, Group 2: mean 13.6 (SD 10.4); n=12; Foot Health Status Questionnaire Function domain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 67.9 (19.6). Baseline usual care: 78.4 (17.1).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, body mass index, symptom duration, side of symptoms, radiographic grades, FPI scores and hallux valgus stage.; Group 1 Number missing: 1, Reason: Reasons not provided; Group 2 Number missing: 3, Reason: Reasons not provided Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 3 months; Group 1: 1/15, Group 2: 3/15; Comments: Reasons not provided. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, height, mass, body mass index, symptom duration, side of symptoms, radiographic grades, FPI scores and hallux valgus stage.; Group 1 Number missing: 1, Reason: Reasons not provided; Group 2 Number missing: 3, Reason: Reasons not provided	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study (subsidiary papers)	Poulsen 2013[228] (Poulsen 2011[227], Poulsen 2013[229])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=118)
Countries and setting	Conducted in Denmark; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks of interventon, 1 year follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Unilateral hip pan for >3 months' duration with radiographic hip osteoarthritis defined as minimal joint space width (JSW) measurement <2.00mm or a side difference in minimal JSW >10%
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Unilateral hip pain >3 months' duration; age 40-80 years; radiographic hip osteoarthritis; ability to speak and read Danish
Exclusion criteria	Other conditions than hip osteoarthritis appearing to be the ca months; previous hip or knee joint replacement surgery; hip osteoarthritis due to hip fracture or infection; rating of worst hip pain during the last week as at least 2 on an 11-box rating scale; hip dysplasia, Center Edge angle <35 and Acetabular Index Angle >10; local knee pain originating from the knee on the same side as the hip osteoarthritis; low back pain dominating over the hip symptoms; inflammatory joint disease; cerebrovascular disease; polyneuropathy or neuromuscular disease; malignant disease; refusal to participate
Recruitment/selection of patients	People were recruited from primary care practices. Information about the project was made available on a closed web site for health care professionals by the Region of Southern Denmark.
Age, gender and ethnicity	Age - Mean (SD): 64.6 (8.6). Gender (M:F): 63:48. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms (mean [SD]): 32 (36) months
Indirectness of population	No indirectness
Interventions	(n=43) Intervention 1: Treatment package - Manual therapy and education programme. Hip school and manual therapy. Hip school involved 5 sessions delivered over 6 weeks consisting of one initial personal interview, three group sessions, and

	one follow-up personal session. The content included information about epidemiology of hip osteoarthritis, anatomy of the hip joint and adjacent functional structures, pain distribution and diagnosis of hip osteoarthritis, recommended activity levels, natural course of the diseas,e and finally information about treatment options. Stretching exercises were taught and instructions were given on how to incorporate these into a daily routine. The manual therapy was a combination of manual soft tissue therapy, stretching and joint manipulation. The soft tissue therapy is trigger point pressure release. The soft tissue stretching is based on muscle energy techniques. The joint manipulation is one of high velocity low amplitude. The purpose of the manual therapy was to improve elasticity of the muscular, ligamentous and capsular tissue of the hip and posterior joints of the pelvis. Combination of treatment modalities was individualised to each person according to examination findings at the discretion of the treating clinician. Treatment sessions lasted 15-20 minutes and was administered twice a week during the 6 weeks. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme (Hip school). 2. Length of package: ≤ 6 weeks (6 weeks). (n=39) Intervention 2: Non-combined active treatment - Education programme. Hip school program only. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme (Hip school). 2. Length of package: ≤ 6 weeks (6 weeks). (n=36) Intervention 3: Standard care (non-organised) or no treatment - Standard care (non-organised). Minimal intervention. People were given a leaflet describing the stretching exercises from hip school and received a short 5 minute instruction in self-care immediately after randomisation. People were advised to live as ususal, not to make any changes to use of possible pain medication or to initiate any other treatment during the following 6 weeks. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Academic or government funding (This study was funded by the Danish Foundation for Chiropractic Research and Postgraduate Education, Region of Southern Denmark, Danish Rheumatism Association and university of Southern Denmark)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EDUCATION PROGRAMME versus EDUCATION

PROGRAMME

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: HOOS hip-related quality of life at 6 weeks; Group 1: mean 12 (SD 18); n=38, Group 2: mean -2 (SD 11); n=37; HOOS hip-related quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 52 (17). Baseline education: 53 (18).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 3, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment.

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: HOOS hip-related quality of life at 12 months; Group 1: mean 10 (SD 20); n=38, Group 2: mean 10 (SD 27); n=37; HOOS hip-related quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 52 (17). Baseline education: 53 (18).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 15, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment. 12 had arthroplasty.

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: HOOS pain at 6 weeks; Group 1: mean 18 (SD 13); n=38, Group 2: mean -1 (SD 11); n=37; HOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 62 (17). Baseline education: 64 (14).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 3, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment.

Protocol outcome 4: Pain at >3 months
- Actual outcome: HOOS pain at 12 months; Group 1: mean 16 (SD 20); n=38, Group 2: mean 11 (SD 23); n=37; HOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 62 (17). Baseline education: 64 (14).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1

bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 15, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment. 12 had arthroplasty.

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: HOOS function in daily living at 6 weeks; Group 1: mean 15 (SD 16); n=38, Group 2: mean 1 (SD 10); n=37; HOOS function of daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 68 (20). Baseline education: 68 (15).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 3, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment.

Protocol outcome 6: Physical function at >3 months
- Actual outcome: HOOS function in daily living at 12 months; Group 1: mean 13 (SD 20); n=36, Group 2: mean 9 (SD 21); n=37; HOOS function in daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 68 (20). Baseline education: 68 (15).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 15, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment. 12 had arthroplasty.

Protocol outcome 7: Discontinuation at ≤3 months
- Actual outcome: Withdrew at 6 weeks; Group 1: 4/38, Group 2: 1/37; Comments: Treatment package: 1 had surgery (not hip), 1 wanted operation, 1 time constraint, 1 got worse from manual therapy. Education: 1 due to lack of commitment.
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 3, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment.

Protocol outcome 8: Discontinuation at >3 months
- Actual outcome: Withdrew at 12 months; Group 1: 7/38, Group 2: 13/37; Comments: Treatment package: 1 had surgery (not hip), 1 wanted operation, 1 time constraint, 1 got worse from manual therapy, 3 had arthroplasty. Education: 1 due to lack of commitment, 12 had arthroplasty.
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 15, Reason: 2 excluded after randomisation (bilateral hip pain). 1 withdrew due to lack of commitment. 12 had arthroplasty.

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: HOOS hip-related quality of life at 6 weeks; Group 1: mean 12 (SD 18); n=38, Group 2: mean 4 (SD 10); n=36; HOOS hip-related quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 52 (17). Baseline minimal care: 46 (12).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 4, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation)

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: HOOS hip-related quality of life at 12 months; Group 1: mean 10 (SD 20); n=38, Group 2: mean 12 (SD 21); n=36; HOOS hip-related quality of life 0-100 Top=High is good outcome; Comments: Baseline treatment package: 52 (17). Baseline minimal care: 46 (12).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 10, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation). 6 had arthroplasty.

Protocol outcome 3: Pain at ≤3 months
- Actual outcome: HOOS pain at 6 weeks; Group 1: mean 18 (SD 13); n=38, Group 2: mean 3 (SD 13); n=36; HOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 62 (17). Baseline minimal care: 58 (14).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 4, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation)

Protocol outcome 4: Pain at >3 months

- Actual outcome: HOOS pain at 12 months; Group 1: mean 16 (SD 20); n=38, Group 2: mean 13 (SD 18); n=36; HOOS pain 0-100 Top=High is good outcome; Comments: Baseline treatment package: 62 (17). Baseline minimal care: 58 (14).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 10, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation). 6 had arthroplasty.

Protocol outcome 5: Physical function at ≤3 months
- Actual outcome: HOOS function in daily living at 6 weeks; Group 1: mean 15 (SD 16); n=38, Group 2: mean 5 (SD 13); n=36; HOOS function in daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 68 (20). Baseline minimal care: 64 (15).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 4, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation)

Protocol outcome 6: Physical function at >3 months
- Actual outcome: HOOS function in daily living at 12 months; Group 1: mean 13 (SD 20); n=38, Group 2: mean 10 (SD 18); n=36; HOOS function in daily living 0-100 Top=High is good outcome; Comments: Baseline treatment package: 68 (20). Baseline minimal care: 64 (15).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 10, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation). 6 had arthroplasty.

Protocol outcome 7: Discontinuation at ≤3 months
- Actual outcome: Withdrew at 6 weeks; Group 1: 4/38, Group 2: 4/36; Comments: Treatment package: 1 had surgery (not hip), 1 wanted operation, 1 time constraint, 1 got worse from manual therapy. Minimal control: 3 disappointed with group, 1 wanted operation.
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 9, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy).; Group 2 Number missing: 4, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation)

Protocol outcome 8: Discontinuation at >3 months

- Actual outcome: Withdrew at 12 months; Group 1: 7/38, Group 2: 10/36; Comments: Treatment package: 1 had surgery (not hip), 1 wanted operation, 1 time constraint, 1 got worse from manual therapy, 3 had arthroplasty. Minimal control: 3 disappointed with group, 1 wanted operation, 6 had arthroplasty. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, BMI, involved side, duration of symptoms, radiographic parameters and baseline values of outcomes; Group 1 Number missing: 12, Reason: 5 excluded after randomisation (1 bilateral hip pain, 2 polyarthritis, 1 no hip osteoarthritis, 1 leg neuropathy), 4 withdrew (1 had surgery [not hip], 1 wanted operation, 1 due to time constraint, 1 got worse from manual therapy). 3 had arthroplasty.; Group 2 Number missing: 10, Reason: 4 withdrew (3 disappointed with group, 1 wanted operation). 6 had arthroplasty.	
Protocol outcomes not reported by the study	Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Quilty 2003[230]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=87)
Countries and setting	Conducted in United Kingdom; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 10 weeks of intervention, 12 months of total follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Chronic knee or hip pain with radiographic evidence of knee osteoarthritis (Kellgren Lawrence grade less than and equal to 2)
Stratum	Overall:
Subgroup analysis within study	Not applicable
Inclusion criteria	Participants from the original SASH cohort study who reported chronic knee or hip pain to a postal questionnaire who had radiographic evidence of patellofemoral joint osteophytes in the absence of advanced radiographic changes of hip or tibiofemoral joint osteoarthritis (grade 3 Kellgren Lawrence score and above).
Exclusion criteria	Previous major knee surgery; fractures involving the knee joint or rheumatoid arthritis
Recruitment/selection of patients	People were recruited from a large community cohort study
Age, gender and ethnicity	Age - Mean (SD): 66.8 (10.4). Gender (M:F): Not stated. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=43) Intervention 1: Treatment package - Combination and education programme. Physiotherapy and patellar taping, postural, footwear and weight reduction advice delivered in 9 sessions over 10 weeks lasting half an hour each. Physiotherapy exercises included: vastus medialis oblique muscle contractions in sitting position (squeezing a rolled-up towel between the knees); exercise 1 with gluteal muscle contractions at the same time; controlled sitting to standing squeezing a rolled-up towel between the knees to encourage contraction of the VMO muscle; controlled small knee bends squeezing a rolled-up towel; controlled stepping up and down steps emphasizing contraction of the VMO muscle and correct posture; 10 maximal isometric quadriceps contractions in mid-range (roughly 70 degrees) using a resistive

	rubber band; controlled balancing on one leg for as long as possible. All exercises were tailored to a person's ability to perform them without pain. All exercises were to be pain-free and performed 10 times each, 5 times a day, except for exercise 6, which was to be performed once each day. Medial patellar taping was applied during an activity that produced their pain to see if this provided benefit. The tape was adjusted to ensure there was at least 50% improvement. If there was no improvement in pain, the tape was not used. At subsequent sessions people were taught how to apply the tape and prevent skin problems developing. They were told to wear the tape only if it was effective in reducing their pain. Posture correction emphasized the correct alignment of the lower limb in standing and during activity. Footwear advice concentrated on wearing shoes that provided shock absorption and supported the medial arches. Weight reduction was advised for overweight patients. Duration 10 weeks. Concurrent medication/care: All people were given an information sheet and encouraged to continue with the exercises after the formal period of supervised therapy. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme (Education about footwear and weight loss weekly over the period of time). 2. Length of package: > 6 weeks (10 weeks). (n=44) Intervention 2: Non-combined active treatment - Exercise. Physiotherapy exercise only. Duration 10 weeks. Concurrent medication/care: All people were given an information sheet and encouraged to continue with the exercises after the formal period of supervised therapy. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (10 weeks).
Funding	Academic or government funding (Supported by the NHS Research and Development programme (Physical and Complex Disabilities PCD|A1|123))

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus EXERCISE

Protocol outcome 1: Pain at >3 months
- Actual outcome: VAS pain index knee at 12 months; Group 1: mean 48.1 (SD 25.7); n=43, Group 2: mean 54.1 (SD 22.5); n=44; VAS pain index knee 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 51.0 (29.3). Baseline exercise: 53.4 (25.9).
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI and baseline values of outcomes; Group 1 Number missing: 10, Reason: 5 did not receive the intervention as allocated. 1 withdrawn. 3 lost to follow up. 1 other.; Group 2 Number missing: 1, Reason: 1 lost to follow up

Protocol outcome 2: Physical function at >3 months - Actual outcome: WOMAC function at 12 months; Group 1: mean 29.7 (SD 11.2); n=43, Group 2: mean 28.3 (SD 11.3); n=44; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 27.4 (12.2). Baseline exercise: 27.8 (10.1). Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI and baseline values of outcomes; Group 1 Number missing: 10, Reason: 5 did not receive the intervention as allocated. 1 withdrawn. 3 lost to follow up. 1 other.; Group 2 Number missing: 1, Reason: 1 lost to follow up Protocol outcome 3: Discontinuation at >3 months - Actual outcome: Discontinuation at 12 months; Group 1: 5/43, Group 2: 1/44; Comments: Treatment package: 1 withdrawn, 3 lost to follow up, 1 other. Exercise: 1 lost to follow up. Risk of bias: All domain - High, Selection - Low, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, height, weight, BMI and baseline values of outcomes; Group 1 Number missing: 10, Reason: 5 did not receive the intervention as allocated. 1 withdrawn. 3 lost to follow up. 1 other.; Group 2 Number missing: 1, Reason: 1 lost to follow up	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	**Rezende 2021[236]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=222)
Countries and setting	Conducted in Brazil; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 24 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Knee osteoarthritis according to the American College of Rheumatology clinical and radiological definitions with Kellgren & Lawrence stages 1-3
Stratum	Overall
Subgroup analysis within study	Not applicable

Inclusion criteria	Outpatients aged 40 years or older with knee osteoarthritis (according to the American College of Rheumatology clinical and radiological definitions with Kellgren & Lawrence stages 1-3) with indications for clinical treatments for osteoarthritis and the ability to understand and provide informed consent
Exclusion criteria	Rheumatologic diseases (other than osteoarthritis); neurological problems or instability that would prevent them from exercises; participating in another program with nutritional guidance.
Recruitment/selection of patients	Volunteers from a waiting list of knee osteoarthritis clinical treatment. People were either patients from the knee group of the institution with osteoarthritis but without indication for surgery or people referred by employees of the hospital and by people in a pilot program.
Age, gender and ethnicity	Age - Mean (SD): 63.5 (9.1). Gender (M:F): 36:155. Ethnicity: Not stated/unclear
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren and Lawrence grace 1-3 (median grade 2). Duration of symptoms: Not stated/unclear
Indirectness of population	No indirectness
Interventions	(n=111) Intervention 1: Treatment package - Exercise and behaviour change intervention. Two days of a structured educational and exercise-based self-management program that were held two months apart. The people received written and video educational material on the first intervention day, with the material describing what was taught in the interventions and including directions for all of the community centers and primary and secondary care centers of the city of Sao Paulo, where the participants could continue the program near their home (in case they did not want to exercise alone at home). The directions were compiled by the social workers. The interventions were conducted over 2 separate days of classes: the first day was conducted from 8:00 a.m. to 5:00 p.m. each of the lecturers lasted approximately 40 minutes to an hour and were provided by a total of seven different teams. The orthopedic surgeons explained the anatomy, joints, the osteoarthritis disease, the risk factors and the treatment modalities to the participants. The psychologists discussed personality characteristics that exist from childhood to adulthood as well as the difference between having a disease and being sick, the importance of their choices (and not their conditions or feelings) and coping skills. The nutritionist emphasised the importance of a well-balanced diet (reduced quantity of food as well as the importance of colourful whole grains and low-calorie meals). During the intervention, patients were given a break every 3 hours and were provided with meals that followed the instructions that were given by the nutritionist; thus, the patients personally experienced a day's worth of the options of the proposed diet. The physical therapists re-enforced the importance of the previously mentioned nutritional options as well as the importance of hydration. Additionally, they introduced the benefits of regularly performing physical exercises (at least three times a week as part of a group and with comfortable clothing), types of exercise (stretching, isometric and isotonic strengthening), adequate posture when performing exercises and the importance of controlled load (with respect to personal limitations). The therapists clarified to the patients that the benefits of the exercises are in the prevention and control of preexisting diseases. They also explained the difference between physical activity and physical exercise. The occupational therapists introduced the importance of protecting the joints during daily activities by optimising ergonomics and by alternating among different levels of energy expenditure. The physical education professional (similarly to the physical therapists) also lectured about the health-related benefits of physical exercise

	and its role in knee osteoarthritis management as well as on the differences between physical activity and exercise and the importance and methods of how to improve physical fitness. After these lectures the patients participated in the following 3 workshops that lasted 50 minutes each: physical therapy with stretching, isometric exercises and isotonic execises, instructions on how to use weights to progressively increase the weight loads and instruction to perform exercises at least three times per week; the physical fitness workshop which focused on performing exercise (resistance and aerobic types) at home by using low-cost alternative tools as well as how to exercise at the appropriate intensity; the occupational therapy team instructed people in a simulated safe house how to protect their joints in daily living activities. The second day of the intervention included social workers who asked about habits, the nutritionist who reviewed the slides from their previous lecture, and workshops similar to those from before with the introduction of a psychology workshop that focused on psychological educational and therapeutic group sessions with the patients, with a focus on what the patients had done since the time of the first intervention, in respect to their arthritis. People continued exercise at home.. Duration 24 months. Concurrent medication/care: People in both groups were seen by the orthopaedic surgeons at inclusion, six, 12 and 24 months. At inclusion people were already receiving diacerhein and/or analgesics such as paracetamol, codeine and/or dipyrone that were prescribed by the physicians when people were first seen. At each visit, the medical team explained the disease and its forms of treatment based on international guidelines and prescribed whatever services they considered appropriate including the need to diet and exercise, orthotics and medications to each patient.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (24 months). (n=111) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care only. Duration 24 months. Concurrent medication/care: People in both groups were seen by the orthopaedic surgeons at inclusion, six, 12 and 24 months. At inclusion people were already receiving diacerhein and/or analgesics such as paracetamol, codeine and/or dipyrone that were prescribed by the physicians when people were first seen. At each visit, the medical team explained the disease and its forms of treatment based on international guidelines and prescribed whatever services they considered appropriate including the need to diet and exercise, orthotics and medications to each patient.. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (24 months).
Funding	Study funded by industry (This study was funded by the Department of Orthopedics (Hospital das Clinicas), Faculdade de Medicina da Universidade de Sao Paulo, with the participation of TRB Pharma Brasil. The Department of Orthopedics (Hospital das Clinicas), Faculdade de Medicina da Universidade de Sao Paulo, supplied the laboratory and imaging exams, the physical structure, the human participants (both professional individuals and patients) and the medications. TRB Pharma Brasil funded the logistical and audio-visual material of the program as well as the statistics and meeting presentations of the program.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Pain at >3 months
- Actual outcome: WOMAC pain at 24 months; Group 1: mean 8.1 (SD 3.9); n=95, Group 2: mean 9.4 (SD 4.5); n=96; WOMAC pain 0-20 Top=High is poor

outcome; Comments: Baseline treatment package: 10.3 (3.9). Baseline standard care: 10.7 (4.1). Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, years of schooling, physical activity, gender, race, Kellgren Lawrence scale and baseline values of outcomes; Group 1 Number missing: 17, Reason: Treatment package: 1 death, 15 missed interventions, 1 lost to follow-up.; Group 2 Number missing: 15, Reason: Standard care: 15 withdrawal. Protocol outcome 2: Physical function at >3 months - Actual outcome: WOMAC function at 24 months; Group 1: mean 30 (SD 13.8); n=95, Group 2: mean 33.6 (SD 14.2); n=96; WOMAC function 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 37.3 (13.8). Baseline standard care: 38.2 (14.3). Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, years of schooling, physical activity, gender, race, Kellgren Lawrence scale and baseline values of outcomes; Group 1 Number missing: 17, Reason: Treatment package: 1 death, 15 missed interventions, 1 lost to follow-up.; Group 2 Number missing: 15, Reason: Standard care: 15 withdrawal. Protocol outcome 3: Discontinuation at >3 months - Actual outcome: Discontinuation at 24 months; Group 1: 17/111, Group 2: 15/111; Comments: Treatment package: 1 death, 15 missed interventions, 1 lost to follow-up. Standard care: 15 withdrawal. Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, years of schooling, physical activity, gender, race, Kellgren Lawrence scale and baseline values of outcomes; Group 1 Number missing: 17, Reason: Treatment package: 1 death, 15 missed interventions, 1 lost to follow-up.; Group 2 Number missing: 15, Reason: Standard care: 15 withdrawal.	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Saw 2016[244]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=71)
Countries and setting	Conducted in South Africa; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks of intervention, 6 months of follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: People diagnosed with osteoarthritis who had been placed on the waiting list to receive a hip/knee arthroplasty
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Willingness to commit to the study; aged between 50-70 years; diagnosed with osteoarthritis of the hip/knee; literate in English, Afrikaans, isiXhosa or isiZulu
Exclusion criteria	Any cognitive impairment, as reported in the medical recorders; previous trauma/surgery to the unaffected leg; deemed not eligible for exercise as per the American College of Sports Medicine guidelines for exercise prescription. Reasons for exclusion according to the ACSM included previous cardiac conditions or surgery, uncontrolled diabetes or asthma; those who had previously taken part in a six-week program aimed at improving self-efficacy and management.
Recruitment/selection of patients	People were contacted from waiting lists at the Tygerberg and Helen Joseph Hospital
Age, gender and ethnicity	Age - Mean (SD): 60.72 (5.54). Gender (M:F): 14:60. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Mixed (Hip or knee).
Extra comments	Severity: Not stated Duration of symptoms: Not stated.
Indirectness of population	No indirectness
Interventions	(n=35) Intervention 1: Treatment package - Exercise and education programme. Exercise and education component. The exercise component allowed people to apply what they learnt. It comprised of various stretching, light aerobic exercise and different lower limb muscle group strengthening exercises. People are required to set exercise goals. The exercise component commenced at low repetitions and intensity and was progressed weekly from 20 minutes in duration, increasing time by 10% and intensity as appropriate, depending on each participant's individual ability. The intervention

	concluded with a relaxation session led by the physiotherapist facilitating various relaxation visualisations. The educational component was aimed at increasing knowledge and understanding of osteoarthritis; pain neuroscience; activity and related topics affected by their condition. Important topics such as self-management skills, problem solving, goal setting, coping mechanisms, stress management, and pacing were discussed to enable the participant in self-management. Each person received a "living with osteoarthritis" workbook (in their preferred language). Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (6 weeks). (n=39) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Academic or government funding (Funding received from South African society of Physiotherapy, Margaret Roper Scholarship and UCT PG funding)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: EQ-5D at 12 weeks; Group 1: mean 0.6 (SD 0.32); n=35, Group 2: mean 0.36 (SD 0.35); n=39; EQ-5D -0.11-1.0 Top=High is good outcome; Comments: Baseline treatment package: 0.36 (0.34). Baseline control: 0.38 (0.32).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported overall values for most baseline values, baseline values of outcomes were similar; Group 1 Number missing: 6, Reason: 2 withdrew from study, 2 falling ill, 1 receiving surgery, 1 work responsibility; Group 2 Number missing: 11, Reason: 2 withdrawing from study, 1 falling ill, 4 forgetting, 2 receiving surgery, 2 work responsibility

Protocol outcome 2: Quality of life at >3 months
- Actual outcome: EQ-5D at 6 months; Group 1: mean 0.55 (SD 0.34); n=35, Group 2: mean 0.37 (SD 0.34); n=39; EQ-5D -0.11-1 Top=High is good outcome; Comments: Baseline treatment package: 0.36 (0.34). Baseline control: 0.38 (0.32).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported overall values for most baseline values, baseline values of outcomes were similar; Group 1 Number missing: 9, Reason: 2 withdrew from study, 3 falling ill, 3 receiving surgery, 1 funeral; Group 2 Number missing: 10, Reason: 2 withdrawing from study, 1 falling ill, 3 forgetting, 2 receiving surgery, 2 transport

Protocol outcome 3: Pain at ≤3 months - Actual outcome: Brief pain inventory - severity at 12 weeks; Group 1: mean 4.34 (SD 2.86); n=35, Group 2: mean 6.05 (SD 2.34); n=39; BPI severity 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 6.71 (2.32). Baseline control: 6.37 (2.16). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported overall values for most baseline values, baseline values of outcomes were similar; Group 1 Number missing: 6, Reason: 2 withdrew from study, 2 falling ill, 1 receiving surgery, 1 work responsibility; Group 2 Number missing: 11, Reason: 2 withdrawing from study, 1 falling ill, 4 forgetting, 2 receiving surgery, 2 work responsibility Protocol outcome 4: Pain at >3 months - Actual outcome: Brief pain inventory - severity at 6 months; Group 1: mean 4.49 (SD 2.85); n=35, Group 2: mean 6.39 (SD 2.3); n=39; BPI severity 0-10 Top=High is poor outcome; Comments: Baseline treatment package: 6.71 (2.32). Baseline control: 6.37 (2.16). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported overall values for most baseline values, baseline values of outcomes were similar; Group 1 Number missing: 9, Reason: 2 withdrew from study, 3 falling ill, 3 receiving surgery, 1 funeral; Group 2 Number missing: 10, Reason: 2 withdrawing from study, 1 falling ill, 3 forgetting, 2 receiving surgery, 2 transport Protocol outcome 5: Discontinuation at ≤3 months - Actual outcome: Attrition rate at 12 weeks; Group 1: 6/35, Group 2: 11/39; Comments: Treatment package: 2 withdrawing from study, 2 falling ill, 1 transport, 1 receiving surgery. Standard care: 2 withdrawing from study, 1 falling ill, 4 forgetting, 2 receiving surgery, 2 work responsibility. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported overall values for most baseline values, baseline values of outcomes were similar; Group 1 Number missing: 6, Reason: 2 withdrew from study, 2 falling ill, 1 receiving surgery, 1 work responsibility; Group 2 Number missing: 11, Reason: 2 withdrawing from study, 1 falling ill, 4 forgetting, 2 receiving surgery, 2 work responsibility Protocol outcome 6: Discontinuation at >3 months - Actual outcome: Attrition rate at 6 months; Group 1: 9/35, Group 2: 10/39; Comments: Treatment package: 2 withdrawing from study, 3 falling ill, 3 receiving surgery, 1 other (funeral). Standard care: 2 withdrawing from study, 1 falling ill, 3 forgetting, 2 transport, 2 receiving surgery. Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported overall values for most baseline values, baseline values of outcomes were similar; Group 1 Number missing: 9, Reason: 2 withdrew from study, 3 falling ill, 3 receiving surgery, 1 funeral; Group 2 Number missing: 10, Reason: 2 withdrawing from study, 1 falling ill, 3 forgetting, 2 receiving surgery, 2 transport	
Protocol outcomes not reported by the study	Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study (subsidiary papers)	SMOotH trial: Dziedzic 2015[94] (Dziedzic 2011[96], Oppong 2014[213])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=257)
Countries and setting	Conducted in United Kingdom; Setting: Primary care
Line of therapy	Unclear
Duration of study	Intervention + follow up: 4 weeks of intervention, 12 months of follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Meeting the American College of Rheumatology criteria for features of hand osteoarthritis, or had unilateral or bilateral thumb base osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People aged 50 years or over who reported hand pain in the last 12 months; reported hand pain, aching or stiffness on 'some days', 'most days' or 'all days' in the last month; had an AUSCAN pain score of at least 5 or an AUSCAN function score of at least 9
Exclusion criteria	Reported that they had seen an occupational therapist or physiotherapist for their hand problem in the last 6 months; had a hand operation, injection or injured their hands badly enough to see a doctor in the previous 6 months; had other members of their household participating in the trial
Recruitment/selection of patients	People were registered with five general practices in Central Cheshire and North Staffordshire, UK
Age, gender and ethnicity	Age - Mean (SD): 65.8 (9.1). Gender (M:F): 87:170. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hand
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=65) Intervention 1: Treatment package - Exercise and behaviour change intervention. Joint protection instruction and hand exercises. Hand exercises including stretching exercises; wrist flexion and extension, pronation and supination, tendon gliding, radial finger walking, making an 'O' with the thumb and index finger, thumb extension, abduction and opposition to the base of the 5th finger and strengthening exercises using an elastic band and play-doh. The joint protection principles included:

	distributing the weight of what you lift over several joints; avoiding putting strain on the thumb and repetitive thumb movements; avoiding prolonged grips in one position; using as large a grip as possible; reducing the effort needed to do a task; and energy conservation. These were delivered over 4 weekly sessions lasting 1.5 hour (individual components were allowed to last up to a maximum of 1 hour). Duration 4 weeks. Concurrent medication/care: All people were given standardised written information on self-management approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using, and were given advice to consult their general practitioner if symptoms continued to be troublesome. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Joint protection). 2. Length of package: ≤ 6 weeks (4 weeks). (n=65) Intervention 2: Non-combined active treatment - Exercise. Hand exercises only. Sessions lasted for at most 1 hour. Duration 4 weeks. Concurrent medication/care: All people were given standardised written information on self-management approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using, and were given advice to consult their general practitioner if symptoms continued to be troublesome. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (4 weeks). (n=62) Intervention 3: Non-combined active treatment - Behaviour change intervention. Joint protection sessions only. Each session could last up to a maximum of 1 hour. Duration 4 weeks. Concurrent medication/care: All people were given standardised written information on self-management approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using, and were given advice to consult their general practitioner if symptoms continued to be troublesome. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention (Joint protection). 2. Length of package: ≤ 6 weeks (4 weeks).

	(n=65) Intervention 4: Standard care (non-organised) or no treatment - No treatment. No additional treatments. Duration 4 weeks. Concurrent medication/care: All people were given standardised written information on self-management approaches for hand osteoarthritis (including information on looking after hand joints and using analgesia). People were advised to continue with any self-management approaches they were currently using, and were given advice to consult their general practitioner if symptoms continued to be troublesome. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (4 weeks).
Funding	Academic or government funding (The trial was funded b the Arthritis Research UK ISRCTN 33870549)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus EXERCISE

Protocol outcome 1: Discontinuation at ≤3 months
- Actual outcome: Withdrew at 3 months; Group 1: 2/65, Group 2: 1/65; Comments: Treatment package: 2 did not want to take part. Exercise: 1 work commitments.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, marital status, routine or manual occupation, currently working, age when left school, left school to go to full-time education or university, gained qualifications through study as an adult and baseline values of outcomes; Group 1 Number missing: 2; Group 2 Number missing: 1

Protocol outcome 2: Discontinuation at >3 months
- Actual outcome: Withdrew at 12 months; Group 1: 6/65, Group 2: 6/65; Comments: Treatment package: 4 did not want to take part, 1 recent bereavement, 1 ill health in the family. Exercise: 1 work commitments, 4 did not want to take part, 1 ill health in the family.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, marital status, routine or manual occupation, currently working, age when left school, left school to go to full-time education or university, gained qualifications through study as an adult and baseline values of outcomes; Group 1 Number missing: 6; Group 2 Number missing: 6

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus BEHAVIOUR CHANGE INTERVENTION

Protocol outcome 1: Discontinuation at ≤3 months
- Actual outcome: Withdrew at 3 months; Group 1: 2/65, Group 2: 3/62; Comments: Treatment package: 2 did not want to take part. Joint protection: 1 family problems, 2 did not want to take part.
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, marital status, routine or manual

occupation, currently working, age when left school, left school to go to full-time education or university, gained qualifications through study as an adult and baseline values of outcomes; Group 1 Number missing: 2; Group 2 Number missing: 3 Protocol outcome 2: Discontinuation at >3 months - Actual outcome: Withdrew at 12 months; Group 1: 6/65, Group 2: 8/62; Comments: Treatment package: 4 did not want to take part, 1 recent bereavement, 1 ill health in the family. Joint protection: 1 family problems, 4 did not want to take part, 1 felt unable to help further, 1 had no time to participate, 1 incorrect address details. Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, marital status, routine or manual occupation, currently working, age when left school, left school to go to full-time education or university, gained qualifications through study as an adult and baseline values of outcomes; Group 1 Number missing: 6; Group 2 Number missing: 8 RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus NO TREATMENT Protocol outcome 1: Discontinuation at ≤3 months - Actual outcome: Withdrew at 3 months; Group 1: 2/65, Group 2: 0/65; Comments: Treatment package: 2 did not want to take part. Leaflet and advice: 0. Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, marital status, routine or manual occupation, currently working, age when left school, left school to go to full-time education or university, gained qualifications through study as an adult and baseline values of outcomes; Group 1 Number missing: 2; Group 2 Number missing: 0 Protocol outcome 2: Discontinuation at >3 months - Actual outcome: Withdrew at 12 months; Group 1: 6/65, Group 2: 5/65; Comments: Treatment package: 4 did not want to take part, 1 recent bereavement, 1 ill health in the family. Leaflet and advice: 1 felt unable to help further, 3 did not want to take part, 3 ill health (people withdrew for multiple reasons in this arm). Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, marital status, routine or manual occupation, currently working, age when left school, left school to go to full-time education or university, gained qualifications through study as an adult and baseline values of outcomes; Group 1 Number missing: 6; Group 2 Number missing: 5	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Skou 2015[253]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=100)
Countries and setting	Conducted in Denmark; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 week intervention, 12 months follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Symptomatic and radiographically-confirmed knee osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Symptomatic and radiographically-confirmed knee osteoarthritis, found not eligible for total knee replacement by an orthopedic surgeon (decision among other factors based on pain, function and radiographic severity), but experiencing more than mild limitations
Exclusion criteria	Less than mild limitations (a score above 75 on a 0-100 worst to best scale in the self-report questionnaire Knee Injury and Osteoarthritis Outcome Score, defined as the average score for the subscale scores); previous ipsilateral knee replacement; mean knee pain intensity in the previous week greater than 60mm on a 100mm visual analog scale.
Recruitment/selection of patients	People were recruited from patients referred to one of two specialised, public outpatient clinics by their general practitioner in the North Denmark Region.
Age, gender and ethnicity	Age - Mean (SD): 66.0 (9.0). Gender (M:F): 49:51. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Low morbidity score (Charlson comorbidity index, 0->3. Median: 1.). 4. Site of osteoarthritis: Knee
Extra comments	Severity: Kellgren Lawrence grade 1-4, median grade 3 Duration of symptoms: 0 months - more than 10 years, median 2-5 years. In outcomes: the discontinuation outcome was no included as it was unclear exactly how many people discontinued the study at any moment and so it wasn't extracted to avoid double counting
Indirectness of population	No indirectness
Interventions	(n=50) Intervention 1: Treatment package - Combination and education programme. The MEDIC treatment, consisting of five components: education, exercise and insoles

	were prescribed to everyone, weight loss and/or pain medication were prescribed if indicated. The education consisted of two 60-minute sessions focusing on disease characteristics, treatment and assistance to support self-help by actively engaging the patients in the sessions and in the treatment. The neuromuscular exercise program included sessions twice a week for 12 weeks for 60 minutes per session including neuromuscular and biomechanical principles in the exercises selected. Dietary advice was given to people with a BMI of at least 25 at baseline, including a 12-week program including four 60 minute sessions to discuss weight loss using motivational interviewing. Medial arch support insoles were provided. Additionally people with a kneelateral-to-foot position had a four degree lateral wedge added to the insole. People requiring pain medication were prescribed (if no contraindications were evident) 1 gram paracetamol four times, 400mg ibuprofen three times a day, and 20mg pantoprazol daily. The prescription was reassessed every 3 weeks. People were contacted for booster sessions somewhere between 20 weeks and 52 weeks.. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme (Contains components of both, but the education program is given to everyone while the behaviour intervention is only given to some). 2. Length of package: > 6 weeks (12 weeks). (n=50) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care, involving two standardized information leaflets (also given to the MEDic group) discussing knee symptoms, etiology, functional limitations, recommended treatments and general advice, and where to seek advice and general healthy lifestyle advice.. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (This trial is partially funded by The Danish Rheumatism Association and The Association of Danish Physiotherapists Research Fund)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at >3 months
- Actual outcome: EQ-5D index at 12 months; Group 1: mean 0.14 (SD 0.16); n=50, Group 2: mean 0.075 (SD 0.21); n=50; EQ-5D -0.11-1 Top=High is good

outcome; Comments: Reported mean change and 95% confidence intervals. Reported treatment package: 0.140 (0.095-0.186). Reported usual care: 0.075 (0.018-0.132). Baseline treatment package: 0.660 (0.160). Baseline standard care: 0.689 (0.145). Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, weight, BMI, study knee, bilateral knee pain, duration of knee symptoms, radiographic knee severity, Charlson comorbidity index, college education or equivalent, employment status, prior treatment, and baseline values of outcomes; Group 1 Number missing: 3, Reason: In total 47 people attended the 12 month follow up, but 2 people did not receive the allocated treatment at the start and it is unclear as to whether those participants were included in the analysis. At 12 months, 3 did not attend, 1 dead, 1 cancellation or no contact, 1 no longer interested; Group 2 Number missing: 6, Reason: 6 did not attend. 1 dead, 2 no longer interested, 1 cancellation or no contact, 1 unhappy with group allocation, 1 personal or health issues Protocol outcome 2: Pain at >3 months - Actual outcome: KOOS pain at 12 months; Group 1: mean 18.7 (SD 21.1); n=50, Group 2: mean 9.3 (SD 22.9); n=50; KOOS pain 0-100 Top=High is good outcome; Comments: Reported mean change and 95% confidence intervals. Reported treatment package: 18.7 (12.9-24.6). Reported usual care: 9.3 (2.9-15.6). Baseline treatment package: 51.6 (14.3). Baseline standard care: 53.6 (13.7). Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, weight, BMI, study knee, bilateral knee pain, duration of knee symptoms, radiographic knee severity, Charlson comorbidity index, college education or equivalent, employment status, prior treatment, and baseline values of outcomes; Group 1 Number missing: 3, Reason: In total 47 people attended the 12 month follow up, but 2 people did not receive the allocated treatment at the start and it is unclear as to whether those participants were included in the analysis. At 12 months, 3 did not attend, 1 dead, 1 cancellation or no contact, 1 no longer interested; Group 2 Number missing: 6, Reason: 6 did not attend. 1 dead, 2 no longer interested, 1 cancellation or no contact, 1 unhappy with group allocation, 1 personal or health issues Protocol outcome 3: Physical function at >3 months - Actual outcome: KOOS activities of daily living at 12 months; Group 1: mean 18.7 (SD 22); n=50, Group 2: mean 5.9 (SD 24.2); n=50; KOOS activities of daily living 0-100 Top=High is good outcome; Comments: Reported mean change and 95% confidence intervals. Reported treatment package: 18.7 (12.6-24.8). Reported usual care: 5.9 (-0.8-12.6). Baseline treatment package: 55.5 (17.1). Baseline standard care: 60.4 (16.4). Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age, weight, BMI, study knee, bilateral knee pain, duration of knee symptoms, radiographic knee severity, Charlson comorbidity index, college education or equivalent, employment status, prior treatment, and baseline values of outcomes; Group 1 Number missing: 3, Reason: In total 47 people attended the 12 month follow up, but 2 people did not receive the allocated treatment at the start and it is unclear as to whether those participants were included in the analysis. At 12 months, 3 did not attend, 1 dead, 1 cancellation or no contact, 1 no longer interested; Group 2 Number missing: 6, Reason: 6 did not attend. 1 dead, 2 no longer interested, 1 cancellation or no contact, 1 unhappy with group allocation, 1 personal or health issues	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Pain at ≤3 months; Physical function at ≤3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study	Stener-victorin 2004[261]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=45)
Countries and setting	Conducted in Sweden; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 5 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Radiographic changes consistent with osteoarthritis in the hip and pain related to motion and/or pain on load and/or ache during rest
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with radiographic changes consistent with osteoarthritis in the hip and pain related to motion and/or pain on load and/or ache during rest
Exclusion criteria	People with a pacemaker, hepatitis B, epilepsy, or rheumatoid diseases
Recruitment/selection of patients	People were preselected by orthopedics at Sahlgrenska University Hospital, Molndal and by general practitioners at the outpatient department in Molndal, Sweden. All people were on the waiting list for total hip arthroplasty.
Age, gender and ethnicity	Age - Range: 42-86. Gender (M:F): 18:27. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): Mixed 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms (range): 4 months - 15 years
Indirectness of population	No indirectness
Interventions	(n=30) Intervention 1: Treatment package - Combination and education programme. Hydrotherapy performed in small groups, 1-3 each, in an Arjo pool with 34 degrees centigrade warm water. The program consisted of warming up, mobility, and strengthening exercises for the muscles around the pelvis and stretching exercises. All people went through patient education. The education consisted of 2 group meetings of 2 hours each. They were taught about hip anatomy and the disease process. Instructions and advice about load-unload, activity-inactivity and pain relief, as well as information about total hip arthroplasty surgery. They were also given information about aid facilities and instructions for a program of home exercise, which included 10 exercises, aiming to improve the muscle strength, joint stability and range of motion in

	the hip. They were taught to train once a day with intensity below pain. People were treated 10 times during 5 weeks, 2 times per week. Each treatment lasted 30 minutes. Or Electroacupuncture placed locally in the most painful area of the hip and distally in points according to the segmental innervation of the hip joint (L3-5). Locally in the pain area, for of the following points were selected: BL54, 36, GB 29, 30, 31 and ST31. The distal points were always the same, GB34 and BL60 ipsilateral, both in the same segmental innervation as the hip joint. The needles were made of stainless steel for single use and were inserted intramuscularly to a depth of 15-35mm. Needle sizes were 0.32 x 30mm and 0.40 x 50mm. They were then rotated manually to evoke needle sensation, reflecting activation of muscle-nerve afferents, in total 4 times during treatment. All needles were attached to an electrical stimulator and stimulated with continuous square wave pulses with alternating polarity. The frequency used was low burst frequency of 2Hz (each pulse has a duration of 180 microseconds, a burst length of 0.1 seconds, and a burst frequency of 80Hz). The intensity was sufficient to cause non-painful local muscular contractions and was optimized for each person in an attempt to activate both the segmental pain control systems and the central descending pain inhibitory systems. All people went through patient education. The education consisted of 2 group meetings of 2 hours each. They were taught about hip anatomy and the disease process. Instructions and advice about load-unload, activity-inactivity and pain relief, as well as information about total hip arthroplasty surgery. They were also given information about aid facilities and instructions for a program of home exercise, which included 10 exercises, aiming to improve the muscle strength, joint stability and range of motion in the hip. They were taught to train once a day with intensity below pain. People were treated 10 times for 5 weeks, 2 times per week. Each treatment lasted for 30 minutes. Duration 5 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (5 weeks). Comments: The exercise treatment package and electroacupuncture treatment package were combined due to class effect (to avoid double counting of the control group) (n=15) Intervention 2: Non-combined active treatment - Education programme. Education only. Duration 5 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (5 weeks).
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: COMBINATION AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Discontinuation at ≤3 months
- Actual outcome: Lost to follow up at 3 months; Group 1: 11/30, Group 2: 8/15; Comments: Treatment package: 2 did not receive electroacupuncture and went on to have operation, 2 did not receive hydrotherapy but went on to have operation, 4 lost to follow up for electroacupuncture package, 3 lost to follow up for hydrotherapy package. Education: 8 lost to follow up.
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age and baseline values of outcomes; Group 1 Number missing: 11, Reason: Treatment package: 2 did not receive electroacupuncture and went on to have operation, 2 did not receive hydrotherapy but went on to have operation, 4 lost to follow up for electroacupuncture package, 3 lost to follow up for hydrotherapy package.; Group 2 Number missing: 8, Reason: Education: 8 lost to follow up.

Protocol outcome 2: Discontinuation at >3 months
- Actual outcome: Lost to follow up at 6 months; Group 1: 12/30, Group 2: 8/15; Comments: Treatment package: 2 did not receive electroacupuncture and went on to have operation, 2 did not receive hydrotherapy but went on to have operation, 4 lost to follow up for electroacupuncture package, 4 lost to follow up for hydrotherapy package. Education: 8 lost to follow up.
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported gender, age and baseline values of outcomes; Group 1 Number missing: 12, Reason: Treatment package: 2 did not receive electroacupuncture and went on to have operation, 2 did not receive hydrotherapy but went on to have operation, 4 lost to follow up for electroacupuncture package, 4 lost to follow up for hydrotherapy package.; Group 2 Number missing: 8, Reason: Education: 8 lost to follow up.

Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Pain at ≤3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months

Study	Tak 2005[266]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=109)
Countries and setting	Conducted in Netherlands; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 8 weeks of intervention, 3 months follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: The diagnosis of osteoarthritis of the hip had been made by the general practitioner and clinical symptoms, evaluated by physical therapists at baseline, meeting criteria for osteoarthritis of the hip of the American College of Rheumatology (pain in the hip together with endorotation of at least 15 degrees, pain present at endorotation of the hip, morning stiffness for no more than 60 minutes after rising, age >50 years.
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Older adults with complaints of osteoarthritis of the hip who were 55 years or older with a clinical diagnosis of osteoarthritis of the hip and living independently.
Exclusion criteria	People on the waiting list for hip replacement (or who had a hip replacement in the past year); serious disorders or impairments that jeopardized safe use of fitness equipemnt, such as neurological or cardiovascular problems; serious depression or dementia (as judged by general practitioners); regular treatment by a physical therapist (more than once a week)
Recruitment/selection of patients	Participants were recruited by means of announcements placed in regional newspapers, health centers, offices of general practitioners and local television
Age, gender and ethnicity	Age - Mean (SD): Intervention: 67.4 (7.6). Control: 68.9.. Gender (M:F): 30:64. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Not stated / Unclear 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Hip
Extra comments	Severity: Not stated Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=55) Intervention 1: Treatment package - Exercise and education programme. Hop with the Hip program, consisting of 8, 1 hour weekly group sessions of strength training using fitness equipment under supervision of a physical therapists. People

	were also offered a home exercise program, personal ergonomic advice (given by an occupational therapist), and dietary advice (given by a dietician). Each session started with group warm-up exercises, followed by instructions on and individual use of group warm-up exercises, followed by instruction on and individual use of fitness equipment and exercises: leg press, leg raise, rotation in sitting position, leaping squat, pull down, treadmill, home trainer, pulleys, bow flex and walking. The training session ended with group cool-down exercises. Intensity was progressed. The home exercise program included warm-up/cool-down and specific exercises for the lower extremities. Separate education on dietary aspects (healthy eating and drinking habits) in relation to body mass was given by a dietician. People with a BMI >30 were invited for a personal consultation. All people could get further information via a special phone line. An occupational therapist visited all people at home for individual counseling regarding activity restrictions caused by osteoarthritis and ways to deal with them.. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (8 weeks). (n=54) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). No additional treatment apart from appointments organised by the individual. Duration 8 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (8 weeks).
Funding	Academic or government funding (Supported by a grant from The Netherlands Health Research and Development Council (Preaventiefonds))

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: Health related quality of life at 12 weeks; Group 1: mean 28.6 (SD 3.6); n=55, Group 2: mean 27.3 (SD 2.7); n=54; Health related quality of life (scale not provided) 7-39 Top=High is good outcome; Comments: Baseline treatment package: 28.2 (3.1). Baseline control: 27.3 (2.4).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, marital status, education, BMI, general health, pain and baseline values of outcomes; Group 1 Number missing: 10, Reason: Treatment package: 2 moved, 1 private reasons, 1 illness, 2 pain back/hip, 1 physiotherapist, 2 surgery hip/knee, 1 not satisfied.; Group 2 Number missing: 5, Reason: Standard care: 1 withdrew from testing, 2 ineligible (no osteoarthritis), 1 illness, 1 surgery hip.

Protocol outcome 2: Pain at ≤3 months - Actual outcome: Harris Hip Score pain scale at 12 weeks; Group 1: mean 29.6 (SD 10.4); n=55, Group 2: mean 26.9 (SD 9.8); n=54; Harris Hip Score pain subscale 0-44 Top=High is good outcome; Comments: Baseline treatment package: 27.9 (8.1). Baseline control: 28.8 (9.0). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, marital status, education, BMI, general health, pain and baseline values of outcomes; Group 1 Number missing: 10, Reason: Treatment package: 2 moved, 1 private reasons, 1 illness, 2 pain back/hip, 1 physiotherapist, 2 surgery hip/knee, 1 not satisfied.; Group 2 Number missing: 5, Reason: Standard care: 1 withdrew from testing, 2 ineligible (no osteoarthritis), 1 illness, 1 surgery hip. Protocol outcome 3: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 12 weeks; Group 1: 10/55, Group 2: 5/54; Comments: Treatment package: 2 moved, 1 private reasons, 1 illness, 2 pain back/hip, 1 physiotherapist, 2 surgery hip/knee, 1 not satisfied. Standard care: 1 withdrew from testing, 2 ineligible (no osteoarthritis), 1 illness, 1 surgery hip. Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported sex, age, marital status, education, BMI, general health, pain and baseline values of outcomes; Group 1 Number missing: 10, Reason: Treatment package: 2 moved, 1 private reasons, 1 illness, 2 pain back/hip, 1 physiotherapist, 2 surgery hip/knee, 1 not satisfied.; Group 2 Number missing: 5, Reason: Standard care: 1 withdrew from testing, 2 ineligible (no osteoarthritis), 1 illness, 1 surgery hip.	
Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study	Talbot 2003[267]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=34)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of intervention with 12 weeks of additional follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Pain in one or both knees on most days, difficulty performing at least one functional task because of pain, and radiographic evidence of osteoarthritis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 60 and older; pain in one or both knees on most days; difficulty performing at least one functional task because of pain; radiographic evidence of osteoarthritis
Exclusion criteria	Current participation in an exercise research study; a medical condition for which exercise is contraindicated, such as unstable angina pectoris or recent myocardial infarction; a score of less than 24 on the Mini-Mental State Examination
Recruitment/selection of patients	People were recruited through senior centers and advertisements in local newspapers
Age, gender and ethnicity	Age - Mean (SD): 70.2 (5.8). Gender (M:F): 8:26. Ethnicity: 30 people were Caucasian, no other information given
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Grades 1-4, median grade 2 Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=17) Intervention 1: Treatment package - Exercise and education programme. Walk+ program and education (self-management) program. Each individuals' daily steps were modified to the individuals' baseline step count and increased by 10% every 4 weeks. By the end of the program they would be walking 30% above their baseline step count. During brief individual counseling, the pedometer logs were reviewed and feedback provided. In addition, people were given a booklet explaining the principles of exercise, including warm-up, cool-down, stretching, and such arthritis principles as the 2-hour pain rule and balancing rest with activity. The Arthritis self-management program teaches techniques for coping with arthritis including exercise

	as a component of management. This was delivered as 12x 1 hour sessions. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks). (n=17) Intervention 2: Non-combined active treatment - Education programme. Arthritis self-management program only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (Funded by the Fund for Geriatric Medicine and Nursing, Johns Hopkins University, and the Intramural Research Program of the National Institute on Aging)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME Protocol outcome 1: Pain at ≤3 months - Actual outcome: Pain rating index at 12 weeks; Group 1: mean 12.41 (SD 9.77); n=17, Group 2: mean 10.12 (SD 4.64); n=17; Pain rating index Scale range unclear Top=High is poor outcome; Comments: Baseline treatment package: 11.65 (11.52). Baseline education: 13.94 (10.64). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnciity, marital status, annual income, college graduate status, class attendance, grade of osteoarthritis, total knee replacement, BMI and baseline values of outcomes; Group 1 Number missing: -; Group 2 Number missing: - Protocol outcome 2: Pain at >3 months - Actual outcome: Pain rating index at 24 weeks; Group 1: mean 12.95 (SD 11.41); n=17, Group 2: mean 10.9 (SD 9.69); n=17; Pain rating index Scale range unclear Top=High is poor outcome; Comments: Baseline treatment package: 11.65 (11.52). Baseline education: 13.94 (10.64). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, gender, ethnciity, marital status, annual income, college graduate status, class attendance, grade of osteoarthritis, total knee replacement, BMI and baseline values of outcomes; Group 1 Number missing: -; Group 2 Number missing: -	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months; Discontinuation at >3 months

Study	Talbot 2003[268]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=38)
Countries and setting	Conducted in USA; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks of intervention, and additional 12 weeks of follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Pain in one or both knees; self reported difficulty in walking, stair climbing or rising from a chair; radiographic evidence of knee osteoarthritis (At least grade 1) based on the criteria of Kellgren and Lawrence
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age 60 years or older; pain in one or both knees; self reported difficulty in walking, stair climbing or rising from a chair; radiographic evidence of knee osteoarthritis (At least grade 1) based on the criteria of Kellgren and Lawrence
Exclusion criteria	Recent participation in an exercise program to increase strength; medical condition in which NMES training is contraindicated i.e. reduced sensory perception in the lower extremity; cognitive impairment that precluded the provision of informed consent; implanted cardiac pacemaker or defibrillator
Recruitment/selection of patients	People were recruited from local senior centers or responded to advertisements in the local newspaper
Age, gender and ethnicity	Age - Mean (SD): 70.5 (5.3). Gender (M:F): 7:27. Ethnicity: Caucasian = 29, no additional information
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Radiographic grade 1-4, median grade 2 Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Treatment package - Electrotherapy and education programme. NMES and Arthritis Self-Help course. NMES was delivered by stimulating the quadriceps femoris muscle of the knee with the greatest disease using a portable electrical muscle stimulator with preset parameters for home use. The contralateral leg was the opposing lower extremity. People performed this at home 3 training sessions

	per week for 12 weeks. At 4 week intervals, the intensity of the stimulator was increased a maximum of 10% of MVC or to a current that could be tolerated by each participant. The electrical impulse was generated by a baterry-operated device that delivered a pulsed current with symmetrical biphasic rectangular waves. Two 4x5 inch high-impedance stimulation electrodes were placed over the quadriceps femoris muscle group of the index leg. The phase width was 300 microseconds at 50% amplitude. Electrical pulse rate was maintained at 50pps. The pulsed current was delivered with a ramp-up time of 3s and a ramp-down time of 1.5s. The duty cycle was set to 10s on and 50s off during stimulation. The current intensity was adjusted and maintained at the appropriate percentage of mVC or to tolerance during each contraction. The treatment protocol was for 15 minute sessions of 15 stimulations to the index leg, 3 times per week. Every 4 weeks the intensity was increased. The arthritis self-help course was delivered over 12 weeks with 1 session per week. It taught disease etiology, self-management of symptoms, and techniques of problem solving, goal setting, contracts and feedback to accomplish individual goals. Leaders for the education program were 2 registered nurses with 16 hours of training. During these weekly sessions, people were asked about their activities in the week and given a time to discuss any difficulties. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks). (n=18) Intervention 2: Non-combined active treatment - Education programme. Arthritis education only. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: > 6 weeks (12 weeks).
Funding	Academic or government funding (Supported by the Fund for Geriatric medicine and Nursing, Johns Hopkins University and the Intramural Research Program of the National Institute on Aging)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: ELECTROTHERAPY AND EDUCATION PROGRAMME versus EDUCATION PROGRAMME

Protocol outcome 1: Pain at ≤3 months
- Actual outcome: Pain rating index - total at 12 weeks; Group 1: mean 16.33 (SD 13.35); n=20, Group 2: mean 11.12 (SD 8); n=18; McGill Pain questionnaire pain rating index 0-78 Top=High is poor outcome; Comments: Baseline treatment package: 20.26 (11.08). Baseline education: 13.81 (10.79). Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low,

Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in pain rating index at baseline; Group 1 Number missing: 2, Reason: 2 incomplete data; Group 2 Number missing: 2, Reason: 2 incomplete data Protocol outcome 2: Pain at >3 months - Actual outcome: Pain rating index - total at 24 weeks; Group 1: mean 16.14 (SD 12.03); n=20, Group 2: mean 12.42 (SD 9.66); n=18; McGill pain questionnaire pain rating scale 0-78 Top=High is poor outcome; Comments: Baseline treatment package: 20.26 (11.08). Baseline education: 13.81 (10.79). Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in pain rating index at baseline; Group 1 Number missing: 2, Reason: 2 incomplete data; Group 2 Number missing: 2, Reason: 2 incomplete data Protocol outcome 3: Discontinuation at >3 months - Actual outcome: Disqualified due to incomplete data at 24 weeks; Group 1: 2/20, Group 2: 2/18 Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in pain rating index at baseline; Group 1 Number missing: 2, Reason: 2 incomplete data; Group 2 Number missing: 2, Reason: 2 incomplete data	
Protocol outcomes not reported by the study	Quality of life at ≤3 months; Quality of life at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at ≤3 months

Study	Wallis 2017[279]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=46)
Countries and setting	Conducted in Australia; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 13 weeks (12 week program)
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Severe knee osteoarthritis rating grade III or IV affecting at least one of the tibiofemoral compartments determined radiographically
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with severe osteoarthritis of the knee referred to a clinic to assess eligibility for total knee replacement. Age at least 50 years and living independently in the community; diagnosed with severe knee osteoarthritis rated as grade III or IV affecting at least one of the tibiofemoral compartments determined radiographically; a cardiovascular risk profile with at least 2 total risk factors using stage 2 of the Adult Exercise Screening Tool; able to participate safely in the moderate-intensity physical activity trial using stage 1 of the Adult Exercise Screening Tool; able to communicate in English
Exclusion criteria	Lived in supported accommodation such as a nursing home; reported daily resting level of pain to be 9 or 10 on a 0 (no pain) to 10 (worst possible pain) Numerical Pain Rating Scale as this level of pain may be indicative of a more serious pathology; had high levels of psychological distress as measured by the Kessler 10 questionnaire with a K10 score >29; had a cognitive impairment measured by the Short Portable mental Status Questionnaire with a score of 8 or less; had a systemic arthritic condition such as rheumatoid arthritis; had a neurological condition that affected walking; had knee surgery or intra-articular corticosteroid injection within the past 6 months; had used oral corticosteroids within 4 weeks.
Recruitment/selection of patients	People were recruited from a metropolitan health service's osteoarthritis hip and knee clinic
Age, gender and ethnicity	Age - Mean (SD): 67.5 (7.5). Gender (M:F): 26:20. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed with imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee

Extra comments	Severity: Radiographic grade III-IV, median grade IV Duration of symptoms: Not stated
Indirectness of population	No indirectness
Interventions	(n=23) Intervention 1: Treatment package - Exercise and behaviour change intervention. A walking dose of 70 minutes per week, of at least moderate intensity, in bouts of at least 10 minutes. The weekly dose was completed for 12 weeks in the community. The participant was instructed to walk at a moderate level of intensity (determined by the Rate of Perceived Exertion Scal). No formal instructions on warming up or stretching were provided. The weekly dose was 70 minutes. In separate sessions provided each session was at least 10 minutes duration. To increase the likelihood of adherence to the intervention, the following behavioural change techniques and strategies were used. First, each person had a planning session with a physiotherapist for up to 30 minutes to plan the location, day and time of day for each walk, and reinforce that each walk was moderate intensity in at least a 10 minute bout. Second, there was regular physiotherapy supervision and monitoring each week, including one-to-one supervised walking sessions or group supervised walking sessions based on patient rrpreference, and regular phone calls or smS reminders. Third, each person wore a pedometer and recorded the number of steps taken and time spent walking during each session in a logbook. Fourth, participants were encouraged to engage social supports such as walking with a friend, family member or other research participants. Duration 12 weeks. Concurrent medication/care: People continued taking their usual medications and other non-surgical treatments to manage their knee osteoarthritis, and used normal assistive devices such as a cane. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Behaviour change intervention 2. Length of package: > 6 weeks (12 weeks). (n=23) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Usual care was non-operative management to manage pain and symptoms including pharmacological and non-pharmacological interventions. They (and their healthcare professionals) were advised not to start any new physical activity in the 12 week period. If requested, the control group could have a copy of the walking program after their final assessment. Duration 12 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: > 6 weeks (12 weeks).

Funding	Academic or government funding (The research received $24,704 from La Trobe University's research focus area on Sport, Exercise and Rehabilitation.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND BEHAVIOUR CHANGE INTERVENTION versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: QOL EQ5D utility at 13 weeks; Group 1: mean 0.07 (SD 0.2); n=23, Group 2: mean -0.03 (SD 0.1); n=23; EQ-5D 0-1 Top=High is good outcome; Comments: Baseline treatment package: 0.54 (0.2). Baseline control: 0.64 (0.2).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in EQ-5D and WOMAC pain at baseline; Group 1 Number missing: 8, Reason: 1 did not receive allocated intervention. 1 lost to follow up. 6 discontinued (2 severe knee pain, 3 unrelated medical reason, 1 family reason).; Group 2 Number missing: 0

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: WOMAC pain at 13 weeks; Group 1: mean 0.5 (SD 2.9); n=23, Group 2: mean 0.9 (SD 2.7); n=23; WOMAC pain 0-20 Top=High is poor outcome; Comments: Baseline treatment package: 11 (2.3). Baseline control: 8.8 (3.2).
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in EQ-5D and WOMAC pain at baseline; Group 1 Number missing: 8, Reason: 1 did not receive allocated intervention. 1 lost to follow up. 6 discontinued (2 severe knee pain, 3 unrelated medical reason, 1 family reason).; Group 2 Number missing: 0

Protocol outcome 3: Physical function at ≤3 months
- Actual outcome: WOMAC activity limitation at 13 weeks; Group 1: mean 0.6 (SD 7.4); n=23, Group 2: mean 0 (SD 7.8); n=23; WOMAC activity limitation 0-68 Top=High is poor outcome; Comments: Baseline treatment package: 37 (10). Baseline control: 34 (9).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in EQ-5D and WOMAC pain at baseline; Group 1 Number missing: 8, Reason: 1 did not receive allocated intervention. 1 lost to follow up. 6 discontinued (2 severe knee pain, 3 unrelated medical reason, 1 family reason).; Group 2 Number missing: 0

Protocol outcome 4: Discontinuation at ≤3 months
- Actual outcome: Discontinuation at 13 weeks; Group 1: 8/23, Group 2: 0/23; Comments: Treatment package: 1 withdrew after randomisation (did not receive treatment). 1 lost to follow-up. 6 discontinued (2 severe knee pain, 3 unrelated to medical reason, 1 family reason). Usual care: 0.
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in EQ-5D and WOMAC pain at baseline; Group 1 Number missing: 8, Reason: 1 did not receive allocated intervention. 1 lost to follow up. 6 discontinued (2 severe knee pain, 3 unrelated medical reason, 1 family reason).; Group 2 Number missing: 0

Protocol outcomes not reported by the study	Quality of life at >3 months; Pain at >3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Study (subsidiary papers)	Yip 2007[288] (Yip 2007[287], Yip 2008[289])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=182)
Countries and setting	Conducted in China; Setting: Outpatient follow up
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 weeks intervention, 1 year follow up in total
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Diagnosed based on the clinical criteria of the American College of Rheumatology
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Fulfilling the American College of Rheumatology clinical criteria with pain in the knee and three of the following: aged at least 50 years of age; less than 30 minutes of morning stiffness; crepitus on active motion; bony tenderness; bony enlargment; and no palpable warmth of the synovium.
Exclusion criteria	People who spent the majority of their time in bed; wheelchair users; experienced loss of balance while standing; had knee replacements; could over-exert in exercise compliance e.g. those currently undergoing active physiotherapy; those currently receiving acupuncture treatments.
Recruitment/selection of patients	People were recruited from the outpatient clinic of the Orthopaedic Department of a local hospital, the general outpatient clinic of a local hospital and the Telehealth clinic.
Age, gender and ethnicity	Age - Other: Mean (SE): Intervention: 65.60 (1.03). Control: 64.02 (1.06).. Gender (M:F): 29:153. Ethnicity: Not stated
Further population details	1. Age (≤/> 75 years): ≤ 75 years 2. Diagnosis with or without imaging: Diagnosed without imaging 3. Multimorbidity: Not stated / Unclear 4. Site of osteoarthritis: Knee
Extra comments	Severity: Not stated Duration of symptoms (mean [SE]): Intervention: 8.31 (0.78). Control: 7.85 (0.65).
Indirectness of population	No indirectness
Interventions	(n=88) Intervention 1: Treatment package - Exercise and education programme. The Arthritis Self Management Program intervention. Consisting of 6x 2 hour classes held once a week, with 10-15 participants trained in small group leadership and basic principles of self-management. The programme focused on the use of an action plan and on teaching participants how to cope with, and manage, common knee osteoarthritic consequences, such as arthritis pain, fatigue, daily activity limitations

	and stress. The topics covered were: an overview of self-management principles; medical aspects and pain management; joint protection, physical activity and exercise; available treatments; managing stress; nutrition; and communication skills and the availability of community resources. The participants were asked to set their goal on exercise practice and received positive feedback by a nurse every week. The three types of exercises were stretching, walking and Tai Chi types of movement aimed at enhancing exercise on the affected joints. This was taught by a lay person tutor. In addition, a pedometer was given to the intervention group for 3 days to act as a reinforcer for walking. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Education programme 2. Length of package: ≤ 6 weeks (6 weeks). (n=94) Intervention 2: Standard care (non-organised) or no treatment - Standard care (non-organised). Routine orthopaedic treatment (treatment prescribed by orthopaedic doctors or outpatient clinic) with no other treatment. Duration 6 weeks. Concurrent medication/care: No additional information. Indirectness: No indirectness Further details: 1. Behaviour change interventions or education programme: Not stated / Unclear 2. Length of package: ≤ 6 weeks (6 weeks).
Funding	Academic or government funding (This study was supported by The Hong Kong Polytechnic University, School of Nursing)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: EXERCISE AND EDUCATION PROGRAMME versus STANDARD CARE (NON-ORGANISED)

Protocol outcome 1: Quality of life at ≤3 months
- Actual outcome: Health Assessment Questionnaire at 7 weeks; Group 1: mean 4.63 (SD 3.8); n=88, Group 2: mean 4.46 (SD 3.63); n=94; Health Assessment Questionnaire 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 5.06 (4.48). Baseline control: 5.07 (3.96).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, duration of symptoms, gender, martial status, education level, occupational status before retirement and number of affected joints; Group 1 Number missing: 9, Reason: 2 lost to follow up, 7 discontinued ; Group 2 Number missing: 24, Reason: 4 lost to follow up, 20 discontinued

Protocol outcome 2: Pain at ≤3 months
- Actual outcome: Current pain rating (VAS) at 7 weeks; Group 1: mean 37.33 (SD 21.06); n=88, Group 2: mean 44.41 (SD 23.23); n=94; VAS 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 50.45 (20.81). Baseline control: 44.26 (24.42).
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, duration of symptoms, gender,

martial status, education level, occupational status before retirement and number of affected joints; Group 1 Number missing: 9, Reason: 2 lost to follow up, 7 discontinued ; Group 2 Number missing: 24, Reason: 4 lost to follow up, 20 discontinued Protocol outcome 3: Pain at >3 months - Actual outcome: Current pain rating (VAS) at 1 year; Group 1: mean -33.5 (SD 23.65); n=88, Group 2: mean -11.97 (SD 24.68); n=94; VAS 0-100 Top=High is poor outcome; Comments: Baseline treatment package: 57.00 (21.77). Baseline control: 41.65 (26.42). Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, duration of symptoms, gender, martial status, education level, occupational status before retirement and number of affected joints; Group 1 Number missing: 53, Reason: Reported that the intervention group only had 45 people, and the control group had only 50 people. 10 people did not attend at 12 months.; Group 2 Number missing: 55, Reason: Reported that the intervention group only had 45 people, and the control group had only 50 people. 11 people did not attend at 12 months. Protocol outcome 4: Discontinuation at ≤3 months - Actual outcome: Discontinuation at 7 weeks; Group 1: 9/88, Group 2: 24/94; Comments: Treatment packages: 2 lost to follow up (can't contact), 7 discontinued (busy; not interested; with walking problems). Control: 4 lost to follow up (passed away + can't contact), 20 discontinued (busy; not interested; with walking problems). Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low, Other 1 - Low; Indirectness of outcome: No indirectness ; Baseline details: Reported age, duration of symptoms, gender, martial status, education level, occupational status before retirement and number of affected joints; Group 1 Number missing: 9, Reason: 2 lost to follow up, 7 discontinued ; Group 2 Number missing: 24, Reason: 4 lost to follow up, 20 discontinued	
Protocol outcomes not reported by the study	Quality of life at >3 months; Physical function at ≤3 months; Physical function at >3 months; Psychological distress at ≤3 months; Psychological distress at >3 months; Osteoarthritis flares at ≤3 months; Osteoarthritis flares at >3 months; Discontinuation at >3 months

Appendix E – Forest plots

E.1 Treatment packages compared to exercise alone

Figure 2: Quality of life (AQOL II, -0.11-1, high is good, change score) at ≤3 months

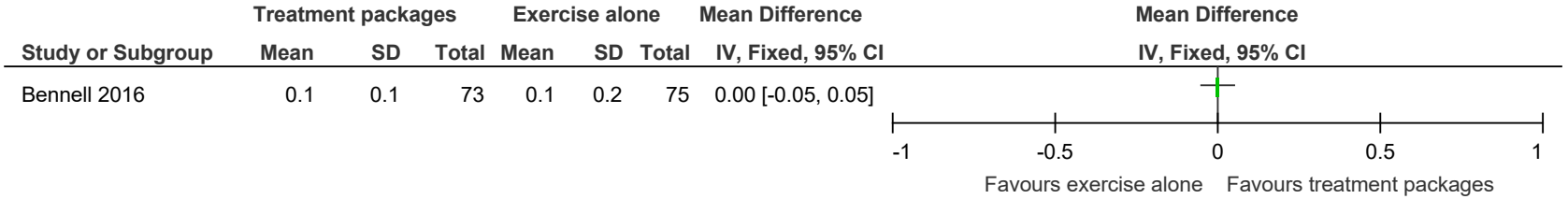


Figure 3: Quality of life (AIMS pain, 0-10, high is poor, final value) at ≤3 months

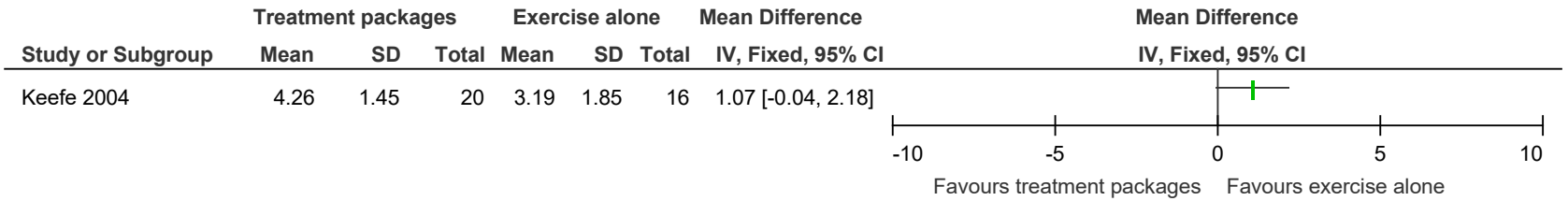


Figure 4: Quality of life (AIMS psychological disability, 0-10, high is poor, final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Exercise alone Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Keefe 2004	2.21	1.21	20	1.88	0.87	16	0.33 [-0.35, 1.01]

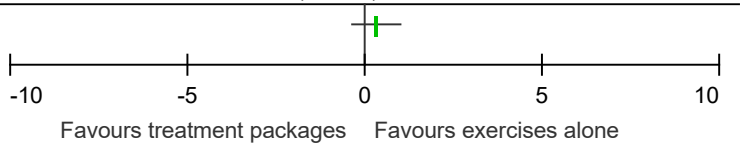
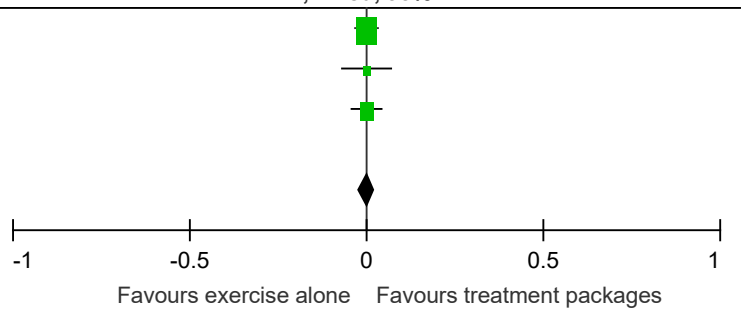


Figure 5: Quality of life (AQOL II, -0.11-1, high is good, change score) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Exercise alone Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Bennell 2016	0.1	0.1	74	0.1	0.1	75	56.0%	0.00 [-0.03, 0.03]
Bennell 2017	0	0.2	66	0	0.2	62	12.0%	0.00 [-0.07, 0.07]
Bennell 2018 (HOPE)	0.02	0.13	73	0.02	0.13	71	32.0%	0.00 [-0.04, 0.04]
Total (95% CI)			**213**			**208**	**100.0%**	**0.00 [-0.02, 0.02]**

Heterogeneity: Chi² = 0.00, df = 2 (P = 1.00); I² = 0%

Test for overall effect: Z = 0.00 (P = 1.00)



Note: Baseline values for Bennell 2016 are significantly different (0.74 [0.12] for treatment packages, 0.71 [0.14] for exercise alone.)

Figure 6: Quality of life (KOOS quality of life, 0-100, high is good, change score) at >3 months

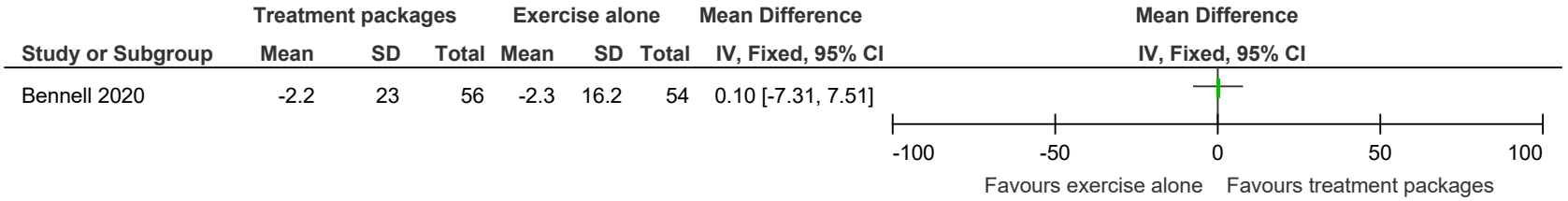


Figure 7: Quality of life (SF-36 physical component, 0-100, high is good, final values) at >3 months

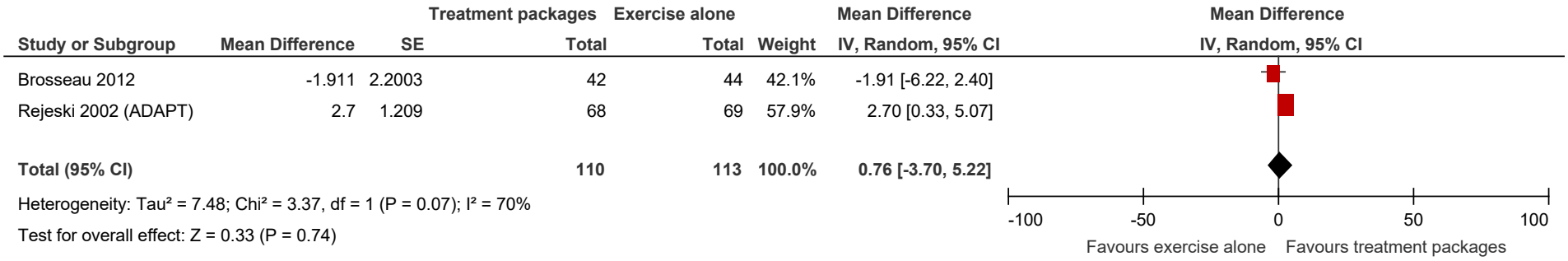


Figure 8: Quality of life (SF-36 mental component, 0-100, high is good, final values) at >3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	Exercise alone Total	Weight	Mean Difference IV, Fixed, 95% CI
Brosseau 2012	1.929	2.1653	42	44	22.1%	1.93 [-2.31, 6.17]
Rejeski 2002 (ADAPT)	-0.22	1.1526	68	69	77.9%	-0.22 [-2.48, 2.04]
Total (95% CI)			**110**	**113**	**100.0%**	**0.25 [-1.74, 2.25]**

Heterogeneity: Chi² = 0.77, df = 1 (P = 0.38); I² = 0%

Test for overall effect: Z = 0.25 (P = 0.80)

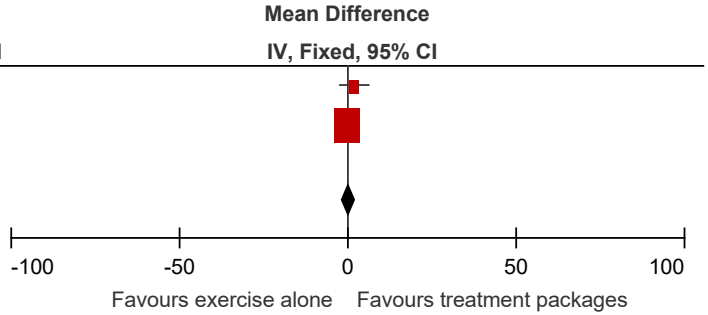


Figure 9: Pain (WOMAC, 0-20, high is poor, change scores) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Exercise alone Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Bennell 2016	-4.4	3	73	-3.3	3.1	75	39.5%	-1.10 [-2.08, -0.12]
Hsu 2021	-2.95	1.12	21	-1.9	1.48	21	60.5%	-1.05 [-1.84, -0.26]
Total (95% CI)			**94**			**96**	**100.0%**	**-1.07 [-1.69, -0.45]**

Heterogeneity: Chi² = 0.01, df = 1 (P = 0.94); I² = 0%

Test for overall effect: Z = 3.40 (P = 0.0007)

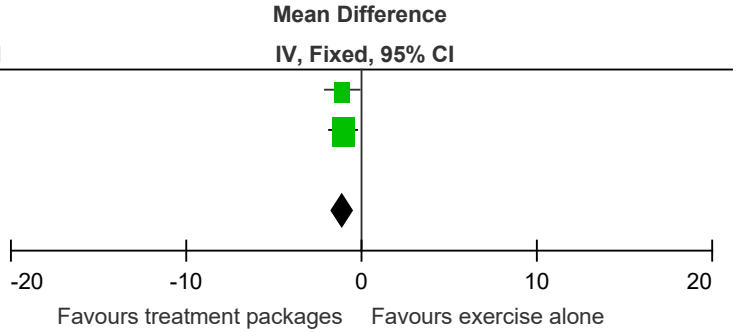


Figure 10: Pain (WOMAC, NRS [different scale ranges], high is poor, final values) at ≤3 months

	Treatment packages			Exercise alone				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Alasfour 2020	3.56	2.1	20	5.18	2.43	20	29.7%	-0.70 [-1.34, -0.06]
Alfieri 2020	41.8	28	22	43.5	21.1	17	29.9%	-0.07 [-0.70, 0.57]
Farr 2010	67.1	68.8	100	47.6	50.9	95	40.4%	0.32 [0.04, 0.60]
Total (95% CI)			**142**			**132**	**100.0%**	**-0.10 [-0.71, 0.51]**

Heterogeneity: Tau² = 0.22; Chi² = 8.53, df = 2 (P = 0.01); I² = 77%

Test for overall effect: Z = 0.32 (P = 0.75)

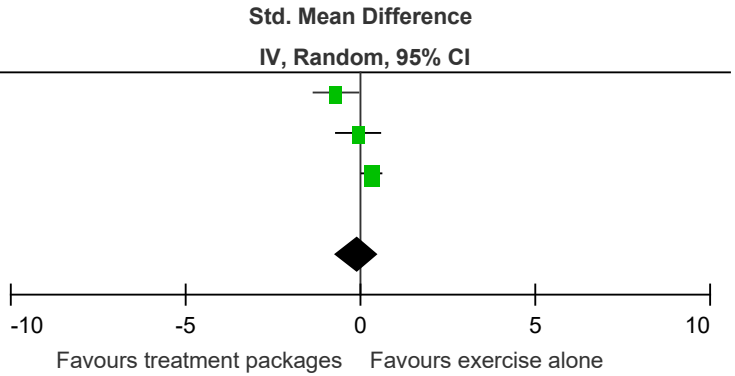


Figure 11: Pain (KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months

	Treatment packages			Exercise alone				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-3.8	3.4	73	-3.2	3.7	75	21.6%	-0.17 [-0.49, 0.15]
Bennell 2017	-3.5	3.9	66	-3.7	5.4	62	18.8%	0.04 [-0.30, 0.39]
Bennell 2018 (HOPE)	-2.9	4.6	73	-3.3	5.4	71	21.1%	0.08 [-0.25, 0.41]
Bennell 2020	0.8	14.9	56	2.6	14.1	54	16.1%	-0.12 [-0.50, 0.25]
Focht 2005	-2.2	4.1	76	-0.4	4.3	80	22.4%	-0.43 [-0.74, -0.11]
Total (95% CI)			**344**			**342**	**100.0%**	**-0.13 [-0.28, 0.02]**

Heterogeneity: Chi² = 5.92, df = 4 (P = 0.21); I² = 32%

Test for overall effect: Z = 1.65 (P = 0.10)

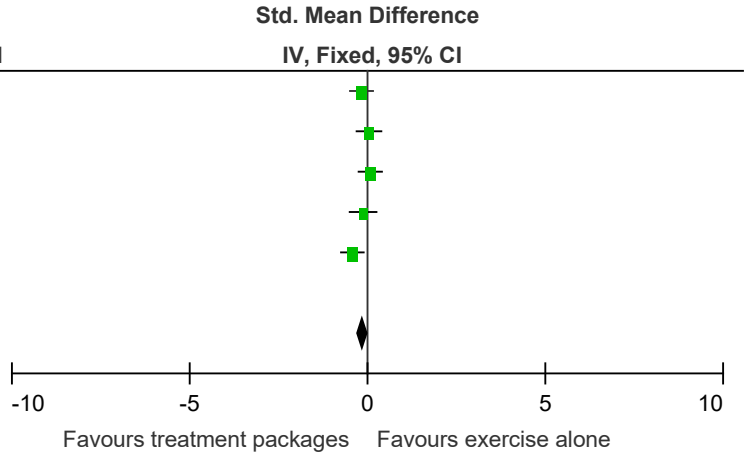


Figure 12: Pain (WOMAC, VAS [different scale ranges], high is poor, final values) at >3 months

	Treatment packages			Exercise alone				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Brosseau 2012	26.16	17.97	42	23.6	15.09	43	23.2%	0.15 [-0.27, 0.58]
Farr 2010	56.2	75.3	100	48.6	61.3	95	53.2%	0.11 [-0.17, 0.39]
Quilty 2003	48.1	25.7	43	54.1	22.5	44	23.6%	-0.25 [-0.67, 0.18]
Total (95% CI)			**185**			**182**	**100.0%**	**0.04 [-0.17, 0.24]**

Heterogeneity: Chi² = 2.28, df = 2 (P = 0.32); I² = 12%

Test for overall effect: Z = 0.34 (P = 0.73)

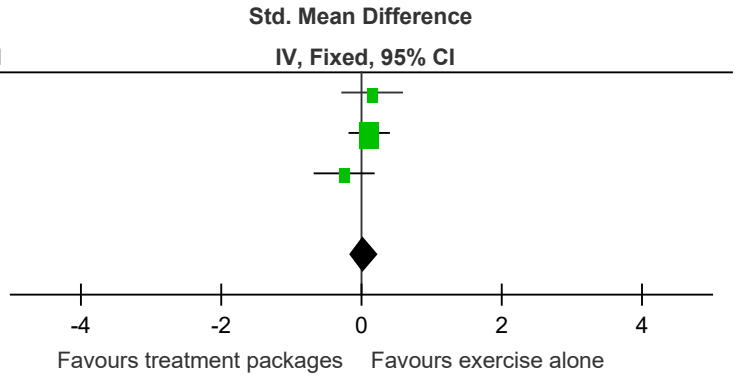


Figure 13: Physical function (WOMAC, 0-68, high is poor, change score) at ≤3 months

	Treatment packages			Exercise alone				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-19.9	9.1	73	-15.1	10.9	75	21.6%	-4.80 [-8.03, -1.57]
Hsu 2021	-8.62	3.58	21	-5.1	1.7	21	78.4%	-3.52 [-5.22, -1.82]
Total (95% CI)			**94**			**96**	**100.0%**	**-3.80 [-5.30, -2.30]**

Heterogeneity: Chi² = 0.47, df = 1 (P = 0.49); I² = 0%

Test for overall effect: Z = 4.96 (P < 0.00001)

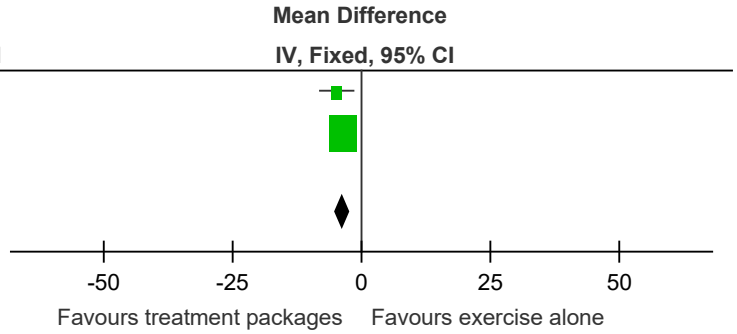


Figure 14: Physical function (WOMAC [different scale ranges], high is poor, final value) at ≤3 months

	Treatment packages			Exercise alone				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Alasfour 2020	3	1.91	20	5.18	3.24	20	49.7%	-0.80 [-1.45, -0.16]
Alfieri 2020	38.4	30.9	22	35.2	18.6	17	50.3%	0.12 [-0.51, 0.75]
Total (95% CI)			**42**			**37**	**100.0%**	**-0.34 [-1.24, 0.56]**

Heterogeneity: Tau² = 0.32; Chi² = 3.99, df = 1 (P = 0.05); I² = 75%

Test for overall effect: Z = 0.74 (P = 0.46)

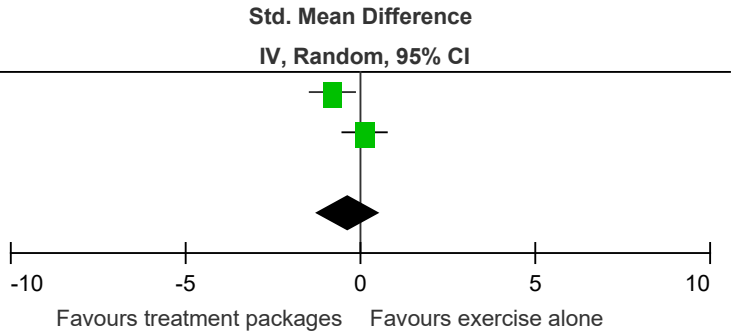


Figure 15: Physical function (KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months

	Treatment packages			Exercise alone				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-19.1	10.1	73	-15.9	12.5	75	27.8%	-0.28 [-0.60, 0.04]
Bennell 2017	-14.5	12.9	66	-12.6	15.1	62	24.2%	-0.13 [-0.48, 0.21]
Bennell 2018 (HOPE)	-8.8	16.4	73	-10.7	17	71	27.2%	0.11 [-0.21, 0.44]
Bennell 2020	0	18.5	56	0.5	14	54	20.8%	-0.03 [-0.40, 0.34]
Total (95% CI)			**268**			**262**	**100.0%**	**-0.09 [-0.26, 0.08]**

Heterogeneity: Chi² = 2.96, df = 3 (P = 0.40); I² = 0%

Test for overall effect: Z = 0.98 (P = 0.32)

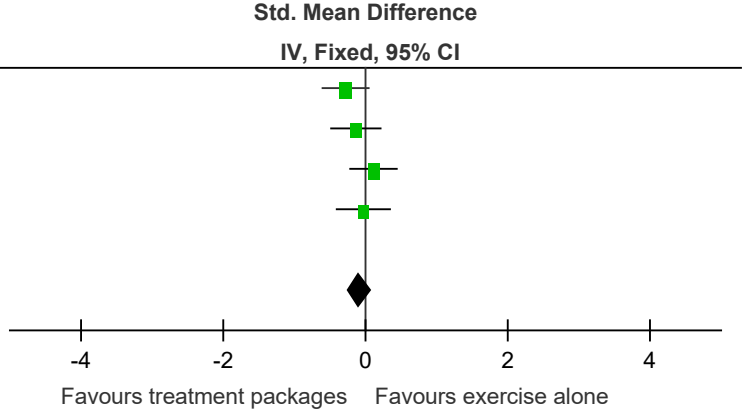


Figure 16: Physical function (WOMAC [different scale ranges], high is poor, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Exercise alone Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Brosseau 2012	24.15	17.24	42	18.2	14.63	43	49.0%	0.37 [-0.06, 0.80]
Quilty 2003	29.7	11.2	43	28.3	11.3	44	51.0%	0.12 [-0.30, 0.54]
Total (95% CI)			**85**			**87**	**100.0%**	**0.24 [-0.06, 0.54]**

Heterogeneity: Chi² = 0.64, df = 1 (P = 0.42); I² = 0%

Test for overall effect: Z = 1.59 (P = 0.11)

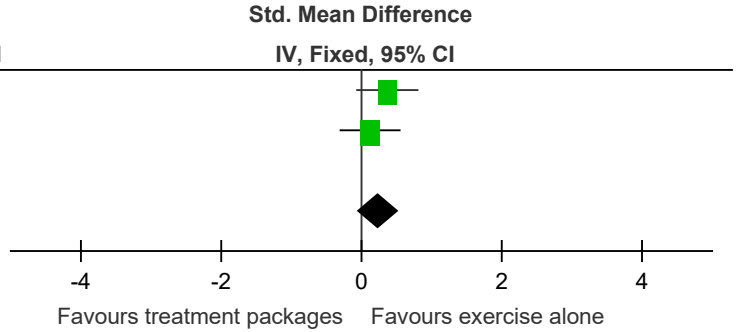


Figure 17: Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Exercise alone Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Bennell 2016	-0.9	4.1	73	-1.1	3.9	75	0.20 [-1.09, 1.49]

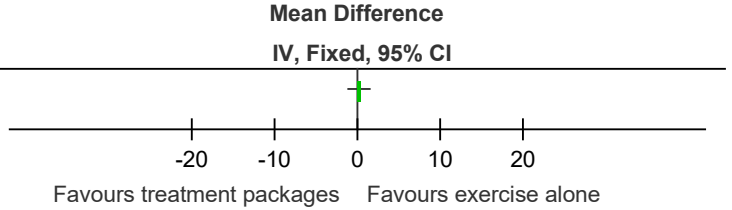


Figure 18: Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Exercise alone Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Bennell 2016	-0.9	4.8	73	-0.7	5.8	75	-0.20 [-1.91, 1.51]

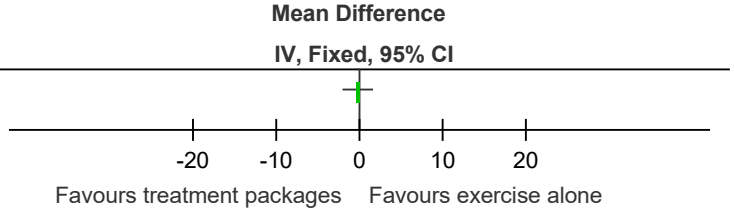


Figure 19: Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at ≤3 months

	Treatment packages			Exercise alone			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-0.5	5.6	73	-1.5	7.6	75	1.00 [-1.15, 3.15]

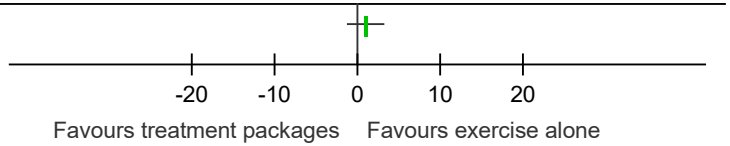


Figure 20: Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at >3 months

	Treatment packages			Exercise alone				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-2	4.9	73	-0.7	6.2	75	4.6%	-1.30 [-3.10, 0.50]
Bennell 2018 (HOPE)	-0.2	1.1	73	-0.1	1.3	71	95.4%	-0.10 [-0.49, 0.29]
Total (95% CI)			**146**			**146**	**100.0%**	**-0.15 [-0.54, 0.23]**

Heterogeneity: Chi² = 1.63, df = 1 (P = 0.20); I² = 39%

Test for overall effect: Z = 0.79 (P = 0.43)

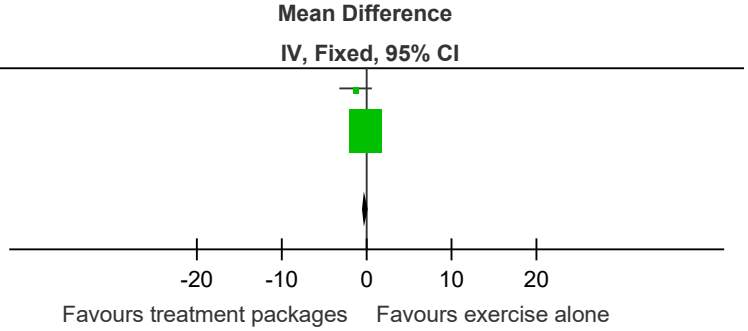


Figure 21: Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at >3 months

	Treatment packages			Exercise alone				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-1.4	6	73	-0.5	5.8	75	6.2%	-0.90 [-2.80, 1.00]
Bennell 2018 (HOPE)	-0.2	1.5	73	-0.1	1.5	71	93.8%	-0.10 [-0.59, 0.39]
Total (95% CI)			**146**			**146**	**100.0%**	**-0.15 [-0.62, 0.32]**

Heterogeneity: Chi² = 0.64, df = 1 (P = 0.42); I² = 0%

Test for overall effect: Z = 0.62 (P = 0.54)

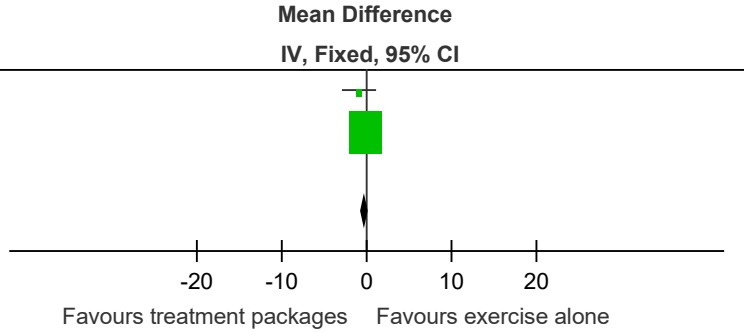


Figure 22: Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at >3 months

	Treatment packages			Exercise alone				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-2.1	6.3	73	-0.8	8.6	75	3.9%	-1.30 [-3.72, 1.12]
Bennell 2018 (HOPE)	-0.4	1.5	73	-0.2	1.5	71	96.1%	-0.20 [-0.69, 0.29]
Total (95% CI)			**146**			**146**	**100.0%**	**-0.24 [-0.72, 0.24]**

Heterogeneity: Chi² = 0.76, df = 1 (P = 0.38); I² = 0%

Test for overall effect: Z = 0.99 (P = 0.32)

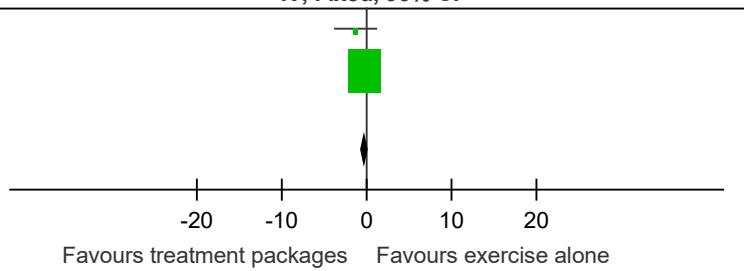


Figure 23: Discontinuation at ≤3 months

Study or Subgroup	Treatment packages Events	Total	Exercise alone Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Alasfour 2020	2	20	3	20	5.6%	0.67 [0.12, 3.57]
Alfieri 2020	7	29	15	32	26.6%	0.51 [0.24, 1.08]
Arnold 2010	5	28	8	27	15.2%	0.60 [0.23, 1.61]
Bennell 2016	5	73	8	75	14.7%	0.64 [0.22, 1.87]
Brosseau 2012	10	69	10	79	17.4%	1.14 [0.51, 2.59]
Dziedzic 2015 (SMOotH)	2	65	1	65	1.9%	2.00 [0.19, 21.52]
Focht 2014	7	40	9	40	16.8%	0.78 [0.32, 1.88]
Hsu 2021	1	22	1	22	1.9%	1.00 [0.07, 15.00]
Total (95% CI)		**346**		**360**	**100.0%**	**0.75 [0.52, 1.08]**
Total events	39		55			

Heterogeneity: Chi² = 3.01, df = 7 (P = 0.88); I² = 0%

Test for overall effect: Z = 1.55 (P = 0.12)

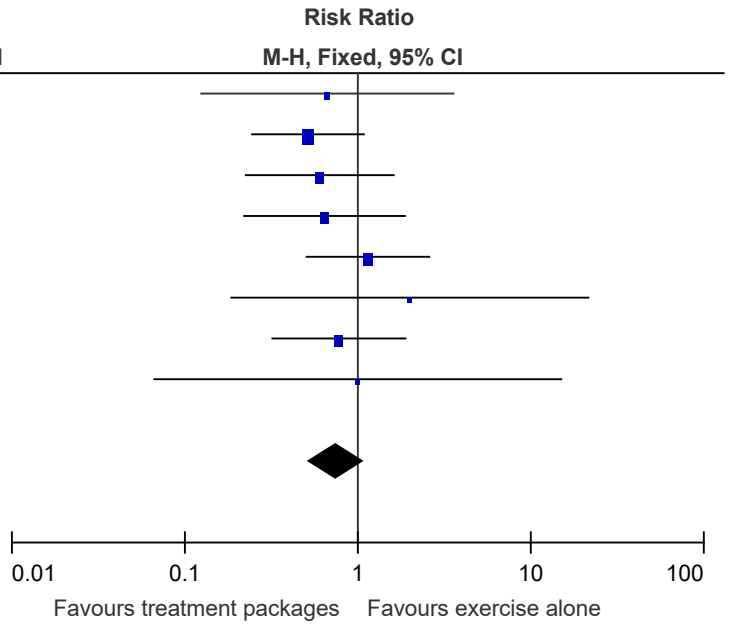
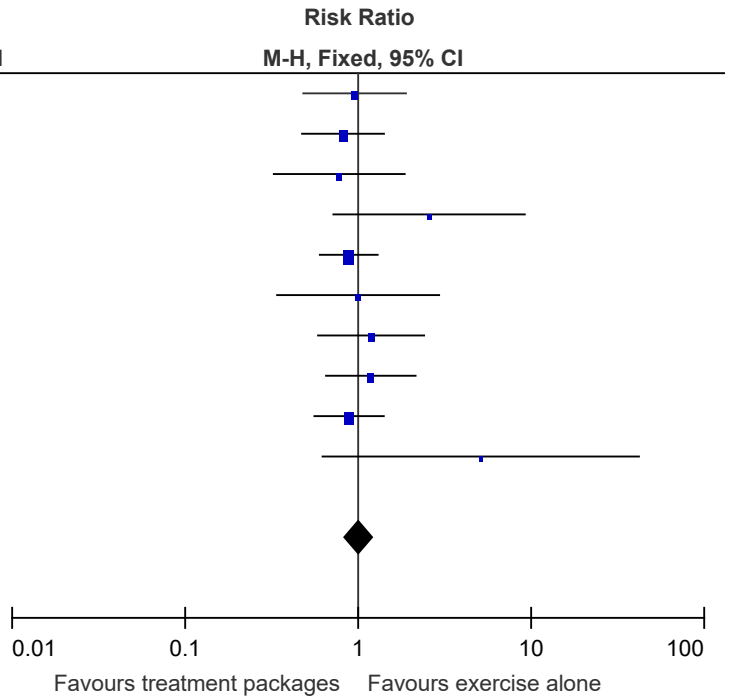


Figure 24: Discontinuation at >3 months

Study or Subgroup	Treatment packages Events	Total	Exercise alone Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Bennell 2016	13	73	14	75	9.6%	0.95 [0.48, 1.89]
Bennell 2017	18	84	22	84	15.3%	0.82 [0.47, 1.41]
Bennell 2018 (HOPE)	8	73	10	71	7.0%	0.78 [0.33, 1.86]
Bennell 2020	8	56	3	54	2.1%	2.57 [0.72, 9.19]
Brosseau 2012	27	69	35	79	22.6%	0.88 [0.60, 1.30]
Dziedzic 2015 (SMOotH)	6	65	6	65	4.2%	1.00 [0.34, 2.94]
Farr 2010	15	100	12	95	8.5%	1.19 [0.59, 2.40]
Focht 2005	18	76	16	80	10.8%	1.18 [0.65, 2.15]
Mcknight 2010	25	95	27	91	19.1%	0.89 [0.56, 1.41]
Quilty 2003	5	43	1	44	0.7%	5.12 [0.62, 42.01]
Total (95% CI)		**734**		**738**	**100.0%**	**1.00 [0.82, 1.22]**
Total events	143		146			

Heterogeneity: Chi² = 6.49, df = 9 (P = 0.69); I² = 0%

Test for overall effect: Z = 0.02 (P = 0.99)



E.2 Treatment packages compared to manual therapy alone

Figure 25: Pain (WOMAC, 0-500, high is poor, final value) at ≤3 months

	Treatment packages			Manual therapy alone			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Dwyer 2015	97.7	86.8	28	102.3	88.9	27	-4.60 [-51.06, 41.86]

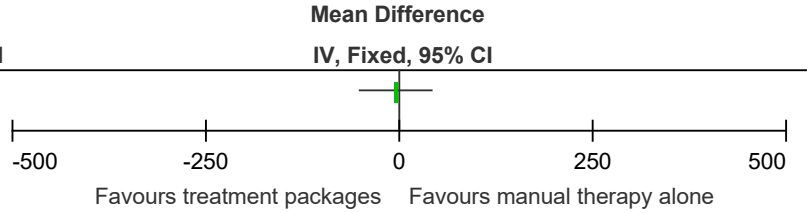


Figure 26: Physical function (WOMAC, 0-1800, high is poor, final value) at ≤3 months

	Treatment packages			Manual therapy alone			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Dwyer 2015	378.9	261.4	28	389.7	293.1	27	-10.80 [-157.76, 136.16]

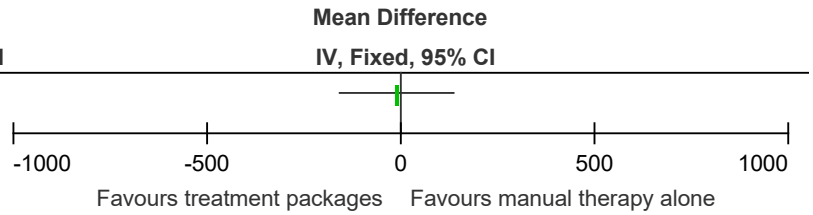
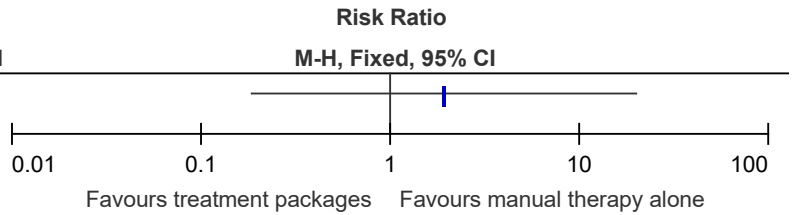


Figure 27: Discontinuation at ≤3 months

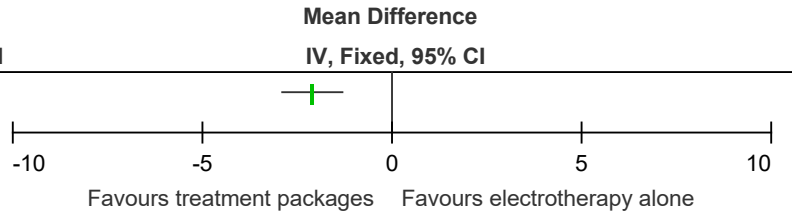
	Treatment packages		Manual therapy alone		Risk Ratio
Study or Subgroup	Events	Total	Events	Total	M-H, Fixed, 95% CI
Dwyer 2015	2	28	1	27	1.93 [0.19, 20.05]



E.3 Treatment packages compared to electrotherapy alone

Figure 28: Pain (VAS, 0-10, high is poor, change score) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Electrotherapy alone Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Huang 2000	-4	2	42	-1.9	1.7	42	-2.10 [-2.89, -1.31]



E.4 Treatment packages compared to behaviour change interventions alone

Figure 29: Quality of life (AQOL II, -0.04-1, high is good, change score) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Bennell 2016	0.1	0.1	73	0.1	0.1	74	0.00 [-0.03, 0.03]

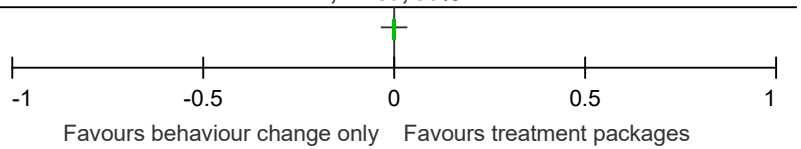


Figure 30: Quality of life (AIMS pain, 0-10, high is poor, final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Keefe 2004	4.26	1.45	20	4	1.56	18	0.26 [-0.70, 1.22]

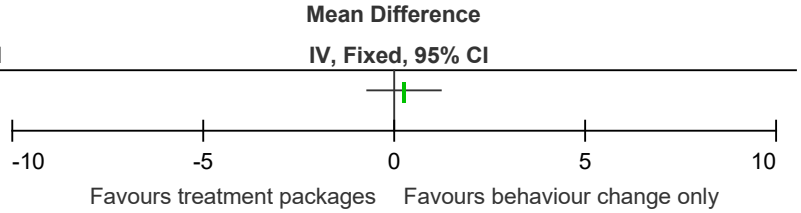


Figure 31: Quality of life (AIMS psychological distress, 0-10, high is poor, final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Keefe 2004	2.21	1.21	20	2.38	1.38	18	-0.17 [-1.00, 0.66]

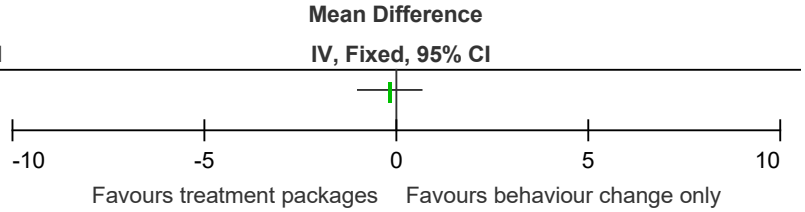


Figure 32: Quality of life (AQOL II, -0.04-1, high is good, change score) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Bennell 2016	0.1	0.1	73	0.1	0.1	74	0.00 [-0.03, 0.03]

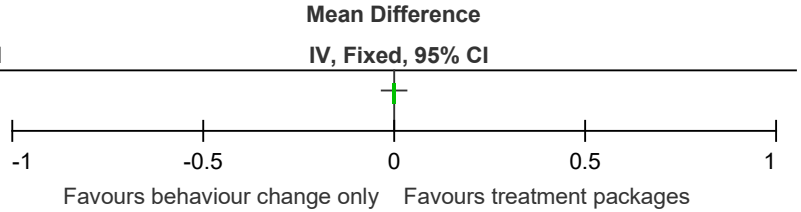


Figure 33: Quality of life (SF-36 physical composite, 0-100, high is good, final value) at >3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	Behaviour change only Total	Mean Difference IV, Fixed, 95% CI
Rejeski 2002 (ADAPT)	2.16	1.1812	68	73	2.16 [-0.16, 4.48]

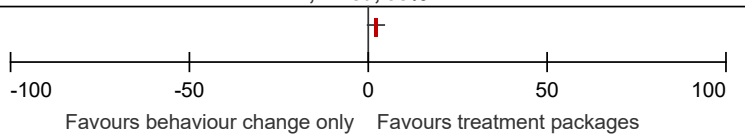


Figure 34: Quality of life (SF-36 mental composite, 0-100, high is good, final value) at >3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	Behaviour change only Total	Mean Difference IV, Fixed, 95% CI
Rejeski 2002 (ADAPT)	-0.55	1.1312	68	73	-0.55 [-2.77, 1.67]

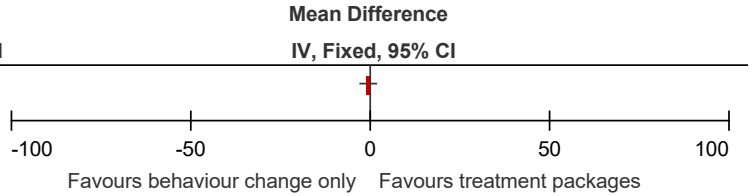


Figure 35: Pain (WOMAC, 0-20, high is poor, change scores) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Bennell 2016	-4.4	3	73	-2.6	3.6	74	41.8%	-1.80 [-2.87, -0.73]
Hsu 2021	-2.95	1.12	21	-2.14	1.28	21	58.2%	-0.81 [-1.54, -0.08]
Total (95% CI)			**94**			**95**	**100.0%**	**-1.22 [-2.18, -0.27]**

Heterogeneity: Tau² = 0.27; Chi² = 2.25, df = 1 (P = 0.13); I² = 55%

Test for overall effect: Z = 2.51 (P = 0.01)

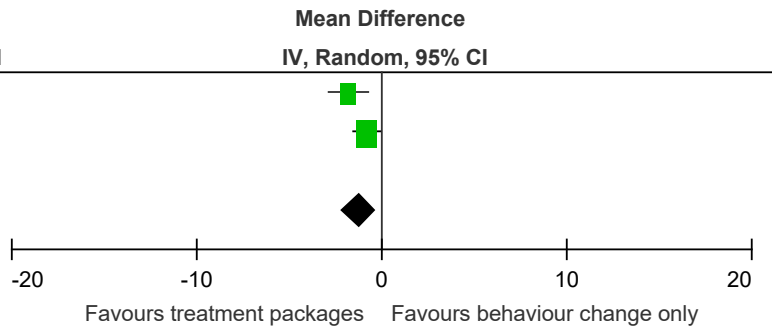


Figure 36: Pain (WOMAC, 0-500, high is poor, final value) at ≤3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Farr 2010	67.1	68.8	100	72	66.3	98	-4.90 [-23.72, 13.92]

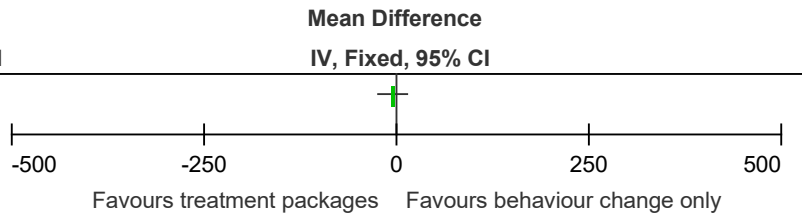


Figure 37: Pain (WOMAC, 0-20, high is poor, change scores) at >3 months

	Treatment packages			Behaviour change only				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Bennell 2016	-3.8	3.4	73	-2.6	3.3	74	58.2%	-1.20 [-2.28, -0.12]
Focht 2005	-2.2	4.1	76	-1.07	4.1	82	41.8%	-1.13 [-2.41, 0.15]
Total (95% CI)			**149**			**156**	**100.0%**	**-1.17 [-2.00, -0.34]**

Heterogeneity: Chi² = 0.01, df = 1 (P = 0.93); I² = 0%

Test for overall effect: Z = 2.78 (P = 0.006)

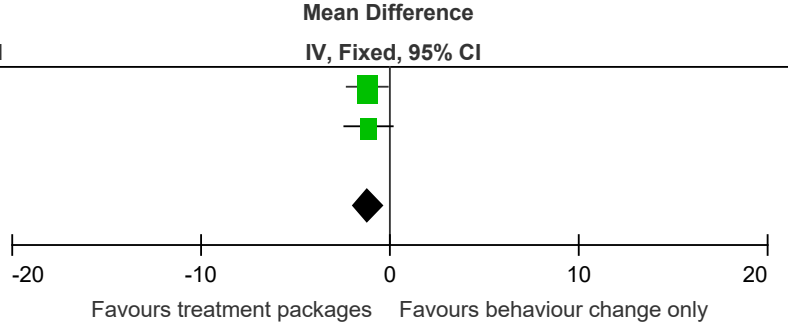


Figure 38: Pain (WOMAC, 0-500, high is poor, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Farr 2010	56.2	75.3	100	62.9	81	98	-6.70 [-28.49, 15.09]

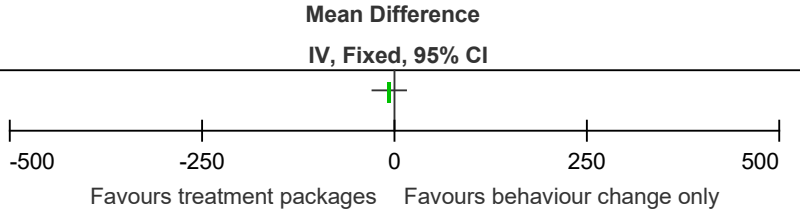


Figure 39: Physical function (WOMAC, 0-68, high is poor, change scores) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Bennell 2016	-19.9	9.1	73	-11.2	10.3	74	47.7%	-8.70 [-11.84, -5.56]
Hsu 2021	-8.62	3.58	21	-5.76	2.84	21	52.3%	-2.86 [-4.81, -0.91]
Total (95% CI)			**94**			**95**	**100.0%**	**-5.65 [-11.36, 0.07]**

Heterogeneity: Tau² = 15.27; Chi² = 9.57, df = 1 (P = 0.002); I² = 90%

Test for overall effect: Z = 1.94 (P = 0.05)

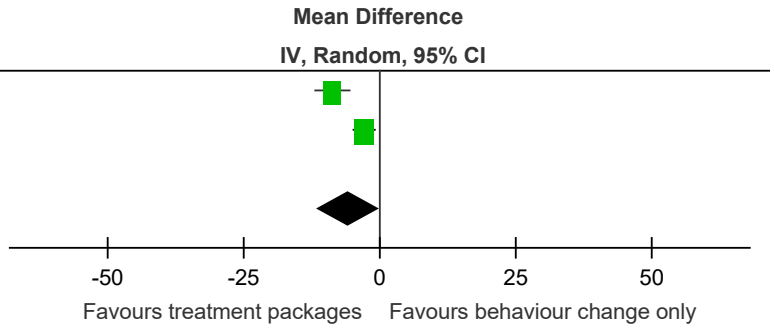


Figure 40: Physical function (WOMAC, 0-68, high is poor, change score) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Behaviour change only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Bennell 2016	-19.1	10.1	73	-12.3	10.7	74	-6.80 [-10.16, -3.44]

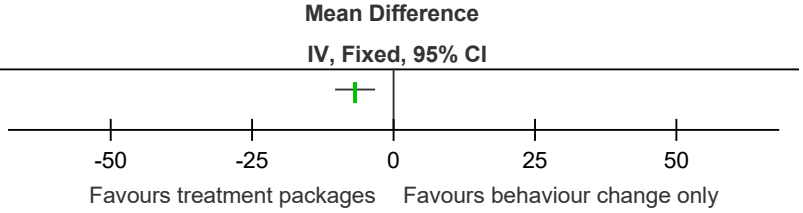


Figure 41: Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at ≤3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-0.9	4.1	73	-1.9	4.1	74	1.00 [-0.33, 2.33]

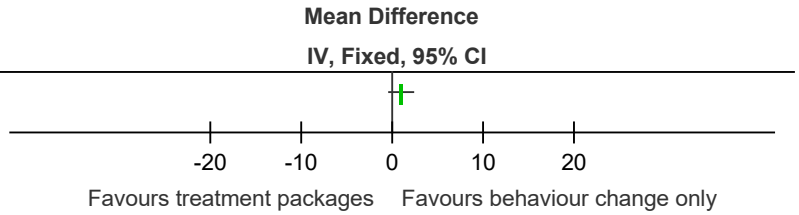


Figure 42: Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at ≤3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-0.9	4.8	73	-0.6	6.3	74	-0.30 [-2.11, 1.51]

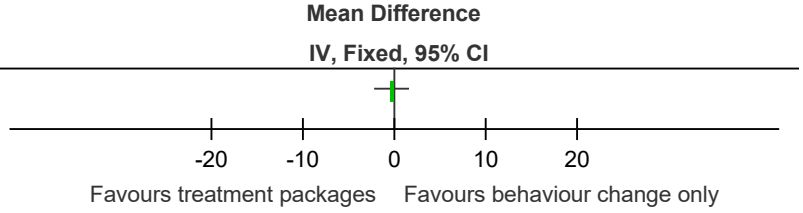


Figure 43: Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at ≤3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-0.5	5.6	73	-0.3	6.1	74	-0.20 [-2.09, 1.69]

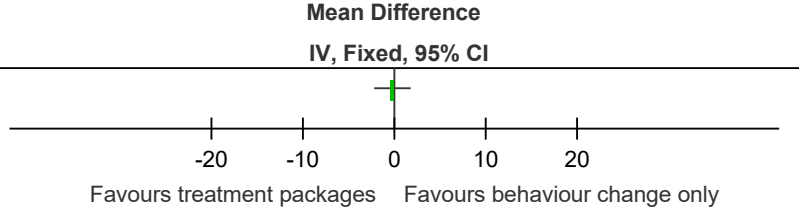


Figure 44: Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at >3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-2	4.9	73	-2.1	4	74	0.10 [-1.35, 1.55]

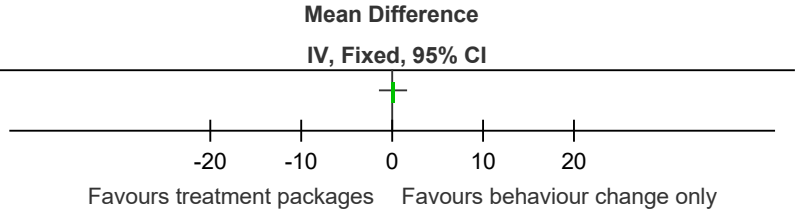


Figure 45: Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at >3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-1.4	6	73	-0.9	4.2	74	-0.50 [-2.18, 1.18]

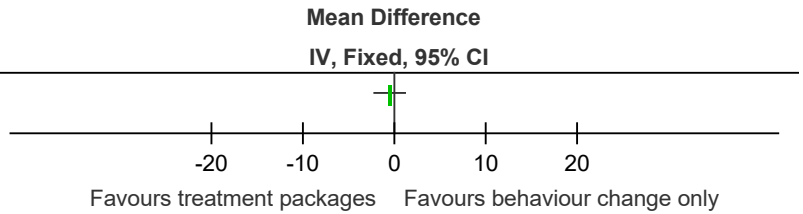


Figure 46: Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at >3 months

	Treatment packages			Behaviour change only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Bennell 2016	-2.1	6.3	73	-1.7	6.7	74	-0.40 [-2.50, 1.70]

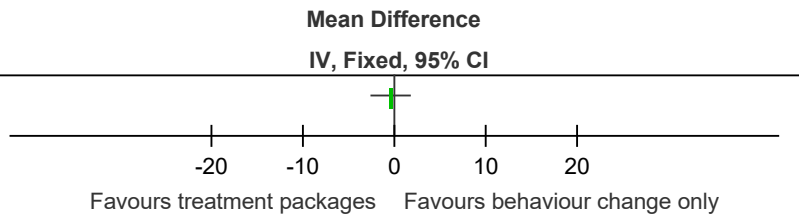


Figure 47: Discontinuation at ≤3 months

Study or Subgroup	Treatment packages Events	Total	Behaviour change only Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Bennell 2016	5	73	8	74	66.1%	0.63 [0.22, 1.85]
Dziedzic 2015 (SMOotH)	2	65	3	62	25.6%	0.64 [0.11, 3.68]
Hsu 2021	1	22	1	22	8.3%	1.00 [0.07, 15.00]
Total (95% CI)		**160**		**158**	**100.0%**	**0.66 [0.28, 1.58]**
Total events	8		12			

Heterogeneity: Chi² = 0.10, df = 2 (P = 0.95); I² = 0%

Test for overall effect: Z = 0.93 (P = 0.35)

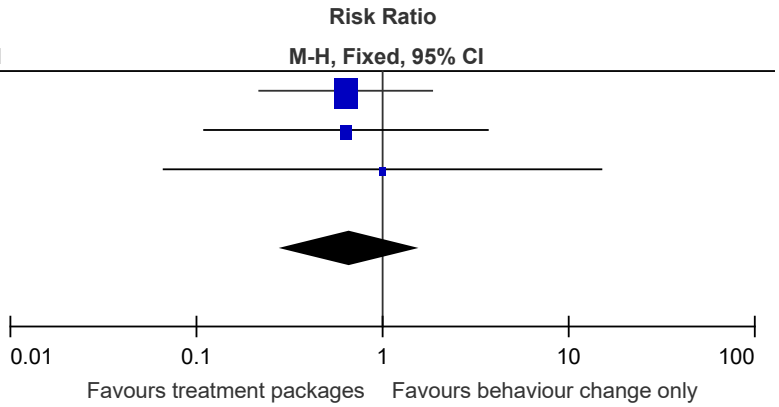
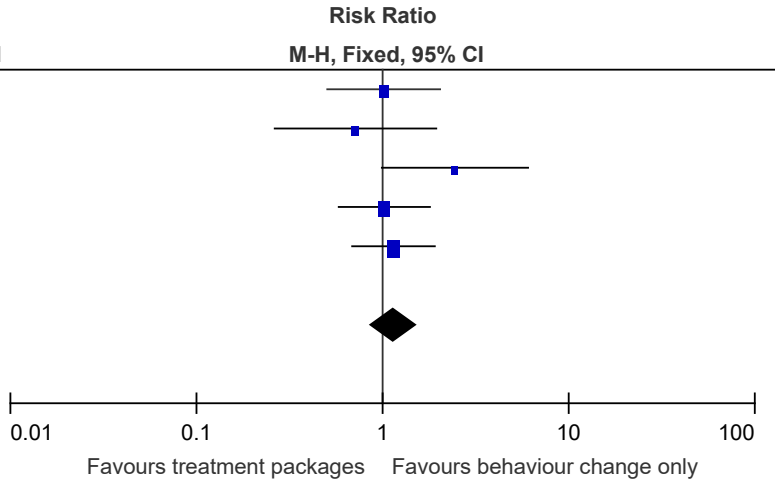


Figure 48: Discontinuation at >3 months

Study or Subgroup	Treatment packages Events	Total	Behaviour change only Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Bennell 2016	13	73	13	74	19.5%	1.01 [0.50, 2.04]
Dziedzic 2015 (SMOotH)	6	65	8	62	12.3%	0.72 [0.26, 1.94]
Farr 2010	15	100	6	98	9.1%	2.45 [0.99, 6.06]
Focht 2005	18	76	19	82	27.6%	1.02 [0.58, 1.80]
Mcknight 2010	25	95	20	87	31.5%	1.14 [0.69, 1.91]
Total (95% CI)		**409**		**403**	**100.0%**	**1.15 [0.86, 1.55]**
Total events	77		66			

Heterogeneity: Chi² = 3.85, df = 4 (P = 0.43); I² = 0%

Test for overall effect: Z = 0.94 (P = 0.35)



E.5 Treatment packages compared to education programmes alone

Figure 49: Quality of life (EQ-5D 5L, -0.11-1, high is good, final value) at ≤3 months

	Treatment packages			Education programme only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Adams 2021	0.63	0.22	84	0.61	0.24	83	0.02 [-0.05, 0.09]

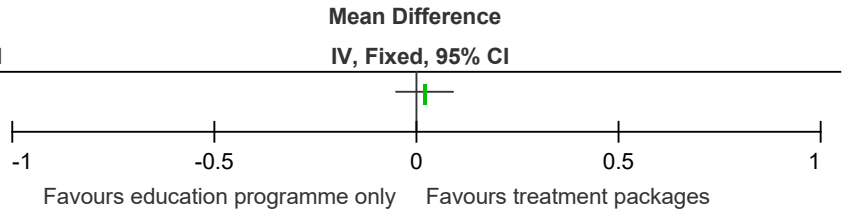


Figure 50: Quality of life (HOOS, KOOS, 0-100, high is good, change scores and final value) at ≤3 months

	Treatment packages			Education programme only				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Crossley 2015	54.7	20	39	49.8	13.8	42	40.7%	4.90 [-2.64, 12.44]
Kemp 2018	3	16	10	-5	18	7	8.4%	8.00 [-8.62, 24.62]
Poulsen 2013	12	18	38	-2	11	37	51.0%	14.00 [7.27, 20.73]
Total (95% CI)			**87**			**86**	**100.0%**	**9.80 [4.99, 14.60]**

Heterogeneity: Chi² = 3.16, df = 2 (P = 0.21); I² = 37%

Test for overall effect: Z = 4.00 (P < 0.0001)

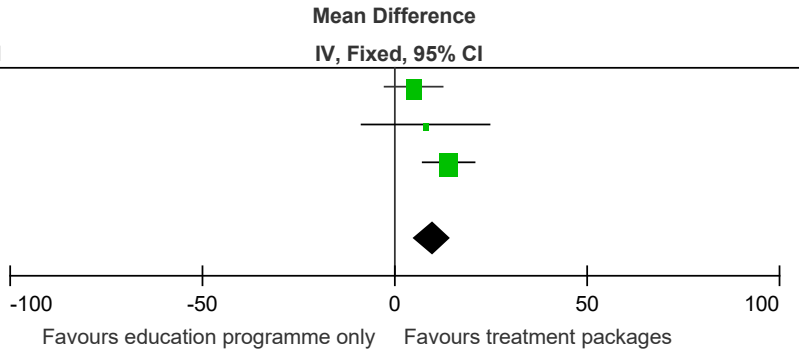


Figure 51: Quality of life (AIMS-2 pain subscale, 0-10, high is good, final value) at ≤3 months

	Treatment packages			Education programme only			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Gaines 2004	5.18	2.11	20	5.99	2.4	18	-0.81 [-2.25, 0.63]	

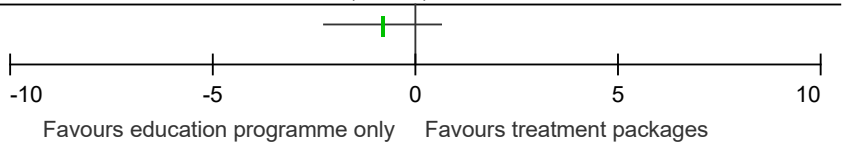


Figure 52: Quality of life (HOOS, KOOS, 0-100, high is good, change score and final value) at >3 months

	Treatment packages			Education programme only				Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Crossley 2015	56	19.6	35	52	15.2	34	63.0%	4.00 [-4.26, 12.26]	
Poulsen 2013	10	20	38	10	27	37	37.0%	0.00 [-10.78, 10.78]	
Total (95% CI)			**73**			**71**	**100.0%**	**2.52 [-4.04, 9.08]**	

Heterogeneity: Chi² = 0.33, df = 1 (P = 0.56); I² = 0%

Test for overall effect: Z = 0.75 (P = 0.45)

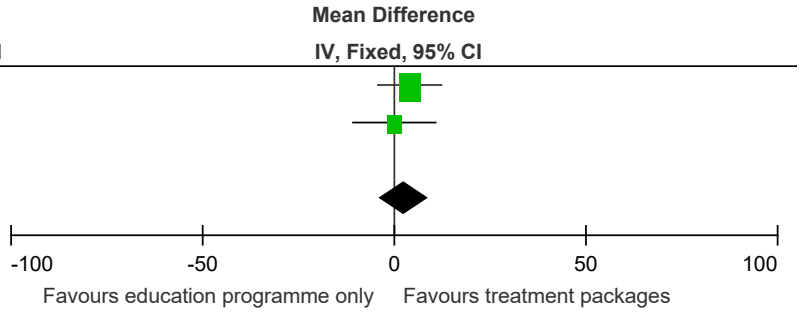


Figure 53: Quality of life (SF-36 physical function, 0-100, high is good, final value) at >3 months

	Treatment packages			Education programme only			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Fernandes 2010	75.5	20.5	40	71.3	20.8	35	4.20 [-5.17, 13.57]	

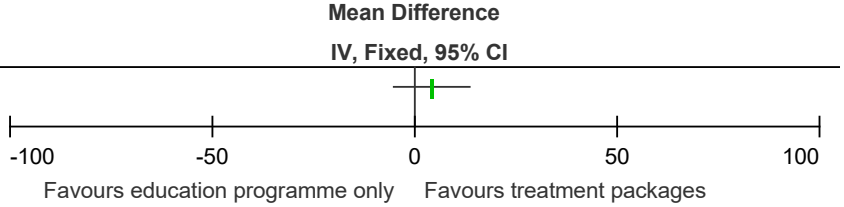


Figure 54: Quality of life (SF-36 bodily pain, 0-100, high is good, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Fernandes 2010	70.5	18.6	41	61.4	24.3	37	9.10 [-0.58, 18.78]

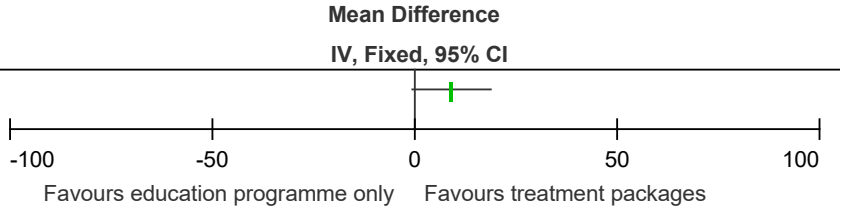


Figure 55: Quality of life (SF-36 role physical, 0-100, high is good, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Fernandes 2010	82.3	25.5	41	75.7	29	37	6.60 [-5.58, 18.78]

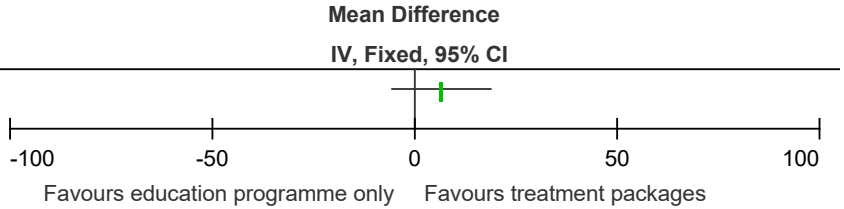


Figure 56: Quality of life (SF-36 vitality, 0-100, high is good, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Fernandes 2010	59	21	41	61.7	20.6	37	-2.70 [-11.94, 6.54]

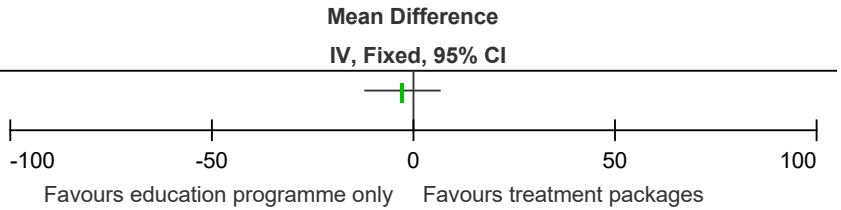


Figure 57: Quality of life (SF-36 general health, 0-100, high is good, final value) at >3 months

	Treatment packages			Education programme only			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Fernandes 2010	71.3	20.7	38	67.6	22.1	36	3.70 [-6.07, 13.47]	

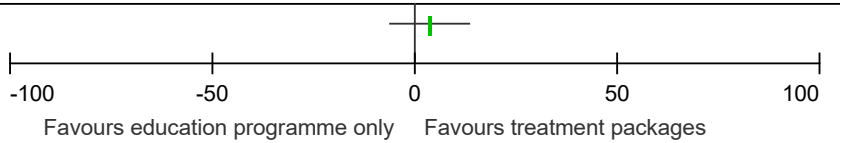


Figure 58: Quality of life (SF-36 mental health, 0-100, high is good, final value) at >3 months

	Treatment packages			Education programme only			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Fernandes 2010	81.8	14.9	40	82.8	15.4	37	-1.00 [-7.78, 5.78]	

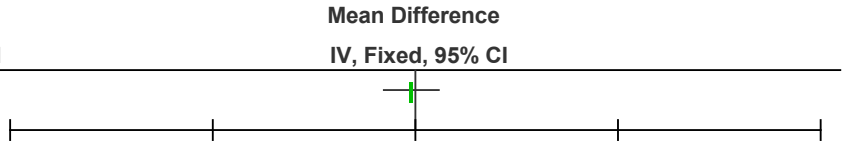


Figure 59: Quality of life (SF-36 role emotional, 0-100, high is good, final value) at >3 months

	Treatment packages			Education programme only			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Fernandes 2010	90.7	15.5	41	90.5	21.7	37	0.20 [-8.25, 8.65]	

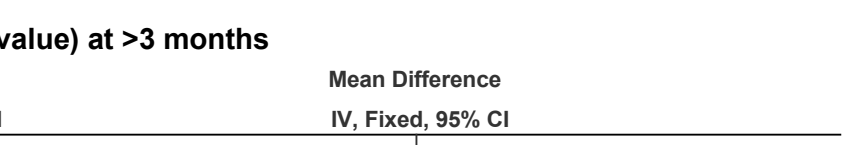


Figure 60: Quality of life (SF-36 social function, 0-100, high is good, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Fernandes 2010	91.2	15.9	41	84.1	26.9	37	7.10 [-2.84, 17.04]

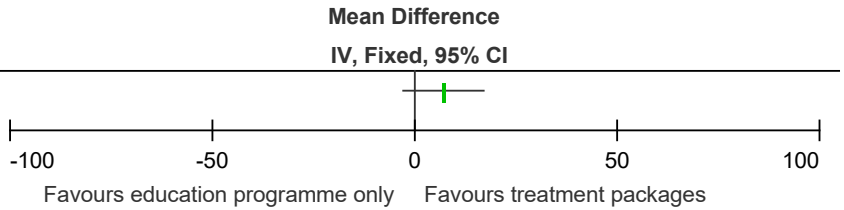


Figure 61: Pain (HOOS, KOOS, WOMAC, VAS, 0-100, high is good, change scores and final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Crossley 2015	76.3	13.4	39	69.4	14.2	42	24.0%	6.90 [0.89, 12.91]
Deveza 2021	-35.5	22.1	96	-43	23.5	98	23.0%	7.50 [1.08, 13.92]
Dias 2017	13.33	16.23	37	2.3	15.1	36	21.3%	11.03 [3.84, 18.22]
Kemp 2018	10	19	10	-2	21	7	6.3%	12.00 [-7.51, 31.51]
Poulsen 2013	18	13	38	-1	11	37	25.3%	19.00 [13.56, 24.44]
Total (95% CI)			**220**			**220**	**100.0%**	**11.31 [5.87, 16.74]**

Heterogeneity: Tau² = 22.62; Chi² = 11.03, df = 4 (P = 0.03); I² = 64%

Test for overall effect: Z = 4.08 (P < 0.0001)

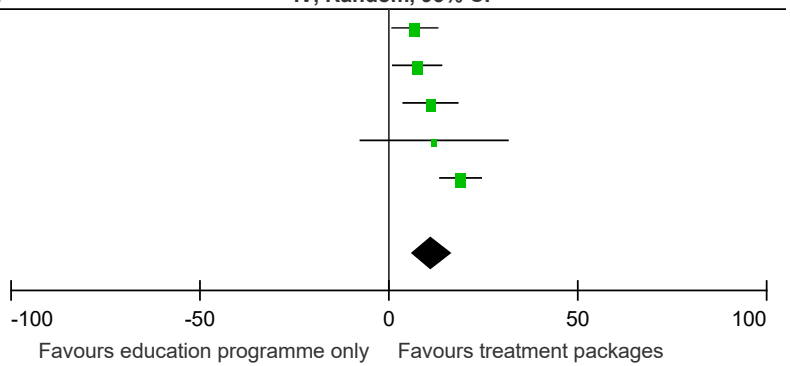


Figure 62: Pain (KOOS, AUSCAN, McGill Pain Questionnaire, pain rating index [different scale ranges], high is poor, final values) at ≤3 months

	Treatment packages			Education programme only				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Adams 2021	-9.7	3.9	91	-9.7	4	90	62.8%	0.00 [-0.29, 0.29]
Gaines 2004	14.95	13.07	20	10.63	4.84	18	12.8%	0.42 [-0.22, 1.06]
Talbot 2003A	12.41	9.77	17	10.12	4.64	17	11.6%	0.29 [-0.38, 0.97]
Talbot 2003B	16.33	13.35	20	11.12	8	18	12.8%	0.46 [-0.19, 1.10]
Total (95% CI)			**148**			**143**	**100.0%**	**0.15 [-0.08, 0.38]**

Heterogeneity: Chi² = 2.73, df = 3 (P = 0.43); I² = 0%

Test for overall effect: Z = 1.24 (P = 0.21)

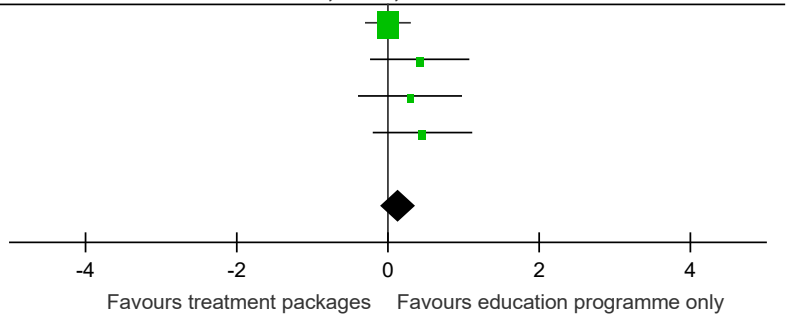


Figure 63: Pain (HOOS, KOOS, WOMAC, 0-100, high is poor, change score and final values) at >3 months

	Treatment packages			Education programme only				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Crossley 2015	-75.5	16.5	35	-73.5	14.4	34	39.7%	-2.00 [-9.30, 5.30]
Fernandes 2010	17.3	14.5	42	22.3	18.4	36	38.2%	-5.00 [-12.44, 2.44]
Poulsen 2013	-16	20	38	-11	23	37	22.2%	-5.00 [-14.77, 4.77]
Total (95% CI)			**115**			**107**	**100.0%**	**-3.81 [-8.41, 0.79]**

Heterogeneity: Chi² = 0.39, df = 2 (P = 0.82); I² = 0%

Test for overall effect: Z = 1.62 (P = 0.10)

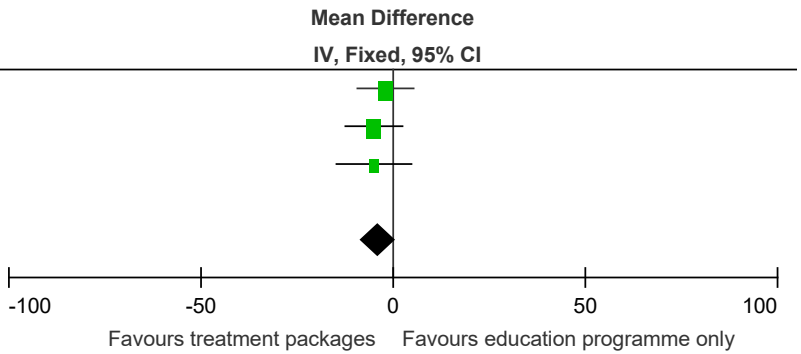


Figure 64: Pain (WOMAC, McGill Pain Questionnaire, pain rating index [different scale ranges], high is poor, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Gaines 2004	19.38	13.66	20	10.44	5.25	18	25.4%	0.83 [0.16, 1.50]
Oh 2020	5.06	4.39	21	10.33	5.22	11	23.6%	-1.10 [-1.88, -0.31]
Talbot 2003A	12.95	11.41	17	10.9	9.69	17	25.3%	0.19 [-0.48, 0.86]
Talbot 2003B	16.14	12.03	20	12.42	9.66	18	25.8%	0.33 [-0.31, 0.97]
Total (95% CI)			**78**			**64**	**100.0%**	**0.09 [-0.66, 0.83]**

Heterogeneity: Tau² = 0.45; Chi² = 14.03, df = 3 (P = 0.003); I² = 79%

Test for overall effect: Z = 0.22 (P = 0.82)

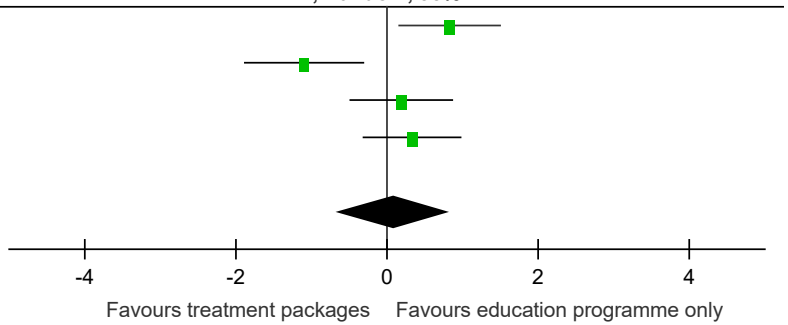


Figure 65: Physical function (HOOS, KOOS, WOMAC, 0-100, high is good, change scores and final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	Education programme only Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Crossley 2015	83.8	12.8	39	76.6	14.6	42	32.8%	7.20 [1.23, 13.17]
Dias 2017	16.4	17.5	37	5.1	9.6	36	28.1%	11.30 [4.85, 17.75]
Kemp 2018	8	13	10	-7	14	7	6.8%	15.00 [1.87, 28.13]
Poulsen 2013	15	16	38	1	10	37	32.3%	14.00 [7.98, 20.02]
Total (95% CI)			**124**			**122**	**100.0%**	**11.08 [7.66, 14.50]**

Heterogeneity: Chi² = 2.87, df = 3 (P = 0.41); I² = 0%

Test for overall effect: Z = 6.35 (P < 0.00001)

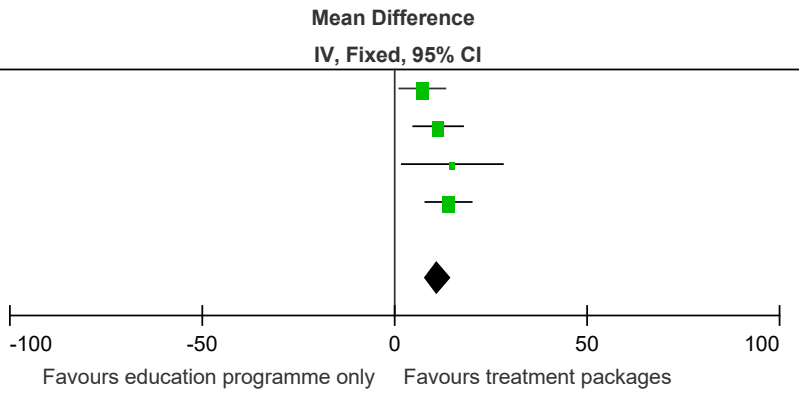


Figure 66: Physical function (AUSCAN, Functional Index of Hand Osteoarthritis [different scale ranges], high is good, final values) at ≤3 months

	Treatment packages			Education programme only				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Adams 2021	17.3	7.9	85	18.2	75	84	49.4%	-0.02 [-0.32, 0.28]
Deveza 2021	-7.6	4.4	96	-9.5	4.4	98	50.6%	0.43 [0.15, 0.71]
Total (95% CI)			**181**			**182**	**100.0%**	**0.21 [-0.23, 0.65]**

Heterogeneity: Tau² = 0.08; Chi² = 4.46, df = 1 (P = 0.03); I² = 78%

Test for overall effect: Z = 0.94 (P = 0.35)

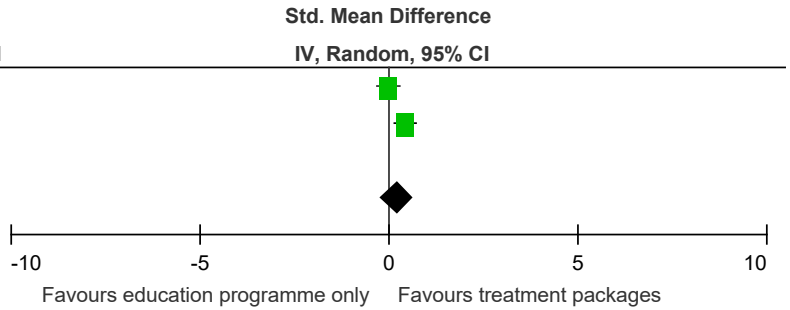


Figure 67: Physical function (HOOS, WOMAC, 0-100, high is poor, change score and final value) at >3 months

	Treatment packages			Education programme only				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI
Crossley 2015	-82.1	14.8	35	-77.7	16	34	39.8%	-4.40 [-11.68, 2.88]
Fernandes 2010	15.1	13.7	41	22.8	18.6	36	38.6%	-7.70 [-15.08, -0.32]
Poulsen 2013	-13	20	36	-9	23	37	21.6%	-4.00 [-13.88, 5.88]
Total (95% CI)			**112**			**107**	**100.0%**	**-5.59 [-10.18, -1.00]**

Heterogeneity: Chi² = 0.52, df = 2 (P = 0.77); I² = 0%

Test for overall effect: Z = 2.39 (P = 0.02)

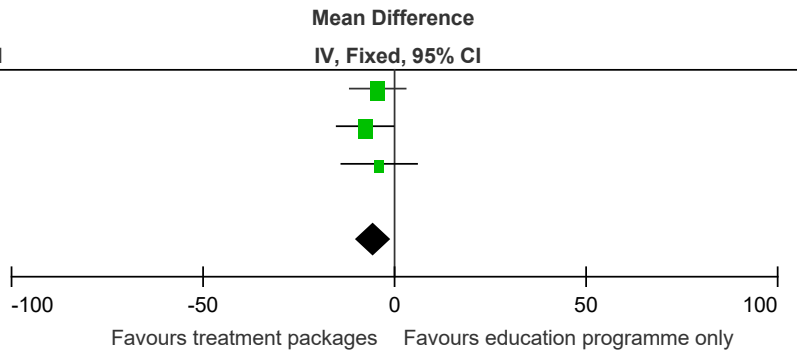


Figure 68: Physical function (WOMAC, 0-68, high is poor, final value) at >3 months

	Treatment packages			Education programme only			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Oh 2020	16.22	10.87	21	30.89	14.09	11	-14.67 [-24.21, -5.13]

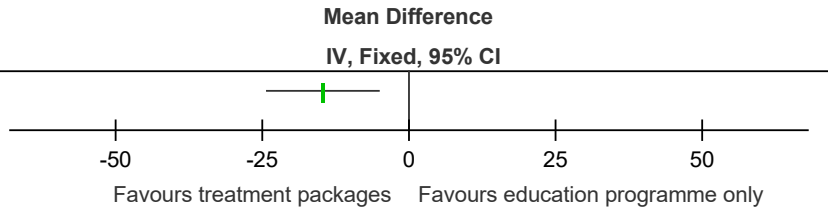


Figure 69: Discontinuation at ≤3 months

	Treatment packages		Education programme only			Risk Difference
Study or Subgroup	Events	Total	Events	Total	Weight	M-H, Fixed, 95% CI
Adams 2021	25	116	26	116	31.7%	-0.01 [-0.12, 0.10]
Crossley 2015	5	44	6	48	12.5%	-0.01 [-0.14, 0.12]
Deveza 2021	12	102	7	102	27.9%	0.05 [-0.03, 0.13]
Dias 2017	4	37	4	36	10.0%	-0.00 [-0.15, 0.14]
Kemp 2018	0	10	0	7	2.2%	0.00 [-0.21, 0.21]
Poulsen 2013	4	38	1	37	10.2%	0.08 [-0.03, 0.19]
Stener-victorin 2004	11	30	8	15	5.5%	-0.17 [-0.47, 0.14]
Total (95% CI)		**377**		**361**	**100.0%**	**0.01 [-0.04, 0.06]**
Total events	61		52			

Heterogeneity: Chi² = 4.02, df = 6 (P = 0.67); I² = 0%

Test for overall effect: Z = 0.31 (P = 0.75)

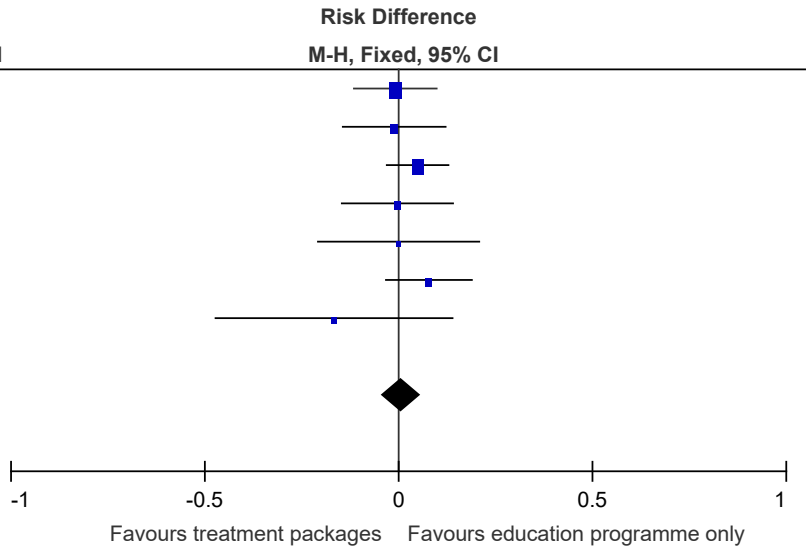
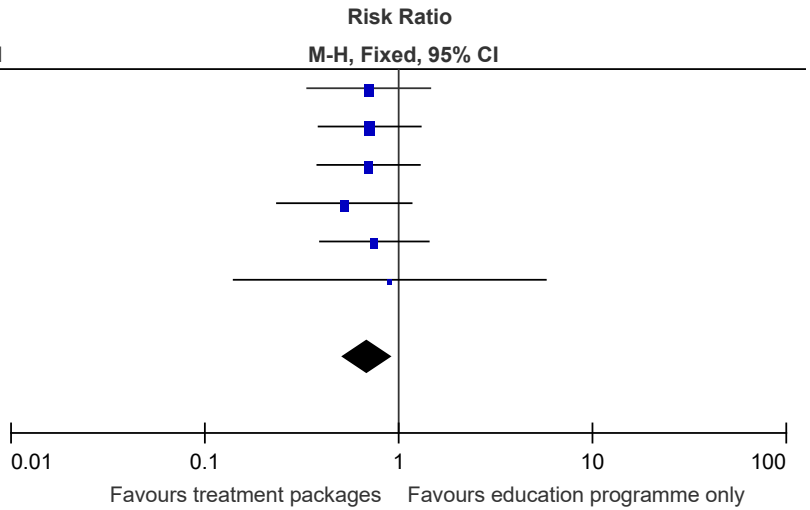


Figure 70: Discontinuation at >3 months

	Treatment packages		Education programme only			Risk Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Fixed, 95% CI**
Crossley 2015	9	44	14	48	18.9%	0.70 [0.34, 1.46]
Fernandes 2010	13	55	18	54	25.6%	0.71 [0.39, 1.30]
Oh 2020	14	40	10	20	18.8%	0.70 [0.38, 1.29]
Poulsen 2013	7	38	13	37	18.6%	0.52 [0.24, 1.17]
Stener-victorin 2004	12	30	8	15	15.1%	0.75 [0.39, 1.43]
Talbot 2003B	2	20	2	18	3.0%	0.90 [0.14, 5.74]
Total (95% CI)		**227**		**192**	**100.0%**	**0.68 [0.51, 0.92]**
Total events	57		65			

Heterogeneity: Chi² = 0.61, df = 5 (P = 0.99); I² = 0%

Test for overall effect: Z = 2.49 (P = 0.01)



E.6 Treatment packages compared to standard care (non-organised) or no treatment

Figure 71: Quality of life (EQ-5D, AQoL-2, -0.11-1, high is good, change scores and final values) at ≤3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Mean Difference IV, Random, 95% CI
Allen 2021	0.02	0.0219	162	96	28.6%	0.02 [-0.02, 0.06]
Bennell 2017 (IMPACT)	0.1	0.0339	70	69	22.8%	0.10 [0.03, 0.17]
Dziedzic 2018	-0.016	0.1924	4	4	1.9%	-0.02 [-0.39, 0.36]
Jessep 2009	0.04	0.0395	29	35	20.2%	0.04 [-0.04, 0.12]
Saw 2016	0.24	0.0779	35	39	9.1%	0.24 [0.09, 0.39]
Wallis 2017	0.1	0.0466	23	23	17.4%	0.10 [0.01, 0.19]
Total (95% CI)			**323**	**266**	**100.0%**	**0.08 [0.02, 0.13]**

Heterogeneity: Tau² = 0.00; Chi² = 11.16, df = 5 (P = 0.05); I² = 55%
Test for overall effect: Z = 2.76 (P = 0.006)

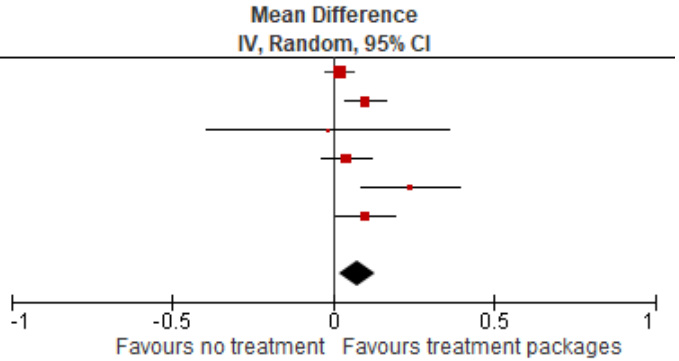


Figure 72: Quality of life (KOOS, HOOS, VAS quality of life, health assessment questionnaire, 0-100, high is good, change score and final values) at ≤3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Mean Difference IV, Random, 95% CI
Hopman-rock 2000	6.7	3.6705	56	49	19.0%	6.70 [-0.49, 13.89]
Kloek 2018	-3.9	4.7096	87	87	14.3%	-3.90 [-13.13, 5.33]
Li 2017	2.9	6.6542	17	17	8.8%	2.90 [-10.14, 15.94]
Poulsen 2013	8	3.3622	38	36	20.6%	8.00 [1.41, 14.59]
Yip 2007	-0.17	0.5516	88	94	37.3%	-0.17 [-1.25, 0.91]
Total (95% CI)			**286**	**283**	**100.0%**	**2.56 [-1.86, 6.97]**

Heterogeneity: Tau² = 13.34; Chi² = 9.88, df = 4 (P = 0.04); I² = 60%
Test for overall effect: Z = 1.13 (P = 0.26)

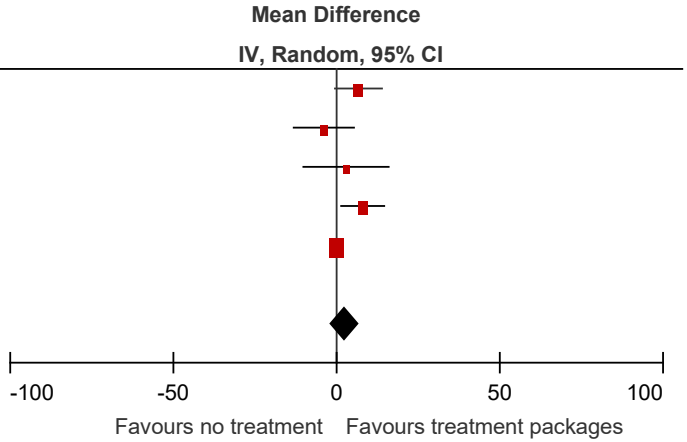


Figure 73: Quality of life (Health related quality of life, 7-39, high is good, final value) at ≤3 months

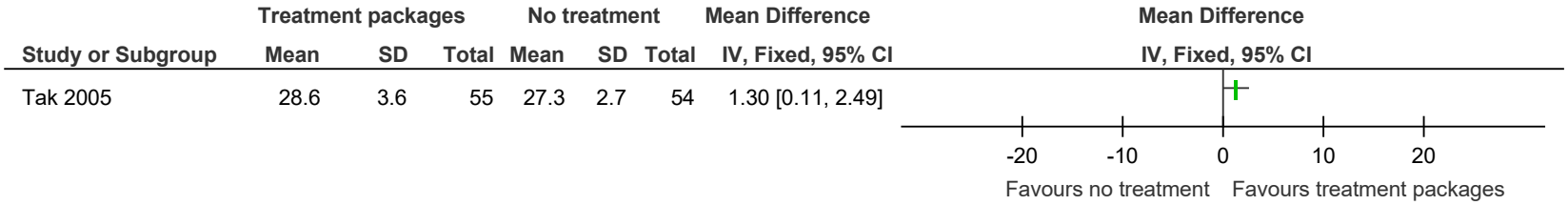


Figure 74: Quality of life (SF-36 physical component, 0-100, high is good, change scores) at ≤3 months

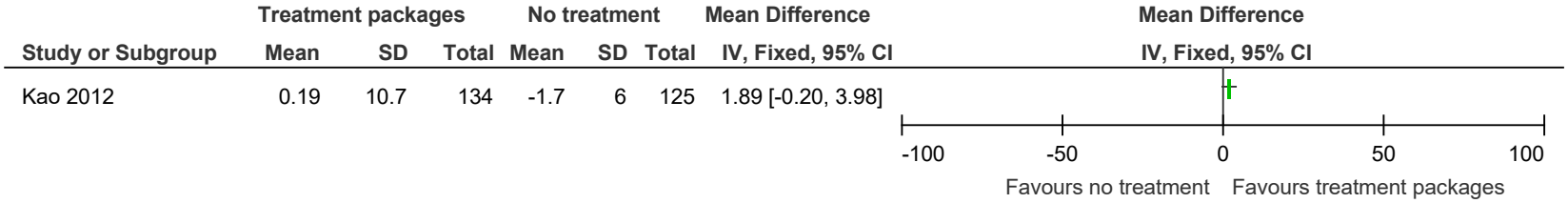


Figure 75: Quality of life (SF-36 mental component, 0-100, high is good, change scores) at ≤3 months

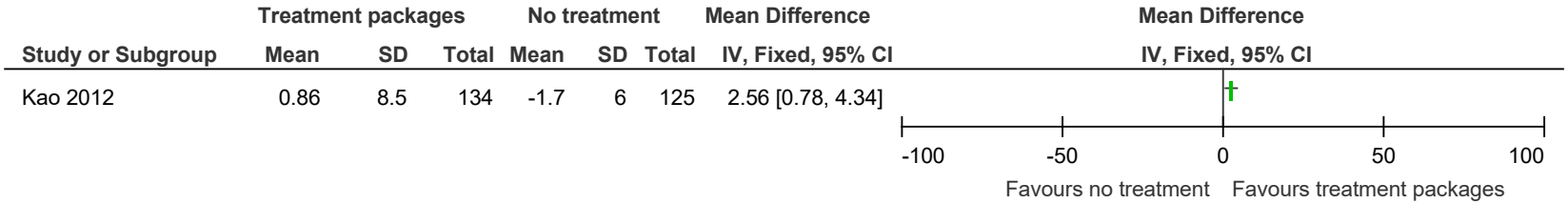


Figure 76: Quality of life (Geri-AIMS pain subscale, AIMS pain, 0-10, high is good, final values) at ≤3 months

	Treatment packages			No treatment				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Hughes 2006	4.67	0.85	115	4.65	0.86	100	49.9%	0.02 [-0.21, 0.25]
Keefe 2004	4.26	1.45	20	4.03	2.08	18	20.1%	0.23 [-0.92, 1.38]
Kovar 1992	-3.77	1.73	47	-4.77	2.12	45	29.9%	1.00 [0.21, 1.79]
Total (95% CI)			**182**			**163**	**100.0%**	**0.36 [-0.30, 1.01]**

Heterogeneity: Tau² = 0.21; Chi² = 5.47, df = 2 (P = 0.06); I² = 63%

Test for overall effect: Z = 1.06 (P = 0.29)

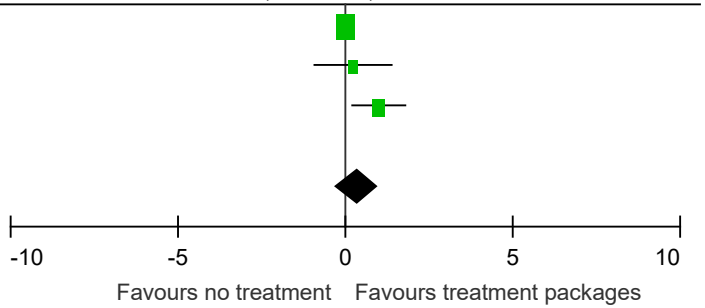


Figure 77: Quality of life (AIMS psychological disability, 0-10, high is poor, final value) at ≤3 months

	Treatment packages			No treatment			Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI
Keefe 2004	2.21	1.21	20	1.8	1.04	18	0.41 [-0.31, 1.13]

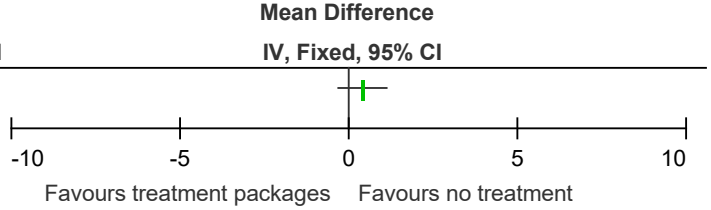


Figure 78: Quality of life (AIMS arthritis impact, 0-10, high is poor, final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	2.86	1.88	47	3.06	1.91	45	-0.20 [-0.97, 0.57]

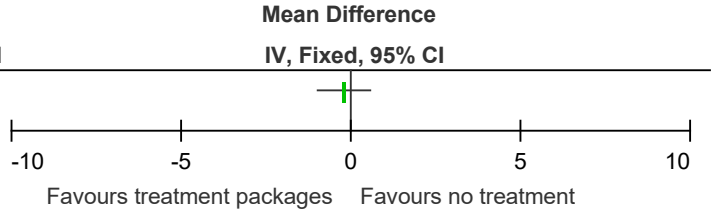


Figure 79: Quality of life (AIMS physical activity, 0-10, high is poor, final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	3.74	2.69	47	5.96	2.32	45	-2.22 [-3.25, -1.19]

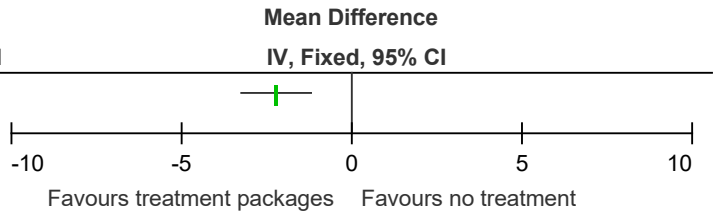


Figure 80: Quality of life (AIMS medications use, 0-6, high is good, final value) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	3.64	1.92	47	2.9	2.02	45	0.74 [-0.07, 1.55]

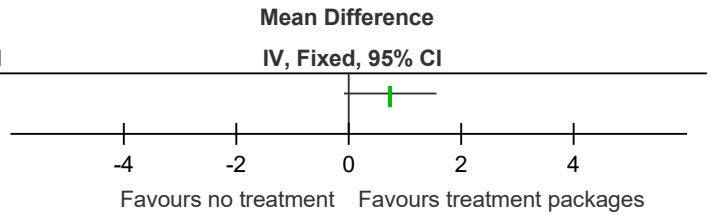


Figure 81: Quality of life (SF-36 physical function, 0-100, high is good, change scores) at ≤3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Da silva 2015	65.33	11.57	15	51.33	21.25	15	14.00 [1.76, 26.24]	

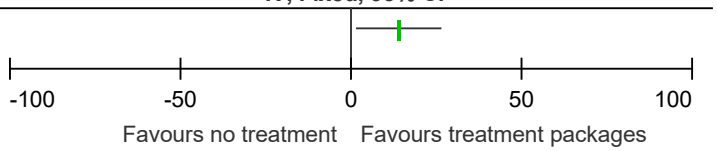


Figure 82: Quality of life (SF-36 bodily pain, 0-100, high is good, change scores) at ≤3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Da silva 2015	57.6	12.48	15	42.8	21.52	15	14.80 [2.21, 27.39]	

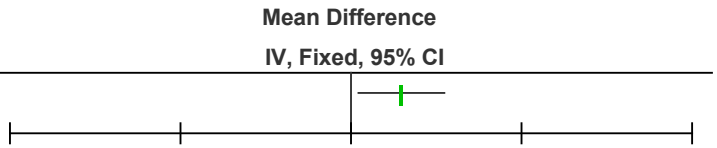


Figure 83: Quality of life (SF-36 role physical, 0-100, high is good, change scores) at ≤3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Da silva 2015	88.33	20.85	15	35	39.87	15	53.33 [30.56, 76.10]	

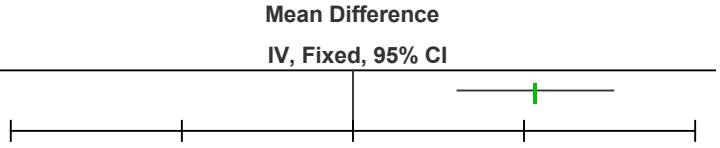


Figure 84: Quality of life (SF-36 vitality, 0-100, high is good, change scores) at ≤3 months

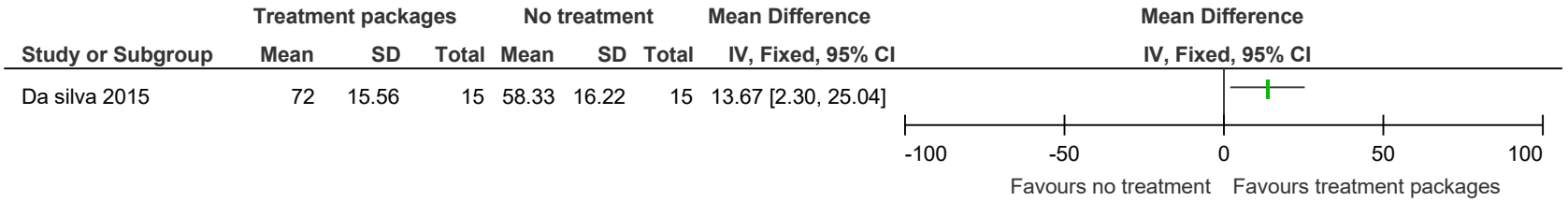


Figure 85: Quality of life (SF-36 general health, 0-100, high is good, change scores) at ≤3 months

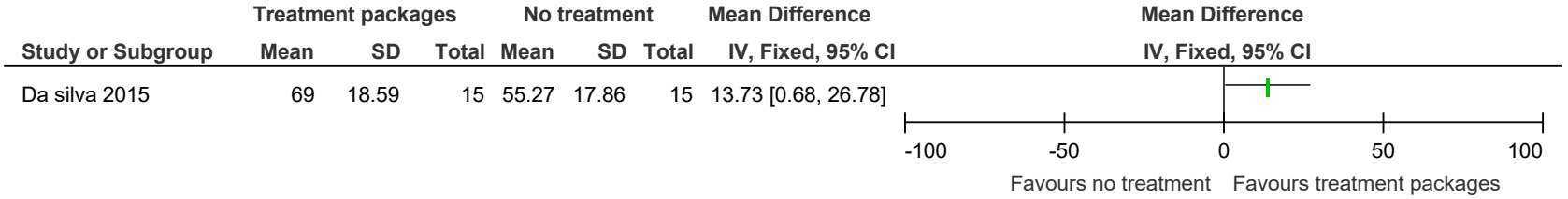


Figure 86: Quality of life (SF-36 mental health, 0-100, high is good, change scores) at ≤3 months

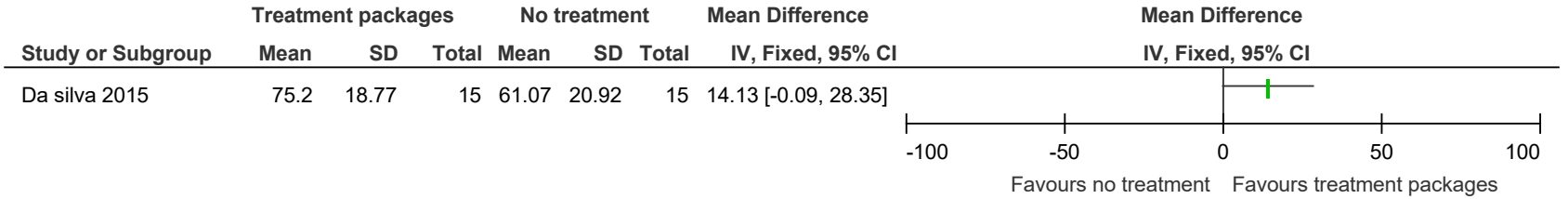


Figure 87: Quality of life (SF-36 role emotional, 0-100, high is good, change scores) at ≤3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Da silva 2015	86.67	30.37	15	53.2	32.99	15	33.47 [10.78, 56.16]	

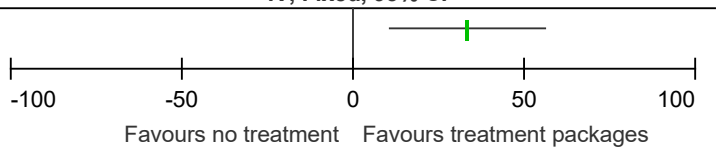


Figure 88: Quality of life (SF-36 social function, 0-100, high is good, change scores) at ≤3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Da silva 2015	91.67	12.2	15	90.83	13.75	15	0.84 [-8.46, 10.14]	

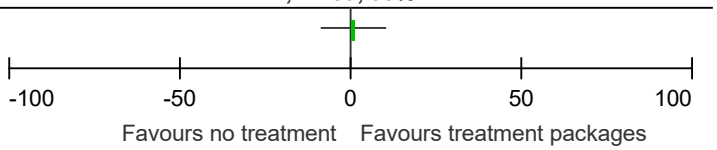


Figure 89: Quality of life (EQ-5D, AQoL-2, -0.11-1, high is good, change scores and final values) at >3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Mean Difference IV, Random, 95% CI
Allen 2021	0.04	0.0207	163	90	24.5%	0.04 [-0.00, 0.08]
Bennell 2017 (IMPACT)	0.1	0.0246	66	67	22.3%	0.10 [0.05, 0.15]
Dziedzic 2018	-0.023	0.1724	4	4	1.4%	-0.02 [-0.36, 0.31]
Hurley 2007	-0.02	0.031	229	140	18.9%	-0.02 [-0.08, 0.04]
Jessep 2009	0.05	0.0506	29	35	11.2%	0.05 [-0.05, 0.15]
Saw 2016	0.18	0.0792	35	39	5.7%	0.18 [0.02, 0.34]
Skou 2015	0.065	0.0373	50	50	15.9%	0.07 [-0.01, 0.14]
Total (95% CI)			**576**	**425**	**100.0%**	**0.05 [0.01, 0.10]**

Heterogeneity: Tau² = 0.00; Chi² = 12.46, df = 6 (P = 0.05); I² = 52%

Test for overall effect: Z = 2.59 (P = 0.009)

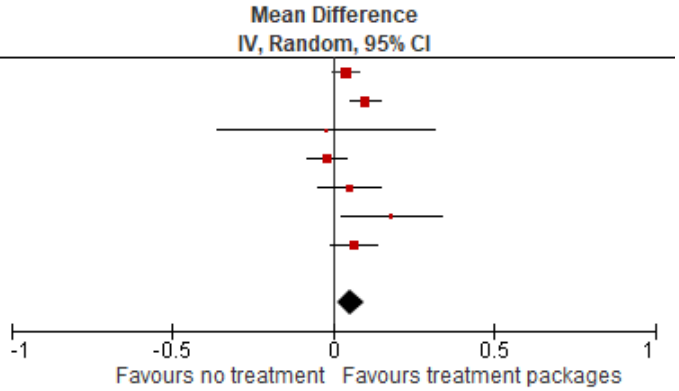


Figure 90: Quality of life (KOOS, HOOS, VAS quality of life, 0-100, high is good, change score and final values) at >3 months

Study or Subgroup	Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Mean Difference IV, Fixed, 95% CI
Hopman-rock 2000	-0.9	3.5837	56	49	53.5%	-0.90 [-7.92, 6.12]
Kloek 2018	-3.6	6.4791	65	69	16.4%	-3.60 [-16.30, 9.10]
Poulsen 2013	-2	4.7725	38	36	30.2%	-2.00 [-11.35, 7.35]
Total (95% CI)			**159**	**154**	**100.0%**	**-1.67 [-6.81, 3.46]**

Heterogeneity: Chi² = 0.14, df = 2 (P = 0.93); I² = 0%

Test for overall effect: Z = 0.64 (P = 0.52)

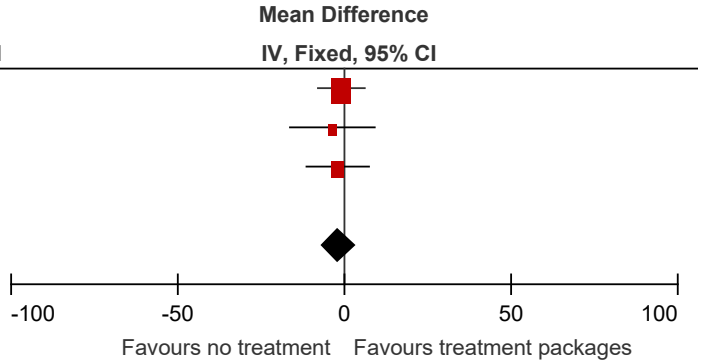


Figure 91: Quality of life (SF-36 physical component, 0-100, high is good, change scores) at >3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Brosseau 2012	40.909	11.038	42	45.149	8.93	36	-4.24 [-8.67, 0.19]	

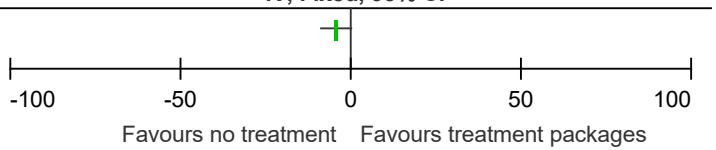


Figure 92: Quality of life (SF-36 mental component, 0-100, high is good, change scores) at >3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Brosseau 2012	53.922	9.023	42	53.101	9.914	36	0.82 [-3.41, 5.06]	

Figure 93: Quality of life (Geri-AIMS pain subscale, AIMS pain, 0-10, high is good, final values) at >3 months

	Treatment packages			No treatment				Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Hughes 2006	4.77	0.82	115	4.61	0.91	100	96.5%	0.16 [-0.07, 0.39]	
Kovar 1992	-4.59	2.4	29	-5.5	2.07	23	3.5%	0.91 [-0.31, 2.13]	
Total (95% CI)			**144**			**123**	**100.0%**	**0.19 [-0.04, 0.42]**	

Heterogeneity: Chi² = 1.41, df = 1 (P = 0.24); I² = 29%

Test for overall effect: Z = 1.60 (P = 0.11)

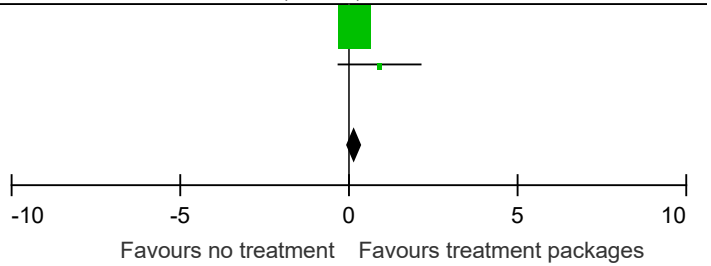


Figure 94: Quality of life (AIMS arthritis impact, 0-10, high is poor, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	3.25	2.6	29	3.8	2.06	23	-0.55 [-1.82, 0.72]

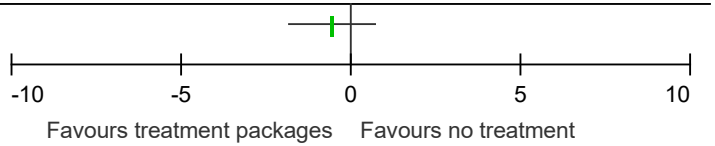


Figure 95: Quality of life (AIMS physical activity, 0-10, high is poor, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	6.07	2.95	29	6.18	2.75	23	-0.11 [-1.66, 1.44]

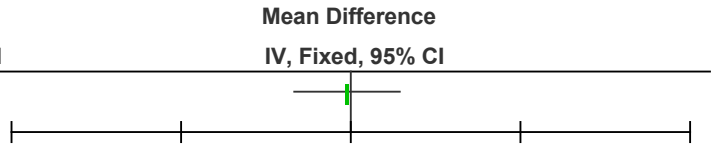


Figure 96: Quality of life (AIMS general health perception, 0-10, high is poor, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	3.71	2.8	29	3.26	1.87	23	0.45 [-0.82, 1.72]

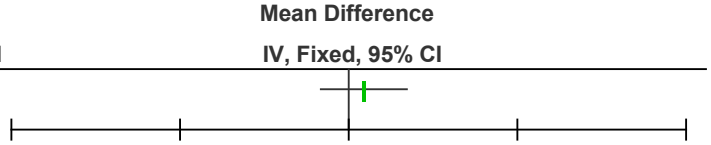


Figure 97: Quality of life (AIMS medications use, 1-6, high is good, final value) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Kovar 1992	3.34	2.16	29	3.6	2.25	23	-0.26 [-1.47, 0.95]

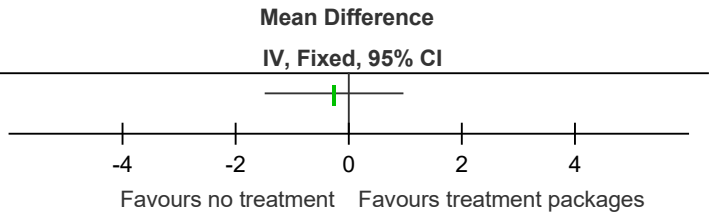


Figure 98: Quality of life (SF-36 physical function, 0-100, high is good, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Nunez 2006	27.2	15.49	43	23.47	18.97	37	3.73 [-3.94, 11.40]

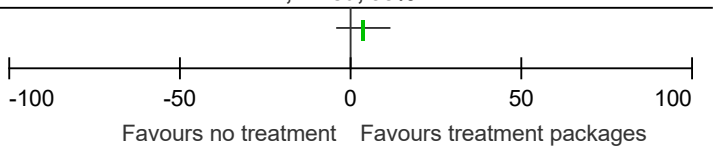


Figure 99: Quality of life (SF-36 bodily pain, 0-100, high is good, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Nunez 2006	38.61	21.93	43	30.33	24.62	37	8.28 [-2.01, 18.57]

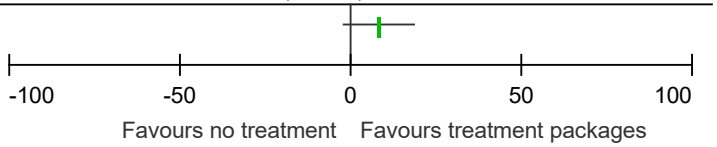


Figure 100: Quality of life (SF-36 role physical, 0-100, high is good, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Nunez 2006	34.76	44.68	43	49.31	42.04	37	-14.55 [-33.57, 4.47]

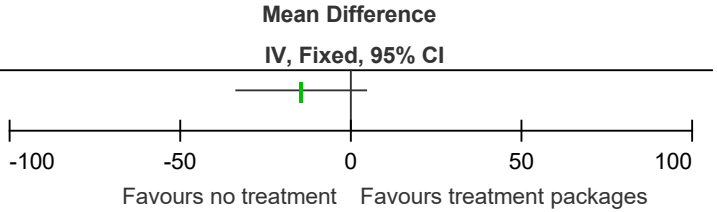


Figure 101: Quality of life (SF-36 vitality, 0-100, high is good, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Nunez 2006	51.34	23.8	43	54.58	25.11	37	-3.24 [-14.01, 7.53]

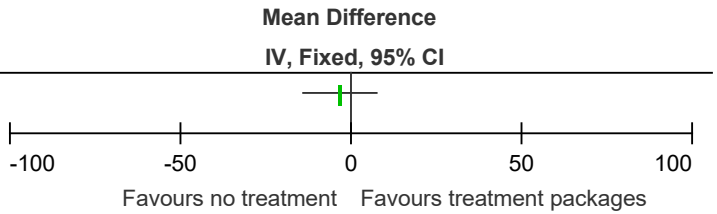


Figure 102: Quality of life (SF-36 general health, 0-100, high is good, final values) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Mean Difference IV, Fixed, 95% CI
Nunez 2006	50.12	22.52	43	56.42	20.88	37	-6.30 [-15.82, 3.22]

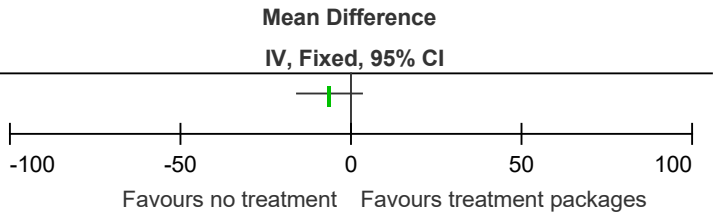


Figure 103: Quality of life (SF-36 mental health, 0-100, high is good, final values) at >3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Nunez 2006	57.27	23.82	43	62.03	47.26	37	-4.76 [-21.57, 12.05]	

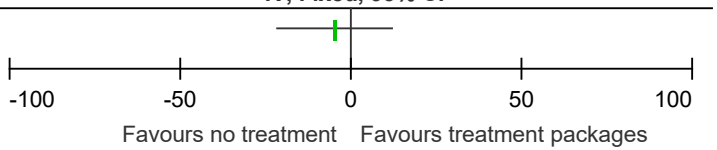


Figure 104: Quality of life (SF-36 role emotional, 0-100, high is good, final values) at >3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Nunez 2006	57.71	47.16	43	63.81	25.94	37	-6.10 [-22.49, 10.29]	

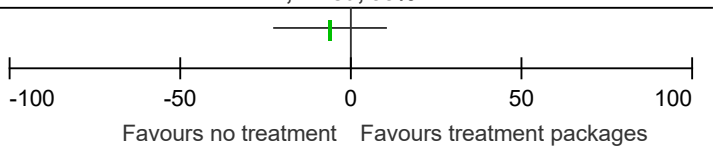


Figure 105: Quality of life (SF-36 social function, 0-100, high is good, final values) at >3 months

	Treatment packages			No treatment			Mean Difference	Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	IV, Fixed, 95% CI	IV, Fixed, 95% CI
Nunez 2006	61.24	30.81	51	62.53	32.15	37	-1.29 [-14.66, 12.08]	

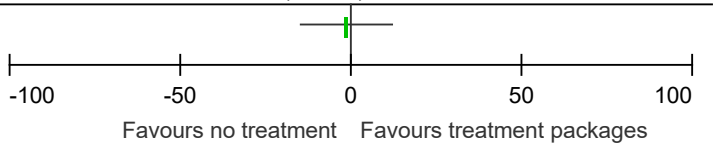


Figure 106: Pain (HOOS, WOMAC, Foot Health Status Questionnaire Pain Domain, VAS [different scale ranges], high is poor, change scores) at ≤3 months

			Treatment packages	No treatment		Std. Mean Difference
Study or Subgroup	Std. Mean Difference	SE	Total	Total	Weight	IV, Random, 95% CI
Allen 2021	-0.242	0.1146	230	115	20.6%	-0.24 [-0.47, -0.02]
Bennell 2017 (IMPACT)	-0.702	0.1749	70	69	19.0%	-0.70 [-1.04, -0.36]
Paterson 2021	0.0838	0.3936	14	12	12.2%	0.08 [-0.69, 0.86]
Poulsen 2013	-1.1418	0.2518	38	36	16.5%	-1.14 [-1.64, -0.65]
Wallis 2017	0.1403	0.2953	23	23	15.1%	0.14 [-0.44, 0.72]
Yip 2007	-1.1418	0.2518	38	36	16.5%	-1.14 [-1.64, -0.65]
Total (95% CI)			**413**	**291**	**100.0%**	**-0.53 [-0.93, -0.13]**

Heterogeneity: Tau² = 0.19; Chi² = 26.20, df = 5 (P < 0.0001); I² = 81%

Test for overall effect: Z = 2.58 (P = 0.010)

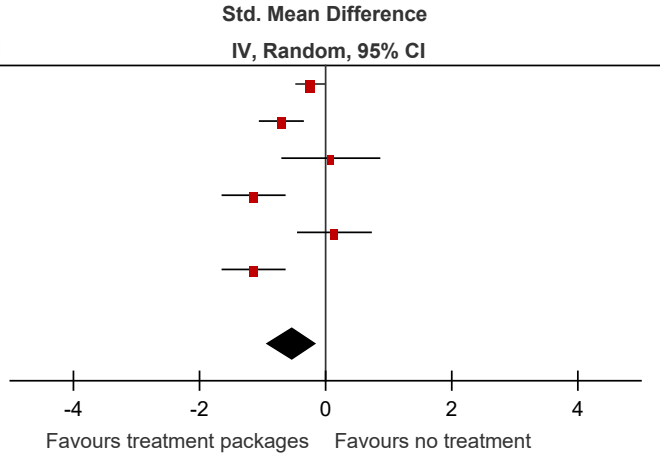


Figure 107: Pain (KOOS, WOMAC, Lequesne index pain subscale, Harris Hip score pain subscale, BPI severity, VAS, Arthritis Self Efficacy Pain subscale [different scale ranges], high is poor, final values) at ≤3 months

Study or Subgroup	Std. Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bearne 2011	-0.0434	0.2887	24	24	6.1%	-0.04 [-0.61, 0.52]
Da silva 2015	-0.876	0.3848	15	15	4.9%	-0.88 [-1.63, -0.12]
Dziedzic 2018	0	0.7071	4	4	2.4%	0.00 [-1.39, 1.39]
Hopman-rock 2000	0.0886	0.1957	56	49	7.3%	0.09 [-0.29, 0.47]
Hughes 2004	-0.3797	0.1966	68	43	7.3%	-0.38 [-0.77, 0.01]
Hughes 2006	-0.4134	0.1382	115	100	8.0%	-0.41 [-0.68, -0.14]
Hurley 2007	-1.0924	0.117	237	128	8.2%	-1.09 [-1.32, -0.86]
Isaramalai 2018	-1.0836	0.2613	50	25	6.4%	-1.08 [-1.60, -0.57]
Jessep 2009	-0.2451	0.2521	29	35	6.6%	-0.25 [-0.74, 0.25]
Kloek 2018	0.1681	0.1519	87	87	7.8%	0.17 [-0.13, 0.47]
Li 2017	-0.0119	0.343	17	17	5.4%	-0.01 [-0.68, 0.66]
Mecklenburg 2018	-0.4704	0.1643	101	61	7.7%	-0.47 [-0.79, -0.15]
Saw 2016	-0.6512	0.2392	35	39	6.7%	-0.65 [-1.12, -0.18]
Tak 2005	0.2653	0.1924	55	54	7.3%	0.27 [-0.11, 0.64]
Yip 2007	-0.3175	0.1493	88	94	7.9%	-0.32 [-0.61, -0.02]
Total (95% CI)			**981**	**775**	**100.0%**	**-0.35 [-0.60, -0.10]**

Heterogeneity: Tau² = 0.18; Chi² = 80.69, df = 14 (P < 0.00001); I² = 83%
Test for overall effect: Z = 2.74 (P = 0.006)

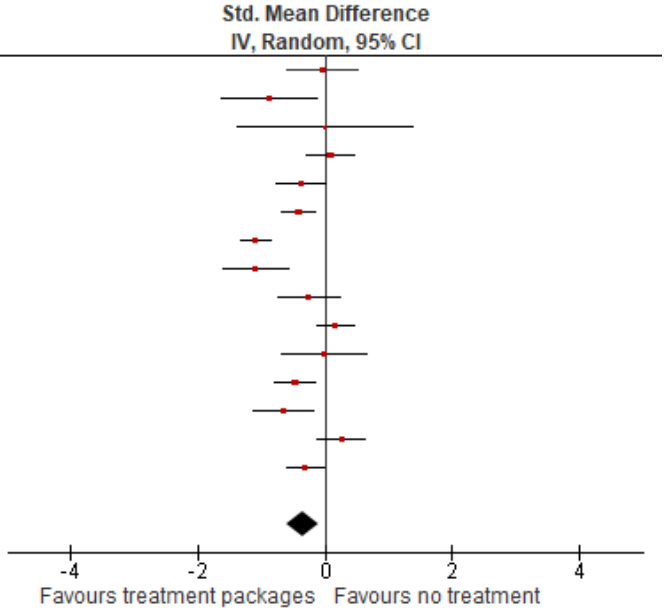


Figure 108: Pain (HOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months

	Treatment packages			No treatment				Std. Mean Difference
Study or Subgroup	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Weight**	**IV, Fixed, 95% CI**
Allen 2021	-1	3.9	230	0.4	3.8	115	39.9%	-0.36 [-0.59, -0.14]
Bennell 2017 (IMPACT)	-3.7	3.3	66	-2.3	3.6	67	17.2%	-0.40 [-0.75, -0.06]
Focht 2005	-2.2	4.1	76	-1.23	4	78	20.2%	-0.24 [-0.56, 0.08]
Poulsen 2013	-16	20	38	-13	18	36	9.7%	-0.16 [-0.61, 0.30]
Skou 2015	-18.7	21.1	50	-9.3	22.9	50	12.9%	-0.42 [-0.82, -0.03]
Total (95% CI)			**460**			**346**	**100.0%**	**-0.33 [-0.47, -0.19]**

Heterogeneity: Chi² = 1.34, df = 4 (P = 0.85); I² = 0%

Test for overall effect: Z = 4.56 (P < 0.00001)

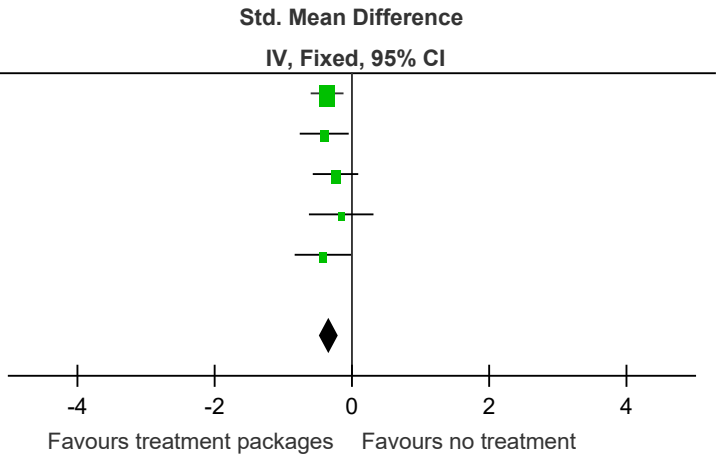


Figure 109: Pain (KOOS, WOMAC, BPI severity, VAS, Arthritis Self Efficacy Pain subscale [different scale ranges], high is poor, final values) at >3 months

Study or Subgroup	Std. Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Bearne 2011	0.1814	0.2893	24	24	3.5%	0.18 [-0.39, 0.75]
Brosseau 2012	0.1472	0.2292	42	35	5.5%	0.15 [-0.30, 0.60]
Dziedzic 2018	0.0789	0.7076	4	4	0.6%	0.08 [-1.31, 1.47]
Hopman-rock 2000	-0.1544	0.1959	56	49	7.6%	-0.15 [-0.54, 0.23]
Hughes 2004	-0.4204	0.2131	60	36	6.4%	-0.42 [-0.84, -0.00]
Hughes 2006	0.0196	0.1367	115	100	15.6%	0.02 [-0.25, 0.29]
Hurley 2007	-0.2062	0.1265	189	94	18.2%	-0.21 [-0.45, 0.04]
Jessep 2009	-0.267	0.2523	29	35	4.6%	-0.27 [-0.76, 0.23]
Kloek 2018	-0.0876	0.1729	65	69	9.7%	-0.09 [-0.43, 0.25]
Kovar 1992	-0.1561	0.2797	29	23	3.7%	-0.16 [-0.70, 0.39]
Nunez 2006	-0.2307	0.225	43	37	5.8%	-0.23 [-0.67, 0.21]
Rezende 2021	-0.3074	0.1456	95	96	13.7%	-0.31 [-0.59, -0.02]
Saw 2016	-0.7303	0.2409	35	39	5.0%	-0.73 [-1.20, -0.26]
Total (95% CI)			**786**	**641**	**100.0%**	**-0.18 [-0.28, -0.07]**

Heterogeneity: Chi² = 13.64, df = 12 (P = 0.32); I² = 12%
Test for overall effect: Z = 3.28 (P = 0.001)

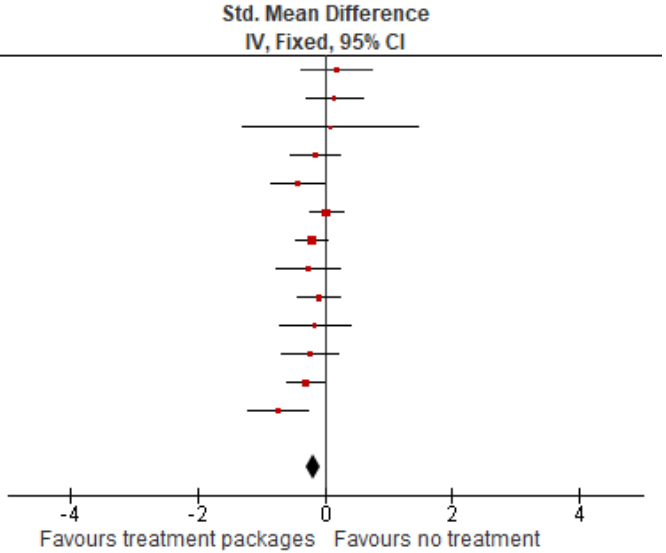


Figure 110: Physical function (HOOS, WOMAC, Foot Health Status Questionnaire Function domain [different scale ranges], high is poor, change scores) at ≤3 months

			Treatment packages	No treatment		Std. Mean Difference
Study or Subgroup	**Std. Mean Difference**	**SE**	**Total**	**Total**	**Weight**	**IV, Random, 95% CI**
Allen 2021	-0.3176	0.1149	230	115	29.4%	-0.32 [-0.54, -0.09]
Bennell 2017 (IMPACT)	-0.855	0.1774	70	69	24.3%	-0.85 [-1.20, -0.51]
Paterson 2021	-0.3348	0.3966	14	12	11.0%	-0.33 [-1.11, 0.44]
Poulsen 2013	-0.6769	0.2395	38	36	19.4%	-0.68 [-1.15, -0.21]
Wallis 2017	0.0776	0.295	23	23	15.8%	0.08 [-0.50, 0.66]
Total (95% CI)			**375**	**255**	**100.0%**	**-0.46 [-0.77, -0.15]**

Heterogeneity: Tau² = 0.07; Chi² = 10.72, df = 4 (P = 0.03); I² = 63%

Test for overall effect: Z = 2.87 (P = 0.004)

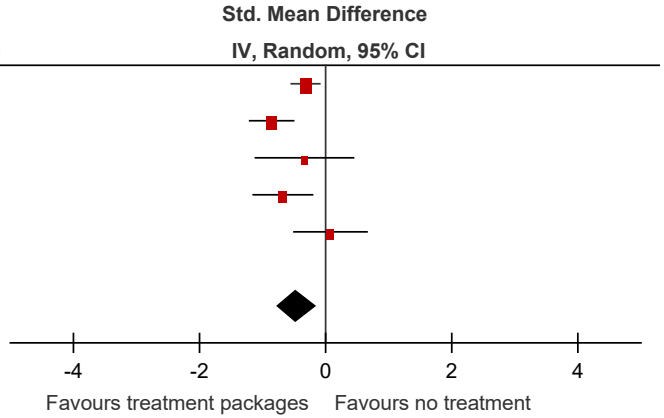


Figure 111: Physical function (KOOS, HOOS, WOMAC, Lequesne index function subscale, Physical Activity Scale for the Elderly [different scale ranges], high is poor, final values) at ≤3 months

Study or Subgroup	Std. Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bearne 2011	-0.2841	0.2903	24	24	8.2%	-0.28 [-0.85, 0.28]
Da silva 2015	-0.5745	0.3737	15	15	6.5%	-0.57 [-1.31, 0.16]
Dziedzic 2018	0.271	0.7135	4	4	2.7%	0.27 [-1.13, 1.67]
Hughes 2004	-0.3917	0.1967	68	43	10.5%	-0.39 [-0.78, -0.01]
Hughes 2006	-0.4171	0.1382	115	100	11.8%	-0.42 [-0.69, -0.15]
Hurley 2007	-0.973	0.1124	237	140	12.4%	-0.97 [-1.19, -0.75]
Isaramalai 2018	-1.1122	0.2621	50	25	8.9%	-1.11 [-1.63, -0.60]
Jessep 2009	0.0179	0.2511	29	35	9.1%	0.02 [-0.47, 0.51]
Kloek 2018	0.0175	0.1516	87	87	11.5%	0.02 [-0.28, 0.31]
Li 2017	-0.2023	0.344	17	17	7.1%	-0.20 [-0.88, 0.47]
Mecklenburg 2018	-0.4761	0.1644	101	61	11.2%	-0.48 [-0.80, -0.15]
Total (95% CI)			**747**	**551**	**100.0%**	**-0.43 [-0.69, -0.17]**

Heterogeneity: Tau² = 0.13; Chi² = 42.26, df = 10 (P < 0.00001); I² = 76%
Test for overall effect: Z = 3.23 (P = 0.001)

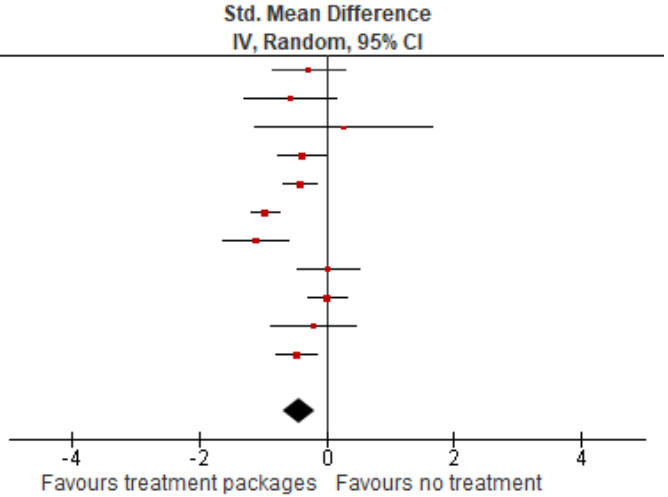


Figure 112: Physical function (HOOS, KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Allen 2021	-3.7	13.2	230	1	12.3	115	50.3%	-0.36 [-0.59, -0.14]
Bennell 2017 (IMPACT)	-13.9	11.4	69	-6.6	11.1	67	21.5%	-0.65 [-0.99, -0.30]
Poulsen 2013	-13	20	38	-9	21	36	12.2%	-0.19 [-0.65, 0.26]
Skou 2015	-18.7	22	50	-5.9	24.2	50	16.0%	-0.55 [-0.95, -0.15]
Total (95% CI)			**387**			**268**	**100.0%**	**-0.43 [-0.59, -0.27]**

Heterogeneity: Chi² = 3.20, df = 3 (P = 0.36); I² = 6%
Test for overall effect: Z = 5.30 (P < 0.00001)

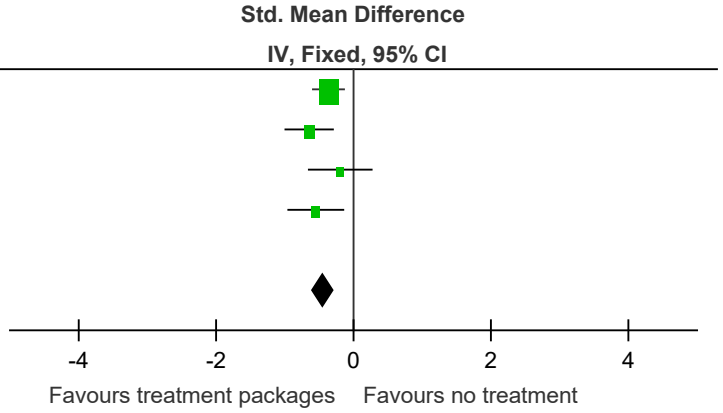


Figure 113: Physical function (KOOS, WOMAC, Physical Activity Scale for the Elderly [different scale ranges], high is poor, final values) at >3 months

Study or Subgroup	Std. Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Bearne 2011	0	0.2887	24	24	4.2%	0.00 [-0.57, 0.57]
Brosseau 2012	0.2739	0.23	42	35	6.6%	0.27 [-0.18, 0.72]
Dziedzic 2018	0.1648	0.7095	4	4	0.7%	0.16 [-1.23, 1.56]
Hughes 2004	-0.43	0.2132	60	36	7.7%	-0.43 [-0.85, -0.01]
Hughes 2006	-0.1804	0.137	115	100	18.6%	-0.18 [-0.45, 0.09]
Hurley 2007	-0.1873	0.1265	189	94	21.8%	-0.19 [-0.44, 0.06]
Jessep 2009	-0.0532	0.2512	29	35	5.5%	-0.05 [-0.55, 0.44]
Kloek 2018	0.051	0.1729	65	69	11.7%	0.05 [-0.29, 0.39]
Nunez 2006	-0.4837	0.2276	43	37	6.7%	-0.48 [-0.93, -0.04]
Rezende 2021	-0.2561	0.1453	95	96	16.5%	-0.26 [-0.54, 0.03]
Total (95% CI)			**666**	**530**	**100.0%**	**-0.16 [-0.28, -0.04]**

Heterogeneity: Chi² = 9.88, df = 9 (P = 0.36); I² = 9%
Test for overall effect: Z = 2.71 (P = 0.007)

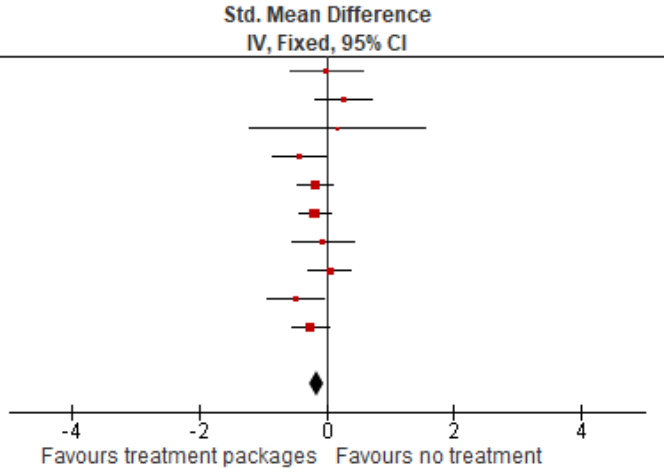


Figure 114: Psychological distress (HADS anxiety, GAD-7 [different scale range], high is poor, final values) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Bearne 2011	4.6	2.6	24	4.1	3	24	40.1%	0.18 [-0.39, 0.74]
Dziedzic 2018	3.16	4.32	4	2.9	4.6	4	6.7%	0.05 [-1.34, 1.44]
Jessep 2009	4.4	3.5	29	4.2	3.3	35	53.2%	0.06 [-0.43, 0.55]
Total (95% CI)			**57**			**63**	**100.0%**	**0.10 [-0.25, 0.46]**

Heterogeneity: Chi² = 0.10, df = 2 (P = 0.95); I² = 0%
Test for overall effect: Z = 0.57 (P = 0.57)

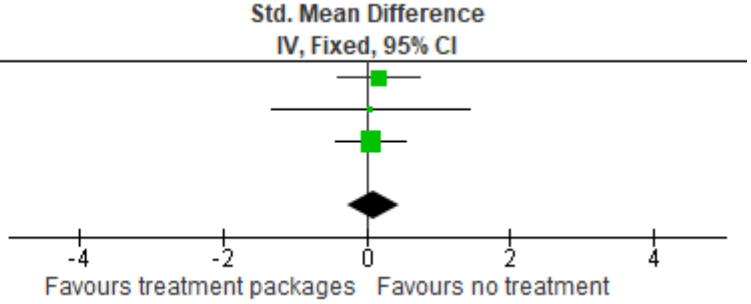


Figure 115: Psychological distress (HADS depression, PHQ-8 [different scale ranges], high is poor, final values) at ≤3 months

Study or Subgroup	Treatment packages Mean	SD	Total	No treatment Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Bearne 2011	2.4	1.8	24	2.9	2.1	24	40.2%	-0.25 [-0.82, 0.32]
Dziedzic 2018	4.36	4.5	4	3.85	4.64	4	6.7%	0.10 [-1.29, 1.48]
Jessep 2009	2.4	1.3	29	3	2.4	35	53.0%	-0.30 [-0.79, 0.20]
Total (95% CI)			**57**			**63**	**100.0%**	**-0.25 [-0.61, 0.11]**

Heterogeneity: Chi² = 0.28, df = 2 (P = 0.87); I² = 0%
Test for overall effect: Z = 1.38 (P = 0.17)

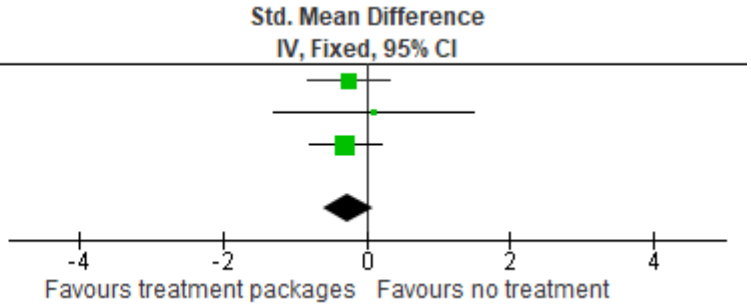


Figure 116: Psychological distress (HADS anxiety, GAD-7 [different scale range], high is poor, final values) at >3 months

Study or Subgroup	Std. Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Bearne 2011	-0.1639	0.2892	24	24	11.4%	-0.16 [-0.73, 0.40]
Dziedzic 2018	0.032	0.7072	4	4	1.9%	0.03 [-1.35, 1.42]
Hurley 2007	0.2325	0.1153	229	113	71.6%	0.23 [0.01, 0.46]
Jessep 2009	0.1135	0.2513	29	35	15.1%	0.11 [-0.38, 0.61]
Total (95% CI)			**286**	**176**	**100.0%**	**0.17 [-0.03, 0.36]**

Heterogeneity: Chi² = 1.71, df = 3 (P = 0.63); I² = 0%
Test for overall effect: Z = 1.70 (P = 0.09)

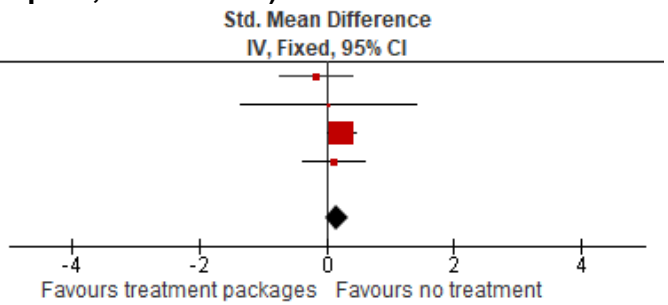


Figure 117: Psychological distress (HADS depression, PHQ-8 [different scale ranges], high is poor, final values) at >3 months

Study or Subgroup	Std. Mean Difference	SE	Treatment packages Total	No treatment Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Bearne 2011	-0.0555	0.2887	24	24	11.4%	-0.06 [-0.62, 0.51]
Dziedzic 2018	0.0285	0.7072	4	4	1.9%	0.03 [-1.36, 1.41]
Hurley 2007	0.1566	0.1151	229	113	71.7%	0.16 [-0.07, 0.38]
Jessep 2009	-0.2307	0.252	29	35	15.0%	-0.23 [-0.72, 0.26]
Total (95% CI)			**286**	**176**	**100.0%**	**0.07 [-0.12, 0.26]**

Heterogeneity: Chi² = 2.18, df = 3 (P = 0.54); I² = 0%
Test for overall effect: Z = 0.74 (P = 0.46)

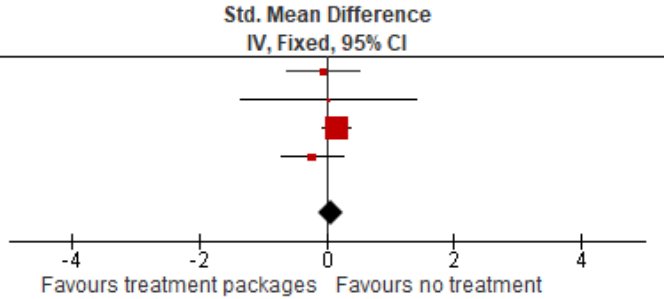


Figure 118: Discontinuation at ≤3 months

Study or Subgroup	Treatment packages Events	Total	No treatment Events	Total	Weight	Risk Difference M-H, Random, 95% CI
Allen 2021	66	230	15	115	6.1%	0.16 [0.07, 0.24]
Arnold 2010	5	28	8	27	3.2%	-0.12 [-0.34, 0.11]
Bearne 2011	2	24	6	24	3.5%	-0.17 [-0.37, 0.04]
Bennell 2017 (IMPACT)	4	74	5	74	6.2%	-0.01 [-0.09, 0.06]
Da silva 2015	4	19	7	22	2.5%	-0.11 [-0.38, 0.16]
Dziedzic 2015 (SMOotH)	2	65	0	65	6.8%	0.03 [-0.02, 0.08]
Hughes 2004	12	80	27	70	4.8%	-0.24 [-0.37, -0.10]
Hughes 2006	32	115	45	100	5.1%	-0.17 [-0.30, -0.04]
Hurley 2007	41	278	12	140	6.5%	0.06 [-0.00, 0.12]
Isaramalai 2018	13	63	20	45	4.0%	-0.24 [-0.41, -0.06]
Jessep 2009	3	29	4	35	4.5%	-0.01 [-0.16, 0.14]
Kao 2012	20	134	34	125	5.8%	-0.12 [-0.22, -0.02]
Kovar 1992	5	52	5	50	5.4%	-0.00 [-0.12, 0.11]
Li 2017	0	17	0	17	5.6%	0.00 [-0.11, 0.11]
Mecklenburg 2018	43	101	25	61	4.4%	0.02 [-0.14, 0.17]
Paterson 2021	1	15	3	15	2.9%	-0.13 [-0.37, 0.11]
Poulsen 2013	4	38	4	36	4.8%	-0.01 [-0.15, 0.14]
Saw 2016	6	35	11	39	3.8%	-0.11 [-0.30, 0.08]
Tak 2005	10	55	5	54	5.1%	0.09 [-0.04, 0.22]
Wallis 2017	8	23	0	23	3.6%	0.35 [0.15, 0.55]
Yip 2007	9	88	24	94	5.5%	-0.15 [-0.26, -0.04]
Total (95% CI)		**1563**		**1231**	**100.0%**	**-0.03 [-0.09, 0.02]**
Total events	290		260			

Heterogeneity: Tau² = 0.01; Chi² = 82.62, df = 20 (P < 0.00001); I² = 76%

Test for overall effect: Z = 1.26 (P = 0.21)

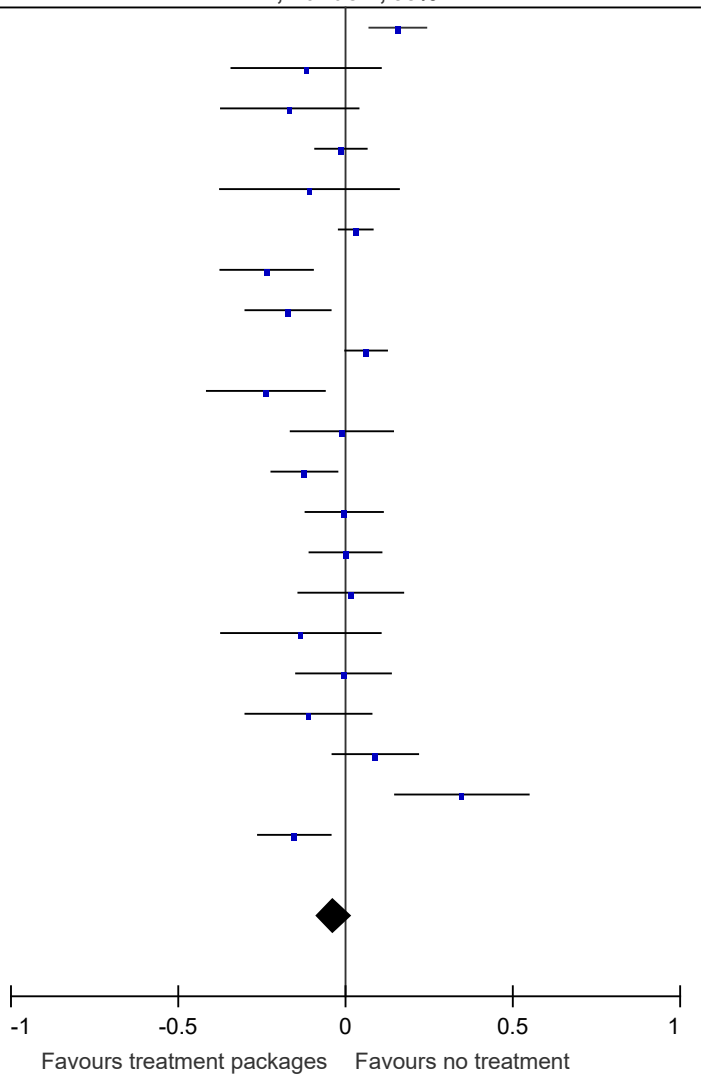
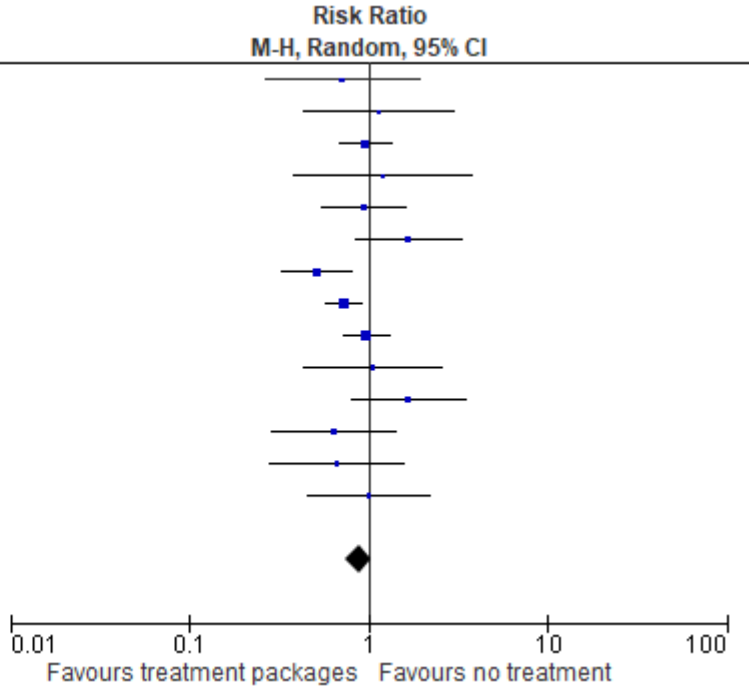


Figure 119: Discontinuation at >3 months

Study or Subgroup	Treatment packages Events	Total	No treatment Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Bearne 2011	5	24	7	24	2.6%	0.71 [0.26, 1.94]
Bennell 2017 (IMPACT)	8	74	7	74	2.8%	1.14 [0.44, 2.99]
Brosseau 2012	34	69	38	74	14.6%	0.96 [0.69, 1.33]
Dziedzic 2015 (SMOotH)	6	65	5	65	2.0%	1.20 [0.39, 3.74]
Dziedzic 2018	21	237	27	288	7.4%	0.95 [0.55, 1.63]
Focht 2005	18	76	11	78	5.1%	1.68 [0.85, 3.32]
Hughes 2004	20	80	34	70	9.7%	0.51 [0.33, 0.81]
Hughes 2006	57	115	68	100	20.4%	0.73 [0.58, 0.92]
Hurley 2007	89	278	46	140	16.4%	0.97 [0.73, 1.31]
Jessep 2009	7	29	8	35	3.2%	1.06 [0.43, 2.56]
Klassbo 2003	17	77	9	68	4.4%	1.67 [0.80, 3.49]
Nunez 2006	8	51	12	49	3.8%	0.64 [0.29, 1.43]
Poulsen 2013	7	38	10	36	3.5%	0.66 [0.28, 1.55]
Saw 2016	9	35	10	39	4.1%	1.00 [0.46, 2.18]
Total (95% CI)		**1248**		**1140**	**100.0%**	**0.88 [0.75, 1.04]**
Total events	306		292			

Heterogeneity: Tau² = 0.02; Chi² = 17.35, df = 13 (P = 0.18); I² = 25%

Test for overall effect: Z = 1.46 (P = 0.14)



Appendix F – GRADE tables

F.1 Treatment packages compared to exercise alone

Table 19: Clinical evidence profile: treatment packages compared to exercise alone

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	exercise alone	Relative (95% CI)	Absolute (95% CI)		
Quality of life (AQOL II, -0.11-1, high is good, change score) at <3 months (follow-up: 12 weeks; Scale from: -0.11 to 1)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	75	-	MD **0** (0.05 lower to 0.05 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Quality of life (AIMS pain, 0-10, high is poor, final value) at <3 months (follow-up: 12 weeks; assessed with: AIMS pain; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	20	16	-	MD **1.07 higher** (0.04 lower to 2.18 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS psychological disability, 0-10, high is poor, final value) at <3 months (follow-up: 12 weeks; assessed with: AIMS psychological disability; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	20	16	-	MD **0.33 higher** (0.35 lower to 1.01 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AQOL II, -0.11-1, high is good, change score) at >3 months (follow-up: mean 14 months; assessed with: AQOL II; Scale from: -0.11 to 1)												
3	randomised trials	serious[a]	not serious	not serious	not serious	none	213	208	-	MD **0** (0.02 lower to 0.02 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Quality of life (KOOS, 0-100, high is good, change score) at >3 months (follow-up: 24 weeks; assessed with: KOOS; Scale from: 0 to 100)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	exercise alone	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	not serious	not serious	not serious	not serious	none	56	54	-	MD **0.1 higher** (7.31 lower to 7.51 higher)	⨁⨁⨁⨁ High	CRITICAL
Quality of life (SF-36 physical component, 0-100, high is good, final values) at >3 months (follow-up: mean 18 months; assessed with: SF-36 physical component; Scale from: 0 to 100)												
2	randomised trials	very serious[a]	serious[c]	not serious	very serious[b]	none	110	113	-	MD **0.76 higher** (3.7 lower to 5.22 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 mental component, 0-100, high is good, final values) at >3 months (follow-up: mean 18 months; assessed with: SF-36 mental component; Scale from: 0 to 100)												
2	randomised trials	very serious[a]	not serious	not serious	not serious	none	110	113	-	MD **0.25 higher** (1.74 lower to 2.25 higher)	⨁⨁◯◯ Low	CRITICAL
Pain (WOMAC, 0-20, high is poor, change score) at <3 months (follow-up: mean 12 weeks; assessed with: WOMAC; Scale from: 0 to 20)												
2	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	94	96	-	MD **1.07 lower** (1.69 lower to 0.45 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (WOMAC, NRS [different scale ranges], high is poor, final value) at <3 months (follow-up: mean 12 weeks; assessed with: WOMAC, NRS)												
3	randomised trials	very serious[a]	very serious[c]	not serious	very serious[b]	none	142	132	-	SMD **0.1 SD lower** (0.71 lower to 0.51 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months (follow-up: mean 57 weeks; assessed with: KOOS, WOMAC)												
5	randomised trials	serious[a]	not serious	not serious	not serious	none	344	342	-	SMD **0.13 SD lower** (0.28 lower to 0.02 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Pain (WOMAC, VAS [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 13 months; assessed with: WOMAC, VAS)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	exercise alone	Relative (95% CI)	Absolute (95% CI)		
3	randomised trials	very serious[a]	not serious	not serious	not serious	none	185	182	-	SMD **0.04 SD higher** (0.17 lower to 0.24 higher)	⨁⨁◯◯ Low	CRITICAL
Physical function (WOMAC, 0-68, high is poor, change scores) at <3 months (follow-up: mean 12 weeks; assessed with: WOMAC; Scale from: 0 to 68)												
2	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	94	96	-	MD **3.8 lower** (5.3 lower to 2.3 lower)	⨁◯◯◯ Very low	CRITICAL
Physical function (WOMAC [different scale ranges], high is poor, final values) at <3 months (follow-up: mean 7 weeks; assessed with: WOMAC)												
2	randomised trials	very serious[a]	very serious[c]	not serious	very serious[b]	none	42	37	-	SMD **0.34 SD lower** (1.24 lower to 0.56 higher)	⨁◯◯◯ Very low	CRITICAL
Physical function (KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months (follow-up: mean 12 months; assessed with: KOOS, WOMAC)												
4	randomised trials	serious[a]	not serious	not serious	not serious	none	268	262	-	SMD **0.09 SD lower** (0.26 lower to 0.08 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Physical function (WOMAC [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 15 months; assessed with: WOMAC)												
2	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	85	87	-	SMD **0.24 SD higher** (0.06 lower to 0.54 higher)	⨁◯◯◯ Very low	CRITICAL
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at <3 months (follow-up: 12 weeks; assessed with: DASS21 Anxiety; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	75	-	MD **0.2 higher** (1.09 lower to 1.49 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at <3 months (follow-up: 12 weeks; assessed with: DASS21 Depression; Scale from: 0 to 42)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	exercise alone	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	75	-	MD **0.2 lower** (1.91 lower to 1.51 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at <3 months (follow-up: 12 weeks; assessed with: DASS21 Stress; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	75	-	MD **1 higher** (1.15 lower to 3.15 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at >3 months (follow-up: mean 12 months; assessed with: DASS21 Anxiety; Scale from: 0 to 42)												
2	randomised trials	serious[a]	not serious	not serious	not serious	none	146	146	-	MD **0.15 lower** (0.54 lower to 0.23 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at >3 months (follow-up: mean 12 months; assessed with: DASS21 Depression; Scale from: 0 to 42)												
2	randomised trials	serious[a]	not serious	not serious	not serious	none	146	146	-	MD **0.15 lower** (0.62 lower to 0.32 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at >3 months (follow-up: mean 12 months; assessed with: DASS21 Stress; Scale from: 0 to 42)												
2	randomised trials	serious[a]	not serious	not serious	not serious	none	146	146	-	MD **0.24 lower** (0.72 lower to 0.24 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Discontinuation at <3 months (follow-up: mean 11 weeks)												
8	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	39/346 (11.3%)	55/360 (15.3%)	**RR 0.75** (0.52 to 1.08)	**38 fewer per 1,000** (from 73 fewer to 12 more)	⨁◯◯◯ Very low	IMPORTANT
Discontinuation at >3 months (follow-up: mean 14 months)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	exercise alone	Relative (95% CI)	Absolute (95% CI)		
10	randomised trials	serious[a]	not serious	not serious	not serious	none	143/734 (19.5%)	146/738 (19.8%)	**RR 1.00** (0.82 to 1.22)	**0 fewer per 1,000** (from 36 fewer to 44 more)	⨁⨁⨁◯ Moderate	IMPORTANT

CI: confidence interval; **MD:** mean difference; **RR:** risk ratio; **SMD:** standardised mean difference

Explanations

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

c. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

F.2 Treatment packages compared to manual therapy alone

Table 20: Clinical evidence profile: treatment packages compared to manual therapy alone

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	manual therapy alone	Relative (95% CI)	Absolute (95% CI)		
Pain (WOMAC, 0-500, high is poor, final value) at ≤3 months (follow up: 5 weeks; assessed with: WOMAC; Scale from: 0 to 500)												
1	randomised trials	serious [a]	not serious	not serious	serious [b]	none	28	27	-	MD **4.6 lower** (51.06 lower to 41.86 higher)	⨁⨁◯◯ LOW	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	manual therapy alone	Relative (95% CI)	Absolute (95% CI)		
Physical function (WOMAC, 0-1800, high is poor, final value) at ≤3 months (follow up: 5 weeks; assessed with: WOMAC; Scale from: 0 to 1800)												
1	randomised trials	serious [a]	not serious	not serious	serious [b]	none	28	27	-	MD **10.8 lower** (157.76 lower to 136.16 higher)	⨁⨁◯◯ LOW	CRITICAL
Discontinuation at ≤3 months (follow up: 5 weeks)												
1	randomised trials	not serious	not serious	not serious	very serious [b]	none	2/28 (7.1%)	1/27 (3.7%)	**RR 1.93** (0.19 to 20.05)	**34 more per 1,000** (from 30 fewer to 706 more)	⨁⨁◯◯ LOW	IMPORTANT

CI: Confidence interval; **MD:** Mean difference; **RR:** Risk ratio

Explanations

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

F.3 Treatment packages compared to electrotherapy alone

Table 21: Clinical evidence profile: treatment packages compared to electrotherapy alone

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	electrotherapy alone	Relative (95% CI)	Absolute (95% CI)		
Pain (VAS, 0-10, high is poor, change score) at ≤3 months (follow up: 12 weeks; assessed with: VAS; Scale from: 0 to 10)												
1	randomised trials	very serious [a]	not serious	not serious	not serious	none	42	42	-	MD **2.1 lower** (2.89 lower to 1.31 lower)	⨁⨁◯◯ LOW	CRITICAL

CI: Confidence interval; **MD:** Mean difference

Explanations

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

F.4 Treatment packages compared to behaviour change interventions alone

Table 22: Clinical evidence profile: treatment packages compared to behaviour change interventions alone

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	behaviour change interventions alone	Relative (95% CI)	Absolute (95% CI)		
Quality of life (AQOL II, -0.04-1, high is good, change score) at <3 months (follow-up: 12 weeks; assessed with: AQOL II; Scale from: -0.04 to 1)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	behaviour change interventions alone	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0** (0.03 lower to 0.03 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Quality of life (AIMS pain, 0-10, high is poor, final value) at <3 months (follow-up: 12 weeks; assessed with: AIMS pain; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	20	18	-	MD **0.26 higher** (0.7 lower to 1.22 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS psychological distress, 0-10, high is poor, final value) at <3 months (follow-up: 12 weeks; assessed with: AIMS psychological distress; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	20	18	-	MD **0.17 lower** (1 lower to 0.66 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AQOL II, -0.04-1, high is good, change score) at >3 months (follow-up: 52 weeks; assessed with: AQOL II; Scale from: -0.04 to 1)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0** (0.03 lower to 0.03 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Quality of life (SF-36 physical composite, 0-100, high is good, final value) at >3 months (follow-up: 18 months; assessed with: SF-36 physical composite; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	68	73	-	MD **2.16 higher** (0.16 lower to 4.48 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 mental composite, 0-100, high is good, final value) at >3 months (follow-up: 18 months; assessed with: SF-36 mental composite; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	not serious	none	68	73	-	MD **0.55 lower** (2.77 lower to 1.67 higher)	⨁⨁◯◯ Low	CRITICAL
Pain (WOMAC, 0-20, high is poor, change scores) at <3 months (follow-up: mean 12 weeks; assessed with: WOMAC; Scale from: 0 to 20)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	behaviour change interventions alone	Relative (95% CI)	Absolute (95% CI)		
2	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	94	95	-	MD **1.22 lower** (2.18 lower to 0.27 lower)	⨁◯◯◯ Very low	CRITICAL
Pain (WOMAC, 0-500, high is poor, final value) at <3 months (follow-up: 12 weeks; assessed with: WOMAC; Scale from: 0 to 500)												
1	randomised trials	very serious[a]	not serious	not serious	not serious	none	100	98	-	MD **4.9 lower** (23.72 lower to 13.92 higher)	⨁⨁◯◯ Low	CRITICAL
Pain (WOMAC, 0-20, high is poor, change scores) at >3 months (follow-up: mean 15 months; assessed with: WOMAC; Scale from: 0 to 20)												
2	randomised trials	serious[a]	not serious	not serious	serious[b]	none	149	156	-	MD **1.17 lower** (2 lower to 0.34 lower)	⨁⨁◯◯ Low	CRITICAL
Pain (WOMAC, 0-500, high is poor, final value) at >3 months (follow-up: 9 months; assessed with: WOMAC; Scale from: 0 to 500)												
1	randomised trials	very serious[a]	not serious	not serious	not serious	none	100	98	-	MD **6.7 lower** (28.49 lower to 15.09 higher)	⨁⨁◯◯ Low	CRITICAL
Physical function (WOMAC, 0-68, high is poor, change scores) at <3 months (follow-up: mean 12 weeks; assessed with: WOMAC; Scale from: 0 to 68)												
2	randomised trials	very serious[a]	very serious[c]	not serious	serious[b]	none	94	95	-	MD **5.65 lower** (11.36 lower to 0.07 higher)	⨁◯◯◯ Very low	CRITICAL
Physical function (WOMAC, 0-68, high is poor, change score) at >3 months (follow-up: 12 months; assessed with: WOMAC; Scale from: 0 to 68)												
1	randomised trials	serious[a]	not serious	not serious	serious[b]	none	73	74	-	MD **6.8 lower** (10.16 lower to 3.44 lower)	⨁⨁◯◯ Low	CRITICAL
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at <3 months (follow-up: 12 weeks; assessed with: DASS21 Anxiety; Scale from: 0 to 42)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	behaviour change interventions alone	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious[a]	not serious	not serious	serious[b]	none	73	74	-	MD **1 higher** (0.33 lower to 2.33 higher)	⨁⨁◯◯ Low	IMPORTANT
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at <3 months (follow-up: 12 weeks; assessed with: DASS21 Depression; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0.3 lower** (2.11 lower to 1.51 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at <3 months (follow-up: 12 weeks; assessed with: DASS21 Stress; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0.2 lower** (2.09 lower to 1.69 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Anxiety, 0-42, high is poor, change score) at >3 months (follow-up: 12 months; assessed with: DASS21 Anxiety; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0.1 higher** (1.35 lower to 1.55 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Depression, 0-42, high is poor, change score) at >3 months (follow-up: 12 months; assessed with: DASS21 Depression; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0.5 lower** (2.18 lower to 1.18 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (DASS21 Stress, 0-42, high is poor, change score) at >3 months (follow-up: 12 months; Scale from: 0 to 42)												
1	randomised trials	serious[a]	not serious	not serious	not serious	none	73	74	-	MD **0.4 lower** (2.5 lower to 1.7 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Discontinuation at <3 months (follow-up: mean 12 weeks)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	behaviour change interventions alone	Relative (95% CI)	Absolute (95% CI)		
3	randomised trials	not serious	not serious	not serious	very serious[b]	none	8/160 (5.0%)	12/158 (7.6%)	**RR 0.66** (0.28 to 1.58)	**26 fewer per 1,000** (from 55 fewer to 44 more)	⨁⨁◯◯ Low	IMPORTANT
Discontinuation at >3 months (follow-up: mean 15 months)												
5	randomised trials	serious[a]	not serious	not serious	serious[b]	none	77/409 (18.8%)	66/403 (16.4%)	**RR 1.15** (0.86 to 1.55)	**25 more per 1,000** (from 23 fewer to 90 more)	⨁⨁◯◯ Low	IMPORTANT

CI: confidence interval; **MD:** mean difference; **RR:** risk ratio

Explanations

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

c. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

F.5 Treatment packages compared to education programmes alone

Table 23: Clinical evidence profile: treatment packages compared to education programmes alone

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	education programmes alone	Relative (95% CI)	Absolute (95% CI)		
Quality of life (EQ-5D 5L, -0.11-1, high is good, final value) at <3 months (follow-up: 12 weeks; assessed with: EQ-5D 5L; Scale from: -0.11 to 1)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	education programmes alone	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	serious[a]	not serious	not serious	very serious[b]	none	84	83	-	MD **0.02 higher** (0.05 lower to 0.09 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (HOOS, KOOS, 0-100, high is good, change scores and final value) at <3 months (follow-up: mean 10 weeks; assessed with: HOOS, KOOS; Scale from: 0 to 100)												
3	randomised trials	serious[a]	not serious	not serious	serious[b]	none	87	86	-	MD **9.8 higher** (4.99 higher to 14.6 higher)	⨁⨁◯◯ Low	CRITICAL
Quality of life (AIMS-2 pain subscale, 0-10, high is good, final value) at <3 months (follow-up: 12 weeks; assessed with: AIMS-2 pain subscale; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	20	18	-	MD **0.81 lower** (2.25 lower to 0.63 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (HOOS, KOOS, 0-100, high is good, change score and final value) at >3 months (follow-up: mean 11 months; assessed with: HOOS, KOOS; Scale from: 0 to 100)												
2	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	73	71	-	MD **2.52 higher** (4.04 lower to 9.08 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 physical function, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 physical function; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	40	35	-	MD **4.2 higher** (5.17 lower to 13.57 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 bodily pain, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 bodily pain; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	41	37	-	MD **9.1 higher** (0.58 lower to 18.78 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 role physical, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 role physical; Scale from: 0 to 100)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	education programmes alone	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	41	37	-	MD **6.6 higher** (5.58 lower to 18.78 higher)	⨁◯◯◯ Very low	CRITICAL

Quality of life (SF-36 vitality, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 vitality; Scale from: 0 to 100)

1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	41	37	-	MD **2.7 lower** (11.94 lower to 6.54 higher)	⨁◯◯◯ Very low	CRITICAL

Quality of life (SF-36 general health, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 general health; Scale from: 0 to 100)

1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	38	36	-	MD **3.7 higher** (6.07 lower to 13.47 higher)	⨁◯◯◯ Very low	CRITICAL

Quality of life (SF-36 mental health, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 mental health; Scale from: 0 to 100)

1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	40	37	-	MD **1 lower** (7.78 lower to 5.78 higher)	⨁◯◯◯ Very low	CRITICAL

Quality of life (SF-36 role emotional, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 role emotional; Scale from: 0 to 100)

1	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	41	37	-	MD **0.2 higher** (8.25 lower to 8.65 higher)	⨁◯◯◯ Very low	CRITICAL

Quality of life (SF-36 social function, 0-100, high is good, final value) at >3 months (follow-up: 16 months; assessed with: SF-36 social function; Scale from: 0 to 100)

1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	41	37	-	MD **7.1 higher** (2.84 lower to 17.04 higher)	⨁◯◯◯ Very low	CRITICAL

Pain (HOOS, KOOS, WOMAC, VAS, 0-100, high is good, change scores and final value) at <3 months (follow-up: mean 10 weeks; assessed with: HOOS, KOOS, WOMAC, VAS; Scale from: 0 to 100)

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	education programmes alone	Relative (95% CI)	Absolute (95% CI)		
6	randomised trials	serious[a]	serious[c]	not serious	serious[b]	none	220	220	-	MD **11.31 higher** (5.87 higher to 16.74 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (KOOS, AUSCAN, McGill Pain Questionnaire, pain rating index [different scale ranges], high is poor, final values) at <3 months (follow-up: mean 12 weeks; assessed with: KOOS, AUSCAN, McGill Pain Questionnaire, pain rating index)												
4	randomised trials	serious[a]	not serious	not serious	not serious	none	148	143	-	SMD **0.15 SD higher** (0.08 lower to 0.38 higher)	⨁⨁⨁◯ Moderate	CRITICAL
Pain (HOOS, KOOS, WOMAC, 0-100, high is poor, change score and final values) at >3 months (follow-up: mean 12 months; assessed with: HOOS, KOOS, WOMAC; Scale from: 0 to 100)												
3	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	115	107	-	MD **3.81 lower** (8.41 lower to 0.79 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (WOMAC, McGill Pain Questionnaire, pain rating index [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 21 weeks; assessed with: WOMAC, McGill Pain Questionnaire, pain rating index)												
4	randomised trials	very serious[a]	not serious	not serious	very serious[b]	none	78	64	-	SMD **0.09 SD higher** (0.66 lower to 0.83 higher)	⨁◯◯◯ Very low	CRITICAL
Physical function (HOOS, KOOS, WOMAC, 0-100, high is good, change scores and final value) at <3 months (follow-up: mean 9 weeks; assessed with: HOOS, KOOS, WOMAC; Scale from: 0 to 100)												
4	randomised trials	serious[a]	not serious	not serious	serious[b]	none	124	122	-	MD **11.08 higher** (7.66 higher to 14.5 higher)	⨁⨁◯◯ Low	CRITICAL
Physical function (AUSCAN, Functional Index of Hand Osteoarthritis [different scale ranges], high is good, final values) at <3 months (follow-up: mean 12 weeks; assessed with: AUSCAN, Functional Index of Hand Osteoarthritis)												
2	randomised trials	not serious	very serious[c]	not serious	serious[b]	none	181	182	-	SMD **0.21 SD higher** (0.23 lower to 0.65 higher)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	education programmes alone	Relative (95% CI)	Absolute (95% CI)		
Physical function (HOOS, WOMAC, 0-100, high is poor, change score and final values) at >3 months (follow-up: mean 12 months; assessed with: HOOS, WOMAC; Scale from: 0 to 100)												
3	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	112	107	-	MD **5.59 lower** (10.18 lower to 1 lower)	⨁◯◯◯ Very low	CRITICAL
Physical function (WOMAC, 0-68, high is poor, final value) at >3 months (follow-up: 5 months; assessed with: WOMAC; Scale from: 0 to 68)												
1	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	21	11	-	MD **14.67 lower** (24.21 lower to 5.13 lower)	⨁◯◯◯ Very low	CRITICAL
Discontinuation at <3 months (follow-up: mean 10 weeks)												
7	randomised trials	not serious	serious[d]	not serious	very serious[e]	none	61/377 (16.2%)	52/361 (14.4%)	**RD 0.01** (-0.04 to 0.06)	**10 fewer per 1,000** (from 40 fewer to 60 more)[f]	⨁◯◯◯ Very low	IMPORTANT
Discontinuation at >3 months (follow-up: mean 9 months)												
6	randomised trials	very serious[a]	not serious	not serious	serious[b]	none	57/227 (25.1%)	65/192 (33.9%)	**RR 0.68** (0.51 to 0.92)	**108 fewer per 1,000** (from 166 fewer to 27 fewer)	⨁◯◯◯ Very low	IMPORTANT

CI: confidence interval; **MD:** mean difference; **RR:** risk ratio; **SMD:** standardised mean difference

Explanations

a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

c. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

d. Downgraded for heterogeneity due to conflicting number of events in different studies (zero events in one or more studies)

e. Downgraded by 1 to 2 increments for imprecision due to zero events and small sample size

f. Absolute effect calculated by risk difference due to zero events in at least one arm of one study

F.6 Treatment packages compared to standard care (non-organised) or no treatment

Table 24: Clinical evidence profile: treatment packages compared to standard care (non-organised) or no treatment

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
Quality of life (EQ-5D, AQoL-2, -0.11-1, high is good, change scores and final values) at <3 months (follow-up: mean 11 weeks; assessed with: EQ-5D, AQoL-2; Scale from: -0.11 to 1)												
6	randomised trials	very serious[a]	serious[b]	not serious	serious[c]	none	323	266	-	MD **0.08 higher** (0.02 higher to 0.13 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (KOOS, HOOS, VAS quality of life, health assessment questionnaire, 0-100, high is good, change score and final values) at <3 months (follow-up: mean 7 weeks; assessed with: KOOS, HOOS, VAS quality of life, health assessment questionnaire; Scale from: 0 to 100)												
5	randomised trials	very serious[a]	serious[b]	not serious	not serious	none	286	283	-	MD **2.56 higher** (1.86 lower to 6.97 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (Health related quality of life, 7-39, high is good, final value) at <3 months (follow-up: 12 weeks; assessed with: Health related quality of life; Scale from: 7 to 39)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	55	54	-	MD **1.3 higher** (0.11 higher to 2.49 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 physical component, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 physical component; Scale from: 0 to 100)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	134	125	-	MD **0.95 higher** (1.16 lower to 3.06 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 mental component, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 mental component; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	134	125	-	MD **2.56 higher** (0.78 higher to 4.34 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (Geri-AIMS pain subscale, AIMS pain, 0-10, high is good, final values) at <3 months (follow-up: mean 9 weeks; assessed with: Geri-AIMS pain subscale, AIMS pain; Scale from: 0 to 10)												
3	randomised trials	very serious[a]	serious[b]	not serious	serious[c]	none	182	163	-	MD **0.36 higher** (0.3 lower to 1.01 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS psychological disability, 0-10, high is poor, final value) at <3 months (follow-up: 12 weeks; assessed with: AIMS psychological disability; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	20	18	-	MD **0.41 higher** (0.31 lower to 1.13 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS arthritis impact, 0-10, high is poor, final value) at <3 months (follow-up: 8 weeks; assessed with: AIMS arthritis impact; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	47	45	-	MD **0.2 lower** (0.97 lower to 0.57 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS physical activity, 0-10, high is poor, final value) at <3 months (follow-up: 8 weeks; assessed with: AIMS physical activity; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	47	45	-	MD **2.22 lower** (3.25 lower to 1.19 lower)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
Quality of life (AIMS medications use, 0-6, high is good, final value) at <3 months (follow-up: 8 weeks; assessed with: AIMS medications use; Scale from: 0 to 6)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	47	45	-	MD **0.74 higher** (0.07 lower to 1.55 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 physical function, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 physical function; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	15	15	-	MD **14 higher** (1.76 higher to 26.24 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 bodily pain, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 bodily pain; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	15	15	-	MD **14.8 higher** (2.21 higher to 27.39 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 role physical, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 role physical; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	not serious	none	15	15	-	MD **53.33 higher** (30.56 higher to 76.1 higher)	⨁⨁◯◯ Low	CRITICAL
Quality of life (SF-36 vitality, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 vitality; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	15	15	-	MD **13.67 higher** (2.3 higher to 25.04 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 general health, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 general health; Scale from: 0 to 100)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	15	15	-	MD **13.73 higher** (0.68 higher to 26.78 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 mental health, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 mental health; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	15	15	-	MD **14.13 higher** (0.09 lower to 28.35 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 role emotional, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 role emotional; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	not serious	none	15	15	-	MD **33.47 higher** (10.78 higher to 56.16 higher)	⨁⨁◯◯ Low	CRITICAL
Quality of life (SF-36 social function, 0-100, high is good, change scores) at <3 months (follow-up: 8 weeks; assessed with: SF-36 social function; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	15	15	-	MD **0.84 higher** (8.46 lower to 10.14 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (EQ-5D, AQoL-2, -0.11-1, high is good, change scores and final values) at >3 months (follow-up: mean 9 months; assessed with: EQ-5D, AQoL-2; Scale from: -0.11 to 1)												
7	randomised trials	serious[a]	serious[b]	not serious	serious[c]	none	576	425	-	MD **0.05 higher** (0.01 higher to 0.1 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (KOOS, HOOS, VAS quality of life, 0-100, high is good, change score and final values) at >3 months (follow-up: mean 10 months; assessed with: KOOS, HOOS, VAS quality of life; Scale from: 0 to 100)												
3	randomised trials	very serious[a]	not serious	not serious	not serious	none	159	154	-	MD **1.67 lower** (6.81 lower to 3.46 higher)	⨁⨁◯◯ Low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
Quality of life (SF-36 physical component, 0-100, high is good, change scores) at >3 months (follow-up: 18 months; assessed with: SF-36 physical component; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	42	36	-	MD **4.24 lower** (8.67 lower to 0.19 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 mental component, 0-100, high is good, change scores) at >3 months (follow-up: 18 months; assessed with: SF-36 mental component; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	42	36	-	MD **0.82 higher** (3.41 lower to 5.06 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (Geri-AIMS pain subscale, AIMS pain, 0-10, high is good, final values) at >3 months (follow-up: mean 12 months; assessed with: Geri-AIMS pain subscale, AIMS pain; Scale from: 0 to 10)												
2	randomised trials	very serious[a]	not serious	not serious	not serious	none	144	123	-	MD **0.19 higher** (0.04 lower to 0.42 higher)	⨁⨁◯◯ Low	CRITICAL
Quality of life (AIMS arthritis impact, 0-10, high is poor, final value) at >3 months (follow-up: 12 months; assessed with: AIMS arthritis impact; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	29	23	-	MD **0.55 lower** (1.82 lower to 0.72 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS physical activity, 0-10, high is poor, final value) at >3 months (follow-up: 12 months; assessed with: AIMS physical activity; Scale from: 0 to 10)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	29	23	-	MD **0.11 lower** (1.66 lower to 1.44 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS general health perception, 0-10, high is poor, final value) at >3 months (follow-up: 12 months; assessed with: AIMS general health perception; Scale from: 0 to 10)												

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	29	23	-	MD **0.45 higher** (0.82 lower to 1.72 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (AIMS medications use, 1-6, high is good, final value) at >3 months (follow-up: 12 months; assessed with: AIMS medications use; Scale from: 1 to 6)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	29	23	-	MD **0.26 lower** (1.47 lower to 0.95 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 physical function, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 physical function; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	43	37	-	MD **3.73 higher** (3.94 lower to 11.4 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 bodily pain, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 bodily pain; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	43	37	-	MD **8.28 higher** (2.01 lower to 18.57 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 role physical, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 role physical; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	43	37	-	MD **14.55 lower** (33.57 lower to 4.47 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 vitality, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 vitality; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	43	37	-	MD **3.24 lower** (14.01 lower to 7.53 higher)	⨁◯◯◯ Very low	CRITICAL

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
Quality of life (SF-36 general health, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 general health; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	43	37	-	MD **6.3 lower** (15.82 lower to 3.22 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 mental health, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 mental health; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	43	37	-	MD **6.54 lower** (17.52 lower to 4.44 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 role emotional, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 role emotional; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	43	37	-	MD **4.32 lower** (25.07 lower to 16.43 higher)	⨁◯◯◯ Very low	CRITICAL
Quality of life (SF-36 social function, 0-100, high is good, final values) at >3 months (follow-up: 9 months; assessed with: SF-36 social function; Scale from: 0 to 100)												
1	randomised trials	very serious[a]	not serious	not serious	very serious[c]	none	51	37	-	MD **1.29 lower** (14.66 lower to 12.08 higher)	⨁◯◯◯ Very low	CRITICAL
Pain (HOOS, WOMAC, Foot Health Status Questionnaire, VAS [different scale ranges], high is poor, change scores) at <3 months (follow-up: mean 10 weeks; assessed with: HOOS, WOMAC, VAS)												
6	randomised trials	very serious[a]	very serious[b]	not serious	serious[c]	none	413	291	-	SMD **0.53 SD lower** (0.93 lower to 0.13 lower)	⨁◯◯◯ Very low	CRITICAL

Pain (KOOS, WOMAC, Lequesne index pain subscale, Harris Hip score pain subscale, BPI severity, VAS, Arthritis self-efficacy pain subscale [different scale ranges], high is poor, final values) at <3 months (follow-up: mean 8 weeks; assessed with: KOOS, WOMAC, Lequesne index pain subscale, Harris Hip score pain subscale, BPI severity, VAS, Arthritis self-efficacy pain subscale)

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
15	randomised trials	very serious[a]	very serious[b]	not serious	serious[c]	none	981	775	-	SMD **0.35 SD lower** (0.6 lower to 0.1 lower)	⨁◯◯◯ Very low	CRITICAL

Pain (HOOS, KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months (follow-up: mean 12 months; assessed with: HOOS, KOOS, WOMAC)

5	randomised trials	serious[a]	not serious	not serious	not serious	none	460	346	-	SMD **0.33 SD lower** (0.47 lower to 0.19 lower)	⨁⨁⨁◯ Moderate	CRITICAL

Pain (KOOS, WOMAC, BPI severity, VAS, Arthritis self-efficacy pain subscale [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 12 months; assessed with: KOOS, WOMAC, BPI severity, VAS, Arthritis self-efficacy pain subscale)

13	randomised trials	very serious[a]	not serious	not serious	not serious	none	782	637	-	SMD **0.18 SD lower** (0.28 lower to 0.07 lower)	⨁⨁◯◯ Low	CRITICAL

Physical function (HOOS, WOMAC, Foot Health Status Questionnaire Function Domain [different scale ranges], high is poor, change scores) at <3 months (follow-up: mean 11 weeks; assessed with: HOOS, WOMAC, Foot Health Status Questionnaire Function Domain)

5	randomised trials	very serious[a]	serious[b]	not serious	serious[c]	none	375	255	-	SMD **0.46 SD lower** (0.77 lower to 0.15 lower)	⨁◯◯◯ Very low	CRITICAL

Physical function (KOOS, HOOS, WOMAC, Lequesne index function subscale, Physical Activity Scale for the Elderly [different scale ranges], high is poor, final values) at <3 months (follow-up: mean 8 weeks; assessed with: KOOS, HOOS, WOMAC, Lequesne index function subscale, Physical Activity Scale for the Elderly)

11	randomised trials	very serious[a]	very serious[b]	not serious	serious	none	743	547	-	SMD **0.43 SD lower** (0.69 lower to 0.17 lower)	⨁◯◯◯ Very low	CRITICAL

Physical function (HOOS, KOOS, WOMAC [different scale ranges], high is poor, change scores) at >3 months (follow-up: mean 11 months; assessed with: HOOS, KOOS, WOMAC)

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
4	randomised trials	very serious[a]	not serious	not serious	serious[c]	none	387	268	-	SMD **0.43 SD lower** (0.59 lower to 0.27 lower)	⨁◯◯◯ Very low	CRITICAL
Physical function (KOOS, WOMAC, Physical Activity Scale for the Elderly [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 14 months; assessed with: KOOS, WOMAC, Physical Activity Scale for the Elderly)												
10	randomised trials	very serious[a]	not serious	not serious	not serious	none	666	530	-	SMD **0.16 SD lower** (0.28 lower to 0.04 lower)	⨁⨁◯◯ Low	CRITICAL
Psychological distress (HADS anxiety, GAD-7 [different scale ranges], high is poor, final values) at <3 months (follow-up: mean 8 weeks; assessed with: HADS anxiety, GAD-7; Scale from: 0 to 21)												
3	randomised trials	serious[a]	not serious	not serious	not serious	none	57	63	-	SMD **0.1 SD higher** (0.25 lower to 0.46 higher)	⨁⨁⨁◯ Moderate	IMPORTANT
Psychological distress (HADS depression, PHQ-8 [different scale ranges], high is poor, final values) at <3 months (follow-up: mean 8 weeks; assessed with: HADS depression; Scale from: 0 to 21)												
3	randomised trials	serious[a]	not serious	not serious	serious[c]	none	57	63	-	SMD **0.25 SD lower** (0.61 lower to 0.11 higher)	⨁⨁◯◯ Low	IMPORTANT
Psychological distress (HADS anxiety, GAD-7 [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 9 months; assessed with: HADS anxiety, GAD-7)												
4	randomised trials	very serious[a]	not serious	not serious	not serious	none	286	176	-	SMD **0.17 SD higher** (0.03 lower to 0.36 higher)	⨁⨁◯◯ Low	IMPORTANT
Psychological distress (HADS depression, PHQ-8 [different scale ranges], high is poor, final values) at >3 months (follow-up: mean 9 months; assessed with: HADS depression, PHQ-8)												
4	randomised trials	very serious[a]	not serious	not serious	not serious	none	286	176	-	SMD **0.07 SD higher** (0.12 lower to 0.26 higher)	⨁⨁◯◯ Low	IMPORTANT

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	treatment packages	standard care (non-organised) or no treatment	Relative (95% CI)	Absolute (95% CI)		
Discontinuation at <3 months (follow-up: mean 10 weeks)												
21	randomised trials	serious[a]	very serious[b]	not serious	very serious[c]	none	290/1563 (18.6%)	260/1231 (21.1%)	**RD -0.03** (-0.09 to 0.02)	**30 fewer per 1,000** (from 90 fewer to 20 more)[d]	⨁◯◯◯ Very low	IMPORTANT
Discontinuation at >3 months (follow-up: mean 13 months)												
16	randomised trials	serious[a]	serious[b]	not serious	serious[c]	none	390/1589 (24.5%)	327/1366 (23.9%)	**RR 0.96** (0.80 to 1.15)	**10 fewer per 1,000** (from 48 fewer to 36 more)	⨁◯◯◯ Very low	IMPORTANT

CI: confidence interval; **MD:** mean difference; **RR:** risk ratio; **SMD:** standardised mean difference

Explanations

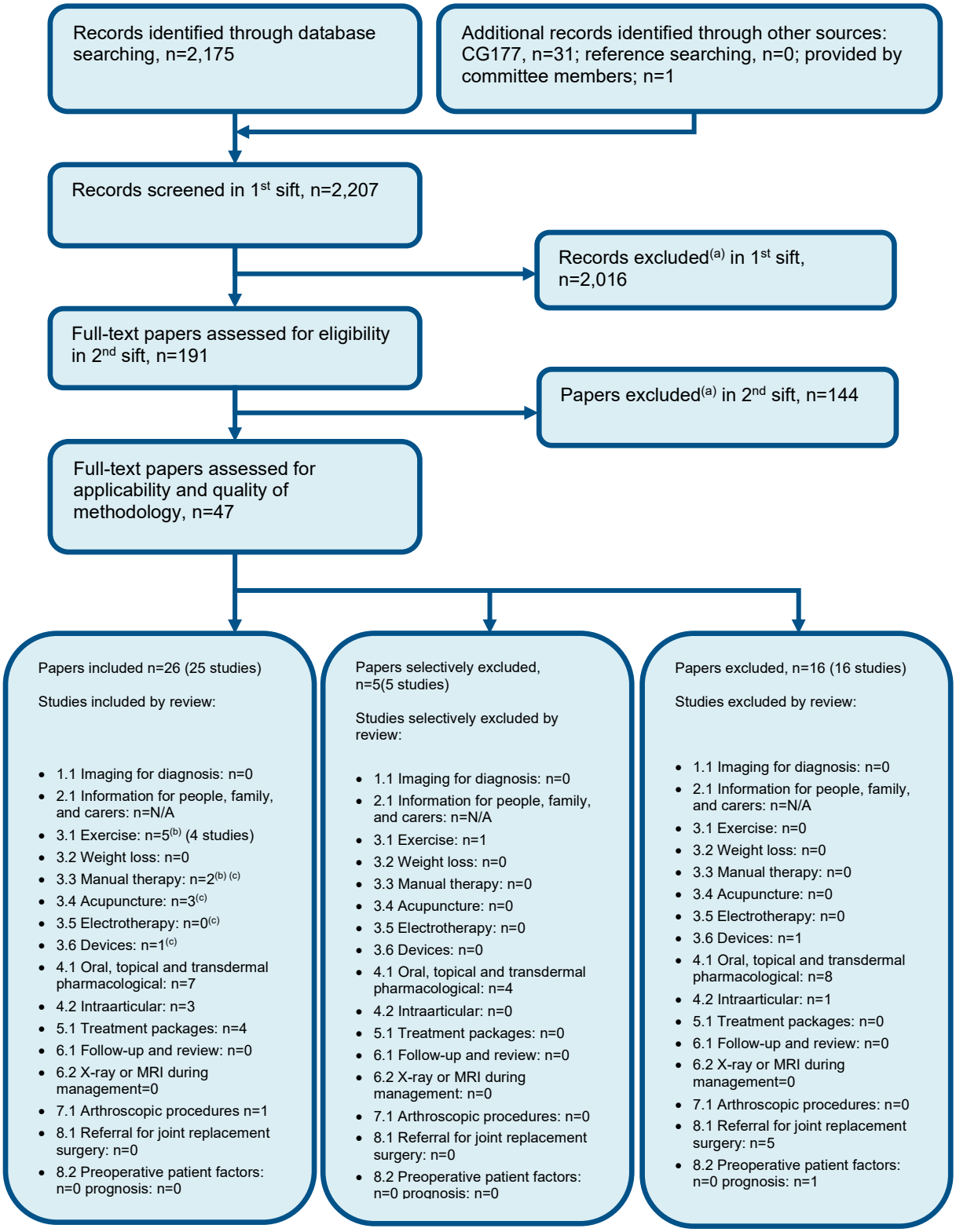
a. Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias

b. Downgraded by 1 or 2 increments because heterogeneity, unexplained by subgroup analysis

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

d. Absolute effect calculated by risk difference due to zero events in at least one arm of one study

Appendix G – Economic evidence study selection



(a) Non-relevant population, intervention, comparison, design or setting; non-English language.
(b) Two articles identified were applicable to Q3.1 and Q3.3, for the purposes of this diagram they have been included under Q3.1 only.
(c) One article identified was applicable to Q3.3, Q3.4, Q3.5 and Q3.6, for the purposes of this diagram it has been included under Q3.3 only.

Appendix H – Economic evidence tables

Study	Bennell 2016[37]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: Cost-utility analysis **Study design:** Within-trial analysis **Approach to analysis:** Costs and QALYs were analysed using mixed linear statistical models of baseline levels and treatment groups, with a random intercept for each physical therapist clustered by site. **Perspective:** Australia **Time horizon:** 52 weeks **Discounting:** n/a	**Population:** Patients with knee osteoarthritis **Cohort settings:** Start age: 63 Male: 40% **Intervention 1:** Exercise (10 individual sessions over 12 weeks lasting 25 minutes each with a physical therapist) **Intervention 2:** PCST (10 individual sessions over 12 weeks lasting 45 minutes each with a physical therapist) **Intervention 3:** PCST/exercise (10 individual sessions over 12 weeks lasting 70 minutes each with a physical therapist)	**Total costs (mean per patient):** Incremental (2−1): £133 (95% CI: NR; p=NR) Incremental (3−1): £285 (95% CI: NR; p=NR) Incremental (3−2): £152 (95% CI: NR; p=NR) **Currency & cost year:** Australian dollars, assumed to be 2012 as this is the period leading to a follow-up (presented here as 2012 UK pounds[(a)]) **Cost components incorporated:** Therapy and other healthcare-related costs, excluding initial fixed cost of physical therapist training and impact on patient incomes or travel/time costs.	**QALYs gained (mean per patient):** Incremental (2−1): 0.01 (95% CI: -0.03 to 0.04; p=NR) Incremental (3−1): 0.03 (95% CI: -0.01 to -0.07; p=NR) Incremental (3−2): 0.03 (95% CI: -0.01 to 0.06; p=NR)	**Cost per QALY (Intervention 2 versus Intervention 1):** £13,300 per QALY gained 95% CI: NR **Cost per QALY (Intervention 3 versus Intervention 1):** £9,500 per QALY gained 95% CI: NR **Cost per QALY (Intervention 3 versus Intervention 2):** £5,067 per QALY gained 95% CI: NR **Analysis of uncertainty:** None reported

Data sources
Health outcomes: QALYs were estimated as the area under the curve of preference-based AQoL-6D scores in the month prior to baseline, and at weeks 12, 32, and 52. **Quality-of-life weights:** The AQoL-6D is a validated preference-based measure of quality of life on a -0.04 (worse than death) to 1 (perfect health) scale. **Cost sources:** The direct cost of treatments was defined as the recorded number of treatment sessions multiplied by the payment rate for physical therapists in the trial. Healthcare-related resource use (hospital inpatient, prescription and non-prescription medications, medical services including hospital outpatient appointments, diagnostic tests, and other health practitioners) was taken from a questionnaire at baseline and at weeks 4, 8, 12, 32 and 52, and valued using prices listed in published studies.
Comments
Source of funding: Australian Health Management, National Health and Medical Research Council. **Limitations:** Patients and physical therapists were not blinded. 17% of patients were lost at follow-up (52 weeks). Results are reflective of the Australian healthcare setting and the training received by physical therapists; it therefore may not be reflective of other healthcare settings. It is unclear how the preference weights for the AQoL-6D were valued. The study did not report final costs per QALYs, so these were calculated from the reported incremental costs (converted to UK pounds first) and QALYs. The incremental QALYs were reported to one significant figure which means the cost per QALY gained is subject to uncertainty. For example. the cost per QALY for intervention 2 vs intervention 1 could feasibly range between £9,500 and £27,000 with the addition of another decimal place. **Other:**
Overall applicability:[b] Partially applicable **Overall quality:**[c] Potentially serious limitations

Abbreviations: AQoL-6D= Assessment of Quality of Life 6 Domains; 95% CI= 95% confidence interval; n/a= not applicable; NR= not reported; PCST= pain coping skills training; QALYs= quality-adjusted life years; UK= United Kingdom

(a) Converted using 2012 purchasing power parities[214]

(b) Directly applicable / Partially applicable / Not applicable

(c) Minor limitations / Potentially serious limitations / Very serious limitations

Study	Health Quality (Ontario HTA) 2018[118]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: Cost-utility analysis **Study design:** Probabilistic decision analytical model **Approach to analysis:** A simple decision analytic model based on an RCT (Skou 2015)[253] comparing group-based structured education and neuromuscular exercise program to usual care. **Perspective:** Canadian healthcare system **Time horizon:** 1 year **Treatment effect duration:**[(a)] 1 year **Discounting:** n/a	**Population:** Adults with knee OA **Cohort settings:** Start age: Intervention: 64.8 Control: 67.1 Male: NR **Intervention 1:** Usual care **Intervention 2:** Structured education and neuromuscular exercise program (two educational sessions and 24 exercise sessions over 12 weeks)	**Total costs (mean per patient):** Intervention 1: £1,874 Intervention 2: £2,436 Incremental (2−1): £407 (95% CI: £232 to £633; p=NR) **Currency & cost year:** 2017 Canadian dollars (presented here as 2017 UK pounds[(b)]) **Cost components incorporated:** Consultations with health care professionals, diagnostic tests and examinations, and hospitalisation	**QALYs (mean per patient):** Intervention 1: 0.73 Intervention 2: 0.76 Incremental (2−1): 0.03 (95% CI: -0.006 to 0.06; p=NR)	**Cost per QALY (Intervention 2 versus Intervention 1):** £13,550 per QALY gained (pa) 95% CI:NR Probability Intervention 2 cost effective (£28k/56K per QALY threshold) [(b)]: 81%/90% **Analysis of uncertainty:** 24- month time horizon: ICER of £6,757 per QALY gained. Reduction in pain medication use among those who participate in a structured education and neuromuscular exercise program: cost per QALY of £10,173 per QALY gained.

Data sources.
Health outcomes: Utilities for the model were taken from an RCT by Skou 2015.[253] **Quality-of-life weights:** EQ-5D was measured at baseline and at 12 months follow up. It was assumed that utility in both arms were identical at baseline with the final score calculated by adding the change in utility at 12 months to the baseline score. **Cost sources:** The cost of structured education and neuromuscular exercise were taken from Skou 2015.[253] Resource use and costs data were obtained over the follow-up period from a study conducted in Ontario but the cost is identical across both arms so is unlikely to impact the cost per QALY valuation. This study surveyed patients from randomly selected family practices to measure their health care service use to manage osteoarthritis and comorbidities.
Comments
Source of funding: NR. **Limitations:** The clinical evidence was derived from a single RCT that measured general health status at 12 months following baseline assessment. The interventional cost estimates were based primarily on expert consultation and currently available (not publicly funded) group-based programmes. Costs and resource for usual care were taken from a study published in 2004. **Other:**
Overall applicability:(c) Partially applicable **Overall quality:**(d) Potentially serious limitations

Abbreviations: 95% CI= 95% confidence interval;; EQ-5D= EuroQol 5 dimensions (scale: 0.0 [death] to 1.0 [full health], negative values mean worse than death); n/a= not applicable; NR= not reported; OA= osteoarthritis; pa= probabilistic analysis; QALYs= quality-adjusted life years; RCT= randomised controlled trial; UK= United Kingdom

(a) For studies where the time horizon is longer than the treatment duration, an assumption needs to be made about the continuation of the study effect. For example, does a difference in utility between groups during treatment continue beyond the end of treatment and if so for how long.

(b) Converted using 2017 purchasing power parities[214]

(c) Directly applicable / Partially applicable / Not applicable

(d) Minor limitations / Potentially serious limitations / Very serious limitations

Study	Jessep 2009 [144]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: Cost utility analysis (health outcome: QALYs) **Study design:** Within-trial analysis (same paper) **Approach to analysis:** Analysis of individual level data EQ5D and resource use. Unit costs applied. **Perspective:** UK NHS **Follow-up:** 1 year **Treatment effect duration:**[(a)] n/a **Discounting:** Costs: n/a; Outcomes: n/a	**Population:** People with mild, moderate or severe non-specific chronic knee pain. **Patient characteristics:** N: 64 Start age: 66.5 Male: 31% **Intervention 1:** Outpatient physiotherapy (usual care, up to a maximum of 10 sessions) **Intervention 2:** ESCAPE (two exercise-based supervised sessions a week lasting 1 hour up to 5 weeks with educational material provided to take home)	**Total costs (mean per patient):** Intervention 1: £583 Intervention 2: £320 Incremental (2−1): saves £263 (95% CI: NR; p=NR) **Currency & cost year:** 2005 UK pounds **Cost components incorporated:** Healthcare utilisations costs included A&E, GP, nurse and outpatient visits, other primary care and medication costs.	**QALYs (mean change per patient from baseline):** Intervention 1: -0.03 Intervention 2: 0.05 Incremental (2−1): 0.08 (95% CI: NR; p=NR)	**ICER (Intervention 2 versus Intervention 1):** Intervention 2 dominates intervention 1 **Analysis of uncertainty:** NR
Data sources				
Health outcomes: This was a single-blind pragmatic RCT where the assessor was unaware of the patient's treatment allocation. The change in score between groups at 12 months was assessed using analysis of covariance to correct for baseline values. **Quality-of-life weights:** EQ-5D UK tariff. **Cost sources:** Healthcare utilisation assessed using the Client Services Receipt Inventory (interview-based questionnaire). Unit costs were taken from the Primary Care Trusts reference costs 2005/06.				
Comments				
Source of funding: Physiotherapy Research Foundation. **Limitations:** Group sessions compared to individual sessions. Small study with only 67 participants were recruited at baseline. No analysis of uncertainty nor sensitivity analysis of results conducted. Health outcomes based on results from a				

single trial. Costs from 2005 may not reflect current UK NHS practice. **Other:** The immediate cost of intervention 2 was nearly half that of intervention 1 and seems to be driven by the assumption that 6 participants will attend the complete programme in a group.	
Overall applicability:[c] Partially applicable	**Overall quality:**[d] Potentially serious limitations

Abbreviations: 95% CI= 95% confidence interval; da= deterministic analysis; ESCAPE: Enabling self-management and coping with arthritic knee pain through exercise; EQ-5D= Euroqol 5 dimensions (scale: 0.0 [death] to 1.0 [full health], negative values mean worse than death); ICER= incremental cost-effectiveness ratio; n/a= not applicable; NR= not reported; QALYs= quality-adjusted life years

(a) For studies where the time horizon is longer than the treatment duration, an assumption needs to be made about the continuation of the study effect. For example, does a difference in utility between groups during treatment continue beyond the end of treatment and if so for how long.

(b) Directly applicable / Partially applicable / Not applicable

(c) Minor limitations / Potentially serious limitations / Very serious limitations

Study	Marra 2014[179]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: Cost-utility analysis **Study design:** Probabilistic decision analytical model **Approach to analysis:** Data from a cluster RCT were used. Missing data (14% of cases for costs, 18% for the PAT-5D and 12% for the HUI3) were assumed to be random and therefore imputed using the Markov Chain Monte Carlo (MCMC) procedure. This produced multiple datasets for incomplete data, so an average of the costs and QALYs was taken and then attributed to the relevant patients. **Perspective:** Canadian healthcare system **Time horizon:** 6 months **Treatment effect duration:**[(a)] 6 months **Discounting:** n/a	**Population:** Patients with newly diagnosed knee OA **Cohort settings:** Start age: NR Male: NR **Intervention 1:** Usual care (educational pamphlet on knee OA care created by The Arthritis Society. **Intervention 2:** Administration of a validated knee OA screening questionnaire by a pharmacist, education, pain medication management by a pharmacist, physiotherapy-guided exercise, and communication with the patient's primary care physician	**Total costs (mean per patient based on HUI3):** Intervention 1: £66 Intervention 2: £71 Incremental (2−1): £5 (95% CI: NR; p=0.35) **Total costs (mean per patient based on PAT-5D):** Intervention 1: £68 Intervention 2: £71 Incremental (2−1): £3 (95% CI: NR; p=0.41) **Currency & cost year:** 2009 Canadian dollars (presented here as 2009 UK pounds[(b)]) **Cost components incorporated:** Physicians visits, treatments/ medications, laboratory tests and imaging.	**QALYs (mean per patient based on HUI3):** Intervention 1: 0.3642 Intervention 2: 0.3863 Incremental (2−1): 0.0221 (95% CI: NR; p<0.01) **QALYs (mean per patient based on PAT-5D):** Intervention 1: 0.4237 Intervention 2: 0.4473 Incremental (2−1): 0.0236 (95% CI: NR; p=<0.01)	**Cost per QALY (Intervention 2 versus Intervention 1 based on HUI3):** £254 per QALY gained (pa) 95% CI: (from cost saving to £1,713) **Cost per QALY (Intervention 2 versus Intervention 1 based on pat-5d):** £137 per QALY gained (pa) 95% CI: (from cost saving to £1,272) Probability Intervention 2 cost effective (£1,200 per QALY threshold): 90% **Analysis of uncertainty:** Results were presented separately by health measures and perspective (societal and ministry of health) but no other sensitivity analysis was conducted.

Data sources
Health outcomes: In the PhIT-OA trial[178], pharmacies were randomly allocated to provide either intervention 1 or intervention 2. The PhIT-OA trial was excluded from the clinical review as the reported clinical outcomes did not fit the protocol. However, it was included in the economic review as the intervention is classified as a treatment package and would provide useful economic data. **Quality-of-life weights:** The Health Utilities Index Mark 3 (HUI3) and the Paper Adaptive Test-5D (PAT-5D) were administered to patients at baseline, 3 months and 6 months. **Cost sources:** Data on healthcare utilisation (physician visits, laboratory tests, hospital admissions, imaging studies, medication, and home care) were collected at 3 and 6 months from patient responses to questionnaires. The costs of healthcare professional visits and the cost of equipment (aids or devices such as braces or canes) were based on the 2009 British Columbia Medical Services Plan. Physiotherapy is not funded in British Columbia, so costs related to physiotherapists were not included here.
Comments
Source of funding: NR. **Limitations:** Patients were not blinded. Short time horizon of 6 months. It is unclear how unit costs were assigned to each component of resource utilisation. It is unclear how the preference weights for utilities were valued and how QALYs were calculated. Results are specific to the Canadian healthcare system and may not be applicable to other settings. **Other:**
Overall applicability:(c) Partially applicable **Overall quality:**(d) Potentially serious limitations

Abbreviations: 95% CI= 95% confidence interval; CUA= cost–utility analysis; HUI3= The Health Utilities Index Mark 3; n/a= not applicable; NR= not reported; pa= probabilistic analysis; PAT-5D= Paper Adaptive Test-5D; PhIT-OA: Pharmacist-Initiated Intervention Trial in Osteoarthritis; QALYs= quality-adjusted life years; UK= United Kingdom

(a) For studies where the time horizon is longer than the treatment duration, an assumption needs to be made about the continuation of the study effect. For example, does a difference in utility between groups during treatment continue beyond the end of treatment and if so for how long.

(b) Converted using 2009 purchasing power parities[214]

(c) Directly applicable / Partially applicable / Not applicable

(d) Minor limitations / Potentially serious limitations / Very serious limitations

Appendix I – Health economic model

No original economic modelling was undertaken.

Appendix J – Excluded studies

Clinical studies

Table 25: Studies excluded from the clinical review

Study	Exclusion reason
Ackerman 2012[1]	Incorrect interventions (education programme only)
Ackerman 2013[2]	Incorrect interventions (education programme only)
Arfaei Chitkar 2021[23]	Wrong comparison (mobile app based instruction and usual care versus usual care (routine medical care, educational content))
Aglamis 2008[5]	Incorrect interventions (exercise only)
Ahern 2018[6]	Incorrect interventions (no education/behaviour change component)
Alfredo 2012[9]	Incorrect interventions (no education/behaviour change component)
Ali 2018[10]	Incorrect study design (qualitative study)
Allegrante 1991[11]	Not available
Allen 2010[15]	Incorrect interventions (education programme only)
Allen 2012[14]	Protocol only
Allen 2016[19]	Not review population (includes people with osteoarthritis and their healthcare providers)
Allen 2016[13]	Incorrect interventions (exercise only)
Allen 2017[16]	Not review population (includes people with osteoarthritis and their healthcare providers)
Allen 2018[12]	Incorrect interventions (no education/behaviour change component)
Allen 2019[17]	Wrong intervention (pain coping skills programme, waiting list)
Altmis 2018[20]	Not review population (includes healthy people without osteoarthritis)
Anonymous 2004[21]	Abstract only
Anwer 2016[22]	Incorrect interventions (exercise only)
Aunger 2020[25]	Wrong intervention (behavioural change intervention
Aunger 2019[26]	Protocol only
Axford 2008[27]	Incorrect interventions (education programme only)
Azma 2018[28]	Inappropriate comparison (compared office-based physical therapy to tele-rehabilitation)
Bandak 2021[29]	Inappropriate comparison (comparing a treatment package to an intraarticular injection of saline, which is not an intervention considered as an active treatment for osteoarthritis in this guideline)
Barker 2021[30]	Wrong population (post-operative patients)
Barker 2020[31]	Wrong population (post-operative patients)
Barlow 2000[32]	Incorrect interventions (education programme only)
Beavers 2014[34]	Incorrect interventions (dietary intervention including specific weight loss products)
Bendrik 2021[35]	Wrong intervention (individually tailored physical activity recommendations, advice only)
Bennell 2005[42]	Incorrect interventions (no education/behaviour change component)

Bennell 2010[41]	Inappropriate comparison (compares a treatment package to inactive ultrasound and an inert gel when the treatment package does not include ultrasound as a component)
Bennell 2014[40]	Inappropriate comparison (compares a treatment package to inactive ultrasound and an inert gel when the treatment package does not include ultrasound as a component)
Bennell 2022[44]	Inappropriate comparison (compares a treatment package containing exercise education and behavioural counselling to another treatment package with an added dietary intervention, which is not specified as a comparison in the protocol)
Bilgici 2005[48]	Not available
Bliddal 2011[49]	Incorrect interventions (dietary intervention including specific weight loss products)
Blixen 2004[50]	Incorrect interventions (education programme only)
Bobos 2018[51]	Incorrect interventions (behaviour change intervention only)
Bossen 2013[52]	Incorrect interventions (exercise only)
Brand 2013[53]	Systematic review: study designs inappropriate (includes only cohort studies)
Broderick 2014[54]	Incorrect interventions (behaviour change intervention only)
Brosseau 2018[55]	Incorrect study design (Delphi study)
Bryant 2014[57]	Not guideline condition. Not review population (physiotherapists)
Buszewicz 2006[58]	Incorrect interventions (education programme only)
Button 2015[59]	Not review population (includes knee osteoarthritis, but also other conditions, such as anterior cruciate ligament pathologies with proportions unclear)
Callaghan 1995[60]	No usable outcomes (reports medians and ranges for continuous outcomes)
Cetin 2008[61]	Incorrect interventions (no education/behaviour change component)
Chang 2014[62]	Protocol only
Chang 2017[63]	Incorrect interventions (no education/behaviour change component)
Cheing 2002[65]	Incorrect interventions (no education/behaviour change component)
Cheing 2004[64]	Incorrect interventions (no education/behaviour change component)
Chen 2013[67]	Incorrect interventions (no education/behaviour change component)
Chen 2019[66]	Incorrect interventions (exercise only)
Chua 2008[68]	No usable outcomes (reports beta-coefficients for continuous outcomes only)
Coelho cde 2014[70]	Incorrect interventions (no education/behaviour change component). Protocol only
Cohen 1986[71]	Incorrect interventions (education programme only)
Coleman 2008[72]	Incorrect interventions (education programme only)
Coleman 2012[73]	Incorrect interventions (education programme only)
Cortes godoy 2014[74]	Incorrect interventions (no education/behaviour change component)
Crotty 2009[77]	Incorrect interventions (behaviour change intervention only)
Cuesta-vargas 2015[78]	Not review population (includes people with osteoarthritis, low back pain and chronic neck pain). Inappropriate comparison

	(compares an intervention delivered 3 times a week to one delivered 2 times a week)
Cuperus 2015[79]	Inappropriate comparison (compares a face-to-face program to a telephone-based treatment)
De jong 2004[81]	Inappropriate comparison (compares a hip osteoarthritis program to a knee osteoarthritis program)
De matos brunelli braghin 2018[82]	Incorrect interventions (no education/behaviour change component)
De rezende 2016[83]	Incorrect interventions (education only)
De rezende 2016[84]	Incorrect interventions (education only)
De vos 2014[85]	Not review population (healthy people at risk of developing osteoarthritis)
Deveza 2017[86]	Protocol only
Devos-comby 2006[88]	Incorrect interventions (the study mostly compared exercise to self-management rather than the combination of the two against the components)
Dincer 2008[90]	Incorrect interventions (no education/behaviour change component)
Dobson 2014[91]	Protocol only
Dunning 2018[92]	Incorrect interventions (no education/behaviour change component)
Ettinger 1997[97]	Incorrect interventions (no education/behaviour change component to exercise intervention)
Fisher 1993[100]	Incorrect interventions (no education/behaviour change component)
Fisken 2015[101]	Incorrect interventions (no education/behaviour change component)
Fitzgibbon, 2020[102]	Wrong comparison (Fit and Strong plus (exercise, education, weight change support) versus Fit and stroke (exercise, education))
Foster 2007[108]	Incorrect interventions (no education/behaviour change component)
Foster 2014[107]	Incorrect interventions (no education/behaviour change component). Protocol only
Ganji 2018[110]	Incorrect interventions (education programme only)
Gay 2018[111]	Protocol only
Ghroubi 2008[112]	Incorrect interventions (dietary intervention including specific weight loss products)
Goff 2021[113]	Wrong intervention (patient education, non-pharmacological comparison)
Gravas 2019[114]	No usable outcomes (only reports likelihood of having surgery)
Hall 2019[115]	Incorrect interventions (dietary intervention including specific weight loss products)
Hansson 2010[116]	Incorrect interventions (education programme only)
Hay 2006[117]	Incorrect interventions (no education component)
Health quality 2018[118]	Systematic review: study designs inappropriate (includes observational studies)
Helminen 2015[119]	Incorrect interventions (behaviour change component only)
Heuts 2005[120]	Incorrect interventions (education programme only)
Higgins 2015[121]	Incorrect interventions (surgical intervention only)
Hinman 2017[122]	Protocol only

Holden 2017[123]	Incorrect interventions (exercise only)
Hoogeboom 2012[124]	No usable outcomes (results presented in graphical form only)
Huang 2005[128]	Incorrect interventions (no education/behaviour change component)
Huang 2017[129]	Incorrect interventions (education programme only)
Hughes 2020[133]	Inappropriate comparison (compares a treatment package to another treatment package)
Hunt 2013[134]	Inappropriate comparison (no education component in the exercise intervention)
Hunter 2015[135]	No usable outcomes (reports radiographic parameters only)
Hurley 2018[136]	Incorrect interventions (Cochrane review, no education component in the exercise interventions)
Ikeda 2018[140]	Incorrect interventions (no education/behaviour change component)
Ismail 2017[142]	Systematic review: quality assessment is inadequate
Jan 1991[143]	Not available
Kars fertelli 2018[147]	Incorrect interventions (no education/behaviour change component)
Keays 2015[149]	Incorrect study design (non-randomised study)
Keogh 2018[152]	Incorrect interventions (exercise only)
Kigozi 2018[153]	Incorrect interventions (exercise only)
Kim 2012[154]	Incorrect study design (non-randomised study)
Kloek 2018[158]	Incorrect study design (non-randomised study)
Kloek cjj phd 2020[157]	Incorrect study design (mixed methods study discussing the qualitative component)
Kroon 2014[160]	Incorrect interventions (education programme only)
Kumar 2013[161]	Incorrect interventions (no education/behaviour change component)
Laufer 2014[162]	Incorrect interventions (no education/behaviour change component)
Lee 2006[165]	Non-English language study
Lee 2017[164]	Incorrect study design (non-randomised study)
Li 2013[166]	Incorrect study design (non-randomised study)
Loeser 2017[169]	Incorrect interventions (dietary intervention including specific weight loss products)
Loew 2017[170]	Incorrect interventions (exercise only)
Lord 1999[171]	Incorrect interventions (education programme only)
Lorig 2008[172]	Not review population. People with conditions that may make them susceptible to osteoarthritis or often occur alongside osteoarthritis (including: crystal arthritis, inflammatory arthritis, septic arthritis, diseases of childhood that may predispose to osteoarthritis, medical conditions presenting with joint inflammation and malignancy)
Magrans-courtney 2011[173]	Incorrect interventions (dietary intervention including specific weight loss products)
Maire 2006[174]	Post-hip arthroplasty
Marconcin 2016[176]	Inappropriate comparison (compares a treatment programme to an education programme that is not a component of the treatment programme being studied)

Marconcin 2018[175]	Inappropriate comparison (compares a treatment programme to an education programme that is not a component of the treatment programme being studied)
Marconcin 2021[177]	Wrong comparison (self- management and exercise versus education only)
Mazzei 2021[180]	Systematic review; references checked
Mazzuca 2004[181]	Incorrect interventions (education programme only)
Mccarthy 2004[182]	Incorrect interventions (exercise only)
Mccarthy 2004[183]	Incorrect interventions (exercise only)
McVeigh, 2021[185]	Wrong comparison (home exercise (supervised strength exercise) and standard conservative therapy (orthoses, education, behaviour change) versus standard conservative therapy only)
Messier 2000[191]	No usable outcomes (report biomechanical outcomes only)
Messier 2007[189]	Incorrect interventions (no education/behaviour change component)
Messier 2013[190]	Incorrect interventions (dietary intervention including specific weight loss products)
Messier 2017[187]	Protocol only
Mihalko 2019[192]	Merge with Messier 2013 [190]
Miller 2006[193]	Incorrect interventions (behaviour change intervention only)
Mizusaki imoto 2013[195]	Incorrect interventions (exercise only)
Moe 2010[197]	Incorrect interventions (education programme only)
Moe 2016[196]	Incorrect interventions (education programme only)
Molgaard 2018[198]	Inappropriate comparison (no education/behaviour change component)
Murphy 2018[199]	Incorrect interventions (behaviour change component only)
Nahayatbin 2018[200]	Incorrect interventions (exercise only)
Nejati 2015[202]	Incorrect interventions (no education/behaviour change component)
Nelligan 2021[203]	Wrong comparison (doesn't compare like with like - the websites were different for each intervention groups, and the intervention group also receives a behaviour change text messaging service)
Nelligan 2019[204]	Wrong comparison (website with education and self-directed strengthening regimen versus website with education)
Ng 2010[205]	Incorrect interventions (exercise only)
Nicklas 2004[206]	No usable outcomes (reported biomarker outcomes only)
Nour 2006[207]	Incorrect interventions (behaviour change intervention only)
O'brien 2018[209]	Incorrect interventions (behaviour change intervention only)
Ogut 2018[210]	Inappropriate comparison (compares a programme with no education/behaviour change component to another programme with one component missing)
Osborne 2006[215]	Incorrect interventions (education programme only)
Østerås 2021[216]	Conference abstract
Ozguclu 2010[217]	Incorrect interventions (no education/behaviour change component)
Palmer 2014[218]	Inappropriate comparison (compares two treatment packages)
Park 2013[220]	Incorrect interventions (no education/behaviour change component)
Park 2014[219]	Incorrect interventions (no education/behaviour change component in the exercise intervention)

Patel 2009[221]	Incorrect interventions (education programme only)
Perez-marmol 2017[223]	Incorrect interventions (no education/behaviour change component)
Peterson 1993[224]	No usable outcomes (reported biomechanical outcomes only)
Pitsillides 2021[225]	Systematic review; references checked
Piyakhachornrot 2011[226]	Inappropriate comparison (compares a treatment package with supervised exercise to a package with unsupervised exercise)
Rafiq, 2021[231]	Abstract only
Rattanachaiyanont 2008[232]	Incorrect interventions (includes sham electrotherapy as a component of a treatment package, comparing this to a package with electrotherapy)
Ravaud 2004[234]	Not review population (rheumatologists providing care for people with osteoarthritis)
Ravaud 2009[233]	Incorrect interventions (education programme only)
Robbins 2017[237]	Protocol only
Rodrigues da silva 2017[238]	Incorrect interventions (education programme only)
Rogind 1998[239]	Incorrect interventions (exercise only)
Rosemann 2007[240]	Incorrect interventions (education programme only)
Runhaar 2016[241]	Not guideline condition. Not review population (people without osteoarthritis)
Saccomanno 2016[242]	Incorrect interventions (no education/behaviour change component)
Sanchez romero 2019[243]	Incorrect interventions (no education/behaviour change component)
Schafer 2018[245]	Systematic review: study designs inappropriate (included non-randomised studies)
Schlenk 2011[247]	Incorrect interventions (education component includes only one session, so does not qualify for a treatment package)
Schlenk, 2020[246]	Wrong comparison (supervised mixed exercise (strength and aerobic) and telephone sessions versus telephone sessions only))
Schrubbe 2016[248]	Protocol only
Sevick 2009[249]	Incorrect interventions (dietary intervention including specific weight loss products)
Sharma 2018[250]	Inappropriate comparison (both interventions include exercise, leading to the comparison being two treatment packages)
Shavianidze 1991[251]	Non-English language study
Skou 2020[254]	No usable outcomes (health economic evidence only)
Smith-ray 2014[255]	Protocol only
Somers 2012[256]	Incorrect interventions (behaviour change intervention only)
Soni 2012[257]	Incorrect interventions (no education/behaviour change component)
Stamm 2002[258]	No usable outcomes (presents results in graphical form only)
Steinhilber 2012[259]	Includes people after having total hip replacement surgery
Steinhilber 2017[260]	Incorrect interventions (no education/behaviour change component)
Stoffer-marx 2018[262]	Incorrect interventions (includes the provision of nutritional supplements)
Taylor 2018[269]	Incorrect interventions (behaviour change intervention only)
Tegiacchi 2018[270]	Erratum only

Teirlinck 2016[271]	Incorrect interventions (education component not stated and not being offered as a formalised package)
Thomas 2002[273]	No usable outcomes (inappropriate pooling of study arms for this protocol)
Thomas 2005[272]	Incorrect study design. Incorrect interventions (no education/behaviour change component)
Umapathy 2015[274]	Protocol only
Vas 2004[276]	Incorrect interventions (no education/behaviour change component)
Victor 2005[277]	Incorrect interventions (education programme only)
Villadsen 2014[278]	Secondary analysis of RCTs
Walsh 2020[280]	Wrong population (mixed hip and knee osteoarthritis and low back pain- unclear numbers)
Wang 2018[282]	Incorrect interventions (education programme only)
Wang 2018[281]	Protocol only
Warsi 2003[283]	People with conditions that may make them susceptible to osteoarthritis or often occur alongside osteoarthritis (including: crystal arthritis, inflammatory arthritis, septic arthritis, diseases of childhood that may predispose to osteoarthritis, medical conditions presenting with joint inflammation and malignancy)
Woods 2017[284]	Incorrect interventions (no education/behaviour change component)
Yan 2013[285]	Protocol only
Yilmaz 2019[286]	Incorrect interventions (no education/behaviour change component)
Yurtkuran 1999[290]	Not available
Zacharias 2014[291]	Incorrect interventions (no education/behaviour change component)
Zammit 2010[292]	Incorrect interventions (Cochrane review, does not include treatment packages by our definition)
Zgibor 2017[293]	People with conditions that may make them susceptible to osteoarthritis or often occur alongside osteoarthritis (including: crystal arthritis, inflammatory arthritis, septic arthritis, diseases of childhood that may predispose to osteoarthritis, medical conditions presenting with joint inflammation and malignancy)
Zhou 2008[294]	Not available
Zhou 2015[295]	Non-English language study

Health Economic studies

Published health economic studies that met the inclusion criteria (relevant population, comparators, economic study design, published 2005 or later and not from non-OECD country or USA) but that were excluded following appraisal of applicability and methodological quality are listed below. See the health economic protocol for more details.

None.