National Institute for Health and Care Excellence

Final

Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain

[E] Evidence review for exercise for chronic primary pain

NICE guideline NG193

Intervention evidence review underpinning recommendations 1.2.1 to 1.2.2 in the NICE guideline

April 2021

This evidence review was developed by the National Guideline Centre based at the Royal College of Physicians

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and, where appropriate, their carer or guardian.

Local commissioners and providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN
978-1-4731-4066-0

Contents

1 Exercise interventions for chronic primary pain

1.1 Review question: What is the clinical and cost effectiveness of exercise interventions for the management of chronic primary pain?

1.2 Introduction

Exercise, or physical activity, is an important part of a healthy lifestyle. Activities associated with daily living such as walking, housework and gardening can be supplemented by activities typically considered to be exercise such as sporting activities and attendance at gyms. Exercise is particularly important for people with a variety of health conditions including musculoskeletal and cardiovascular, and is increasingly seen to be important in managing mental health problems. Increased physical activity is often recommended for people with chronic pain. A challenge for people with pain is to identify the amount and type of exercise that will reduce the impact pain has on their lives, set up healthy exercise habits, and enable them to enjoy the wider health benefits of maintaining an active lifestyle. Remaining motivated to continue exercising can also be more challenging for people living with pain.

Exercise can be carried out alone or as part of social interaction in groups and with teams. Supervised exercise can often be delivered in group settings. The emphasis is usually on encouraging and supporting the person to carry out the exercise independently and regularly.

A growing body of research shows exercise has an impact on many biological systems, including the nervous system, leading to a focus on exercise as a means to pain reduction. Exercise therapy can helpfully be framed in this context.

Although the variety of exercise types is vast, they can broadly be classified into one or more of four categories:

- Cardiovascular/aerobic/conditioning
- Resistance/anaerobic/strength
- Flexibility including stretching
- Proprioceptive including balance and movement awareness.

More recently terms like mind-body have emerged to define exercises that include movement with an emphasis on focussed awareness and often with connection to metaphysical and cultural philosophies. Examples include the various forms of Yoga and Tai Chi. These exercises can also be classified using the existing classification system above. This evidence review will look at the effectiveness of these types of exercise for people with chronic primary pain, including its effects on quality of life and function.

1.3 PICO table

For full details see the review protocol in appendix A.

Table 1: PICO characteristics of review question

Population	People, aged 16 years and over, with chronic primary pain (whose pain management is not addressed by existing NICE guidance) (chronic widespread pain, complex regional pain syndrome, chronic visceral pain, chronic orofacial pain, chronic musculoskeletal pain other than orofacial)

	Chronic pain in one or more anatomical regions that is characterized by significant emotional distress (anxiety, anger/frustration or depressed mood) and functional disability (interference in daily life activities and reduced participation in social roles). The diagnosis is appropriate independently of identified biological or psychological contributors unless another diagnosis would better account for the presenting symptoms.
Interventions	Interventions: • Mind-body exercise (e.g. yoga, Tai Chi) • Biomechanical (e.g. pilates) exercise • Proprioceptive exercise • Strength training • Flexibility • Aerobic (e.g. swimming, walking programme, aerobic exercise) • Graded motor imagery • Mixed modality exercise (aerobics and/or mind-body and/or biomechanical).
Comparisons	Comparators: • Each other • Usual care • Psychological therapies • Other physical therapies (e.g. manual therapy) • Manual therapy + exercise.
Outcomes	CRITICAL: • Pain reduction (any validated scale) • Health related quality of life (including meaningful activity) • Physical function (e.g. 6minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • Psychological distress (depression/anxiety) (preferably Hospital Anxiety and Depression Scale) IMPORTANT: • Use of healthcare services • Sleep • Discontinuation. Outcomes will be extracted at the longest time point up to 3 months and at the longest time point after 3 months.
Study design	Randomised controlled trials (RCTs) and systematic reviews of RCTs Cross-over RCTs will be considered if no non-cross-over RCT evidence is identified.

1.4 Clinical evidence

1.4.1 Included studies

91 studies were included in the review; these are summarised in the tables below. Evidence from these studies is summarised in the clinical evidence summary below.

3 Cochrane reviews that were relevant to this review question were identified and included in the review.[33, 49, 250] These covered the following:

- Mind-body therapy for fibromyalgia
- Aerobic exercise for fibromyalgia

- Strength training for fibromyalgia.

Evidence that had been published since the Cochrane publication dates were added to the original analyses, as were additional populations, interventions, comparisons and outcomes relevant to this review protocol.

Two Cochrane reviews relevant to this review question were identified after this review had been conducted. These reviews were not included, however references were cross-referenced against this review[32, 150].

Evidence was identified for the following populations:

- Fibromyalgia (58 studies)
- Chronic neck pain (31 studies)
- Complex regional pain syndrome (1 study)
- Masticatory pain (1 study)
- Chronic pelvic pain syndrome (1 study)

Evidence was identified for the following comparisons:

1. Aerobic exercise versus usual care
2. Strength training versus usual care
3. Aerobic exercise and strength training versus usual care
4. Aerobic, strength and flexibility versus usual care
5. Strength training and flexibility versus usual care
6. Strength, proprioception and flexibility versus usual care
7. Proprioception versus usual care
8. Mind-body exercise versus usual care
9. Flexibility versus usual care
10. Aerobic exercise versus strength training
11. Aerobic exercise versus flexibility
12. Aerobic exercise versus biomechanical exercise
13. Aerobic exercise and strength training versus aerobic exercise
14. Aerobic exercise and strength training versus flexibility
15. Aerobic exercise and flexibility versus mind-body exercise
16. Aerobic exercise and flexibility versus aerobic exercise
17. Aerobic, strength, mind-body and proprioception versus flexibility
18. Strength training versus mind-body exercise
19. Strength training versus biomechanical exercise
20. Strength training versus flexibility
21. Strength and flexibility versus flexibility
22. Strength and flexibility versus mind-body exercise
23. Strength, flexibility and proprioception versus mind-body exercise
24. Strength versus proprioception
25. Mind-body exercise versus flexibility
26. Mind-body exercise versus biomechanical exercise
27. Flexibility and proprioception versus flexibility
28. Flexibility and relaxation versus aerobic exercise
29. Exercise versus psychological therapies
30. Manual therapy and exercise versus manual therapy
31. Manual therapy and exercise versus exercise

32. Exercise versus manual therapy.

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.

1.4.2 Excluded studies

See the excluded studies list in appendix I.

1.4.3 Summary of clinical studies included in the evidence review

1.4.3.1 Aerobic exercise versus usual care

Table 2: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Andrade 2019[17]	16 week interventions. **Intervention 1: Aerobic exercise (n=27)** 32 aerobic pool sessions, 45 minutes each, twice a week. Conducted in groups of 5 and supervised by three physiotherapists. Progression of exercises was adjusted throughout in order to maintain optimum heart rate and reach the established perceived exertion threshold for each participant. **Intervention 2: Usual care (n=27)** No treatment; no further details	Women with fibromyalgia (n=54) Mean age 47.5(8) years Mean pain duration 7.5 years	At 16 weeks (post-intervention): • Quality of life • Pain reduction • Psychological distress • Sleep • Discontinuation	
Da costa 2005[67]	12 week interventions. **Intervention 1: Aerobic exercise (n=39)** Meeting four times with an exercise physiologist. Visits were 90 minutes with 30 minute follow ups. Exercises were individualised for each participant and following the American college of sports medicine guidelines. Exercise focused mainly on aerobic fitness with exercises at heart rate intensity of 60-70% initially then to 75-85% depending on progress, and duration of exercise depended on the intensity although the guidelines suggested individuals should perform 60-120minutes per week. Stretching and strength exercises were also prescribed with the amount depending on the needs of each participant. Participants were provided with a heart rate monitor.	Women with fibromyalgia (n=80) Mean age 51.2 years Mean pain duration 11 years	At 12 months follow up (including 3 months intervention): • Quality of life	

Study	Intervention and comparison	Population	Outcomes	Comments
	Intervention 2: Usual care (n=41) Usual care control group			
Gowans 2001 [113] (Gowans 2002[110])	23 week interventions. **Intervention 1: Aerobic exercise (n=27)** Water walking/running progressing to land walking/running. Classes for the first 6 weeks were conducted in a warm therapeutic pool; then progressed to 2 walking classes in a gym and 1 pool class. Classes were three times per week for 30 minutes (5 minutes stretching, 20 minutes aerobic activity, and 5 minutes stretching). Designed to generate a heart rate of 60-75% of age adjusted maximum heart rate. **Intervention 2: Usual care (n=23)** Continue ad libitum activity.	Fibromyalgia (n=50) Female:Male: 44:6 Mean age: 44.6 (8.7); 49.8 (7.3) years Duration of pain: 9.6 (8.6); 8.4 (7.6) years	At 23 weeks (post intervention): • Quality of life • Physical function • Psychological distress • Discontinuation	In Cochrane review (Bidonde 2017)
Kayo 2011 [142]	16 week interventions. **Intervention 1: Aerobic exercise (n=30)** Supervised indoor or outdoor walking, three times a week for 60 minutes (5-10 minutes stretching, walking and 5 minutes cool down). **Intervention 3: Usual care (n=30)** Control conditions not specified. Participants in all 3 groups were asked to discontinue tricyclic antidepressants but were allowed to use acetaminophen (paracetamol) for pain.	Fibromyalgia (n=60) All female Mean age: 47.7 (5.3); 46.7 (6.3); 46.1 (6.4) years Duration of pain: 4.0 (3.1); 4.7 (5.7); 5.4 (3.5) years	At 28 weeks (follow up, including 16 weeks intervention): • Quality of life • Pain • Physical function • Discontinuation	In Cochrane review (Bidonde 2017)
King 2002 [152]	12 week interventions. **Intervention 1: Aerobic exercise (n=42)**	Fibromyalgia (n=170; third arm of study reported under exercise versus	At 24 weeks (follow up including 12 week intervention): • Quality of life	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	Walking, aquacise (deep and shallow water), or low impact aerobics. Three times a week starting at 10-15 minutes and progressing to 20-40 minutes. **Intervention 2: Usual care (n=34)** Waitlist control. Participants received written instructions for basic stretches and 5 items related to general coping strategies.	psychological therapy comparison) Females only Mean age: 45.2 (9.4); 44.9 (10); 47.4 (9); 47.3 (7.3) years Duration of pain: 7.8; 10.9; 8.9; 9.6 years	• Physical function • Pain	
Mengshoel 1992 [189]	20 week interventions. **Intervention 1: Aerobic exercise (n=11)** Modified low-impact aerobic dance; exercise for upper extremities performed at intervals between periods of rest; exercises modified to prevent pain, fatigue, and static muscle work. Twice a week for 60 minutes. **Intervention 2: Usual care (n=14)** Participants instructed to not change their habits regarding physical activities.	Fibromyalgia (n=25) All female Mean age: 33.5 (21 to 42); 34 (25 to 38) years Duration of pain: 8.5 (3 to 20), 8 (3 to 23) years	At 20 weeks (post intervention): • Pain • Discontinuation	In Cochrane review (Bidonde 2017)
McBeth 2012[181] (Beasley 2015[28])	6 month intervention **Intervention 1: Aerobic exercise (n=109)** Gym based programme with monthly assessments led by instructors to reassess the programme. Exercise intensity increased until exercise levels achieved 40-85% maximum heart rate; recommended session length 20 to 60 minutes 3-5 times a week). **Intervention 3: Usual care (n=109)**	Chronic widespread pain (n=330; third arm of study reported under exercise versus psychological therapy comparison) Mean age 55.7(12.5) years	At 9 months: • Quality of life • Sleep • Discontinuation (6 months)	Gym sessions were not supervised (70% finished the exercise intervention, those that finished reached the compliance threshold of at least 2 sessions per week. 16.2% didn't complete sessions other than the

Study	Intervention and comparison	Population	Outcomes	Comments
	Usual care from family physician, although precise care delivered, if any, was not recorded	Duration of pain not stated		monthly fitness instructor sessions.
Nichols 1994 [204]	8 week interventions. **Intervention 1: Aerobic exercise (n=10)** Fast paced walking on an indoor track. Each session included a warm up and cool down regimen of stretching exercises, 1 warm up and cool down lap of slow paced walking. Three times a week. **Intervention 2: Usual care (n=9)** Daily activities as usual not involving physical activity.	Fibromyalgia (n=19) Female:Male: 17:2 Mean age: 47.8 (11.1); 50.8 (11.8) years Duration of pain: > 10; > 10 years except for a person who had 4	At 8 weeks (post intervention): • Discontinuation	In Cochrane review (Bidonde 2017)
Norouzi 2019 [206]	12 week interventions. **Intervention 1: Aerobic exercise (n=40)** Half of participants took part in walking on a treadmill. Walking was at an intensity of 60-75% estimated maximum heart rate. The other half of particpants took part in Zumba dancing. Each session consisted of a warm up followed by active upper and lower body movements, followed by a cool down and stretching. Three times a week for 60 minutes. **Intervention 2: Usual care (n=20)** Current daily activity levels were maintained and participants were asked to refrain from additional exercise or sport activities.	Fibromyagia (n=60) All female Mean age: 35.5 (2.42); 35.4 (2.80) years Duration of pain: 2.28 (0.3); 2.83 (0.29) years	At 12 weeks (post intervention) • Psychological distress • Physical function • Discontinuation	3 armed trial; 'aerobic exercise' arm and 'Zumba dancing' arm combined for analysis
Sanudo 2010 [234]	24 week interventions. **Intervention 1: Aerobic exercise (n=22)** Warm-up included slow walks, easy movements of progressive intensity, steady state aerobics included continuous walking with arm movements and jogging, interval	Fibromyalgia (n=64 ; third arm of study reported under aerobic and strength versus aerobic comparison)	At 24 weeks (post intervention): • Pain • Quality of life • Physical function	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	training included aerobic dance and jogging, cool-down included slow walks, easy movements, relaxation training. Twice a week for 45-60 minutes (10 minutes warm-up, 5-10 minutes cool down, 15-20 minutes steady aerobics, 15 minutes interval training). **Intervention 3: Usual care (n=21)** Medical treatment for fibromyalgia and continued normal daily activities, which did not include structured exercise.	Females only Mean age: 55.9 (1.6); 55.9 (1.7); 56.6 (1.9) years Duration of pain: not specified	• Discontinuation (additional outcome)	
Sanudo 2015[232]	24 week interventions **Intervention 1: Aerobic exercise (n=16)** Two sessions per week of 45-60 minutes duration. Each session included 10 minutes of warm up activities (easy movements and slow walking), 15-20 minutes of steady state exercise at 60-65% of predicted maximum heart rate (including continuous walking with arm movements and jogging) and 15 minutes of interval training at 75-80% (six repetitions of 1.5 minutes with 1 minute interpolated rest intervals), and 5-10 minutes of cool-down activities (slow walks, easy movements, relaxation training). Exercise intensity was monitored by a heart rate telemetric system. The intensity progressively increased as participants improved their exercise capacity to maintain the heart rate in the prescribed range. **Intervention 2: Usual care (n=16)** Participants continued their normal daily activities which did not include structured exercise.	Women with fibromyalgia (n=32) Mean age 56.5 years Mean pain duration not stated	At 24 weeks (post-intervention): • Pain reduction • Psychological distress • Sleep • Discontinuation	
Schachter 2003 [239]	16 week interventions. **Intervention 1: Aerobic exercise (n=51)** Home programme of low impact aerobics (long bout) with rhythmical movements designed to use all major muscle groups of the lower extremities performed to music. Three-	Fibromyalgia (n=143) Females only	At 16 weeks (post intervention): • Quality of life • Pain • Physical function	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	five times a week for 10-30 minutes, increasing in intensity over the first 10 weeks. **Intervention 2: Aerobic exercise (n=56)** Home program of low-impact aerobics (short bout) to videotaped instructor and music, rhythmical movements of lower body muscles. Three to five times a week, twice a day for 5-15 minutes, increasing in intensity over the first 10 weeks. *NB Aerobic exercise interventions pooled in the analysis.* **Intervention 3: Usual care (n=36)** Participants were asked to refrain from starting any new regular physical activity or exercise programs or other non-pharmacological interventions.	Mean age: 41.3 (8.7); 41.9 (8.6); 42.5 (6.7) years Duration of pain: not specified	• Psychological distress	
Sencan 2004 [241]	6 week interventions. **Intervention 1: Aerobic exercise (n=20)** Supervision unclear. Cycle ergometry 3 times a week for 40 minutes. **Intervention 2: Usual care (n=20)** Placebo group received sham transcutaneous electrical stimulation 3 times a week for 20 minutes each; electrodes applied on the 2 most painful tender points (with no current)	Women with fibromyalgia (n=60) Mean age 35.4 years Mean duration of pain 4.7 years	At 6 weeks post intervention and 26 weeks follow up: • Pain reduction	
Van eijk-hustings 2013 [264]	12 week interventions. **Intervention 1: Aerobic exercise (n=47)** Sessions twice a week by a trained physiotherapist in a community gym (groups of 9 to 10 participants). Every session started with a 10-min warm up, comprising aerobic and stretching, followed by 30 minutes of aerobic exercise.	Fibromyalgia (n=96*) Mean age 42 years Mean duration of pain not reported	At 12 weeks (post-intervention) and 18 months (follow-up): • Pain reduction • Quality of life • Physical function	*Third arm of RCT included in pain management programme evidence review.

Study	Intervention and comparison	Population	Outcomes	Comments
	The low- intensity aerobic part aimed to reach 55–64 % of the predicted maximum heart rate. Then, resistance training was applied during 15 min to strengthen major muscle groups. Finally, every session was finished with a 5-min cool down. Participants received a digital video disc presenting exercises to do at home, and they were advised to perform these once a week. **Intervention 2: Usual care (n=48)** Usual care involved GP appointments and at least some individualised education about fibromyalgia.		• Psychological distress • Use of healthcare services • Sleep • Discontinuation	
Wigers 1996 [276]	14 week interventions. **Intervention 1: Aerobic exercise (n=20)** Aerobic exercise, focusing on the whole body and aimed at minimizing eccentric muscle strain. Exercise involved movement to music and games. Three times a week for 45 minutes (23 minute music session including warming up and 2 peaks of high intensity training, 15 minutes of aerobic games with 2 high intensity periods). **Intervention 2: Usual care (n=20)** Continued treatments being used at baseline.	Fibromyalgia (n=40) Mean age: 43 (9); 44 (12); 46 (9) years Duration of pain: 9 (5); 11 (10); 11 (9) years	At 14 weeks (post intervention) and 4 years (follow-up): • Pain • Sleep • Psychological distress • Discontinuation (additional outcome)	In Cochrane review (Bidonde 2017)

1.4.3.2 Strength training versus usual care

Table 3: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Assumpcao 2018 [23]	12 week interventions **Intervention 1: Strength training (n=19)** 12 week supervised resistance training programme of 40-minute sessions performed twice a week with progressive overload.	Women with fibromyalgia (n=35) Mean age 47 years	At 12 weeks (post-intervention): • Pain reduction • Physical function • Discontinuation	60% were taking concomitant medication for fibromyalgia (antidepressants, analgesics, anti-

Study	Intervention and comparison	Population	Outcomes	Comments
	Equipment included dumbbells, shin pads. No load was used in the first 2 sessions, after which time 0.5kg was added each week if the patient identified the effort as slightly intense on the Borg scale. 8 repetitions for: triceps, quadriceps, hip adductors and abductors, hip flexors, elbow flexors and extensors, pectoralis major and rhomboids. **Intervention 2: Usual care (n=16)** After 12 weeks patients were reassessed and offered physical therapy based on stretching and resistance training.	Mean pain duration not stated		inflammatories or psychotropic medication)
Borisut 2013[39]	12 week interventions **Intervention 1: Strength exercise (n=25)** A progressive resistance exercise program for the neck muscles, especially the superficial neck flexor and extensor muscles. Neck flexion and extension were performed in the supine and prone positions. The first 4 weeks involved 12-15 repetitions, and the next 8 weeks involved 3 sets of 15 repetitions. **Intervention 2: Strength exercise (n=25)** A craniocervical flexion exercises which consisted of a low load exercise for the cranio-cervical flexor muscles. Participants moved to increase air pressure on a sensor and held for 10 seconds in 15 repetitions. **Intervention 3: Strength exercise (n=25)** A combination of progressive resistance exercise and craniocervical flexion exercise. *NB Strength exercise interventions pooled in the analysis.* **Intervention 2: Usual care (n=25)**	Women with chronic neck pain (n=100) Mean age: 31.1 (3.38); 30.40 (3.54) Mean pain duration not reported	At 12 weeks: • Pain • Physical function	

Study	Intervention and comparison	Population	Outcomes	Comments
	After the data collection period, participants were advised to perform both the strength progressive resistance and cranio-cervical exercises			
Chiu 2005[59]	**Intervention 1: Strength training (n=67)** There were 2 training sessions per week for a period of 6 weeks. The exercise program began with a warm up which involved one set (10 minutes) of activation of the deep neck, then 15 repetitions of flexion and extension of the neck. The resistance used during the warm up was set at approximately 20% of the maximum intensity. After the warm up, dynamic training started, which consisted of 3 sets of variable resistance load allowing 8-12 repetitions of full flexion and extension within pain tolerance. A 5 minute rest between sessions was given. The weight load was increased approximately 5% when a set of 12 or more repetitions had been achieved. **Intervention 2: Usual care (n=78)** The control group received infrared irradiation twice a week for 6 weeks. The irradiation time was 20 minutes.	Chronic neck pain for longer than 3 months (n=145) Mean age 43.3 years 61% had pain for over 12 months	At 6 weeks (post intervention): • Pain reduction • Physical function • Discontinuation	Infrared irradiation was given to both the exercise group and the control group. For the exercise group, irradiation was given before the exercise program.
Falla 2013[90]	8 week interventions **Intervention 1: Strength training (n=23)** Progressive exercise programme for the neck flexors and extensor muscles. Participants received personal instruction and supervision by a physiotherapist for 30 minutes once per week for 8 weeks. The therapist examined the exercises and progressed the participant if appropriate. The programme consisted of 2 stages. The first stage was 6 weeks duration. The principal exercise task during this period was flexion in a relaxed supine lying position and patients were guided by a pressure unit. The second stage was 2 weeks and involved higher load exercise with head weight as the load. During this stage, participants performed up to 15 repetitions of a head lift for flexors and neck extension for the extensor group. Participants	Chronic non-specific neck pain (n=46) Mean age 38.9 years Mean duration of pain 9.1 years	At 8 weeks (post-intervention): • Pain reduction • Quality of life • Physical function • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	practiced twice per day, and the programme was 10-20 minutes/day. **Intervention 2: Usual care (n=23)** The control group did not receive any intervention, however they patients were not asked to refrain from seeking treatment.			
Glasgow 2017[107]	8 week interventions **Intervention 1: Strength training (n=14)** Supervised resistance exercises twice a week for 8 weeks, each lasting 30 minutes. 3 sets of 8-12 repetitions followed by 90 second rest periods between each set. Exercises were chest presses, leg extensions, leg curls and seated rows, initially at a training intensity of 50-60% of maximum. Resistance was increased when participants could complete 12 repetitions on all 3 sets over 2 consecutive training days. **Intervention 2: Usual care (n=12)** Control group (non-exercising, no further details).	Women with fibromyalgia (n=26) Mean age 51 years Mean pain duration not specified	At 8 weeks (post-intervention): • Quality of life • Psychological distress • Discontinuation	
Hakkinen 2001 [119]	21 week interventions. **Intervention 1: Strength training (n=11)** Resistance training including 6-8 dynamic resisted exercises using David 200 dynamometer to upper extremity, lower extremity, and trunk muscle groups. Twice a week. **Intervention 2: Usual care (n=10)** Controls maintained their normal low-intensity recreational physical activities but did not participate in the strength training.	Fibromyalgia (n= 21) All female Mean age: 37 (6) to 39 (6) years Duration of pain: 12 (4) years	At 21 weeks (post intervention): • Pain • Sleep • Physical function • Psychological distress	In Cochrane review (Busch 2013)
Kayo 2011 [142]	16 week interventions. **Intervention 1: Strength training (n=30)**	Fibromyalgia (n=60) All female	At 28 weeks (follow up, including 16 weeks intervention): • Quality of life	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	Supervised exercise protocol consisting of 11 free active exercises for upper and lower limbs and trunk muscles, with free weights and body weight performed in the standing, sitting, and lying positions. Sessions were three times a week for 60 minutes. Exercise load and intensity increased every 2 weeks. **Intervention 2: Usual care (n=30)** Control conditions not specified. Participants in all 3 groups were asked to discontinue tricyclic antidepressants but were allowed to use acetaminophen (paracetamol) for pain.	Mean age: 47.7 (5.3); 46.7 (6.3); 46.1 (6.4) years Duration of pain: 4.0 (3.1); 4.7 (5.7); 5.4 (3.5) years	• Pain	
Kingsley 2005[153]	12 week interventions **Intervention 1: Strength training (n=15)** Twice a week sessions for 30 minutes. Sessions consisted of 11 exercises. Resistance machine exercises included chest press, leg extension, standing leg curl, shoulder press, lumbar extension and abdominal crunch. The cable exercises included low pulley biceps curl, high pulley triceps extension, and the mid pulley standing row. Body weight was used for the standing calf raises and body weight Swiss ball squats. Before and after workouts, participants performed 5 minutes of warm up and cool down that included stretching and walking. Participants began training at 40% of their 1-RM. Once 12 repetitions were performed in proper form, weight was increased by 2.3 to 4.5kg (5-10lb). **Intervention 2: Usual care (n=14)** Participants were asked not to change their activity levels during the 12 week intervention period.	Women with fibromyalgia (n=29) Mean age 46.2 years Mean pain duration 8 years	At 12 weeks (post-intervention) • Quality of life • Physical function • Discontinuation	
Suvarnnato 2019[247]	6 week interventions **Intervention 1: Strength training (n=18)**	Chronic neck pain (n=54)	At 6 weeks (post-intervention) and 16 weeks (follow up):	

Study	Intervention and comparison	Population	Outcomes	Comments
	Semispinalis cervicis-training group. Exercises involved a physical therapist applying resistance to the posterior vertebral arches of the participant's C2 vertebra whilst participants pushed against the resistance. Exercises were held for 10 seconds, 10 times per set, 3 sets per day (30 second rest between sets). Exercises performed twice per week over the 6 week period. **Intervention 2: Strength training (n=18)** Deep cervical flexor-training group. Low-load exercises focused on activating the deep flexor muscles of the cervical region. Exercises performed 10 times per set, 3 sets at a time with a 30 second rest between sets. Performed under supervision twice per week and advised to perform twice per day at home. *NB Strength training interventions pooled in the analysis* **Intervention 3: Usual care (n=18)** Usual care deemed appropriate by physical therapists other than strength exercises, e.g. stretching, manual therapy. 10-12 appointments within 6 weeks.	Mean age 42.94 years Mean duration of pain 12.86 months	• Pain reduction • Physical function	
Valkeinen 2004 [260]	21 week interventions. **Intervention 1: Strength training (n=13)** Resisted dynamic exercise to knee extensors x 2 plus 5-6 exercises for other main muscle groups of body. Twice a week for 60-90 minutes. **Intervention 2: Usual care** Control conditions were treatment as usual and physical activity as usual.	Fibromyalgia (n-26) All females Mean age: 59.1 (3.5) to 60.2 (2.5) years Duration of pain: 8.5 (4.3) to 6.6 (4.1) years	At 21 weeks (post intervention): • Physical function • Discontinuation	In Cochrane review (Busch 2013)
Viljanen 2003[267]	12 week interventions **Intervention 1: Strength training (n=135)**	Chronic non-specific neck pain (n=393; third arm of study reported under	At 12 months follow up (including 12 week intervention):	All participants were office workers

Study	Intervention and comparison	Population	Outcomes	Comments
	Led by trained physiotherapist 3 times a week sessions for 30 minutes each, followed by on week of reinforcement training 6 months after randomisation. Dumbbells were used for dynamic muscle training (weight 1-3kg each according to maximum repetitions with a test weight of 7.5 kg). The Exercises, conducted in the same order in each session, were chosen to activate large muscle groups in the neck and shoulder region. After the 5thweek participants were taught 3 exercises from the program with stretches, after the 9th week they were asked to perform the full training program by themselves. **Intervention 2: Usual care (n=130)** Usual care, no change to physical activity or means of relaxation during the 12 months of follow up.	exercise versus psychological therapy comparison) Mean age 44 years Mean pain duration 10.8 years	• Pain reduction • Discontinuation	
Von trott 2009[271]	12 week interventions **(Intervention 1: Strength and flexibility n=39)** 24 sessions at 45 minutes each held over 12 weeks, with 6-12 participants in each group. A standardised programme for computer and workplace related neck pain. It included repeated active cervical rotations as well and strength and flexibility exercises. Special intention as paid so that the patients' individual pain limits were not exceeded. About 90% of the exercises were repeated in each lesion; some 10% was exchanged regularly **Intervention 2: Usual care (n=40)** Waiting list control participants did not receive Qigong or exercise therapy.	Office workers with chronic neck pain (n=79) Mean age 76 years Mean pain duration 18.6 years	At 12 weeks (post-intervention) and 24 weeks follow up: • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	

1.4.3.3 Aerobic exercise and strength training versus usual care

Table 4: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Espi-lopez 2016[83]	8 week interventions. **Intervention 1: Aerobic, Strength training (n=13)** Low-impact aerobic exercise with low impact strength exercises. Two sessions per week. Each session consisted of 60min and was divided into three parts: warm up (15 min); games, group dynamics and aerobics (30 min); and cool down with stretching for 15 min. The warm up consisted of combined low impact aerobic exercises, free range of motion exercises of limbs and spine, and coordination exercises plus stretching. This was followed by active low load resistance exercises involving arms and legs, followed by a circuit of coordination and agility exercises and then low-impact strength exercises of the trunk. This was followed by a cool down with stretches. **Intervention 2: Usual care (n=9)** No intervention, no further details	Fibromyalgia (n=22) Mean age 53.6(8.1) years Mean pain duration not stated	At 8 weeks (post-intervention): • Quality of life • Psychological distress • Discontinuation	
Etnier 2009[84]	18 week interventions. **Intervention 1: Aerobic, Strength exercise (n=8)** The exercise sessions were 60 minutes in duration 3 days a week. During the sessions, participants walked, performed light resistance exercises, and performed static bridging and stretching exercises. All sessions were conducted and directly supervised by one of the authors. In terms of the walking portion, participants were encouraged to walk a comfortable/brisk pace (55-65% of maximal heart rate reserve) for 15 minutes. Over the course of the intervention, they were encouraged to try to walk a greater distance in the 15 minute period and used this as a self-measure of aerobic	Women with fibromyalgia (n=16) Mean age not reported Mean duration of pain not reported	At 18 weeks (post-intervention): • Quality of life • Physical function • Psychological distress • Discontinuation	Most participants reported having symptoms as teenagers and received a medical diagnosis within the last 1-10 years.

Study	Intervention and comparison	Population	Outcomes	Comments
	fitness. In terms of the light resistance exercises, participants moved through an 8 station light resistance exercise circuit. When subjects were able to easily complete the required number of repetitions for a certain exercise, resistance was increased by 1 pound. Often, this caused participants to reduce the number of repetitions for a short time followed by slowly working back to the required number. Static-bridging exercises require that the exerciser support her body (holding the body very still) in various positions to increase core (abdominal, back and pelvic), muscle strength/endurance. Usually 10 repetitions of approximately 3 seconds were completed in each session. **Intervention 2: Usual care (n=8)** No treatment control condition.			
Izquierdo-Alventosa 2020 [132]	8 week interventions. **Intervention 1: Aerobic, Strength training (n=16)** Low intensity physical exercise combing endurance training (aerobic and low-load resistance exercises aimed at improving endurance) and coordination. Each session consisted of a warm up of walking at a slow pace (10-15 minutes), training which involved 10 exercises (25-40 minutes), and a cool down of walking, stretching, and breathing (10-20 minutes). Twice a week for 60 minutes. **Intervention 2: Usual care (n=16)** No treatment control condition.	Women with fibromyalgia (n=32) Mean age: 53.06 (8.4); 55.13 (7.35) years Mean pain duration not stated	At 8 weeks (post-intervention) • Pain reduction • Physical functioning • Psychological functioning • Quality of life • Discontinuation	
Latorre roman 2015[159]	18 week interventions. **Intervention 1: Aerobic, Strength training (n=20)** Sixty-minute sessions of functional training 3 times a week. Of those 3 weekly sessions, 2 consistent of exercise in water and 1 of exercise on land. A specialist instructed both groups.	Women with fibromyalgia (n=39) Mean age 51.7 years	At 18 weeks (post-intervention) • Pain reduction • Quality of life • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	Each session included a warm up (5 minutes) and exercises of muscular strengthening and balance (40 minutes), and a cool down (5 minutes). Exercise intensity was increased during the whole programme by modifying the number of reps per set, by introducing weights (in on land exercises, 0.5-2kg per exercise) and materials that raised the resistance offered by water. Strength training consisted in 1-3 sets of 8-12 reps per exercise and circuit training. On land, multiple functional exercises were performed individually and on a circuit, for example, climbing stairs using weights as the external load (medicine ball). **Intervention 2: Usual care (n=19)** Participants continued with their daily activities that did not include any kind of physical exercise similar to that of the study group.	Mean pain duration not stated		
Munguia-izquierdo 2007[200] (Munguia-izquierdo 2008[199])	16 week intervention **Intervention 1: Aerobic, Strength training (n=35)** The exercise group trained in a chest-high warm pool (32°C) 3 times a week for 16 weeks. Each session included 10 minutes of warming up with slow walks and mobility exercises, 10 to 20 minutes of strength exercises developed at a slow pace using water and aquatic materials as a means of resistance including a stepped progression during the program, 20 to 30 minutes of aerobic exercises developed progressively at intensity sufficient to achieve 50% to 80% of the age predicted maximum heart rate equation (220 – age), and 10 minutes of cooling down with low-intensity and relaxation exercises. Heart rate was monitored with a pulse meter. **Intervention 2: Usual care (n=25)**	Fibromyalgia (n=60) Mean age 48 years Mean pain duration 14 years	At 16 weeks (post-intervention): • Quality of life • Psychological distress • Sleep • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	The control group was instructed not to change their habits regarding physical activities during the period. Usual activities and medication allowed.			
Sanudo 2011[235]	24 week interventions **Intervention 1: Aerobic, Strength training (n=21)** Twice weekly sessions of combined aerobic and muscle strength training for 24 weeks. 10 minute warm up followed by 10-15 minutes of aerobic exercises at 65-70% of maximum heart rate. Participants were in small groups and performed continuous walking with arm movements and jogging. This was followed by 15-20 minutes of muscle strengthening exercises with a circuit of 8 exercises using multiple muscles. Participants carried out 1 set of 8-10 repetitions and resistance was increased according to the patient's tolerance. This was followed by a cool-down of 10 minutes which consisted of flexibility exercises. Duration 24 weeks. Concurrent medication/care: 81.25% were taking medication for FMS (analgesic or NSAID, antidepressant or other combination). **Intervention 2: Usual care (n=21)** Participants continued their usual treatment and daily activities which did not include any structured exercise.	Fibromyalgia (n=42) Mean age 55.87 years Mean pain duration not specified	At 24 weeks (post-intervention): • Quality of life • Psychological distress • Discontinuation	81.25-84.2% were taking concurrent medication for fibromyalgia
Sanudo 2012[233]	24 week interventions **Intervention 1: Strength training and aerobic exercise (n=21)** Exercise was twice weekly for 45-60 minutes. Each session included 10 minutes of warm up activities (slow walking and gently movements of progressive intensity e.g. arm swinging); 10-15 minutes of aerobic exercise at 65% to 70% of maximal heart rate, 15-20 minutes of muscle strengthening exercises (one set of 8-10 repetitions for 8 different muscle groups, with a load of 1-3kg), and 10 minutes of flexibility	Fibromyalgia (n=41) Mean age not reported Mean pain duration not reported	At 24 weeks (post-intervention): • Physical function • Psychological distress • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	exercises (1 set of 3 repetitions for 8-9 different exercises, maintaining the stretched position for 30 seconds). Strengthening and flexibility exercises focused on the main areas of pain in patients with FM (deltoids, biceps, neck, hips, back and chest). **Intervention 2: Usual care (n=20)** Usual medical treatment of fibromyalgia and continued normal daily activities which did not include structured exercise.			
Tomas-carus 2008[252] (Tomas-carus 2007[254], Tomas-carus 2009[253], [116])	8 month interventions **Intervention 1: Aerobic and strength exercise (n=18)** Supervised training in waist high pool of warm water 3 times per week during an 8 month period. Each session 1 hour, 10 minutes warming up with slow walks and easy movements of progressive intensity, 10 minutes of aerobic exercises (60-65% maximal heart rate), 20 minutes of strength exercises using water resistance (4 sets of 10 repetitions), 10 minutes of cooling down with low intensity exercises. **Intervention 2: Usual care (n=17)** Control group continuing daily activities which did not include any form of physical exercise similar to those in the therapy.	Women with fibromyalgia (n=34) Mean age 50.8 years Mean pain duration 19.8 years	At 3 months and 8 months (post-intervention): • Pain reduction • Quality of life • Psychological distress • Physical function • Psychological distress • Discontinuation	
Waling 2002[273]	10 week interventions **Intervention 1: Aerobic exercise (n=34)** Endurance training of the shoulder muscles consisted of arm-cycling and arm exercises with rubber band resistance on the endurance level (30 RM repetition maximum). **Intervention 2: Strength exercise (n=34)** Strength training consisted of neck and shoulder exercises with individualized loads of 10 to 12 maximal voluntary	Women with work-related trapezius myalgia Mean age: 37.7 (5.6); 31.1 (15.8) years Pain duration: 6.3 (4.0); 7.3 (4.34) years	At post-intervention: • Pain At 3 years (follow up): • Pain • Use of healthcare services	

Study	Intervention and comparison	Population	Outcomes	Comments
	contractions in three sets. *NB Aerobic and strength exercise interventions pooled in the analysis.* **Intervention 2: Usual care (n=27)** Participants, led by an occupational nurse, studied stress management once a week, 2 hours at a time, for 10 weeks. No exercises were performed in this group.			
Ylinen 2003[284] (Ylinen 2007[281], Ylinen 2006[285])	2 week interventions **Intervention 1: Strength training (n=60)** 10 patients in each group, 12 day program with 5 sessions per week, each lasting 45 minutes. Exercises aimed to strengthen neck flexor muscles by using an elastic rubber band to train the muscles at a resistance of 80% of maximum (15 repetitions in each direction). Following this the group performed dynamic exercises for the shoulders and upper extremities, with an individually adjusted single dumbbell, performing only 1 set for each exercise with the highest load possible to perform 15 repetitions. This was followed by exercises for the trunk and leg muscles in the same format, which was then concluded by stretching exercises for 20 minutes. **Intervention 2: Strength training (n=60)** 10 patients in each group, 12 day program with 5 sessions per week, each lasting 45 minutes. Exercises aimed to strengthen neck flexor muscles by lifting head up from the supine position in 3 series of 20 repetitions. Following this the group performed dynamic exercises for the shoulders and upper extremities, at 3 sets of 20 repetitions for each exercise with a pair of dumbbells each weighing 2 kg. This was followed by exercises for the trunk and leg muscles in	Office workers with chronic neck pain (n=180) Mean age 46 years Mean pain duration not stated (but minimum 6 months)	At 12 month follow up: • Use of healthcare services	

Study	Intervention and comparison	Population	Outcomes	Comments
	the same format, which was then concluded by stretching exercises for 20 minutes. *NB: Strength training interventions pooled in the analysis* **Intervention 3: Usual care (n=60)** Performed recreational activities on assessment days. Received written information about the same stretching exercises and were advised to practice these 20 minutes 3 times a week. They were also advised to perform aerobic exercise 3 times a week.			

1.4.3.4 Aerobic exercise, Strength and flexibility versus usual care

Table 5: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Garcia-martinez 2012[99]	12 week interventions **Intervention 1: Aerobic, strength and flexibility exercise (n=14)** 3 times a week sessions for 12 weeks. Each session was 60 min long and included 10 min of warming-up with slow walks and easy movements of progressive intensity, 20 min of aerobic exercise that began at 60–70% of maximal heart rate and was gradually increased to as high as 75–85% maximum, depending on the subjects' adaptation, 20 min of stretching and strength exercise and 10 min of cooling down with low-intensity exercises. **Intervention 2: Usual care (n=14)**	Fibromyalgia (n=28) Mean age 58.9 years Mean duration of pain 10.3 years	Quality of life at 12 weeks (post-intervention)	

Study	Intervention and comparison	Population	Outcomes	Comments
	Subjects continued their daily activities which did not include any physical exercise.			

1.4.3.5 Strength and flexibility combination versus usual care

Table 6: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Acar 2012 [1]	2 week intervention **Intervention 1: Strength and stretching combination (n=20)** Strength exercises for multiple muscles and neck stretching exercises. 10 sessions 5 days a week, supervised by physiotherapists. **Intervention 2: Usual care (n=20)** No details.	Chronic cervical pain (n=40) Mean age 38(11.75) years Mean pain duration 46.5 years	Pain reduction at 2 weeks (post-intervention)	
Rendant 2011[222]	6 month interventions **Intervention 1: Strength and flexibility (n=39)** Exercise therapy was carried out by 6 qualified therapists. The exercises were based on a standard programme for chronic pain. Each lesson started with a warm up using a softball and was followed by repeated active cervical rotations and strengthening and flexibility exercises. The individual's pain level was not exceeded. There were 18 sessions over a period of 6 months (1 session per week in the first 3 months, and biweekly sessions in the following 3 months). **Intervention 2: Usual care (n=41)** Waiting list control participants received no intervention.	Chronic non-specific neck pain (n=123; third arm of study reported under strength and flexibility versus mind-body and mind-body versus usual care comparisons) Mean age 44.6 years Mean pain duration 3.1 years	At 6 months (post-intervention) • Pain reduction • Quality of life • Physical function • Discontinuation	Pain rating of 40 or more required at baseline (VAS 0-100) Third arm of study reported under separate comparisons (Qi-gong).

1.4.3.6 Strength, proprioception and flexibility versus usual care

Table 7: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Lauche 2016[160]	12 week interventions **Intervention 1: Strength, proprioception and flexibility (n=37)** Participants in the neck exercise group met once weekly for a 60- to 75-minute session for 12 weeks in total. This group was instructed in neck exercises, which were similar to those taught in rehabilitation programs containing exercises and education for a healthy back. Classes contained basic training of ergonomic principles (bodily alignment while standing), proprioceptive exercises, and isometric and dynamic mobilization, stretching, and strengthening neck and core exercises. The sessions opened with 5 to 10 minutes of warm-up exercises and ended with relaxation exercises. Participants also received illustrated and written information that covered the most important exercises, and they were asked to execute the exercises for at least 15 minutes each day. **Intervention 2: Usual care (n=39)** Participants in this group were advised to continue their usual activities and therapies, but not to initiate any new therapeutic regimen for symptom management.	Chronic non-specific neck pain (n=114; third arm of study reported under mind-body versus usual care and strength, proprioception and flexibility versus mind-body comparisons) Mean age 48.49 years Mean pain duration not specified	At 12 weeks (post-intervention) and 24 weeks (follow up): • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	VAS score of 45 or higher (0-100) inclusion criteria.

1.4.3.7 Proprioception versus usual care

Table 8: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Altan 2004 [9]	12-week interventions **Intervention 1: Proprioception (pool-based) (n= 24)** All patients were given two educational sessions of 1 h each for 2 days by a physiatrist about the description and available diagnosis and treatment methods of FMS. Next, they were assigned randomly into two groups by the researcher other than the one who performed the evaluation throughout the study. In group 1, a pool-based exercise program was given by a physiotherapist to 25 patients in a therapeutic pool at 37°C for 35 min a day three times a week for 12 weeks. The program included warming (walking back and forth in the pool), activity (jumping in the pool and active joint motion range and stretching of the neck and the extremities), relaxation (lying supine on the water and slow swimming), and out-of-pool exercises (bending back and forth, squatting, and relaxing with deep breaths) for a period of 35 min. **Intervention 2: Usual care (n=22)** Warm balneotherapy pool sessions of 35 minutes 3 times a week for 12 weeks.	Fibromyalgia Mean 43.5 (6.32) years, 43.91 Duration of pain not described	At 12 weeks (post-intervention) and 24 weeks follow up: • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	

1.4.3.8 Mind-body versus usual care

Table 9: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Baptista 2012 [27]	16 week interventions: **Intervention 1: Mind-body exercise (n=40)** 1 hour belly dance class twice a week for 16 weeks. Each class had a maximum of 8 students and was led by physiotherapists. Classes began with warm up, followed by movements for the day, choreography and a cool-down exercise. Participants also received a disc with music and an exercise book with all movements for the programme. From the 4th week a set sequence of movements in the form of choreography was established for training at home. **Intervention 2: Usual care (n=40)** Offered intervention at the end of study.	Women with fibromyalgia (n=80) Mean age 49.3 years Pain duration not stated	At 32 weeks (follow up, including 16 week intervention): • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	
Bojner-Horwitz, 2003 [38]	12 week interventions. **Intervention 1: Mind-body exercise (n=20)** Dance and movement therapy consisted of four main themes including; awareness of the body; movement expressions; movement, feeling, image; and differentiation of feelings and integration 1 hour session, held weekly for 6 months. **Intervention 2: Usual care (n=16)** Participants received the intervention on completion of the study.	Women with fibromyalgia (n=36) Mean age 57 years Duration of pain not stated	Discontinuation at 6 months	
Carson 2010 [53]	8 week interventions. **Intervention 1: Mind-body exercise (n=25)** Yoga consisted of 2 hour sessions, held weekly for 8 weeks in a group based format led by a certified, experienced yoga teacher. The intervention included meditation, breathing exercises, study	Fibromyalgia (n=53) All females Mean age: 53.7 (SD 11.5) years	At 8 weeks (post intervention): • Quality of life • Physical function (additional outcome)	In Cochrane review (Theadom 2015)

Study	Intervention and comparison	Population	Outcomes	Comments
	of the application of yoga principles to optimal coping and gentle stretching poses and group discussions. **Intervention 2: Usual care (n=28)** Wait list.	Duration of pain: not reported	• Discontinuation (additional outcome)	
Carson 2012 [54]	8 week interventions. **Intervention 1: Mind-body exercise (n=25)** Yoga delivered within group sessions by a certified yoga instructor 120 minute sessions, delivered weekly over 8 weeks. **Intervention 2: Usual care (n=28)** Wait list.	Fibromyalgia (n=53) All females Mean age: not reported Duration of pain: not reported	At 8 weeks (post-intervention): • Quality of life • Pain (additional outcome) • Discontinuation (additional outcome)	In Cochrane review (Theadom 2015)
Haak 2008[118]	7 week interventions **Intervention 1: Mind-body exercise – Qigong (n=29)** Total Qigong time 711.5 hours. Participants were instructed to practice Qigong at home with the support of a free instruction tape, twice a day for 20 minutes. Supervisors of the intervention were experienced Qigong masters. The sessions included internal and external methods of Qigong (influenced by oneself and influenced by the Qigong master). **Intervention 2: Usual care (n=28)**	Women with fibromyalgia (n=57) Mean age 53 years Mean duration of symptoms 15 years	At 7 weeks (follow up, including 4 week intervention): • Pain reduction • Quality of life • Psychological	
Holmer 2004 [124]	12 week interventions. **Intervention 1: Mind-body exercise -Yoga (n=11)** Delivered by a certified yoga instructor. No further details **Intervention 2: Usual care (n=17)** No further details.	Fibromyalgia (n=28) Age range 18 to 65 years Pain duration not specified	At 12 weeks (post-intervention): • Pain • Physical function • Psychological distress • Sleep	

Study	Intervention and comparison	Population	Outcomes	Comments
Lauche 2016[160]	12 week interventions **Intervention 1: Mind-body exercise - Tai Chi (n=38)** Participants in the Tai Chi group met once weekly for a 75- to 90-minute session. The Tai Chi intervention was on the basis of a popular and internationally recognized Yang style (13 forms from Mantak Chia). Each session included a warm-up of 5 to 10 minutes, the Tai Chi form practice, and 5 to 10 minutes of relaxation at the end. Tai Chi forms followed explicit protocols outlined in a training manual, as required during teacher training certification. Sessions also included educational units and breathing exercises, and they were accompanied by relaxation music. Participants received illustrated written information that covered movement sequences learned in the previous session. They were asked to practice Tai Chi outside of classes for at least 15 minutes each day. **Intervention 2: Usual care (n=39)** Participants in this group were advised to continue their usual activities and therapies, but not to initiate any new therapeutic regimen for symptom management.	Chronic non-specific neck pain (n=114; third arm of study reported under strength, proprioception and flexibility versus mind-body and strength, proprioception and flexibility versus usual care comparisons) Mean age 50.94 years Mean pain duration not stated.	At 12 weeks (post-intervention) and 24 weeks (follow up): • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	VAS score of 45 or higher (0-100) inclusion criteria. .
Liu 2012 [166]	6 week interventions. **Intervention 1: Mind-body exercise (n=7)** Qi-gong delivered in a group based format with home practice in between sessions 15 to 20 minute sessions, held weekly for 6 weeks. **Intervention 2: Usual care (n=7)** Sham qi-gong delivered in a group based format with no meditation or healing sounds 15 to 20 minute sessions, held weekly for 6 weeks.	Fibromyalgia (n=14) Sex not reported Age: 18-70 years Duration of pain: not reported	At 6 weeks (post-intervention): • Discontinuation	In Cochrane review (Theadom 2015) Query sham qi-gong

Study	Intervention and comparison	Population	Outcomes	Comments
Lynch 2012 [172]	8 week interventions. **Intervention 1: Mind-body exercise (n=53)** Qi-gong delivered by a psychologist in a group based format in the community 3.5 day workshops held weekly with additional refresher sessions. **Intervention 2: Usual care (n=47)** Wait-list control.	Fibromyalgia (n=100) Sex not reported Age: not reported Duration of pain: not reported	At post-intervention (8 weeks) and 6 month follow-up: • Pain • Discontinuation (additional outcome)	In Cochrane review (Theadom 2015)
Mannerkorpi 2004 [174]	14 week interventions. **Intervention 1: Mind-body exercise (n=19)** Qi-gong + relaxation, 14 group sessions of 1.5 hours, were held weekly, delivered by a physiotherapist. The treatment included various breathing, relaxation and concentration techniques conducted in a supine or standing position including qi-gong movements. The movements were individually modified to match the functional limitations of the patients and there was an opportunity for discussion about the movements with the therapist. Participants were encouraged to practice the movements in between sessions. **Intervention 2: Usual care (n=17)** No further details.	Fibromyalgia (n=36) All females Age: 18-65 years Duration of pain: not reported	At 14 weeks (post intervention): • Quality of life • Physical function • Discontinuation (additional outcome)	In Cochrane review (Theadom 2015)
Michalsen 2012[192]	9 week interventions **Intervention 1: Mind-body exercise – Yoga (n=38)** Weekly 90 minute yoga classes using a wide range of postures to enhance flexibility, alignment, stability and mobility in muscles joints and tendons, run by a certified yoga instructor and physician. The exercises specifically addressed neck pain complaints and each class built up on the previous one. Subjects	Chronic non-specific neck pain (n=77) Mean age 47.9 years Mean pain duration 6.55 years	At 10 weeks (post-intervention) • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	Pain score of at least 4 on VAS 0-10 scale.

Study	Intervention and comparison	Population	Outcomes	Comments
	were requested to practice at home for 10-15 minutes, 2 to 3 times a week. **Intervention 2: Usual care (n=39)** Waiting list control. A standard self-care manual about exercise and education for chronic neck pain was given. The manual described exercises that could be carried out to aid chronic neck pain and participants were asked to practice at home for 10-15 minutes at least 3 times a week.			
Rendant 2011[222]	6 month interventions **Intervention 1: Mind-body exercise – Qigong (n=42)** Qigong was performed by three qualified teachers certified by the German Qigong Society. Each session of qigong took 90 minutes. Neiyanggong, a special silent and slow form of qigong was chosen by the therapist in a consensus process. The lessons started with up to 12 neck exercises followed by 9 exercises for the shoulder and finished with breathing and moving exercises. There were 18 sessions over a period of 6 months (1 session per week in the first 3 months, and biweekly sessions in the following 3 months) **Intervention 2: Usual care (n=41)** Waiting list control participants received no intervention.	Chronic non-specific neck pain (n=123; third arm of study reported under strength and flexibility versus usual care and strength and flexibility versus mind-body comparisons) Mean age 44.6 years Mean pain duration 3.1 years	At 6 months (post-intervention) • Pain reduction • Quality of life • Physical function • Discontinuation	Pain rating of 40 or more required at baseline (VAS 0-100)
Von trott 2009[271]	12 week interventions **(n=38) Intervention 1: Mind-body exercise - Qigong.** Twenty-four sessions (each 45 minutes), held over a period of 12 weeks, in groups of 6-12 participants. Qigong lessons started with about 10 minutes of typical qigong 'opening' exercises, continued with up to 4 exercises of Dantian Qigong, and finished with about 10 minutes of 'closing' exercises. **(n=40) Intervention 2: Usual care**	Office workers with chronic neck pain (n=78) Mean age 76 years Mean pain duration 18.6 years	At 12 weeks (post-intervention) and 24 weeks follow up: • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	Waiting list control participants did not receive Qigong or exercise therapy.			
Wong 2018[278]	12 week interventions **Intervention 1: Mind-body exercise - Tai Chi (n=18)** Supervised sessions 3 times a week for 12 weeks. In the first session, the instructor explained the theory behind tai chi and its procedures providing participants with printed materials on its principles and techniques. In subsequent sessions, participants practiced 10 forms from the classic Yang style of tai chi. The sessions lasted approximately 55 minutes and included a 10 minute warm up, 40 minutes of practice and exercise finalising with a final 5 minute cool down period. During the sessions, the participants' heart rate was 40-50% of the HR reserve as they imitated the instructors' motion at the same speed. HR during training sessions was monitored using a polar device. **Intervention 2: Usual care (n=19)** Participants did not participate in any supervised or unsupervised exercise protocol and were asked to maintain their regular lifestyle habits for the duration of the study.	Women with fibromyalgia (n=37) Mean age 51 years Mean pain duration 27.5 years	At 12 weeks (post-intervention): • Pain reduction • Sleep • Discontinuation	
Wu 1999[279]	10 week interventions **Intervention 1: Mind-body exercise – Qigong (n=13)** 6 sessions of qigong training with 2 recognised qigong masters. Sessions included musical compositions and visual images which were coded to represent specific organ systems which qi is believed to stimulate. Each session lasted 40 minutes twice a week for 3 weeks, followed by 7 weeks of home exercises on a daily basis. **Intervention 2: Usual care (n=13)** Involving sham qigong. 6 sessions of simulated qigong training led by a simulated qigong master, in order to maximise nonspecific treatment effects. Participants were shown visual	Complex regional pain syndrome type I (late-stage) (n=26) Mean age 38.5 years Duration of pain not reported	At 10 weeks (post-intervention) • Pain reduction	Participants were required to have failed to achieve 50% pain reduction through drug therapy or palliative physical or chiropractic therapy

Study	Intervention and comparison	Population	Outcomes	Comments
	images and listened to recorded music similar to that in the qigong group. After this time a simulated qi adjustment was performed by the facilitator. Each session lasted for 40 minutes. This was followed by 7 weeks of home exercises.			

1.4.3.9 Flexibility versus usual care

Table 10: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Assumpcao 2018 [23]	12 week interventions **Intervention 2: Flexibility (n=18)** Patients underwent a 12 week supervised exercise program of 40-minute sessions performed twice a week. Segmental active muscle stretching was conducted without therapist assistance. Large muscles were chosen for their role in the muscular chains of global posture. Patients started with three repetitions up to a maximum of 5 by week 9. The stretch was held until the point of moderate discomfort, for 30 seconds **Intervention 3: Usual care (n=16)** Usual medical treatment. After 12 weeks patients were reassessed and offered physical therapy based on stretching and resistance training.	Women with fibromyalgia (n=36) Mean age 47 years Mean pain duration not stated	At 12 weeks (post-intervention): • Pain reduction • Physical function • Discontinuation	60% were taking concomitant medication for fibromyalgia (antidepressants, analgesics, anti-inflammatories or psychotropic medication)

1.4.3.10 Aerobic exercise versus strength training

Table 11: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Bircan 2008 [34]	8 week interventions.	Fibromyalgia (n=30)	At 8 weeks (post intervention):	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	Intervention 1: Aerobic exercise (n=15) Aerobic exercise program comprised walking on treadmill, initially for 20 min and increasing up to 30 min as the patient tolerated. Exercise intensity was adjusted to generate heart rates equivalent to 60–70% of age-adjusted maxi- mum heart rates (220 ¡ age in years). Heart rate monitoring was performed by using a pulse oximeter (Nonin Medical, Inc., MN, USA). At the beginning and end of each session mild stretches were included for 5 min. **Intervention 2: Strength training (n=15)** Patients received a supervised, progressive physical training program in a group setting with muscle strength exercises performed in the standing, sitting, and lying positions. Exercises strengthened the upper and lower limb muscles and trunk muscles, initially with 4–5 repetitions and progressing to 12 repetitions gradually. Free weights and body weight were used for strength. Patients began with resistance levels they could do easily, and weight was increased gradually according to patient's tolerance. Exercise sessions began with a low intensity warm up of marching in place and gentle stretching for 5 min, followed by 30 min of muscle strength, and concluded with 5 min of cool down and stretching.	All female Mean age 47.2 years Mean pain duration 4.2 years	• Pain reduction • Quality of life • Psychological distress • Sleep • Discontinuation (additional outcome)	
Ericsson 2016[80]	12 week interventions **Intervention 1: Aerobic exercise (n=17)** Pool exercise programme. 50 minute sessions in groups of 6-8 participants twice a week for 12 weeks, supervised by a physiotherapist. Sessions included aerobic exercise with endurance, strength, flexibility, coordination and relaxation. Patients were instructed to exercise at their own rhythm and modify exercises with respect to thresholds of pain and fatigue. They were encouraged to increase intensity and resistance with or without water equipment, based on the rate of perceived exertion on the Borg scale.	Fibromyalgia (n=34) All male Mean age 59 years Mean pain duration 5.3 years	At 12 weeks (post intervention): • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	(n=17) Intervention 2: Strength training Twice a week sessions for 12 weeks with free weights and resistance machines in groups of 8-10 patients, supervised by a physiotherapist. The sessions lasted approximately 1 hour and include exercises for multiple main muscle groups. Load was increased from 40% to 80% of one repetition maximum established at baseline. Participants performed 3 sets with 15-20 repetitions of each exercise, when the load increased they performed 2 sets but fewer repetitions. All sessions started with 10 minute warm up on an ergometer bicycle.			
Hooten 2012[125]	3 week interventions. **Intervention 1: Aerobic exercise (n=36)** Stationary bicycle exercises supervised by a physical therapist. Sessions also had a warm up and cool down and intensity of exercises was gradually increased to achieve 70-75% of maximal heart rate based on age. Exercise started at 10 minutes daily during week 1 (5 times a week), 15 minutes in week 2 and up to 20 to 30 minutes daily during week 3. **Intervention 2: Strength training (n=36)** Upper and lower body strengthening exercises were performed daily using resistive techniques, all supervised by a physical therapist with experience in treating patients with fibromyalgia. Each daily strength training session was 25-30 minutes in duration and also involved a warm up and cool down period. Participants were encouraged to train at the maximal amount of load tolerated, using one set of 10 repetitions.	Fibromyalgia (n=72) Mean age 46.5 years Mean pain duration 12.5 years	At 3 weeks (post-intervention): • Pain reduction • Discontinuation	
Kayo 2011 [142]	16 week interventions. **Intervention 1: Aerobic exercise (n=30)**	Fibromyalgia (n=60) All female	At 28 weeks (follow up, including 16 weeks intervention): • Quality of life • Pain	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	Supervised indoor or outdoor walking, three times a week for 60 minutes (5-10 minutes stretching, walking and 5 minutes cool down). **Intervention 2: Strength training (n=30)** Supervised exercise protocol consisting of 11 free active exercises for upper and lower limbs and trunk muscles, with free weights and body weight performed in the standing, sitting, and lying positions. Sessions were three times a week for 60 minutes. Exercise load and intensity increased every 2 weeks. Participants in all groups were asked to discontinue tricyclic antidepressants but were allowed to use acetaminophen (paracetamol) for pain	Mean age: 47.7 (5.3); 46.7 (6.3); 46.1 (6.4) years Duration of pain: 4.0 (3.1); 4.7 (5.7); 5.4 (3.5) years		
Sevimli 2015[242]	**12-week interventions. Intervention 1 and 2 pooled.** **Intervention 1: Aerobic exercise – Swimming (n=25)** Pool based aquatic aerobic exercise programme with group therapy 2 times a week. Duration was 40 minutes in the first month, 45 in the second month and 50 minutes in the final month. **Intervention 2: Aerobic exercise - Other aerobic exercise (n=25)** Gymnastic-based aerobic exercise programme with group therapy 2 times a week. Duration was 40 minutes in the first month, 45 in the second month and 50 minutes in the final month. No further details. *NB Aerobic exercise interventions pooled in the analysis.* **Intervention 3: Strength training (n=25)** Isometric strength and stretching exercise program lasting 15 minutes per day. Three minute loadings with 30 seconds rest between 3 sets of low to moderate intensity were repeated in the first month of the exercise programme, and in the second month this was increased to high intensity loadings of 4 sets, and in the	Women with fibromyalgia (n=75) Mean age 35 years Mean pain duration not specified	At 12 weeks (post-intervention) • Pain reduction • Quality of life • Physical function • Psychological distress	

Study	Intervention and comparison	Population	Outcomes	Comments
	third month rest intervals were reduced to 10 seconds with 5 sets of 3 minute loadings.			

1.4.3.11 Aerobic exercise versus flexibility

Table 12: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Mannerkorpi 2009[176]	20 week interventions **Intervention 1: Aerobic exercise (n=20)** 60 minutes 3 times weekly. After a 10-minute preliminary warm-up exercise, patients were subjected to sustained heart rate elevation training through the use of a bicycle ergometer. Heart rates were maintained in excess of 150 beats per minute for gradually increasing time periods, and were monitored with a Sanyo HRM-97E digital pulse meter. **(n=20) Intervention 2: Flexibility.** Participants met at similar intervals but at different times over the same 20-week observation period. Instruction was administered in a group setting by the same instructors as for CVR training, but consisted only of flexibility manoeuvres, such that sustained heart rate responses greater than 115 beats per minute were not attained.	Women with fibromyalgia (n=40) Mean age 42 years Duration of pain not specified	At 20 weeks post-intervention): • Pain reduction	Medication for pain discontinued at least 3 weeks before entry into the trial (patients receiving amitriptyline within the previous 3 months were excluded). Paracetamol allowed if required.
Mccain 1986[182 183]	20 week interventions **Intervention 1: Aerobic exercise (n=18)** Three times a week programme. Participants had sustained heart rate elevated training via a bicycle ergometer. Heart rates were maintained in excess of 150 beats per minute for gradually incremental durations. **(n=16) Intervention 2: Flexibility**	Fibromyalgia (n=34) Mean age 43 years Duration of pain not specified	At 20 weeks (post-intervention): • Pain reduction	

Study	Intervention and comparison	Population	Outcomes	Comments
	Participants met at similar intervals to the aerobic group. Exercise consisted of flexibility manoeuvres such that sustained heart rate responses were over 115 beats per minute were not attained.			
Valim 2003[259]	20 week interventions **Intervention 1: Aerobic exercise (n=38)** Walking programme monitored and supervised by a physiotherapist 3 times a week, with 45 minute duration for 20 weeks. Speed was determined by the training heart rate Patients cool down after each session consisted of making rhythmic movements to promote cooling off for 5 minutes. **Intervention 2: Flexibility (n=38)** 3 sessions a week of 45 minute duration including 17 stretching exercises using both muscles and joints. Each position sustained for maximum 30 seconds (supervised by physiotherapist).	Women with fibromyalgia (n=76) Mean age 46.8 years Pain duration not specified	At 10 and 20 weeks (post-intervention): • Pain reduction • Quality of life • Psychological distress • Discontinuation	Acetaminophen allowed as rescue treatment.

1.4.3.12 Aerobic exercise versus biomechanical exercise

Table 13: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
de Medeiros 2020 [69]	12 week interventions. **Intervention 1: Aerobic exercise (n=21)** Aquatic aerobics involved six main exercises lasting 30 min with different intensities. Two warm-up exercises and two cool-down exercises were performed before and after the program. Each session lasted 40 minutes. **Intervention 2: Biomechanical exercise (n=21)**	Women with fibromyalgia (n=42) Mean age: 50.7 (9.7); 45.5 (10.6) years Duration of pain not reported	At 12 weeks (post-intervention) • Pain • Quality of life • Psychological distress • Sleep • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	Mat Pilates was used in groups of up to 4 women. The focus of the sessions was on centralization, concentration, control, precision, breathing and flow. Nine exercises were performed for the main muscle groups with progressions each month. The exercises were initially performed in 1 series of 8 repetitions in the first month. Then they were performed in 2 sets of 10 repetitions in the second month. Finally, they were performed in 3 sets of 8 repetitions in the last month. Three Swiss ball relaxation exercises were performed in 1 set of 30 s each (Fig. 2a.10 to a.12) at the end of each session. Each session lasted 50 minutes.			

1.4.3.13 Aerobic and strength versus aerobic exercise

Table 14: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Sanudo 2010 [234]	24 week interventions. **Intervention 2: Mixed modality exercise (n=21)** Combined supervised aerobic exercise and resistance exercise. Resistance included 1 set of 8-10 reps for 8 different muscle groups with a load of 1-3 kg, flexibility included 1 set of 3 reps of 8-9 different exercises, maintaining stretch position for 30 seconds. The exercises focused on main areas of pain in patients with fibromyalgia (deltoids, biceps, neck (trapezius), hops (gluteus, quadriceps), back/chest/torso (latissimus dorsi, pectoralis major, and abdominals)). Twice a week, each session including 10 minutes warm-up, 10-15 minutes aerobic exercise, 15-20 minutes resistance, 10 minutes flexibility. **Intervention 1: Aerobic exercise (n=22)**	Fibromyalgia (n=64 ; third arm of study reported under aerobic versus usual care comparison) Females only Mean age: 55.9 (1.6); 55.9 (1.7); 56.6 (1.9) years Duration of pain: not specified	At 24 weeks (post intervention): • Quality of life • Psychological distress • Discontinuation	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	Warm-up included slow walks, easy movements of progressive intensity, steady state aerobics included continuous walking with arm movements and jogging, interval training included aerobic dance and jogging, cool-down included slow walks, easy movements, relaxation training. Twice a week for 45-60 minutes (10 minutes warm-up, 5-10 minutes cool down, 15-20 minutes steady aerobics, 15 minutes interval training).			

1.4.3.14 Aerobic and Strength versus flexibility

Table 15: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Giubilei 2007[106]	18 week interventions **Intervention 1: Aerobic and Strength exercise (n=52)** 18 week walking program, 3 times per week. Each exercise session included a warm up and cool down regimen of slow paced walking, specific postural muscle and isometric strengthening exercises, and 40 minutes of fast paced walking on in-outdoor track, at 70-80% of maximum heart rate **Intervention 2: Flexibility (n=51)** Participants participated in a flexibility and motion exercise program for the same period of time and frequency as the aerobic group. Patients were instructed about the correct exercise execution and were advised to maintain their heart rate under 110bpm. Exercises were simply stretches with some motion exercises such as leg lifts.	Men with chronic prostatitis/chronic pelvic pain syndrome (n=103) Mean age 36.7 years Mean pain duration 5.72 years	At 6 weeks and18 weeks (post-intervention): • Pain reduction • Quality of life • Psychological distress • Discontinuation	

1.4.3.15 Aerobic and flexibility versus mind-body exercise

Table 16: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Wang 2018[274]	24 week interventions **Intervention 1: Aerobic exercise and flexibility (n=75)** Each session lasted 60 minutes and ran twice a week for 24 weeks. Participants were encouraged to integrate at least 30 minutes of aerobic exercise into their daily routine during the intervention, and to continue this throughout the 52 week follow up. Sessions were closely supervised in a group format and were moderate intensity. Each session consisted of an active warm-up, choreographed aerobic training that progressed gradually from low to moderate intensity and a cool down involving low intensity movements and dynamic and static stretching. During the first week there was a 15 minute warm up, 20 minutes of aerobic training and 25 minutes of cool-down, which increased to 40 minutes of aerobic training by week 10 to (at 60-70% of estimated maximum heart rate). **Intervention 2: Mind-body exercise - Tai Chi (n=36)** Each session lasted 60 minutes and ran twice a week for 24 weeks. Participants were encouraged to integrate at least 30 minutes of tai chi into their daily routine during the intervention, and to continue this throughout the 52 week follow up. Sessions were run by experienced instructors and sessions were recorded to monitor quality and provide feedback to instructors. Participants also received printed materials on tai chi principles and fibromyalgia. The sessions included warm up, meditative movements, breathing techniques and various relaxation methods.	Fibromyalgia (n=111) Mean age 51 years Duration of pain 12.5 years	At 1 year follow up (including 24 week intervention): • Quality of life • Pain reduction • Physical function • Psychological distress • Sleep • Discontinuation	

1.4.3.16 Aerobic exercise and flexibility versus aerobic exercise

Table 17: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Gomez-Hernandez 2020 [108]	12 week interventions. **Intervention 1: Aerobic exercise and stretching (n=32)** Aerobic exercise was identical to intervention 2 (as described). Additionally, 45 minutes of stretching was carried out once per week. Each session consisted of three repetitions of 10 seconds for each trunk muscle and two repetitions of 10 seconds for each extremity muscle. After each repetition, there was a 10-second pause. **Intervention 2: Aerobic exercise (n=32)** Supervised cycling, with each session consisting of 2-minute cycling warm-up and 10 minutes of moderateintensity cycling (50%–70% of predicted maximum heart rate). Three times per week for 12 minutes.	Women with fibromyalgia (n=64) Mean age: 54.27 (6.94) years Duration of pain not reported	At 4 weeks and 12 weeks (post-intervention) • Pain • Quality of life • Sleep • Discontinuation	4 week outcomes are measured before end of intervention.

1.4.3.17 Aerobic, strength, mind-body and proprioception versus flexibility

Table 18: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Carvalho 2020 [55]	7 week interventions **Intervention 1: Aerobic, strength, mind-body and proprioception (n=16)** An exergame programme performed on a Nintendo Wii system. The programme consisted of 6 sub games, which included jogging, a game involving active movement of the upper limbs in isolation from weight and balance training,	Women with fibromyalgia (n=35) Mean age: 55.64 (9.16); 47.70 (15.46) years	At 7 weeks (post-intervention) • Qualtiy of life • Physical function • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	yoga, a Hula Hoop game involving action of the trunk muscles and balance control, a step game involving alternating movements of lower limbs and balance, and a stationary walking game. This was performed three times per week for 1 hour. **Intervention 2: Flexibility (n=19)** Chain muscle stretching technique, which involved 9 stretching positions, held for 4 deep and prolonged breaths. These positions were chosen to include standing, sitting and lying positions, and to engage all muscle groups. The sessions were performed 3 times per week for 1 hour.	Duration of pain: 9.91 (7.29); 14.65 (12.14) years		

1.4.3.18 Strength training versus mind-body exercise

Table 19: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Lansinger 2013[157]	12 week interventions **Intervention 1: Strength training (n=62)** Exercise therapy was performed individually and the training programme was adjusted for each participant. A physiotherapist instructed the participants throughout the training programme, which focused mainly on the cervical and shoulder/thoracic region. Each training session started with a warm up on a stationary bicycle for about 10 minutes, followed by 40 minutes of dynamic exercises. These exercises consisted of active movements aimed to increase range of motion in all neck directions and muscle exercises aimed to maintain/increase circulation, endurance and strength. The amount of load was individualised and was maintained within pain tolerance (aimed not to increase pain). The load at the muscle exercises was to achieve between	Non-specific neck pain for at least 12 weeks (n=122) Mean age 43.8 years Duration of pain: 60% for 1-10 years	At 12 weeks post-intervention): • Discontinuation	Inclusion criteria minimum VAS rating of 20 (0-100 scale) Both groups received verbal ergonomic advice for both work and free time, as well as an information pamphlet on neck pain

Study	Intervention and comparison	Population	Outcomes	Comments
	30% and 70% of maximum muscle capacity and was gradually increased as endurance and strength were gained. The exercises were performed with low resistance, allowing 20-30 repetitions of maximal voluntary contractions in three sets. 12 sessions in 3 months. **Intervention 2: Mind-body exercise – Qigong (n=60)** 10-12 1 hour sessions conducted on a weekly or biweekly basis over 3 months. Qigong was performed according to medical qigong which is a modality of traditional Chinese medicine and is a way of affecting and directing qi (energy) for medical benefit. Each qigong exercise includes body posture and gentle movement, meditation (concentration) and purposeful relaxation, breathing regulation practice and self-administered massage. Qigong was conducted in groups of 10-15 participants.12 sessions in 3 months.			
Ulug 2018[257]	6 week interventions. **Intervention 1: Strength exercise (n=20)** Isometric exercise. In the sitting position, participants were instructed to place their hands firstly on the front (then the other sides) of their heads and push forward, but resist any movement of the head while maintaining the head and neck in the neutral position for 5 s. They were encouraged to do these exercises in 2 sets of 30 repetitions per day. **Intervention 2: Mind-body exercise (n=20)** Four Iyengar yoga exercises were taught to participants. They were told to maintain each yoga posture starting from at least 10–20 s in the following days. They were encouraged to do these exercises in 2 sets of 10 repetitions per day.	Chronic neck pain (>3 months) (n=60) Mean age: 44.6 (4.3); 35.9 (9.8) Duration of pain: 58.8 (63.3); 56 (60.1) months	At 6 weeks (follow-up): • Pain	All participants also received physical therapy 5 days a wweek for 3 weeks, including hot pack, ultrasound and conventional transcutaneous electrical nerve stimulation (TENS)

1.4.3.19 Strength versus biomechanical

Table 20: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Ulug 2018[257]	6 week interventions. **Intervention 1: Strength exercise (n=20)** Isometric exercise. In the sitting position, participants were instructed to place their hands firstly on the front (then the other sides) of their heads and push forward, but resist any movement of the head while maintaining the head and neck in the neutral position for 5 s. They were encouraged to do these exercises in 2 sets of 30 repetitions per day. **Intervention 2: Biomechanical (n=20)** Pilates involved participants being taught how to activate their deep abdominal muscles (transversus abdominis and multifidus) using visual imagery, verbal cueing or demonstrations. Five key elements of Pilates, including lateral costal breathing, centering (pelvic placement), ribcage placement, shoulder blade placement, head and neck placement, were taught. Four Pilates beginner mat exercises, including double-leg stretch level, shoulder bridge level, arm openings level and breast stroke level, were taught and patients were encouraged to perform these exercises in 2 sets of 10 repetitions per day.	Chronic neck pain (>3 months) (n=60) Mean age: 44.6 (4.3); 38.7 (7.9) Duration of pain: 58.8 (63.3); 55.1 (47) months	At 6 weeks (follow-up): • Pain	All participants also received physical therapy 5 days a week for 3 weeks, including hot pack, ultrasound and conventional transcutaneous electrical nerve stimulation (TENS)

1.4.3.20 Strength training versus flexibility

Table 21: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Assumpcao 2018 [23]	12 week interventions	Women with fibromyalgia (n=37)	At 12 weeks (post-intervention):	60% were taking concomitant

Study	Intervention and comparison	Population	Outcomes	Comments
	Intervention 1: Strength training (n=19) 12 week supervised resistance training programme of 40-minute sessions performed twice a week with progressive overload. Equipment included dumbbells, shin pads. No load was used in the first 2 sessions, after which time 0.5kg was added each week if the patient identified the effort as slightly intense on the Borg scale. 8 repetitions for: triceps, quadriceps, hip adductors and abductors, hip flexors, elbow flexors and extensors, pectoralis major and rhomboids. **Intervention 2: Flexibility (n=18)** Patients underwent a 12 week supervised exercise program of 40-minute sessions performed twice a week. Segmental active muscle stretching was conducted without therapist assistance. Large muscles were chosen for their role in the muscular chains of global posture. Patients started with three repetitions up to a maximum of 5 by week 9. The stretch was held until the point of moderate discomfort, for 30 seconds	Mean age 47 years Mean pain duration not stated	• Pain reduction • Physical function • Discontinuation	medication for fibromyalgia (antidepressants, analgesics, anti-inflammatories or psychotropic medication)
Gavi 2014[100]	16 week interventions **Intervention 1: Strength training (n=40)** 45 minute sessions 2 times a week for 16 weeks. Supervised progressive training in standing and sitting positions using weight machines. Moderate intensity with load of 45% the estimated maximum. Multiple muscle groups were trained in 12 different exercises, with 3 sets of 12 repetitions **Intervention 2: Flexibility (n=40)** 45 minute sessions 2 times a week for 16 weeks. Stretching of the major muscles. No further details.	Women with fibromyalgia (n=80) Mean age 47.61 years Mean pain duration not specified	At 16 weeks (post-intervention): • Quality of life • Discontinuation	7% were taking benzodiazepines or amitriptyline concurrently
Jones 2002 [136]	12 week interventions. **Intervention 1: Strength training (n=28)**	Fibromyalgia (n=56) All females	At 12 weeks (post intervention): • Pain • Physical function	In Cochrane review (Busch 2013)

Study	Intervention and comparison	Population	Outcomes	Comments
	Supervised dynamic resistance exercise for lower and upper limbs and trunk using hand weight (1-3 lb (0.45-1.36 kg)) and elastic tubing; minimization of eccentric work (a videotape to guide home practice of the strengthening exercise regimen was provided to participants). Twice a week for 60 minutes, progressing from 4-12 reps. **Intervention 2: Flexibility (n=28)** Supervised static stretches (a videotape to guide home practice of the flexibility exercise regimen was provided to participants). Twice a week for 60 minutes.	Mean age: 46.4 (8.6) to 49.2 (6.3) years Duration of pain: 6.9 (6.6) to 7.7 (5.5) years	• Psychological distress • Sleep	

1.4.3.21 Strength and flexibility versus flexibility

Table 22: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Salo 2012[230]	12-month interventions **Intervention 1: Combined strength training and flexibility (n=49)** Participants used elastic rubber bands attached around the head for the isometric neck strength exercises. During each session they performed a series of 15 repetitions directly forward, obliquely toward the right and left and directly backwards. The aim was to reach the level of resistance that was 80% of the patient's maximum isometric neck strength. In each exercise session, the patients also performed a single series of 15 repetitions of dynamic exercises for the shoulders and upper extremities with an individually adjusted highest load. These exercises involved shrugs, presses, curls, bent over rows, flyers and pullovers using dumbbells. The training programme also involved a single series of squats, sit ups and back extension exercises that used only the patient's own body weight; these exercises were	Chronic non-specific neck pain (n=101) Mean age 40.5 years Duration of pain 62 months	At 12 months post-intervention): • Quality of life • Discontinuation	Both groups were instructed to perform their exercises at home regularly three times a week and to keep a weekly exercise diary throughout the year. Both groups received written information about the exercises.

Study	Intervention and comparison	Population	Outcomes	Comments
	performed until muscle tiredness. The training session included stretching exercises for the neck, shoulder, and upper limb muscles with the exercise for each muscle lasting 30 seconds and repeated 3 times. Supervised meetings were conducted once a week for 6 weeks, then one session was conducted every second month for a total of 10 sessions over the 12 month period. Each group had 6-8 participants. **Intervention 2: Flexibility (n=52)** Those in the stretching group performed the same stretching exercises to the other group.			

1.4.3.22 Strength and flexibility versus mind-body exercises

Table 23: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Cramer 2013[65]	**9 week interventions.** **Intervention 1: Strength and flexibility exercise (manual based) (n=26)** Participants received a self-care manual to relieve neck pain and stiffness. The manual described and depicted a staged seated exercise program for the neck and shoulder region. The program began with taking a proper upright sitting posture, followed by stretching exercises for the neck and shoulders. Then, strength exercises and isometric exercises for the neck-shoulder region were performed. Patients were required to practice at home for 10 minutes each day and to record their practice in a diary. **Intervention 2: Mind-body exercise – Yoga (n=25)** 90 minute weekly classes of 10-15 participants over 9 weeks. Designed for patients with chronic neck pain without previous experience in yoga. Each class consisted of 8 to 11 yoga	Non-specific neck pain for at least the previous 12 weeks (n=51) Mean age 47.8 years Duration of pain 8.1 years	At 9 weeks (post intervention): • Pain reduction • Quality of life • Physical function • Discontinuation	Participants in both groups were allowed to continue their usual pain medication and physical activity. They were asked not to change their treatment regimen during the course of the study and to daily record pain medications and other treatments for neck pain in their diaries.

Study	Intervention and comparison	Population	Outcomes	Comments
	postures chosen from a pool of 14 standing, sitting and supine postures, starting with relatively simple postures and succeeding to more complex ones. The focus of postures was given on lengthening and strength muscles of the neck and shoulder region and to improve stability and posture. Patients were required to practice at home for 10 minutes each day. Patients received a manual describing and depicting 3 basic standing and 3 basic sitting postures.			
Rendant 2011[222]	**6 month interventions** **Intervention 1: Strength and flexibility training (n=39)** Exercise therapy was carried out by 6 qualified therapists. The exercises was based on a standard programme for chronic pain. Each lesson started with a warm up using a softball and was followed by repeated active cervical rotations and strengthening and flexibility exercises. The individual's pain level was not exceeded. There were 18 sessions over a period of 6 months (1 session per week in the first 3 months, and biweekly sessions in the following 3 months). **Intervention 2: Mind-body exercise – Qigong (n=42)** Qigong was performed by three qualified teachers certified by the German Qigong Society. Each session of qigong took 90 minutes. Neiyanggong, a special silent and slow form of qigong was chosen by the therapist in a consensus process. The lessons started with up to 12 neck exercises followed by 9 exercises for the shoulder and finished with breathing and moving exercises. There were 18 sessions over a period of 6 months (1 session per week in the first 3 months, and biweekly sessions in the following 3 months).	Chronic non-specific neck pain (n=123; third arm of study reported under strength and flexibility versus usual care and mind-body versus usual care comparisons) Mean age 44.6 years Mean pain duration 3.3 years	At 6 months (post-intervention) • Pain reduction • Quality of life • Physical function • Discontinuation	Pain rating of 40 or more required at baseline (VAS 0-100)
Von trott 2009[271]	12 week interventions **Intervention 1: Strength and flexibility training (n=39)**	Office workers with chronic neck pain (n=77)	At 12 weeks (post-intervention) and 24 weeks follow up: • Pain reduction	

Study	Intervention and comparison	Population	Outcomes	Comments
	24 sessions (each 45 minutes) at 2 sessions per week with groups of 6-12. A standardised programme for computer and workplace related neck pain. It included repeated active cervical rotations as well and strength and flexibility exercises. Special intention as paid so that the patients' individual pain limits were not exceeded. About 90% of the exercises were repeated in each lesion; some 10% was exchanged regularly. **Intervention 2: Mind-body exercises – Qigong (n=38)** Twenty-four sessions (each 45 minutes), held over a period of 12 weeks, in groups of 6-12 participants. Qigong lessons started with about 10 minutes of typical qigong 'opening' exercises, continued with up to 4 exercises of Dantian Qigong, and finished with about 10 minutes of 'closing' exercises.	Mean age 76 years Mean pain duration 18.6 years	• Quality of life • Physical function • Psychological distress • Discontinuation	

1.4.3.23 Strength, flexibility and proprioception versus mind-body exercises

Table 24: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Lauche 2016[160]	12 week interventions **Intervention 1: Strength, proprioception and flexibility training (n=37)** Participants in the neck exercise group met once weekly for a 60- to 75-minute session for 12 weeks in total. This group was instructed in neck exercises, which were similar to those taught in rehabilitation programs containing exercises and education for a healthy back. Classes contained basic training of ergonomic principles (bodily alignment while standing), proprioceptive exercises, and isometric and dynamic mobilization, stretching, and strengthening neck and core exercises. The sessions opened with 5 to 10 minutes of warm-up exercises and ended	Chronic non-specific neck pain (n=114; third arm of study reported under mind-body versus usual care and strength, proprioception and flexibility versus mind-body comparisons) Mean age 49.53 years	At 12 weeks (post-intervention) and 24 weeks (follow up): • Pain reduction • Quality of life • Physical function • Psychological distress • Discontinuation	VAS score of 45 or higher (0-100) inclusion criteria.

Study	Intervention and comparison	Population	Outcomes	Comments
	with relaxation exercises. Participants also received illustrated and written information that covered the most important exercises, and they were asked to execute the exercises for at least 15 minutes each day. **Intervention 2: Mind-body exercise - Tai Chi (n=38)** Participants in the Tai Chi group met once weekly for a 75- to 90-minute session. The Tai Chi intervention was on the basis of a popular and internationally recognized Yang style (13 forms from Mantak Chia). Each session included a warm-up of 5 to 10 minutes, the Tai Chi form practice, and 5 to 10 minutes of relaxation at the end. Tai Chi forms followed explicit protocols outlined in a training manual, as required during teacher training certification. Sessions also included educational units and breathing exercises, and they were accompanied by relaxation music. Participants received illustrated written information that covered movement sequences learned in the previous session. They were asked to practice Tai Chi outside of classes for at least 15 minutes each day.	Mean pain duration not stated		

1.4.3.24 Strength versus proprioceptive training

Table 25: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Gallego Izquierdo 2016 [97]	8 week interventions **Intervention 1: Strength training (n=14)** Cranio-cervical flexion training led by physiotherapists. Low load training of flexor muscles to target deep flexors and aiming to minimize the activation of the superficial flexor muscles. The patient initially performed CCF to sequentially	Chronic non-specific neck pain for at least 3 months (n=28) Mean age 29.2 years	At 8 weeks (post-intervention): • Pain reduction • Physical function	

Study	Intervention and comparison	Population	Outcomes	Comments
	reach 5 pressure targets in 2 mmHg increments from a baseline of 20 mmHg to the final level of 30 mmHg. Once one set of 10 repetitions of 10 s was achieved at one target level, the exercise was progressed to train at the next target level up to the final target of 10 repetitions of 10 s at 30 mmHg. The exercise load prescribed to each patient was based on their assessment performance. Participants were taught to do exercises at home without biofeedback **Intervention 2: Proprioceptive exercise (n=14)** Participants trained in cervical proprioception following the protocol described by Revel et al. This regime consisted of exercises of head relocation, eye-follow, gaze stability and eye-head coordination. All active movements of the cervical spine (flexion, extension, rotation, lateral flexion) were performed. All exercises were progressed by increasing the speed and range of motion of the target and with participants in a standing position.	Mean duration of pain not specified		

1.4.3.25 Mind-body versus flexibility

Table 26: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Calandre 2009 [50]	6 week interventions. **Intervention 1: Mind-body exercise (n=39)** Tai chi was performed in a pool with water heated at 36 ° and was preceded by a shower with warm water to condition patients' bodies. A trained physiotherapist adjusted the movement intensity to meet individual needs and participants were taught the 16 movements which constitute tai chi therapy. Both groups received 18 sessions of 60 minutes, delivered 3 times per week for 6 weeks.	Fibromyalgia (n=81) Female:Male 73:8 Age: 32 to 69 years Duration of pain: not reported	At 3 months (follow-up): • Quality of life • Psychological distress • Sleep • Discontinuation (additional outcome)	In Cochrane review (Theadom 2015)

Study	Intervention and comparison	Population	Outcomes	Comments
	Intervention 2: Flexibility (n=42) Stretching was facilitated using supportive aids such as long wooden sticks, flexible strings and tubes to stretch muscles in the cervical, upper and lower extremities and trunk. Both groups received 18 sessions of 60 minutes, delivered 3 times per week for 6 weeks.			

1.4.3.26 Mind-body versus biomechanical

Table 27: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Ulug 2018[257]	6 week interventions. **Intervention 1: Mind-body exercise (n=20)** Four Iyengar yoga exercises were taught to participants. They were told to maintain each yoga posture starting from at least 10–20 s in the following days. They were encouraged to do these exercises in 2 sets of 10 repetitions per day. **Intervention 2: Biomechanical (n=20)** Pilates involved participants being taught how to activate their deep abdominal muscles (transversus abdominis and multifidus) using visual imagery, verbal cueing or demonstrations. Five key elements of Pilates, including lateral costal breathing, centering (pelvic placement), ribcage placement, shoulder blade placement, head and neck placement, were taught. Four Pilates beginner mat exercises, including double-leg stretch level, shoulder bridge level, arm openings level and breast stroke level, were taught and patients were encouraged to perform these exercises in 2 sets of 10 repetitions per day.	Chronic neck pain (>3 months) (n=60) Mean age: 35.9 (9.8); 38.7 (7.9) Duration of pain: 56 (60.1); 55.1 (47) months	At 6 weeks (follow-up): • Pain	All participants also received physical therapy 5 days a week for 3 weeks, including hot pack, ultrasound and conventional transcutaneous electrical nerve stimulation (TENS)

1.4.3.27 Flexibility and relaxation versus aerobic exercise

Table 28: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Richards 2002[224]	12 week interventions **Intervention 1: Flexibility and relaxation (n=67)** Comprised of upper and lower limb stretches and relaxation techniques based on the published regimen by Ost. As the classes continued more techniques were introduced progressing through progressive muscle relaxation, release only relaxation and visualisation, cue controlled relaxation, and differential relaxation. This occupied the whole one hour class. The sessions were carried out twice weekly. **Intervention 2: Aerobic exercise (n=69)** Both groups met in hour-long classes of up to 18 individuals twice weekly for 12 weeks. The interventions were carried out by personal trainers. Exercise therapy comprised an individualised aerobic exercise programme, mostly walking on treadmills and cycling on exercise bicycles. Each individual was encouraged to increase the amount of exercise steadily as tolerated. When people first started classes they usually did two periods of exercise per class lasting six minutes. By 12 weeks they were doing two periods of 25 minutes at an intensity that made them sweat slightly while being able to talk comfortably in complete sentences.	Fibromyalgia (n=136) Mean age 46.5 years Duration of pain 5 years (median)	• Quality of life (12 months) • Discontinuation (12 weeks, post-intervention)	Participants continued their medication at entry. They received standardised advice including an explanation of fibromyalgia and encouragement and were told that the exercise offered through prescription would improve their condition. Each week at the classes all individuals received an information leaflet covering an aspect of their condition.

1.4.3.28 Flexibility and proprioception versus flexibility

Table 29: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Kibar 2015[147]	6 week interventions	Fibromyalgia (n=68)	At 6 weeks (post-intervention):	

Study	Intervention and comparison	Population	Outcomes	Comments
	Intervention 1: Flexibility and proprioception exercises (n=35) Balance exercises included postures that gradually reduced the base of support, dynamic movements that disturbed the centre of gravity, exercises that stressed the postural muscle groups and exercises that reduced sensory input (standing with eyes closed). Training was provided by an experienced physiotherapist for 20 sessions over a 4 week period (20 minutes for each session, 5 days/week). The group also received 5 minutes of static and 5 minutes of dynamic balance training with a KAT device 3 days/week. For flexibility, active static exercises were performed in order to enable compliance to exercise and its maintenance without being forced. Exercises were performed in 8 large muscle groups in three 60-second static stretching repetitions. Ten minutes of walking in place was also recommended as warm up. **(n=33) Intervention 2: Flexibility** As per the flexibility section of the combined intervention described above.	Mean age 48.14 years Duration of pain not reported	• Quality of life • Psychological distress • Discontinuation	

1.4.3.29 Exercise versus psychological therapies

Table 30: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Ericsson 2016[81]	15 week interventions **Intervention 1: Strength training (n=67)** Exercise sessions were twice a week for 15 weeks at physiotherapy premises and at a local gym and were supervised by experienced physiotherapists. The exercise program was standardized and performed in groups of five to seven participants but the load was	Fibromyalgia (n=130) Aged 22 to 64 years Mean pain duration not specified	At 15 weeks (post-intervention): • Pain reduction • Quality of life • Physical function • Psychological distress	

Study	Intervention and comparison	Population	Outcomes	Comments
	adjusted individually. The exercise session started with 10 minutes of warm up followed by 50 minutes of resistance exercises focused on large muscle groups in all four extremities and trunk. The resistance exercise was initiated at 40 % of 1 repetition maximum (RM) and progressed up to 80 % of 1 RM during the 15 weeks. Possibilities for progression of loads were evaluated every 3–4 weeks. Forty-two participants (62.7 %) in the resistance exercise group reached exercise loads of 80 % of 1 RM while seven participants (10.4 %) reached exercise loads of 60 % of 1 RMv. This was followed by 10 minutes of stretching exercises **(n=63) Intervention 2: Relaxation therapy** Performed twice a week for 15 weeks, guided by experienced physiotherapists and conducted at physiotherapy premises in groups of five to eight participants. It was performed as autogenic training. which refers to a series of mental exercises including autosuggestion and relaxation. The relaxation therapy lasted for approximately 25 minutes, followed by stretching exercises.		• Discontinuation	
Fontaine 2010 [94]	12 week interventions. **Intervention 1: Aerobic exercise (n=43)** Walking (the most common form of life physical activity) and other forms (e.g., gardening/mowing the lawn) of household activity (e.g., vacuuming); and sports activity (e.g., cycling, swimming, field hockey). Frequency of 5-7 times per week for 60 minutes. **Intervention 2: Education (n=26)**	Fibromyalgia (n=69) All female Mean age: 46.4 (11.6); 49 (10.2) years Duration of pain: 5.9 (5.1); 9.6 (6.8) years	At 12 weeks (post intervention): • Quality of life • Pain • Physical function • Psychological distress • Discontinuation	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	Education, question and answer, and social support. Frequency of once per month for 90-120 minutes.			
Gavish 2006[101]	8 week interventions **Intervention 1: Strength training (n=10)** Chewing exercise. Two units of sugarless chewing gum were chewed three times daily for 10 minutes (weeks 1 and 2), increasing to 15 minutes three times daily (weeks 5 and 6), and 30 minutes 3 times daily (weeks 7 and 8). Patients were instructed to chew at their own rate. All patients received a detailed explanation of their disorder, its cyclic nature and possible aetiology at the initial examination. They then received a detailed description of the chewing exercise protocol (at session 1). Sessions 2, 3, and 4 were to report patient's condition, reassurance, support, and encouragement. They also reported their performance. **Intervention 2: Pain education (n=10)** All patients received a detailed explanation of their disorder, its cyclic nature and possible aetiology at the initial examination. Sessions 2, 3, and 4 were to report patient's condition, reassurance, support, and encouragement.	Masticatory muscle pain for at least 6 months (n=20) Mean age 27.2 years Duration of pain not reported	At 8 weeks (post-intervention): • Pain reduction • Discontinuation	Inclusion criteria of age 20-45 years
Jones 2012 [136]	12 week interventions. **Intervention 1: Mind-body exercise (n=51)** Tai chi delivered in a group based format 90 minute sessions delivered twice weekly for 12 weeks. **Intervention 2: Education (n=50)** Education sessions delivered in a group based format on fibromyalgia, healthy eating, education based CBT	Fibromyalgia (n=101) Mean age 51.4 years Mean duration of pain 18.4 years	At 12 weeks (post-intervention): • Pain • Quality of life • Physical function (additional outcome) • Discontinuation (additional outcome)	In Cochrane review (Theadom 2015)

Study	Intervention and comparison	Population	Outcomes	Comments
	strategies, sleep hygiene and lifestyle management 90 minute sessions delivered twice weekly for 12 weeks.			
King 2002 [152]	12 week interventions. **Intervention 1: Aerobic exercise (n=42)** Walking, aquacise (deep and shallow water), or low impact aerobics. Three times a week starting at 10-15 minutes and progressing to 20-40 minutes. **Intervention 2: Education (n=41)** Educational session provided by a multidisciplinary team. Topics focused on potential causes of fibromyalgia, principles of self-management (goal setting, maximizing energy for household chores or personal activities, pain or fatigue coping strategies, benefits of exercise, evaluating alternative therapies, and barriers to behaviour change). Once a week for 1.5-2 hours.	Fibromyalgia (n=170; third arm of study reported under aerobic versus usual care comparison) Females only Mean age: 45.2 (9.4); 44.9 (10); 47.4 (9); 47.3 (7.3) years Duration of pain: 7.8; 10.9; 8.9; 9.6 years	At 24 weeks (follow up including 12 week intervention): • Quality of life • Physical function • Pain • Discontinuation	In Cochrane review (Bidonde 2017)
Martin 1996[179]	6 week interventions **Intervention 1: Aerobic, Strength training (n=30)** Participants met 3 times a week for 6 weeks and participated in 1 h supervised exercise program. The program included 20 minutes walking at a pace sufficient to raise heart rate to 60-80% of maximum, 20 minutes of flexibility and strength training for multiple muscles. **Intervention 2: Relaxation (n=30)** 3 times per week for 6 week, supervised relaxation program for 1 hour in a quiet room. Patients were taught visualization, yoga and autogenic relaxation by experienced instructors.	Fibromyalgia (n=60) Mean age 44.8 years Duration of pain 9.2 years	At 6 weeks post-intervention): • Quality of life • Discontinuation	
McBeth 2012[181]	6 month intervention	Chronic widespread pain (n=330)	At 9 months: • Quality of life	

Study	Intervention and comparison	Population	Outcomes	Comments
(Beasley 2015[28])	**Intervention 1: Aerobic exercise (n=109)** Gym based programme with monthly assessments led by instructors to reassess the programme. Exercise intensity increased until exercise levels achieved 40-85% maximum heart rate; recommended session length 20 to 60 minutes 3-5 times a week) **Intervention 2: Cognitive behavioural therapy (n=112)** Telephone delivered, 7 weekly sessions (30-45 minutes each) plus initial assessment, followed by 1 session at 3 months and 1 session at 6 months. Delivered by 4 therapists. **Intervention 3: Usual care (n=109)** Usual care from family physician, although precise care delivered, if any, was not record3ed	Duration of pain not stated Mean age 55.7(12.5) years	• Sleep • Discontinuation (6 months)	
Silva 2019 [243]	12 week interventions **Intervention 1: Strength training (n=30)** Resistance training, which consisted of 3 sets of 12 repetitions, alternating lower limbs. Loads were 60% of the 1 rep maximum in the first month, increasing to 80% in the third month. The following muscles were trained: biceps brachial, triceps, pectoralis, trapezius, knee extensors, knee flexors and hip abductors. Twice a week for 40 minutes. **Intervention 2: Relaxation (n=30)** Body relaxation sessions, which involved lying down with relaxing movement. Participants were invited to think about their illness, their life, imagining positive and negative points and to analyze everything. The	Women with fibromyalgia (n=60) Mean age: 44.93 (10.30); 49.40 (8.30) years Duration of pain not reported	At 8 and 12 weeks (end of intervention) • Pain reduction • Physical function • Quality of life • Discontinuation	Only pain reduction reported at 8 weeks. Intervention not finished at 8 weeks so outcome measured before end of intervention.

Study	Intervention and comparison	Population	Outcomes	Comments
	physiotherapist also asked them to focus on the negative aspects and concentrate on these negative points, and they were asked to try to see good aspects of each point. Twice a week for 40 minutes.			
Viljanen 2003[267]	12 week interventions **Intervention 1: Strength training (n=135)** Led by trained physiotherapist 3 times a week sessions for 30 minutes each, followed by on week of reinforcement training 6 months after randomisation. Dumbbells were used for dynamic muscle training (weight 1-3kg each according to maximum repetitions with a test weight of 7.5 kg). The Exercises, conducted in the same order in each session, were chosen to activate large muscle groups in the neck and shoulder region. After the 5thweek participants were taught 3 exercises from the program with stretches, after the 9th week they were asked to perform the full training program by themselves. **Intervention 2: Relaxation (n=128)** Led by trained physiotherapist 3 times a week sessions for 30 minutes each, followed by on week of reinforcement training 6 months after randomisation. Exercises aimed to teach participants to activate only those muscles needed for different daily activities and to relax other muscles. Participants were taught to perform the exercises alone from the 5th week.	Chronic non-specific neck pain (n=393; third arm of study reported under strength versus usual care comparison) Mean age 44 years Mean pain duration 10.8 years	At 12 months follow up (including 12 week intervention): • Pain reduction • Discontinuation	All participants were office workers
Wigers 1996 [276]	14 week interventions. **Intervention 1: Aerobic exercise (n=20)** Aerobic exercise, focusing on the whole body and aimed at minimizing eccentric muscle strain. Exercise involved movement to music and games. Three times a week for 45 minutes (23 minute music session including warming	Fibromyalgia (n=40) Female:Male: 55:5 Mean age: 43 (9); 44 (12); 46 (9) years	At 14 weeks (post intervention) and 4 years (follow-up): • Pain • Sleep • Psychological distress	In Cochrane review (Bidonde 2017)

Study	Intervention and comparison	Population	Outcomes	Comments
	up and 2 peaks of high intensity training, 15 minutes of aerobic games with 2 high intensity periods). **Intervention 2: Stress management training (n=20)** Stress management training with 2 treatment groups of 10, with each totalling 20 sessions and 30 hours of active treatment (twice a week for 6 weeks, and once a week for 8 weeks, each session 90 minutes).	Duration of pain: 9 (5); 11 (10); 11 (9) years	• Discontinuation	

1.4.3.30 Manual therapy and exercise versus exercise

Table 31: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Akhter 2014 [5]	12 week interventions **Intervention 1: Manual therapy, Strength and stretching (n=31)** Manual therapy: (Maitland's approach Grade V, High velocity thrust, low amplitude application, rotation/lateral flexion technique on painful and stiff cervical spinal segments in supine position, maximum 6 sessions in 3 weeks). Exercise: regime included a set of strength exercises consisted of isometric, concentric and eccentric exercises with rest in between and a set of stretching exercises of cervical spine and stretches 10 repetitions each.	People with a history of neck pain for 3 months with no related medical dysfunction (n=62) Mean age 38.8 years Mean duration of pain 4.45 years	At 12 weeks (post intervention): • Pain reduction • Physical function	After 3 weeks intervention both groups taught and practiced a home exercise program. A printed exercise sheet was provided with frequency and repetition details: twice a day, 7 days a week, for 3 months. This home exercise program consisted of strength exercises for neck/scapular stability, stretching exercises and general range of motion exercises for neck with advice regarding posture awareness and correction

Study	Intervention and comparison	Population	Outcomes	Comments
	Intervention 2: Strength and flexibility (n=31) Participants performed supervised exercise regime same as the other group, and also followed the same home exercise programme.			
Bronfort 2001 [43]	11 week interventions **Intervention 1: Aerobic & Strength exercise (n=60)** Warm up of stretching and upper body strength followed by 15 to 20 minutes of aerobic exercise using a stationary bike. Resistance exercises were performed on the MedX cervical extension and rotation machines, and resistance was increased periodically, with patients performing approximately 20 repetitions of each exercise. Duration 11 weeks. **Intervention 2: Manual therapy and strength exercise (n=63)** Spinal manipulation therapy and exercise plus strength exercises for the neck and upper body preceded by a short aerobic warm up of the upper body and light stretching. 2 sets of 15-30 repetitions were conducted and resistance was increased gradually over time.	Mechanical neck pain (no specific identified cause) (n=123) Mean age 44.3 years Mean pain duration 5 years	At 11 weeks post intervention and 12 months follow up: • Pain reduction • Physical function • Discontinuation	
El-Gendy 2019 [78]	4 week interventions **Intervention 1: Manual therapy and stretching (n=20)** Myofacial release therapy applied from sitting position after exact determination of the pain location. Superficial stroke massage was performed for 2-3 minutes on the back region to the neck and shoulders in reciprocating and transverse way. Then the therapist focused on the pain region locally and	Chronic mechanical neck pain (n=40) Gender not reported Mean age: 33.9 (5.51); 33.65 (5.7) years Duration of pain not reported	At 4 weeks (post intervention) • Pain • Physical function • Discontinuation	Three armed trial; third arm electrotherapy not included in the analysis

Study	Intervention and comparison	Population	Outcomes	Comments
	applied myofascial release technique. At the end of the treatment session, about 2-3-minute surface stroke massage was performed again. There were 3 sessions per week for 20 minutes. Stretching was also performed as identical to the stretching group (as described). **Intervention 2: Stretching exercise (n=20)** Stretching involved gentle stretching of the pectoral muscle, trapezius muscle, scaleni muscles, levator scapulae muscle, the suboccipital muscle. Also included some strengthening exercises including cervical flexion and extension, shoulder retraction exercise, upright rowing with resistance tubing and push ups if tolerated. Three sessions per week.			
Evans 2002[85]	12 week interventions **Intervention 1: Manual therapy and Strength exercise (n=64)** Spinal manipulation combined with rehabilitative exercise. Spinal manipulation treatment included manual spinal manipulation with light soft tissue massage as facilitate the spinal manipulative therapy. Rehabilitative exercise began each session with a warm up on a stationary bike with arm levers and light stretching, followed by upper body strengthening exercises including push-ups and dumbbell shoulder exercises. Dynamic neck extension, flexion, and rotation exercises were performed with the patient lying on a therapy table wearing headgear with variable weight attachments. Weights were determined by baseline strength performance and were increased gradually during the treatment phase.	Chronic mechanical neck pain for 12 weeks or more (n=127) Mean age 44.7 years Median pain duration 6 years	At 12 weeks (post-intervention) and 2 years (follow up): • Pain reduction • Quality of life • Physical function • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	Each session was 1 hour and there were 20 sessions. **Intervention 2: Strength exercise (n=63)** 20 sessions. Warm up of stretching and aerobic exercise using a stationary bike, followed by strengthening exercises of the shoulders and upper back using variable resistance equipment. Patients were stabilized with torso restraints to isolate and specifically exercise the cervical musculature. They were encouraged to perform repetitions to volitional muscle fatigue (maximum 20 reps) even if the pain was exacerbated, and resistance was increased periodically.			
Evans 2012[86]	12 week interventions **Intervention 1: Manual therapy and Strength exercise (n=91)** Identical exercises as strength intervention (as described) which was preceded by a 15-20 minute session with a licensed chiropractor who administered spinal manipulation therapy. Sessions focused mainly on manual manipulation to the cervical and thoracic spines using high velocity, low amplitude pressure applied to the joints. Up to 5 minutes of light soft tissue massage was also used **Intervention 2: Strength exercise (n=89)** Predominantly upper body and neck exercises that were partially individualised in terms of intensity, according to the participants' abilities. One-on-one supervision in 20 1 hour sessions. The main focus was cervical strength exercises using low-tech methods performed with the patient lying on a therapy table, wearing headgear with variable weight	Chronic nonspecific neck pain for at least 12 weeks (Grade I or II classification according to the Neck Pain Task Force) (n=180) Mean age 46.3 years Mean duration of pain 9.4 years	At 12 weeks (post-intervention) and 52 weeks (follow up): • Pain reduction • Quality of life • Physical function • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	attachments. 3 sets of 15-25 repetitions were conducted. There was also light aerobic warm up (5 minutes) and stretching before and after strength training.			
Lee 2016[163]	10 week interventions **Intervention 1: Manual therapy and strength exercise (n=16)** Thoracic manipulation for 10 minutes plus deep craniocervical flexors training for 15 minutes, plus self-stretching of the levator scapulae and upper trapezius muscles as cool-down exercises for 10 minutes. **Intervention 2: Strength (n=15)** Deep craniocervical flexors training for 25 minutes, with self-stretching of the levator scapulae and upper trapezius muscles as cool-down exercises for 10 minutes. **Intervention 3: Strength (n=15)** Active ROM self-exercise, including neck flexion, extension, lateral flexion, and rotation without provocation of pain) for 35 minutes. *NB Strength interventions were pooled in the analysis*	People with chronic neck pain (n=46) Mean age not reported Mean pain duration not reported	At 10 weeks (post-intervention) • Pain • Physical function	
Panton 2009[210]	16 week interventions **Intervention 1: Manual therapy and strength exercise (n=12)** Exercise as in the strength group (below), plus manual therapy. Participants met twice a week for	Women with fibromyalgia (n=27) Mean age 48.5 years	At 16 weeks (post-intervention): • Quality of life • Physical function • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	exercise, and twice a week for chiropractic treatment. Chiropractic treatment began with 5 minutes of ischemic compression to tender points on the back of the neck and spine. Pressure was applied with thumbs over tender points until the patient reacted to the pressure. The pressure was sustained for 10 seconds. This technique was continued throughout the 16 weeks with increasing pressure until an application of 4kg of digital pressure was reached. This 4kg of pressure was continued until the completion of the study. The next 5 minutes consisted of diversified chiropractic spinal adjustments. These adjustments consisted of short lever, low amplitude, high velocity thrusts. Cervical, thoracic and lumbar adjustments were performed. Target joints were determined at each visit through static and motion palpitation. **Intervention 2: Strength training (n=15)** Resistance training. Participants met twice a week. Resistance training was chosen to maximise strength gains. Participants performed one set of 8-12 repetitions twice a week on 10 exercises. Participants began training at approximately 50% of their initial 1-RM measurement and were slowly progressed to approximately 100% of their initial 1RM by the end of the 16 weeks. Once 12 repetitions were completed on 2 consecutive workouts, weights were increased by 5-10 pounds for upper and lower body respectively.	Mean pain duration 5.5 years		
Toprak celenay 2017[256]	6 week interventions **Intervention 1: Aerobic & Strength exercise (n=24)** The combined exercise programme was carried out 2 days a week for 6 weeks and took 1 hour. It was	Women with fibromyalgia (n=49) Mean age 41 years	At 6 weeks post-intervention): • Discontinuation	

Study	Intervention and comparison	Population	Outcomes	Comments
	composed of 10 minute warm up exercises, 40 minutes aerobic and strengthening exercises including neck, trunk, upper and lower limb muscles. The aerobic exercise consisted of 20 minutes walking on a treadmill. The target heart rate was initially adjusted to 65-70% of the maximal heart rate and to 75-80% of the maximal heart rate in the advanced programme. Muscle strengthening exercises were then performed with elastic resistive bands for 20 minutes where multiple muscles were strengthened. When they performed 15 repetitions without serious pain or fatigue, they progressed to the next colour resistance band. They had 10 repetitions with a holding period of 10 seconds. **Intervention 2: Manual therapy and exercise (n=25)** Connective tissue massage was applied 2 days per week for a total of 12 sessions. While patients were in a sitting position, starting from the lumbosacral region, the lower thoracic, scapular, interscapular, and cervical regions were included in the treatment, respectively. For creating traction between cutaneous tissues, the middle fingers of both hands were used during the application. Each session lasted around 5-20 minutes. Exercise the same as above.	Duration of pain not specified		

1.4.3.31 Manual therapy and exercise versus manual therapy alone

Table 32: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Evans 2002[85]	12 week interventions **Intervention 1: Manual therapy and Strength training (n=64)**	Chronic mechanical neck pain for 12 weeks or more (n=128)	At 12 weeks (post-intervention) and 2 years (follow up): • Pain reduction	

Study	Intervention and comparison	Population	Outcomes	Comments
	Spinal manipulation combined with rehabilitative exercise. Spinal manipulation treatment included manual spinal manipulation with light soft tissue massage as facilitate the spinal manipulative therapy. Rehabilitative exercise began each session with a warm up on a stationary bike with arm levers and light stretching, followed by upper body strengthening exercises including push-ups and dumbbell shoulder exercises. Dynamic neck extension, flexion, and rotation exercises were performed with the patient lying on a therapy table wearing headgear with variable weight attachments. Weights were determined by baseline strength performance and were increased gradually during the treatment phase. Each session was 1 hour and there were 20 sessions. **Intervention 2: Manual therapy (n=64)** Patients received the same spinal manipulation treatment as in the combined treatment group. Duration 11 weeks. Concurrent medication/care: Patients were also given 45 minutes of micronutrient therapy (sham) to minimise the effects of attention bias.	Mean age 44.7 years Median pain duration 6 years	• Quality of life • Physical function • Discontinuation	

1.4.3.32 Exercise versus manual therapy

Table 33: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Evans 2002[85]	12 week interventions **Intervention 1: Strength training (n=61)** 20 sessions. Warm up of stretching and aerobic exercise using a stationary bike, followed by	Chronic mechanical neck pain for 12 weeks or more (n=125) Mean age 44.7 years	At 12 weeks (post-intervention) and 2 years (follow up): • Pain reduction • Quality of life	

Study	Intervention and comparison	Population	Outcomes	Comments
	strengthening exercises of the shoulders and upper back using variable resistance equipment. Patients were stabilized with torso restraints to isolate and specifically exercise the cervical musculature. They were encouraged to perform repetitions to volitional muscle fatigue (maximum 20 reps) even if the pain was exacerbated, and resistance was increased periodically. **Intervention 2: Manual therapy (n=64)** Patients received the same spinal manipulation treatment as in the combined treatment group. Duration 11 weeks. Concurrent medication/care: Patients were also given 45 minutes of micronutrient therapy (sham) to minimise the effects of attention bias.	Median pain duration 6 years	• Physical function • Discontinuation	

See appendix D for full evidence tables.

1.4.4 Quality assessment of clinical studies included in the evidence review

Table 34: Clinical evidence summary: Aerobic exercise versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	40 (1 study) 6 weeks	⨁⨁⊖ MODERATE1 due to risk of bias		The mean pain score in the control group was 62	The mean pain score at in the intervention groups was 21.5 lower (30.38 to 12.62 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
Pain at >3 months (VAS, FIQ pain subscale, 0-100, final values, high is poor outcome)	528 (9 studies) 12-24 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean pain score in the control groups was 66.5	The mean pain score in the intervention groups was 6.97 lower (10.77 to 3.17 lower)
Pain at >3 months (FIQ pain subscale, 0-100, high is poor outcome)	95 (1 study) 18 months	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean pain score in the control groups was 53	The mean pain score in the intervention groups was 1 lower (10.34 lower to 8.34 higher)
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)	372 (5 studies) 12-24 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, imprecision, inconsistency		The mean quality of life score in the control groups was 56.5	The mean quality of life score in the intervention groups was 7.89 lower (13.23 to 2.55 lower)
Quality of life at >3 months (SF-36 functional capacity subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean quality of life score in the control groups was 38	The mean quality of life score in the intervention groups was 12.5 higher (3.85 to 21.15 higher)
Quality of life at >3 months (SF-36 physical appearance subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 13.8	The mean quality of life score in the intervention groups was 16 higher (2.68 lower to 34.68 higher)
Quality of life at >3 months (SF-36 pain subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 29.2	The mean quality of life score in the intervention groups was 7.5 higher (8.62 lower to 23.62 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 30.2	The mean quality of life score in the intervention groups was 7.7 higher (2.49 lower to 17.89 higher)
Quality of life at >3 months (SF-36 social aspects subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 45.4	The mean quality of life score in the intervention groups was 8.9 higher (3.16 lower to 20.96 higher)
Quality of life at >3 months (SF-36 emotional aspects subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 22.4	The mean quality of life score in the intervention groups was 9.7 higher (10.7 lower to 30.1 higher)
Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)	54 (1 study) 16 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 43.4	The mean quality of life score in the intervention groups was 3.4 higher (7.46 lower to 14.26 higher)
Quality of life at ≤3 months (EQ-5D, -0.594-1, high is good outcome, final values)	95 (1 study) 12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 0.5	The mean quality of life score in the intervention groups was 0.03 lower (0.15 lower to 0.09 higher)
Quality of life at >3 months (EQ-5D, -0.594-1, high is good outcome, final values)	259 (2 studies) 9-18 months	⊕⊖⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 0.57	The mean quality of life score in the intervention groups was 0.06 higher (0.01 lower to 0.13 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
Quality of life at ≤3 months (EQ-5D VAS, 0-100. high is good outcome, final values)	95 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 48.3	The mean quality of life score in the intervention groups was 5.6 higher (2.86 lower to 14.06 higher)
Quality of life at >3 months (EQ-5D VAS, 0-100, high is good outcome, final values)	95 (1 study) 18 months	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean quality of life score in the control groups was 51.9	The mean quality of life score in the intervention groups was 1.4 higher (8.17 lower to 10.97 higher)
Physical function at ≤3 months (Final values, timed up and go, seconds, high is good outcome)	60 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 9.99	The mean physical function score in the intervention groups was 0.62 lower (1.40 lower to 0.16 higher)
Physical function at ≤3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)	95 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 40	The mean physical function score in the intervention groups was 3 lower (11.32 lower to 5.32 higher)
Physical function at >3 months (6 minute walking test, final values, metres, high is good outcome)	169 (3 studies) 12-24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 449.8	The mean physical function score in the intervention groups was 56.18 higher (27.8 to 84.56 higher)
Physical function at >3 months (FIQ and SF-36 physical function subscales, 0-100, final values, high is poor outcome)	246 (3 studies) 16-24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 49.9	The mean physical function score in the intervention groups was 10.16 lower (15.39 to 4.94 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
Physical function at >3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)	95 (1 study) 18 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 39	The mean physical function score in the intervention groups was 3 lower (16.14 lower to 10.14 higher)
Psychological distress at >3 months (Change scores and final values, beck depression inventory, 0-21, high is poor outcome)	123 (3 studies) 16-24 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		-	The mean psychological distress score in the intervention groups was 3.36 lower (6.16 to 0.56 lower)
Psychological distress at >3 months (Final values, VAS and FIQ depression scale, 0-10, high is poor outcome)	306 (4 studies) 12-24 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean psychological distress score in the control groups was 4.9	The mean psychological distress in the intervention groups was 0.39 lower (1.05 lower to 0.28 higher)
Psychological distress at >3 months (Final values, VAS and FIQ anxiety scale, Beck anxiety inventory, final values, high is poor outcome)	320 (4 studies) 12-24 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		-	The mean psychological distress score in the intervention groups was 0.28 standard deviations lower (0.51 lower to 0.04 higher)
Psychological distress at >3 months (Change scores, STAI anxiety total scores, high is poor outcome)	50 (1 study) 23 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress change score in the control groups was 4.8	The mean psychological distress score in the intervention groups was 9.7 lower (23.6 lower to 4.2 higher)
Psychological distress at >3 months (final values, FIQ depression scale, 0-10, high is poor outcome)	95 (1 study) 18 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 4.2	The mean psychological distress score in the intervention groups was 0.8 higher (0.46 lower to 2.06 higher)
Psychological distress at >3 months (final values, FIQ anxiety scale, 0-10, high is poor outcome)	95 (1 study) 18 months	⨁⨁⊖⊖ LOW1		The mean psychological distress score in the	The mean psychological distress score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
		due to risk of bias		control groups was 4.8	0.2 higher (1.06 lower to 1.46 higher)
Psychological distress at ≤3 months (Final values, BDI depression scale, high is poor outcome)	60 (1 study) 12 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean psychological distress score in the control groups was 30.14	The mean psychological distress score in the intervention groups was 12.77 lower (14.65 to 10.88 lower)
Use of healthcare services at ≤3 months (Number of GP contacts)	95 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean use of healthcare services in the control groups was 0.5	The mean use of healthcare services in the intervention groups was 1 higher (0.11 lower to 2.11 higher)
Use of healthcare services at >3 months (Number of GP contacts)	95 (1 study) 18 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean use of healthcare services in the control groups was 0.7	The mean use of healthcare services in the intervention groups was 0.3 higher (0.68 lower to 1.28 higher)
Use of healthcare services at ≤3 months (Number of medical specialist contacts)	95 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean use of healthcare services in the control groups was 0.2	The mean use of healthcare services in the intervention groups was 0.1 higher (0.18 lower to 0.38 higher)
Use of healthcare services at >3 months (Number of medical specialist contacts)	95 (1 study) 18 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean use of healthcare services in the control groups was 0.2	The mean use of healthcare services in the intervention groups was 0.2 higher (0.08 lower to 0.48 higher)
Use of healthcare services at ≤3 months (Number of physiotherapist contacts)	95 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of		The mean use of healthcare services in the control groups	The mean use of healthcare services in the intervention groups was 3.1 lower (4.49 to 1.17 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus control (95% CI)
		bias, imprecision		was 3.4	
Use of healthcare services at >3 months (Number of physiotherapist contacts)	95 (1 study) 18 months	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean use of healthcare services in the control groups was 4.8	The mean use of healthcare services in the intervention groups was 4.4 lower (5.79 to 3.01 lower)
Sleep at >3 months (VAS sleep scale, PSQI, FIQ sleep subscale, final values, high is poor outcome)	414 (5 studies) 12-40 weeks	⨁⊖⊖⊖ VERY LOW1,3 due to risk of bias, inconsistency		-	The mean sleep score in the intervention groups was 0.16 standard deviations lower (0.43 lower to 0.1 higher)
Discontinuation	607 (9 studies) 8-24 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision	RD 0.11 (-0.04 to 0.27)	113 per 1000	110 more per 1000 (from 40 fewer to 270 more)

1 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
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 35: Clinical evidence summary: Strength training versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus control (95% CI)
Pain reduction at ≤3 months (final values, VAS, high is poor outcome)	251 (3 studies) 6-12 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain change in the control groups was 54.44	The mean pain score reduction in the intervention groups was 18.85 lower (34.50 to 3.21 lower)
Pain reduction at ≤3 months (change scores and final values, VAS, NRS, 0-100, high is poor outcome)	156 (3 studies) 6-8 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		-	The mean pain score reduction in the intervention groups was 15.76 lower (22.79 to 8.72 lower)
Pain reduction at >3 months (VAS, NRS, 0-100, final values and change scores, high is poor outcome)	449 (4 studies) 21-52 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain change in the control groups was 32	The mean pain score reduction in the intervention groups was 16.06 lower (36.93 lower to 4.82 higher)
Quality of life at ≤3 months (SF-36 physical component summary, 0-100, change scores, high is good outcome)	42 (1 study) 8 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean change quality of life change score in the control groups was 2	The mean quality of life score at 8 in the intervention groups was 7.6 higher (0.25 lower to 15.45 higher)
Quality of life at ≤3 months (SF-36 mental component summary, 0-100, change scores, high is good outcome)	102 (2 studies) 8-16 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life change in the control groups was 8.37	The mean quality of life score at 8-16 in the intervention groups was 3.39 higher (2.43 lower to 9.21 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus control (95% CI)
Quality of life at ≤3 months (FIQ scale, 0-100, final values, high is poor outcome)	52 (2 studies) 8-12 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean quality of life change in the control groups was 62.85	The mean quality of life in the intervention groups was 14.91 lower (45.78 lower to 15.96 higher)
Physical function at ≤3 months (Neck disability index, change scores and final values, 0-100, high is poor outcome)	146 (3 studies) 6-12 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		-	The mean physical function score in the intervention groups was 9.89 lower (23.15 lower to 3.37 higher)
Physical function at ≤3 months (final values, FIQ physical function subscale, Northwick Park Questionnaire, high is poor outcome)	151 (2 studies) 6-12 weeks	⊕⊕⊖⊖ VERY LOW1 due to risk of bias		-	The mean physical function score in the intervention groups was 0 standard deviations higher (0.33 lower to 0.32 higher)
Physical function at ≤3 months (6 minute walking test, metres, final values, high is good outcome)	20 (1 study) 12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 538.3m	The mean physical function score in the intervention groups was 8.4m lower (89.59 lower to 72.79 higher)
Physical function at >3 months months (final values, Northwick Park Questionnaire, Neck Disability Index, high is poor outcome)	163 (2 studies) 16-24 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, imprecision, inconsistency		-	The mean physical function score in the intervention groups was 0.23 standard deviations lower (0.68 lower to 1.14 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus control (95% CI)
Physical function at >3 months (change scores, SF-36 physical function subscale, HAQ, 0-100, high is poor outcome)	105 (3 studies) 16-21 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function change score in control groups was -0.56	The mean physical function score in the intervention groups was 6.2 lower (10.41 to 2 lower)
Psychological distress at ≤3 months (Pain Catastrophising Scale, 0-100, high is poor outcome)	25 (1 study) 8 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress in the control group was +20	The mean psychological distress score in the intervention groups was 9 lower (19.70 lower to 1.70 higher)
Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)	21 (1 study) 21 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean psychological distress change score in the control groups was +0.9	The mean psychological distress score in the intervention groups was 3.7 lower (6.37 to 1.03 lower)
Use of health care services at >3 months	179 (1 study) 52 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision	RR 0.68 (0.42 to 1.11)	333 per 1000	107 fewer per 1000 (from 193 fewer to 37 more)
Sleep at >3 months (VAS sleep, 0-100, change scores, high is poor outcome)	21 (1 study) 21 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean sleep change score in the control groups was -3	The mean sleep score at 21 in the intervention groups was 7 lower (20.9 lower to 6.9 higher)
Discontinuation at ≤3 months	133 (4 studies) 8-12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	OR 2.27 (0.77 to 6.73)	65 per 1000	71 more per 1000 (from 14 fewer to 254 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus control (95% CI)
Discontinuation at >3 months	252 (4 studies) 16-24 weeks	⨁⨁⨁⊖ MODERATE1,2 due to risk of bias	RD 0.08 (-0.02 to 0.17	33 per 1000	33 fewer per 1000 (from 27 fewer to 34 fewer)

1 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
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 36: Clinical evidence summary Aerobic and strength versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus control (95% CI)
Pain at ≤3 months (VAS, 0-100, change scores, high is poor outcome)	129 (2 studies) 10-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain change score in the control groups was 0.5	The mean pain score in the intervention groups was 2.45 lower (34.16 lower to 29.27 higher)
Pain at >3 months (VAS, FIQ pain subscale, 0-100, final values, high is poor outcome)	161 (3 studies) 18 weeks - 3 years	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean pain final values in the control groups was 56.83	The mean pain score in the intervention groups was 13.74 lower (22.11 to 5.37 lower)
Quality of life at ≤3 months (EQ-5D, -0.594 to 1, final values, high is poor outcome)	30 (1 study) 3 months	⨁⨁⊖⊖ LOW1,2 due to risk of		The mean quality of life in the control	The mean quality of life score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus control (95% CI)
		bias, imprecision		groups was 0.334	0.25 higher (0.05 to 0.45 higher)
Quality of life at ≤3 months (Fibromyalgia impact questionnaire, 0-100, final values, high is poor outcome)	54 (2 studies) 8 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 62.9	The mean quality of life score in the intervention groups was 3.42 lower (12.66 lower to 5.82 higher)
Quality of life at >3 months (Fibromyalgia impact questionnaire, 0-100, final values and change scores, high is poor outcome)	171 (4 studies) 16-52 weeks	⨁⨀⨀⨀ VERY LOW1,2,3 due to risk of bias, imprecision, inconsistency		-	The mean quality of life score in the intervention groups was 9.05 lower (15.43 to 2.68 lower)
Quality of life at >3 months (EQ-5D, -0.594 to 1, final values, high is poor outcome)	30 (1 study) 8 months	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life in the control groups was 0.334	The mean quality of life score in the intervention groups was 0.19 higher (0.00 to 0.39 higher)
Quality of life at >3 months (SF-36 physical functioning subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 45.2	The mean quality of life score in the intervention groups was 11.6 higher (2.02 to 21.18 higher)
Quality of life at >3 months (SF-36 physical role subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 19.4	The mean quality of life score in the intervention groups was 1.9 higher (14.93 lower to 18.73 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus control (95% CI)
Quality of life at >3 months (SF-36 emotional role subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 52.1	The mean quality of life score in the intervention groups was 19 higher (6.96 lower to 44.96 higher)
Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 28.6	The mean quality of life score in the intervention groups was 12.7 higher (2.73 to 22.67 higher)
Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 44.2	The mean quality of life in the intervention groups was 15.8 higher (3.75 to 27.85 higher)
Quality of life at >3 months (SF-36 social role subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 52.2	The mean quality of life in the intervention groups was 11.7 higher (1.9 lower to 25.3 higher)
Quality of life at >3 months (SF-36 bodily pain subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 19.5	The mean quality of life in the intervention groups was 10.4 higher (0.16 lower to 20.96 higher)
Quality of life at >3 months (SF-36 general health subscale, 0-100, final values, high is good outcome)	42 (1 study) 24 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 33.5	The mean quality of life score in the intervention groups was 9.6 higher (2.82 to 16.38 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus control (95% CI)
Physical function at >3 months (seconds, quarter mile walk test, final values, high is poor outcome)	16 (1 study) 18 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 320.15	The mean physical function score in the intervention groups was 37.3 lower (63.19 to 11.41 lower)
Physical function at >3 months (metres, 6-minute walk test, final values, high is good outcome)	37 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 459.07	The mean physical function score in the intervention groups was 54.8 higher (0.54 lower to 110.14 higher)
Physical function at >3 months (FIQ physical function subscale, 0-10, final values, high is poor outcome)	30 (1 study) 32 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 3.7	The mean physical function score in the intervention groups was 1.3 lower (2.63 lower to 0.03 higher)
Physical function at ≤3 months (metres, 6-minute walk test, high is good outcome)	32 (1 study) 8 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 12.21	The mean physical function score in the intervention groups was 15.69 higher (33.37 lower to 64.75 higher)
Psychological distress at ≤3 months (BDI, 0-30, final values, high is poor outcome)	54 (2 studies) 8 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 21.03	The mean psychological distress score in the intervention groups was 1.44 lower (6.85 lower to 3.97 higher)
Psychological distress at ≤3 months (State anxiety inventory, 0-10, change scores, high is poor outcome)	58 (1 study) 8 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of		The mean psychological distress change	The mean psychological distress score in the intervention groups was 0.1 higher (5.12 lower to 5.32 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus control (95% CI)
		bias, imprecision		in the control groups was -0.4	
Psychological distress at ≤3 months (HADS anxiety, 0-21, high is poor outcome)	32 (1 study) 8 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 11.9	The mean psychological distress score in the intervention groups was 1.25 lower (3.77 lower to 1.27 higher)
Psychological distress at >3 months (CES-D, BDI, FIQ depression subscale, final values, high is poor outcome)	125 (4 studies) 18-32 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		-	The mean psychological distress score in the intervention groups was 0.45 standard deviations lower (0.81 to 0.09 lower)
Psychological distress at >3 months (State anxiety inventory, 20-80, final values and change scores, high is poor outcome)	83 (2 studies) 16-32 weeks	⨁⨀⨀⨀ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		-	The mean psychological distress score in the intervention groups was 2.95 lower (9.75 lower to 3.85 higher)
Sleep at >3 months (Pittsburgh sleep quality index, high is poor outcome, change scores, 0-21)	58 (1 study) 16 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean sleep change in the control groups was +0.5	The mean sleep score in the intervention groups was 2.2 lower (3.39 to 1.01 lower)
Healthcare utilisation at >3 months	78 (1 study) 3 years	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision	RR 0.85 (0.49 to 1.47)	476 per 1000	71 fewer per 1000 (from 243 fewer to 224 more)
Discontinuation at ≤3 months	125 (4 studies) 8-12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of	RD 0 (-0.01 to 0.17)	17 per 1000	0 more per 1000 (from 10 fewer to 170 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus control (95% CI)
		bias, imprecision			
Discontinuation at >3 months	230 (7 studies) 16-32 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RD 0.02 (-0.05 to 0.09)	49 per 1000	49 more per 1000 (from 43 fewer to 50 more)

1 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
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 37: Clinical evidence summary: Aerobic, strength and flexibility versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic, strength and flexibility versus control (95% CI)
Quality of life at ≤3 months (SF-36 mental component, 0-100, final values, high is good outcome)	25 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 32.9	The mean quality of life score in the intervention groups was 12.1 higher (2.14 to 22.06 higher)
Quality of life at ≤3 months (SF-36 physical component, 0-100, final values, high is good outcome)	25 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 31.3	The mean quality of life score in the intervention groups was 5.1 higher (3.18 lower to 13.38 higher)

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

Table 38: Clinical evidence summary: Strength and flexibility versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus control (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	110 (2 studies) 2-12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 52.8	The mean pain score at 2-12 in the intervention groups was 11.71 lower (21.49 to 1.92 lower)
Pain at >3 months (VAS, SF-36 pain score, final values, 0-100, high is poor outcome)	144 (2 studies) 24 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 50.45	The mean pain score in the intervention groups was 13.19 lower (20.33 to 6.05 lower)
Quality of life at ≤3 months (SF-36 mental component, 0-100, final values, high is poor outcome)	70 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 49.8	The mean quality of life score in the intervention groups was 0.6 lower (6.12 lower to 4.92 higher)
Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is poor outcome)	144 (2 studies) 24 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 45	The mean quality of life score in the intervention groups was 1.78 higher (1.35 lower to 4.91 higher)
Quality of life at ≤3 months (SF-36 physical component, 0-100, final values, high is poor outcome)	70 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 28.6	The mean quality of life score in the intervention groups was 1.7 higher (2.42 lower to 5.82 higher)
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is poor outcome)	144 (2 studies) 24 weeks	⊕⊕⊕⊖ LOW1,3 due to risk of		The mean quality of life score in the	The mean quality of life score in the intervention groups was 0.16 lower (3.87 lower to 3.56 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus control (95% CI)
		bias, inconsistency		control groups was 37.3	
Physical function at ≤3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)	70 (1 study) 12 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias, imprecision		The mean physical function score in the control groups was 39.1	The mean physical function score in the intervention groups was 5.5 lower (16.59 lower to 5.59 higher)
Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)	144 (2 studies) 24 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean physical function score in the control groups was 39.7	The mean physical function score in the intervention groups was 6.7 lower (12.3 to 1.1 lower)
Psychological distress at ≤3 months (ADS depression scale, 0-60, final values, high is poor outcome)	70 (1 study) 12 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 18.6	The mean psychological distress score in the intervention groups was 1.6 higher (2.59 lower to 5.79 higher)
Psychological distress at >3 months (ADS depression scale, 0-60, final values, high is poor outcome)	70 (1 study) 24 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 19.8	The mean psychological distress score in the intervention groups was 1.1 higher (3.41 lower to 5.61 higher)
Discontinuation at >3 months	157 (2 studies) 9-24 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	OR 0.88 (0.32 to 2.4)	117 per 1000	13 fewer per 1000 (from 76 fewer to 124 more)

1 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
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 39: Clinical evidence summary: Strength, proprioception and flexibility versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength, proprioception and flexibility versus control (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	76 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 41.8	The mean pain score in the intervention groups was 16.6 lower (25.8 to 7.4 lower)
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)	76 (1 study) 24 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 44.6	The mean pain score in the intervention groups was 11.5 lower (20.71 to 2.29 lower)
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	76 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 42.9	The mean quality of life score in the intervention groups was 2.3 higher (0.13 lower to 4.73 higher)
Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	76 (1 study) 24 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 42	The mean quality of life score in the intervention groups was 2 higher (1.48 lower to 5.48 higher)
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	76 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 46.1	The mean quality of life score in the intervention groups was 1.6 higher (2.73 lower to 5.93 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength, proprioception and flexibility versus control (95% CI)
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	76 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 46.4	The mean quality of life score in the intervention groups was 0.5 higher (3.82 lower to 4.82 higher)
Psychological distress at ≤3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	76 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 6.7	The mean psychological distress score in the intervention groups was 1.2 lower (2.68 lower to 0.28 higher)
Psychological distress at >3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	76 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 6.7	The mean psychological distress score in the intervention groups was 1.2 lower (2.66 lower to 0.26 higher)
Psychological distress at ≤3 months (HADS: depression, 0-21, final values, high is poor outcome)	76 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 4.9	The mean psychological distress score in the intervention groups was 1.1 lower (2.4 lower to 0.2 higher)
Psychological distress at >3 months (HADS: depression, 0-21, final values, high is poor outcome)	76 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 5.4	The mean psychological distress score in the intervention groups was 1.3 lower (2.85 lower to 0.25 higher)
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	76 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 27.5	The mean physical function in the intervention groups was 4.8 lower (9.47 to 0.13 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength, proprioception and flexibility versus control (95% CI)
Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)	76 (1 study) 24 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 29.4	The mean physical function in the intervention groups was 4.3 lower (10.06 lower to 1.46 higher)
Discontinuation at ≤3 months	76 (1 study) 12 weeks	⨁⨁⊖⊖ LOW2 due to imprecision	RR 1.37 (0.69 to 2.73)	256 per 1000	95 more per 1000 (from 79 fewer to 443 more)

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

Table 40: Clinical evidence summary: Proprioception versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Proprioception versus control (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	46 (1 study) 12 weeks	⨁⨁⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 5.63	The mean pain score in the intervention groups was 0.18 higher (1.09 lower to 1.45 higher)
Pain at >3 months (VAS, 0-10, final values, high is poor outcome)	46 (1 study) 24 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 6.36	The mean pain score in the intervention groups was 0.97 lower (2.47 lower to 0.53 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Proprioception versus control (95% CI)
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	46 (1 study) 12 weeks	⨁⨁⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 50.17	The mean quality of life score in the intervention groups was 1.88 lower (11.11 lower to 7.35 higher)
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)	46 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 52.96	The mean quality of life score in the intervention groups was 3.59 lower (14.37 lower to 7.19 higher)
Physical function at ≤3 months (sit to stand test, final values, high is good outcome)	46 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 28.59	The mean physical function score in the intervention groups was 4.38 lower (14.37 lower to 7.19 higher)
Physical function at >3 months (sit to stand test, final values, high is good outcome)	46 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 25.77	The mean physical function score in the intervention groups was 0.86 lower (3.18 lower to 1.46 higher)
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	46 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 13.95	The mean psychological distress score in the intervention groups was 4.74 lower (8.43 to 1.05 lower)
Psychological distress at >3 months (BDI, 0-61, final values, high is poor outcome)	46 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control	The mean psychological distress score in the intervention groups was 4.86 lower (9.84 lower to 0.12 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Proprioception versus control (95% CI)
				groups was 14.86	
Discontinuation at >3 months	50 (1 study) 24 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision	RR 0.33 (0.04 to 2.99)	120 per 1000	80 fewer per 1000 (from 115 fewer to 239 more)

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

Table 41: Clinical evidence summary: Mind-body exercise versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body exercises versus control (95% CI)
Pain at ≤3 months (VAS, Visual numeric scale, FIQ pain subscale, 0-100, final values and change scores, high is poor outcome)	393 (8 studies) 7-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain score in the control groups was 50.3	The mean pain score in the intervention groups was 11.17 lower (17.32 to 5.02 lower)
Pain improvement at ≤3 months (30% improvement on NRS)	117 (1 study) 8 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias	RR 3.19 (1.56 to 6.52)	159 per 1000	348 more per 1000 (from 89 more to 878 more)
Pain improvement at >3 months (30% improvement on NRS)	117 (1 study) 24 weeks	⨁⊖⊖⊖ VERY LOW1,3 due to risk of	RR 2.11 (1.06 to 4.21)	182 per 1000	202 more per 1000 (from 11 more to 584 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body exercises versus control (95% CI)
		bias, imprecision			
Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome) - Fibromyalgia	80 (1 study) 32 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean pain score in the control groups was 73	The mean pain score in the intervention groups was 26 lower (35.63 to 16.37 lower)
Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome) - Chronic neck pain	221 (3 studies) 24 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean pain score in the control groups was 48.5	The mean pain score in the intervention groups was 11.29 lower (174219.52 to 5.17 lower)
Quality of life at ≤3 months (WHOQOL-BREF, 0-5, final values, high is good outcome)	57 (1 study) 7 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 2.79	The mean quality of life score in the intervention groups was 0.58 higher (0.16 to 1 higher)
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	106 (3 studies) 8-14 weeks	⨁⨀⨀⨀ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean quality of life score in the control groups was 49.3	The mean quality of life score in the intervention groups was 1.55 lower (13.36 lower to 10.25 higher)
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	220 (3 studies) 10-12 weeks	⨁⨁⨁⨀ MODERATE1,3 due to risk of bias		The mean quality of life score in the control groups was 37.3	The mean quality of life score in the intervention groups was 4.14 higher (2.15 to 6.12 higher)
Quality of life at ≤3 months (SF-36 mental component summary score,	220 (3 studies)	⨁⨀⨀⨀ VERY		The mean quality of life	The mean quality of life score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body exercises versus control (95% CI)
0-100, final values, high is good outcome)	10-12 weeks	LOW1,2,3 due to risk of bias, inconsistency, imprecision		score in the control groups was 45.6	2.33 higher (2.57 lower to 7.24 higher)
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is poor outcome)	221 (3 studies) 24 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean quality of life score in the control groups was 43.3	The mean quality of life score in the intervention groups was 1.64 lower (11.62 lower to 8.33 higher)
Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is poor outcome)	221 (3 studies) 24 weeks	⊕⊕⊖⊖ LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 34.2	The mean quality of life score in the intervention groups was 0.69 higher (2.05 lower to 3.43 higher)
Quality of life at >3 months (SF-36, 0-100, functional capacity scale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 39.1	The mean quality of life score in the intervention groups was 17.2 higher (8.01 to 26.39 higher)
Quality of life at >3 months (SF-36, 0-100, physical aspects subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 13.8	The mean quality of life score in the intervention groups was 22.7 higher (9.73 to 35.67 higher)
Quality of life at >3 months (SF-36, 0-100, pain subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 29.1	The mean quality of life score in the intervention groups was 16.9 higher (9.19 to 24.61 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body exercises versus control (95% CI)
Quality of life at >3 months (SF-36, 0-100, vitality subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 37.1	The mean quality of life score in the intervention groups was 10.5 higher (0.5 to 20.5 higher)
Quality of life at >3 months (SF-36, 0-100, general health subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 41.5	The mean quality of life score in the intervention groups was 3.4 higher (4.81 lower to 11.61 higher)
Quality of life at >3 months (SF-36, 0-100, social subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 51.3	The mean quality of life score in the intervention groups was 5.9 higher (5.61 lower to 17.41 higher)
Quality of life at >3 months (SF-36, 0-100, emotional subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 31.5	The mean quality of life score in the intervention groups was 20.4 higher (3.24 to 37.56 higher)
Quality of life at >3 months (SF-36, 0-100, mental health subscale, final values, high is good outcome)	80 (1 study) 32 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 46.2	The mean quality of life score in the intervention groups was 6.1 higher (3.42 lower to 15.62 higher)
Physical function at >3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)	363 (7 studies) 32 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias,		-	The mean physical function score in the intervention groups was 0.40 standard deviations lower (0.84 to 0.04 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body exercises versus control (95% CI)
		inconsistency, imprecision			
Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)	225 (3 studies) 32 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean physical function score in the control groups was 36.3	The mean physical function score in the intervention groups was 6.79 lower (10.57 to 3.01 lower)
Physical function at >3 months (6 minute walk test, metes, final values, high is good outcome)	80 (1 study) 32 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean physical function score in the control groups was 343	The mean physical function score in the intervention groups was 88 higher (51.42 to 124.58 higher)
Psychological distress at ≤3 months (HADS:D, Beck depression inventory, CES-D, ADS depression scale, final values, high is poor outcome)	306 (5 studies) 7-12 weeks	⨁⨀⨀⨀ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		-	The mean psychological distress score in the intervention groups was 0.51 standard deviations lower (0.96 to 0.05 lower)
Psychological distress at ≤3 months (State trace anxiety inventory, final values, high is poor outcome) - Fibromyalgia	57 (1 study) 7 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 51.7	The mean psychological distress score in the intervention groups was 9.91 lower (15.59 to 4.23 lower)
Psychological distress at ≤3 months (HADS:A, final values, high is poor outcome) - Chronic neck pain	77 (2 study) 12 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 6.7	The mean psychological distress score in the intervention groups was 0.2 lower (2 lower to 1.6 higher)
Psychological distress at >3 months (Beck depression inventory,	223 (3 studies)	⨁⨁⨁⨀ MODERATE1		-	The mean psychological distress score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body exercises versus control (95% CI)
HADS:D, final values, high is poor outcome)	24-32 weeks	due to risk of bias			0.02 standard deviations lower (0.29 lower to 0.24 higher)
Psychological distress at >3 months (HADS:A, 0-21, final values, high is poor outcome)	77 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 6.7	The mean psychological distress score in the intervention groups was 0.6 lower (2.38 lower to 1.18 higher)
Sleep at ≤3 months (VAS sleep outcome, pittsburgh sleep quality index, final values, high is poor outcome)	60 (2 studies) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, inconsistency, imprecision		-	The mean sleep score in the intervention groups was 0.43 standard deviations lower (1.58 lower to 0.72 higher)
Discontinuation at >3 months	784 (12 studies) 8-32 weeks	⨁⨀⨀⨀ VERY LOW1,2, 3 due to risk of bias, imprecision, inconsistency	RD 0.03 (-0.03 to 0.10)	77 per 1000	30 more per 1000 (from 30 fewer to 100 more)

1 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
2 Downgraded for heterogeneity, unexplained by subgroup analysis
3 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 42: Clinical evidence summary: Flexibility versus usual care

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Flexibility versus control (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	28 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 64	The mean pain score in the intervention groups was 18 lower (37.89 lower to 1.89 higher)
Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)	28 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 10.5	The mean physical function score in the intervention groups was 1.5 lower (5.39 lower to 2.39 higher)
Discontinuation at ≤3 months	34 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	OR 8.41 (0.81 to 86.84)	0 per 1000	180 more per 1000 (from 20 more to 370 more)

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

Table 43: Clinical evidence summary: Aerobic exercise versus strength

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus strength (95% CI)
Pain at ≤3 months (VAS, FIQ pain subscale, MDPI, 0-100, final values and change scores, high is poor outcome)	199 (4 studies) 3-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias,		-	The mean pain score in the intervention groups was 4.47 lower (20.48 lower to 11.54 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus strength (95% CI)
		inconsistency, imprecision			
Pain at >3 months (VAS, 0-100, change scores, high is poor outcome)	60 (1 study) 16 weeks	⨁⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean pain change score in the control groups was -27.7	The mean pain score in the intervention groups was 6.7 lower (16.22 lower to 2.82 higher)
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values and change scores, high is good outcome)	127 (3 studies) 8-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		-	The mean quality of life score in the intervention groups was 4.29 higher (8.4 lower to 16.98 higher)
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values and change scores, high is good outcome)	127 (3 studies) 8-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		-	The mean quality of life score in the intervention groups was 4.69 higher (6.6 lower to 15.97 higher)
Physical function at ≤3 months (Multidimensional fatigue inventory-20 reduced activity subscale, change scores, 0-20, high is poor outcome)	26 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision		The mean physical function change score in the control groups was -1.3	The mean physical function score in the intervention groups was 1 higher (1.18 lower to 3.18 higher)
Physical function at ≤3 months (6 minute walking test, metres, final values, high is good outcome)	75 (1 study) 12 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean physical function score in the control groups was 628.8	The mean physical function score at 12 weeks (6 minute walking test, metres, high is good outcome) in the intervention groups was 88.4 lower (114.7 to 62.1 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus strength (95% CI)
Physical function at ≤3 months (Final values and change scores, SF-36 physical functioning subscale, 0-100, high is good outcome)	86 (2 studies) 8-16 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		-	The mean physical function score in the intervention groups was 1.85 higher (3.79 lower to 7.49 higher)
Psychological distress at ≤3 months (Hospital anxiety and depression anxiety score, 0-21, final values and change scores, high is poor outcome)	52 (2 studies) 8-12 weeks	⨁⨀⨀⨀ VERY LOW1,3 due to risk of bias, imprecision		-	The mean psychological distress score in the intervention groups was 0.93 lower (2.46 lower to 0.61 higher)
Psychological distress at ≤3 months (Final values and change scores, Hospital anxiety and depression scale, depression score, 0-21, high is poor outcome)	52 (2 studies) 8-12 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		-	The mean psychological distress score in the intervention groups was 0.04 higher (1.37 lower to 1.46 higher)
Psychological distress at ≤3 months (Final values, BDI, 0-60, high is poor outcome)	75 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,3 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 9.9	The mean psychological distress score in the intervention groups was 12.7 higher (9.01 to 16.39 higher)
Sleep at ≤3 months (VAS Sleep scale, 0-100, final values, high is poor outcome)	26 (1 study) 8 weeks	⨁⨀⨀⨀ VERY LOW1,3 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 25.8	The mean sleep score in the intervention groups was 13.3 lower (31.93 lower to 5.33 higher)
Discontinuation at ≤3 months (due to other diagnoses, transportation problems)	196 (4 studies) 3-16 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision	RR 0.67 (0.32 to 1.4)	150 per 1000	49 fewer per 1000 (from 102 fewer to 60 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus strength (95% CI)
1 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 2 Downgraded for heterogeneity, unexplained by subgroup analysis 3 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.					

Table 44: Clinical evidence summary: Aerobic exercise versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus flexibility (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	60 (1 study) 10 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 47	The mean pain score in the intervention groups was 3 higher (10.19 lower to 16.19 higher)
Pain at >3 months (VAS, 0-100, final values and change scores, high is poor outcome)	94 (2 studies) 20 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		-	The mean pain score in the intervention groups was 12.65 lower (22.45 to 2.84 lower)
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	60 (1 study) 10 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 42.55	The mean quality of life score in the intervention groups was 2.82 higher (1.29 lower to 6.93 higher)
Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	60 (1 study) 20 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 42.82	The mean quality of life score in the intervention groups was 2.55 higher (2.08 lower to 7.18 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus flexibility (95% CI)
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	60 (1 study) 10 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 39.87	The mean quality of life score in the intervention groups was 4.26 higher (1.69 lower to 10.21 higher)
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	60 (1 study) 20 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 40.09	The mean quality of life score in the intervention groups was 7.91 higher (2.43 to 13.39 higher)
Psychological distress at ≤3 months (BDI, 0-21, high is poor outcome)	60 (1 study) 10 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 13.56	The mean psychological distress score in the intervention groups was 0.44 higher (6.83 lower to 7.71 higher)
Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)	60 (1 study) 20 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 12.15	The mean psychological distress score in the intervention groups was 0.74 lower (4.53 lower to 3.05 higher)
Psychological distress at ≤3 months (State trace anxiety inventory, 0-100, final values, high is poor outcome)	60 (1 study) 10 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 47.4	The mean psychological distress score in the intervention groups was 1.83 lower (6.33 lower to 2.67 higher)
Psychological distress at >3 months (State trace anxiety inventory, 0-100, final values, high is poor outcome)	60 (1 study) 20 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of		The mean psychological distress score in the control	The mean psychological distress score in the intervention groups was 4.83 lower (9.22 to 0.44 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic exercise versus flexibility (95% CI)
		bias, imprecision		groups was 45.04	
Discontinuation at >3 months	76 (1 study) 20 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.67 (0.67 to 4.13)	158 per 1000	106 more per 1000 (from 52 fewer to 495 more)

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

Table 45: Clinical evidence summary: Aerobic exercise versus biomechanical exercise

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus aerobic (95% CI)
Pain at ≤3 months (VAS, 0-10, high score is poor outcome)	42 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was **6.2**	The mean pain score in the intervention groups was **0.6 lower** (1.79 lower to 0.59 higher)
Psychological distress at ≤3 months (Scale of Catastropic Thoughts on Pain, 0-5, high score is poor outcome)	42 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 2.5	The mean pain score in the intervention groups was 0.2 lower (1.08 lower to 0.68 higher)
Quality of life at ≤3 months (SF36 role social subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of		The mean quality of life score in the control groups	The mean quality of life score in the intervention groups was 10.6 lower (27.34 lower to 6.14 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus aerobic (95% CI)
		bias, imprecision		was **64.2**	
Quality of life at ≤3 months (SF36 general health status subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 39	The mean quality of life score in the intervention groups was 2 lower (15.89 lower to 11.89 higher)
Quality of life at ≤3 months (SF36 vitality subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 43.8	The mean quality of life score in the intervention groups was 1.2 lower (12.43 lower to 10.03 higher)
Quality of life at ≤3 months (SF36 functional capacity subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 43.5	The mean quality of life score in the intervention groups was 9.6 lower (21.76 lower to 2.56 higher)
Quality of life at ≤3 months (SF36 role physical subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 36.2	The mean quality of score in the intervention groups was 14.3 lower (35.85 lower to 7.25 higher)
Quality of life at ≤3 months (SF36 emotional aspects subscale, 0-100, high score is good outcome)	42 (1 study)	⨁⊖⊖⊖ VERY LOW1,2 due to risk of		The mean quality of life score in the	The mean quality of life score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus aerobic (95% CI)
	12 weeks	bias, imprecision		control groups was 43.6	9 lower (34.66 lower to 16.66 higher)
Quality of life at ≤3 months (SF36 pain subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 44.9	The mean quality of life score in the intervention groups was 7 lower (18.72 lower to 4.72 higher)
Quality of life at ≤3 months (SF36 mental health subscale, 0-100, high score is good outcome)	42 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 65.9	The mean quality of life score in the intervention groups was 10.9 lower (25.37 lower to 3.57 higher)
Sleep at ≤3 months (Pittsburgh Sleep Quality Index, 0-21, high score is poor outcome)	42 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean sleep score in the control groups was 9.9	The mean sleep score in the intervention groups was 0.4 lower (2.64 lower to 1.84 higher)
Discontinuation at ≤3 months	42 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision	RR 0.50 (0.10 to 2.44)	190 per 1000	95 fewer per 1000 (from 171 fewer to 274 more)

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

Table 46: Clinical evidence summary: Aerobic and strength versus aerobic exercise

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus aerobic (95% CI)
Quality of life at >3 months (FIQ, 0-100, change scores, high is poor outcome)	43 (1 study) 24 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score change in the control groups was -8.8	The mean quality of life in the intervention groups was 0 higher (7.78 lower to 7.78 higher)
Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)	43 (1 study) 24 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean psychological distress change in the control groups was -8.5	The mean psychological distress in the intervention groups was 2.1 higher (1.66 lower to 5.86 higher)
Discontinuation at >3 months	43 (1 study) 24 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.05 (0.3 to 3.66)	182 per 1000	9 more per 1000 (from 127 fewer to 484 more)

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

Table 47: Clinical evidence summary: Aerobic and strength versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus flexibility (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	85 (1 study) 6 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 47	The mean pain score in the intervention groups was 4 lower (9.96 lower to 1.96 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus flexibility (95% CI)
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)	76 (1 study) 18 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 42	The mean pain score in the intervention groups was 8 lower (13.89 to 2.11 lower)
Quality of life at ≤3 months (NIH CPSI quality of life subscale, 0-12, final values, high is poor outcome)	85 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 6.9	The mean quality of life score in the intervention groups was 1.8 lower (2.69 to 0.91 lower)
Quality of life at >3 months (NIH CPSI quality of life subscale, 0-12, final values, high is poor outcome)	76 (1 study) 18 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 6.2	The mean quality of life score in the intervention groups was 1.8 lower (2.68 to 0.92 lower)
Psychological distress at ≤3 months (BDI, 0-21, final values, high is poor outcome)	85 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 9.3	The mean psychological distress score in the intervention groups was 0.5 higher (1.33 lower to 2.33 higher)
Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)	76 (1 study) 18 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 7.8	The mean psychological distress score in the intervention groups was 0.5 higher (0.97 lower to 1.97 higher)
Discontinuation at ≤3 months	103 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.96 (0.72 to 5.34)	98 per 1000	94 more per 1000 (from 27 fewer to 425 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and strength versus flexibility (95% CI)
1 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 2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.					

Table 48: Clinical evidence summary: Aerobic and flexibility versus mind-body exercise

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and flexibility versus mind-body (95% CI)
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)	111 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean quality of life change in the control groups was +3.3	The mean quality of life score in the intervention groups was 1.5 lower (4.65 lower to 1.65 higher)
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)	111 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life change in the control groups was +3.8	The mean quality of life score in the intervention groups was 3.2 lower (6.38 to 0.02 lower)
Quality of life at >3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)	111 (1 study) 12 months	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean quality of life change in the control groups was +5.4	The mean quality of life score in the intervention groups was 2.8 lower (6.65 lower to 1.05 lower)
Quality of life at >3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)	111 (1 study) 12 months	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean quality of life change in the control groups was +5.4	The mean quality of life score in the intervention groups was 2.4 lower (7.88 lower to 3.08 higher)
Physical function at ≤3 months (6 minute walking test change scores,	111 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1		The mean physical function change in	The mean physical function score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and flexibility versus mind-body (95% CI)
metres, change scores, high is good outcome)		due to risk of bias		the control groups was +7.4	1.9 higher (25.15 lower to 28.95 higher)
Physical function at >3 months (6 minute walking test change scores, metres, change scores, high is good outcome)	111 (1 study) 12 months	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean physical function change in the control groups was +30.2	The mean physical function score in the intervention groups was 22.2 lower (60.46 lower to 16.06 higher)
Psychological distress at ≤3 months (HADS: depression, 0-21, change scores, high is poor outcome)	111 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean psychological distress change in the control groups was -1.7	The mean psychological distress score in the intervention groups was 1.2 higher (0.68 lower to 3.08 higher)
Psychological distress at ≤3 months (HADS: anxiety, 0-21, change scores, high is poor outcome)	111 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress change in the control groups was -1.6	The mean psychological distress score in the intervention groups was 1.8 higher (0.4 to 3.2 higher)
Psychological distress at >3 months (HADS: anxiety, 0-21, change scores, high is poor outcome)	111 (1 study) 12 months	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress change in the control groups was -2.2	The mean psychological distress score in the intervention groups was 1.8 higher (0.12 lower to 3.48 higher)
Psychological distress at >3 months (HADS: depression, 0-21, change scores, high is poor outcome)	111 (1 study) 12 months	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean psychological distress change in the control groups was -2.2	The mean psychological distress score in the intervention groups was 1.6 higher (0.86 lower to 4.06 higher)
Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)	111 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean sleep change in the control groups was -1.6	The mean sleep score in the intervention groups was 0.7 higher (0.74 lower to 2.14 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Aerobic and flexibility versus mind-body (95% CI)
Sleep at >3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)	111 (1 study) 12 months	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean sleep change in the control groups was -2	The mean sleep score in the intervention groups was 0.8 higher (1.14 lower to 2.74 higher)
Discontinuation at ≤3 months	111 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.35 (0.71 to 2.57)	227 per 1000	79 more per 1000 (from 66 fewer to 356 more)

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

Table 49: Clinical evidence summary: Aerobic exercise and flexibility versus aerobic exercise

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus mind-body exercises (95% CI)
Pain perception at <3 months (Final score; VAS)	64 (1 study) 4 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean pain score in the control groups was 7.33	The mean pain perception score in the intervention groups was 0.65 lower (0.86 to 0.44 lower)
Pain perception at >3 months (Final score; VAS)	64 (1 study) 12 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean pain score in the control groups was 6.74	The mean pain perception score in the intervention groups was 0.94 lower (1.14 to 0.74 lower)
Quality of life at <3 months (final score; FIQ)	64 (1 study)	⨁⨁⨁⊖ MODERATE1		The mean quality of life	The mean quality of life score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus mind-body exercises (95% CI)
	4 weeks	due to risk of bias		score in the control groups was 69.81	5.49 lower (7.46 to 3.52 lower)
Quality of life at >3 months (final score; FIQ)	64 (1 study) 12 weeks	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 66.1	The mean quality of life score in the intervention groups was 10.62 lower (12.34 to 8.9 lower)
Sleep quality at <3 months (final score; Pittsburgh Sleep Quality Index)	64 (1 study) 4 weeks	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean sleep quality score in the control groups was 12.39	The mean sleep quality score in the intervention groups was 3.94 lower (4.62 to 3.26 lower)
Sleep quality at >3 months (final score; Pittsburgh Sleep Quality Index)	64 (1 study) 12 weeks	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean sleep quality score in the control groups was 10.45	The mean sleep quality score in the intervention groups was 5.03 lower (5.51 to 4.55 lower)
Discontinuation at >3 months	64 (1 study) 12 weeks	⊕⊕⊕⊖ MODERATE1 due to risk of bias	RD 0.00 (-0.06 to 0.06)	-	0 fewer per 1000 (from 6 fewer to 6 more)

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

Table 50: Clinical evidence summary: Aerobic, strength, mind-body and proprioception versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus mind-body exercises (95% CI)
Quality of life at ≤3 months (FIQ total score, high is poor outcome)	21 (1 study) 7 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 46.44	The mean quality of life score in the intervention groups was 13.04 lower (21.92 to 4.16 lower)
Physical function at ≤3 months (number of steps, high is good outcome)	21 (1 study) 7 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 103.39	The mean physical function score in the intervention groups was 9.19 higher (11.24 lower to 29.62 higher)
Discontinuation at ≤3 months	35 (1 study) 7 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision	RR 0.66 (0.28 to 1.57)	474 per 1000	161 fewer per 1000 (from 341 fewer to 270 more)

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

Table 51: Clinical evidence summary: Strength versus mind-body

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus mind-body exercises (95% CI)
Pain (VAS, <3 months) Scale from: 0 to 10.	36 (1 study) 6 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of		The mean pain in the control	The mean pain in the intervention groups was 1.1 higher (0.31 lower to 2.51 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus mind-body exercises (95% CI)
		bias, imprecision		groups was 1.4	
Quality of life (Nottingham health profile, <3 months) Scale from: 0 to 600.	36 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life in the control groups was 89.8	The mean quality of life in the intervention groups was 56.1 higher (13.21 lower to 125.41 higher)
Physical function (NDI, <3 months) Scale from: 0 to 100.	36 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function in the control groups was 8.2	The mean physical function in the intervention groups was 3.1 higher (0.56 lower to 6.76 higher)
Psychological distress (BDI, <3 months) Scale from: 0 to 63.	36 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress in the control groups was 6.2	The mean psychological distress in the intervention groups was 3.3 higher (1.24 lower to 7.84 higher)
Discontinuation at ≤3 months	122 (1 study) 12 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.55 (0.68 to 3.52)	129 per 1000	71 more per 1000 (from 41 fewer to 325 more)

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

Table 52: Clinical evidence summary: Strength versus biomechanical

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus biomechanical exercises (95% CI)
Pain (VAS, <3 months) Scale from: 0 to 10.	38 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean pain in the control groups was 1.7	The mean pain in the intervention groups was 0.8 higher (0.52 lower to 2.12 higher)
Quality of life (Nottingham health profile, <3 months) Scale from: 0 to 600.	38 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life in the control groups was 118.2	The mean quality of life in the intervention groups was 27.7 higher (44.07 lower to 99.47 higher)
Physical function (NDI, <3 months) Scale from: 0 to 100.	38 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function in the control groups was 10	The mean physical function in the intervention groups was 1.3 higher (2.29 lower to 4.89 higher)
Psychological distress (BDI, <3 months) Scale from: 0 to 63.	38 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress in the control groups was 8.5	The mean psychological distress in the intervention groups was 1.2 higher (3.36 lower to 5.76 higher)

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

Table 53: Clinical evidence summary: Strength versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus flexibility (95% CI)
Pain reduction at ≤3 months (VAS, 0-100, change scores and final values, high is poor outcome)	86 (2 studies) 12 weeks	⊕⊕⊕⊖ MODERATE1 due to risk of bias		-	The mean pain score reduction in the intervention groups was 8.09 lower (14.58 to 1.59 lower)
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is good outcome)	66 (1 study) 16 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 9.2	The mean quality of life score in the intervention groups was 1.5 higher (2.64 lower to 5.64 higher)
Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is good outcome)	66 (1 study) 16 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 44.55	The mean quality of life score in the intervention groups was 5.39 lower (11.75 lower to 0.97 higher)
Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)	30 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 9.5	The mean physical function score in the intervention groups was 6 higher (2.34 to 9.66 higher)
Psychological distress at ≤3 months (BDI, 0-61, change scores, high is poor outcome)	56 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean psychological distress change score in the control groups was -1.84	The mean psychological distress score in the intervention groups was 1.83 lower (3.99 lower to 0.33 higher)
Psychological distress at ≤3 months (BAI, 0-61, change scores, high is poor outcome)	56 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1 due to risk of bias, imprecision		The mean psychological distress change score in the	The mean psychological distress score in the intervention groups was 3.2 lower (6.42 lower to 0.02 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus flexibility (95% CI)
				control groups was +0.7	
Sleep at ≤3 months (FIQ sleep subscale, 0-10, change scores, high is poor outcome)	56 (1 study) 12 weeks	⨁⨁⨁⨀ MODERATE1 due to risk of bias		The mean sleep change score in the control groups was -0.53	The mean sleep score in the intervention groups was 1.77 lower (2.62 to 0.92 lower)
Discontinuation at >3 months	157 (3 studies) 12-16 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision	RR 0.68 (0.36 to 1.28)		
				214 per 1000	68 fewer per 1000 (from 137 fewer to 60 more)

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

Table 54: Clinical evidence summary: Strength and flexibility versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus flexibility (95% CI)
Quality of life at >3 months (SF-36 physical functioning subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 92.4	The mean quality of life score in the intervention groups was 0.4 lower (4.92 lower to 4.12 higher)
Quality of life at >3 months (SF-36 role physical subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⨁⨁⨁⨀ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 79.4	The mean quality of life score in the intervention groups was 1.1 lower (15.9 lower to 13.7 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus flexibility (95% CI)
Quality of life at >3 months (SF-36 role emotional subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 87	The mean quality of life score in the intervention groups was 2.1 higher (9.7 lower to 13.9 higher)
Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 63.4	The mean quality of life score in the intervention groups was 5.2 higher (2.96 lower to 13.36 higher)
Quality of life at >3 months (SF-36 emotional wellbeing subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 75.9	The mean quality of life score in the intervention groups was 3.6 higher (3.43 lower to 10.63 higher)
Quality of life at >3 months (SF-36 social functioning subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 88.7	The mean quality of life score in the intervention groups was 1.7 higher (5.28 lower to 8.68 higher)
Quality of life at >3 months (SF-36 bodily pain subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 70.9	The mean quality of life score in the intervention groups was 1.7 lower (10.14 lower to 6.74 higher)
Quality of life at >3 months (SF-36 general health subscale, 0-100, final values, high is good outcome)	86 (1 study) 12 months	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 71.4	The mean quality of life score in the intervention groups was 0.7 higher (6.41 lower to 7.81 higher)
Discontinuation at >3 months	101 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of	RR 0.71 (0.27 to 1.84)	173 per 1000	50 fewer per 1000 (from 126 fewer to 145 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus flexibility (95% CI)
		bias, imprecision			

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

Table 55: Clinical evidence summary: Strength and flexibility versus mind-body

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus mind-body (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	117 (2 studies) 9-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain score in the control groups was 42.2	The mean pain score in the intervention groups was 10.4 lower (23.66 lower to 2.85 higher)
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)	140 (2 studies) 24 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean pain score in the control groups was 39.9	The mean pain score in the intervention groups was 0.78 lower (8.05 lower to 6.49 higher)
Quality of life at ≤3 months (SF-36 mental component, 0-100, final values, high is good outcome)	117 (2 studies) 9-12 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 46.95	The mean quality of life score in the intervention groups was 2.88 higher (0.8 lower to 6.55 higher)
Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is good outcome)	140 (2 studies) 24 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 45.45	The mean quality of life score in the intervention groups was 1.05 higher (2.28 lower to 4.38 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus mind-body (95% CI)
Quality of life at ≤3 months (SF-36 physical component, 0-100, final values, high is good outcome)	117 (2 studies) 9-12 weeks	⨁⨁⨁⨀ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 37.3	The mean quality of life score in the intervention groups was 1.04 higher (1.9 lower to 3.99 higher)
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is good outcome)	140 (2 studies) 24 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean quality of life score in the control groups was 39.2	The mean quality of life score in the intervention groups was 2.21 lower (4.81 lower to 0.38 higher)
Physical function at ≤3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)	117 (2 studies) 9-12 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		-	The mean physical function score in the intervention groups was 0.22 standard deviations lower (0.59 lower to 0.14 higher)
Physical function at >3 months (Neck pain disability scale, final values, high is poor outcome)	140 (2 studies) 24 weeks	⨁⨁⨁⨀ MODERATE1 due to risk of bias		The mean physical function score in the control groups was 19.9	The mean physical function score in the intervention groups was 0.22 higher (5.02 lower to 5.46 higher)
Psychological distress at ≤3 months (Depression scale ADS, 0-60, final values, high is poor outcome)	66 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 19.7	The mean psychological distress score in the intervention groups was 0.5 higher (3.66 lower to 4.66 higher)
Psychological distress at >3 months (Depression scale ADS, 0-60, final values, high is poor outcome)	66 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,3 due to risk of bias, The mean quality of life score in the control		The mean psychological distress score in the control groups was 22.7	The mean psychological distress score in the intervention groups was 1.8 lower (6.07 lower to 2.47 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength and flexibility versus mind-body (95% CI)
		groups was imprecision			
Discontinuation at >3 months	209 (3 studies) 9-24 weeks	⊕⊖⊖⊖ VERY LOW1,3 due to risk of bias, imprecision	OR 0.87 (0.35 to 2.14)	103 per 1000	12 fewer per 1000 (from 64 fewer to 94 more)

1 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
2 Downgraded for heterogeneity, unexplained by subgroup analysis
3 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 56: Clinical evidence summary: Strength, flexibility and proprioception versus mind-body

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength, flexibility and proprioception versus mind-body (95% CI)
Pain reduction at ≤3 months (VAS, 0-100, final values, high is poor outcome)	75 (1 study) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 32.4	The mean pain score in the intervention groups was 7.2 lower (16.72 lower to 2.32 higher)
Pain reduction at >3 months (VAS, 0-100, final values, high is poor outcome)	75 (1 study) 24 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 35	The mean pain score reduction in the intervention groups was 1.9 lower (12.99 lower to 9.19 higher)
Quality of life at ≤3 months (SF-36 physical component summary score,	75 (1 study)	⊕⊕⊖⊖ LOW1,2		The mean quality of life score in the	The mean quality of life score in the intervention groups was 2.1 lower

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength, flexibility and proprioception versus mind-body (95% CI)
final values, 0-100, high is good outcome)	12 weeks	due to risk of bias, imprecision		control groups was 47.3	(5.48 lower to 1.28 higher)
Quality of life at >3 months (SF-36 physical component summary score, final values, 0-100, high is good outcome)	75 (1 study) 24 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 46.5	The mean quality of life score in the intervention groups was 2.5 lower (6.22 lower to 1.22 higher)
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	75 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 46.8	The mean quality of life score in the intervention groups was 0.9 higher (3.77 lower to 5.57 higher)
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	75 (1 study) 24 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 47	The mean quality of life score in the intervention groups was 0.1 lower (4.96 lower to 4.76 higher)
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	75 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 21.5	The mean physical function score in the intervention groups was 1.2 higher (3.7 lower to 6.1 higher)
Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)	75 (1 study) 24 weeks	⨁⨁⨁⨀ MODERATE 1 due to risk of bias		The mean physical function score in the control groups was 24.3	The mean physical function score in the intervention groups was 0.8 higher (5.31 lower to 6.91 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength, flexibility and proprioception versus mind-body (95% CI)
Psychological distress at ≤3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	75 (1 study) 12 weeks	⨁⨁◯◯ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 6.5	The mean psychological distress score in the intervention groups was 1 lower (2.8 lower to 0.8 higher)
Psychological distress at >3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	75 (1 study) 24 weeks	⨁⨁◯◯ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 6.1	The mean psychological distress score in the intervention groups was 0.6 lower (2.34 lower to 1.14 higher)
Psychological distress at ≤3 months (HADS: depression, 0-21, final values, high is poor outcome)	75 (1 study) 12 weeks	⨁⨁⨁◯ MODERATE 1 due to risk of bias		The mean psychological distress score in the control groups was 3.9	The mean psychological distress score in the intervention groups was 0.1 lower (1.52 lower to 1.32 higher)
Psychological distress at >3 months (HADS: depression, 0-21, final values, high is poor outcome)	75 (1 study) 24 weeks	⨁⨁⨁◯ MODERATE 1 due to risk of bias		The mean psychological distress score in the control groups was 4.1	The mean psychological distress score in the intervention groups was 0 higher (1.51 lower to 1.51 higher)
Discontinuation at ≤3 months	75 (1 study) 12 weeks	⨁⨁⨁⨁ HIGH	RR 4.45 (1.38 to 14.35)	79 per 1000	273 more per 1000 (from 30 more to 1000 more)

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

Table 57: Clinical evidence summary: Strength versus proprioception

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Strength versus proprioception (95% CI)
Physical function at ≤3 months (Neck disability index, 0-50, final values, high is poor outcome)	26 (1 study) 8 weeks	⨁⨁⨁⨀ MODERATE 1 due to risk of bias		The mean physical function score in the control groups was 4.14	The mean physical function score in the intervention groups was 0.32 higher (1.47 lower to 2.11 higher)
1 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					

Table 58: Clinical evidence summary: Mind-body versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body versus flexibility (95% CI)
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	55 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 69	The mean pain score in the intervention groups was 2 higher (9.65 lower to 13.65 higher)
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	49 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 77.6	The mean quality of life score in the intervention groups was 22.9 lower (33.4 to 12.4 lower)
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	81 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of		The mean psychological distress score in	The mean psychological distress score in the intervention groups was 0.5 higher (3.55 lower to 4.55 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body versus flexibility (95% CI)
		bias, imprecision		the control groups was 17.8	
Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, final values, high is poor outcome)	81 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias		The mean sleep score in the control groups was 13.7	The mean sleep score in the intervention groups was 0 higher (1.92 lower to 1.92 higher)
Discontinuation at ≤3 months	62 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision	RR 1.83 (0.83 to 4.02)	Moderate	
				219 per 1000	182 more per 1000 (from 37 fewer to 661 more)

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

Table 59: Clinical evidence summary: Mind-body versus biomechanical

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body versus biomechanical (95% CI)
Pain (VAS, <3 months) Scale from: 0 to 10.	38 (1 study) 6 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean pain in the control groups was 1.7	The mean pain in the intervention groups was 0.3 lower (1.51 lower to 0.91 higher)
Quality of life (Nottingham health profile, <3 months) Scale from: 0 to 600.	38 (1 study) 6 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of		The mean quality of life in the control groups	The mean quality of life in the intervention groups was 28.4 lower (84.68 lower to 27.88 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Mind-body versus biomechanical (95% CI)
		bias, imprecision		was 118.2	
Physical function (NDI, <3 months)	38 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function in the control groups was 10	The mean physical function in the intervention groups was 1.8 lower (4.86 lower to 1.26 higher)
Psychological distress (Depression, BDI, <3 months) Scale from: 0 to 63.	38 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress in the control groups was 8.5	The mean psychological distress in the intervention groups was 2.1 lower (6.11 lower to 1.91 higher)

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

Table 60: Clinical evidence summary: Flexibility and proprioception versus flexibility

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Flexibility and proprioception versus flexibility (95% CI)
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	57 (1 study) 6 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 65.55	The mean quality of life score in the intervention groups was 12.7 lower (21.27 to 4.13 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Flexibility and proprioception versus flexibility (95% CI)
Psychological distress at ≤3 months (BDI, 0-63, final values, high is poor outcome)	57 (1 study) 6 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 13.79	The mean psychological distress score in the intervention groups was 3.88 higher (0.46 lower to 8.22 higher)
Discontinuation at ≤3 months	68 (1 study) 6 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.65 (0.53 to 5.12)	Moderate	
				121 per 1000	79 more per 1000 (from 57 fewer to 499 more)

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

Table 61: Clinical evidence summary: Flexibility and relaxation versus aerobic exercise

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Flexibility and relaxation versus aerobic (95% CI)
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)	133 (1 study) 12 months	⊕⊕⊕⊖ MODERATE1 due to risk of bias		The mean quality of life score in the control groups was 55.6	The mean quality of life score in the intervention groups was 0.4 higher (4.64 lower to 5.44 higher)
Discontinuation at ≤3 months	136 (1 study) 12 months	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 0.97 (0.47 to 2.01)		
				30 per 1000	10 fewer per 1000 (from 130 fewer to 120 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Flexibility and relaxation versus aerobic (95% CI)
1 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 2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs					

Table 62: Clinical evidence summary: Exercise versus psychological therapies

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with	Risk difference with Exercise versus psychological therapies (95% CI)
Pain at ≤3 months (VAS, FIQ pain scale, 0-100, high is poor outcome, final values and change scores) - Fibromyalgia	251 (4 studies) 8-12 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain score in the control groups was 31.35	The mean pain score in the intervention groups was 1.61 lower (15.09 lower to 11.87 higher)
Pain at >3 months (VAS, NRS, 0-100, high is poor outcome, final values)	468 (4 studies) 12-52 weeks	⨁⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency, imprecision		The mean pain score in the control groups was 50.35	The mean pain score in the intervention groups was 7.19 lower (13.98 to 0.41 lower)
Quality of life at ≤3 months (FIQ, 0-100, high is poor outcome, final values and change scores)	292 (4 studies) 6-12 weeks	⨁⨁⨁⊖ MODERATE1 due to risk of bias		-	The mean quality of life score in the intervention groups was 6.7 lower (10.88 to 2.52 lower)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with	Risk difference with Exercise versus psychological therapies (95% CI)
Quality of life at >3 months (EQ-5D, high is good outcome, final values)	152 (1 study) 9 months	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 0.754	The mean quality of life score in the intervention groups was 0.05 lower (0.12 lower to 0.02 higher)
Quality of life at >3 months (SF36 social aspects subscale, 0-100, high score is good outcome	60 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 63.9	The mean quality of life score outcome in the intervention groups was 3.4 higher (9.27 lower to 16.07 higher)
Quality of life at >3 months (SF36 general health status aspects subscale, 0-100, high score is good outcome	60 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 44.6	The mean quality of life score in the intervention groups was 2.6 higher (8.08 lower to 13.28 higher)
Quality of life at >3 months (SF36 funcitonal capacity aspects subscale, 0-100, high score is good outcome	60 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score outcome in the control groups was 40	The mean quality of life score in the intervention groups was 13.1 higher (2.72 to 23.48 higher)
Quality of life at >3 months (SF36 limitations due to physical aspects subscale, 0-100, high score is good outcome	60 (1 study) 12 weeks	⨁⨀⨀⨀ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 38.1	The mean quality of score in the intervention groups was 17.2 higher (2.83 lower to 37.23 higher)
Quality of life at >3 months (SF36 limitations due to emotional aspects	60 (1 study)	⨁⨀⨀⨀ VERY		The mean quality of life score in the	The mean quality of life score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with	Risk difference with Exercise versus psychological therapies (95% CI)
subscale, 0-100, high score is good outcome	12 weeks	LOW1,2 due to risk of bias, imprecision		control groups was 37.5	11.9 higher (8.74 lower to 32.54 higher)
Quality of life at >3 months (SF36 pain subscale, 0-100, high score is good outcome)	60 (1 study) 12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 29.9	The mean quality of life score in the intervention groups was 5 higher (5.39 lower to 15.39 higher)
Quality of life at >3 months (SF36 mental health subscale, 0-100, high score is good outcome)	60 (1 study) 12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life score in the control groups was 58.6	The mean quality of life score in the intervention groups was 0.9 higher (11.04 lower to 12.84 higher)
Physical function at ≤3 months (FIQ physical function subscale, 0-10, high is poor outcome, change scores)	98 (1 study) 12 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean physical function change in the control groups was -0.5	The mean physical function score in the intervention groups was 0.7 lower (2.75 lower to 1.35 higher)
Physical function at ≤3 months (6 minute walk test, metres, high is good outcome, final values)	139 (2 studies) 12 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 429.4	The mean physical function score in the intervention groups was 26.42 higher (0.85 lower to 53.69 higher)
Physical function at >3 months (6 minute walking test, metres, high is good outcome, final values)	165 (2 studies) 12-5 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 474.5	The mean physical function score in the intervention groups was 49.05 higher (25.45 to 72.65 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with	Risk difference with Exercise versus psychological therapies (95% CI)
Psychological distress at ≤3 months (CES-D, 0-100, high is poor outcome, final values)	62 (1 study) 12 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress score in the control groups was 67	The mean psychological distress score in the intervention groups was 10.3 lower (20.07 to 0.53 lower)
Psychological distress at >3 months (Hospital anxiety and depression scale, depression subscale, 0-21, high is poor outcome, change scores)	104 (1 study) 15 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress change in the control groups was +0.3	The mean psychological distress score in the intervention groups was 1 lower (2.25 lower to 0.25 higher)
Psychological distress at >3 months (Hospital anxiety and depression scale, anxiety subscale, 0-21, high is poor outcome, change scores)	105 (1 study) 15 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean psychological distress change in the control groups was +0.5	The mean psychological distress score in the intervention groups was 0.8 lower (2.01 lower to 0.41 higher)
Sleep at >3 months (the sleep scale, 0-30, final values, high is poor outcome)	190 (1 study) 9 months	⨁⨁⨁⨀ MODERATE1 due to risk of bias		The mean sleep in the control groups was 12.4	The mean sleep score in the intervention groups was 0.3 higher (1.22 lower to 1.82 higher)
Sleep at >3 months (Pittsburgh sleep quality index, 0-21, high is poor outcome, change scores)	105 (1 study) 15 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias, imprecision		The mean sleep change in the control groups was +0.5	The mean sleep score in the intervention groups was 1.1 lower (2.32 lower to 0.12 higher)
Discontinuation at >3 months (due to increased pain, personal reasons, lost to follow up)	1062 (10 studies) 8-52 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision	RD -0.03 (-0.07 to 0.02)	172 per 1000	30 fewer per 1000 (from 70 fewer to 20 more)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with	Risk difference with Exercise versus psychological therapies (95% CI)
1 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 2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.					

Table 63: Clinical evidence summary: Manual therapy and exercise versus manual therapy

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Manual therapy and exercise versus manual therapy (95% CI)
Pain at ≤3 months (NRS, high is poor outcome, final values, 0-10, final values)	101 (1 study) 11 weeks	⨁⨁⨀⨀ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 3.7	The mean pain score in the intervention groups was 0.8 lower (1.66 lower to 0.06 higher)
Pain at >3 months (NRS, high is poor outcome, final values, 0-10, final values)	101 (1 study) 52 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias, imprecision		The mean pain score in the control groups was 3.9	The mean pain score in the intervention groups was 0.5 lower (1.42 lower to 0.42 higher)
Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50, final values)	101 (1 study) 11 weeks	⨁⨁⨀⨀ LOW1 due to risk of bias, imprecision		The mean physical function score in the control groups was 18.7	The mean physical function score in the intervention groups was 5.1 lower (9.65 to 0.55 lower)
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)	101 (1 study)	⨁⨁⨀⨀ LOW1,2 due to risk		The mean physical function score in the	The mean physical function score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Manual therapy and exercise versus manual therapy (95% CI)
	24 months	of bias, imprecision		control groups was 20.5	4.9 lower (9.85 lower to 0.05 higher)
Discontinuation at ≤3 months	127 (1 study) 11 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 0.91 (0.47 to 1.79)		
				222 per 1000	20 fewer per 1000 (from 118 fewer to 175 more)

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

Table 64: Clinical evidence summary: Manual therapy and exercise versus exercise

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Manual therapy and exercise versus exercise (95% CI)
Pain at ≤3 months (VAS, NRS, high is poor outcome, final values, 0-100, final values)	542 (6 studies) 4-12 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, imprecision, inconsistency		The mean pain score in the control groups was 30.9	The mean pain score in the intervention groups was 6.34 lower (13.82 lower to 1.13 higher)
Pain at >3 months (NRS, VAS, high is poor outcome, final values, 0-100)	394 (3 studies) 52 weeks	⊕⊕⊕⊖ MODERATE 1 due to risk of bias		The mean pain score in the control groups was 32	The mean pain score in the intervention groups was 0.95 higher (3.51 lower to 5.4 higher)
Quality of life at >3 months (Fibromyalgia impact questionnaire,	21 (1 study)	⊕⊖⊖⊖ VERY		The mean quality of life score in the	The mean quality of life score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Manual therapy and exercise versus exercise (95% CI)
0-100, final values, high is poor outcome)	16 weeks	LOW1,2 due to risk of bias, imprecision		control groups was 46.9	1 lower (13.87 lower to 11.87 higher)
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	180 (1 study) 12 weeks	⊕⊕⊕⊖ MODERATE 1 due to risk of bias		The mean quality of life score in the control groups was 50.1	The mean quality of life score in the intervention groups was 0.6 higher (1.34 lower to 2.54 higher)
Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)	180 (1 study) 52 weeks	⊕⊕⊕⊖ MODERATE 1 due to risk of bias		The mean quality of life score in the control groups was 49.8	The mean quality of life score in the intervention groups was 0.2 higher (1.79 lower to 2.19 higher)
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	180 (1 study) 12 weeks	⊕⊕⊕⊖ MODERATE 1 due to risk of bias		The mean quality of life score in the control groups was 54.6	The mean quality of life score in the intervention groups was 0.7 lower (3.55 lower to 2.15 higher)
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)	180 (1 study) 52 weeks	⊕⊕⊕⊖ MODERATE 1 due to risk of bias		The mean quality of life score in the control groups was 54.8	The mean quality of life score in the intervention groups was 1.8 lower (4.34 lower to 0.74 higher)
Physical function at >3 months (Neck disability index, functional performance scale, final values, high is poor outcome, 0-100)	477 (5 studies) 11-16 weeks	⊕⊖⊖⊖ VERY LOW1,2,3 due to risk of bias, inconsistency , imprecision		-	The mean physical function score in the intervention groups was 0.29 standard deviations lower (0.62 lower to 0.04 higher)

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Manual therapy and exercise versus exercise (95% CI)
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-100)	394 (3 studies) 24 months	⨁⨁⨁⊖ MODERATE 1 due to risk of bias		The mean physical function score in the control groups was 16.7	The mean physical function score in the intervention groups was 0.17 lower (2.6 lower to 2.25 higher)
Physical function at ≤3 months (Neck disability index, high is poor outcome, 0-100)	86 (2 studies) 4-10 weeks	⨁⨁⊖⊖ LOW1 due to risk of bias		The mean physical function score in the control groups was 18.68	The mean physical function score in the intervention groups was 8.14 lower (9.92 to 6.35 lower)
Discontinuation	542 (6 studies) 6-16 weeks	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RD 0 (-0.05 to 0.06)	127 per 1000	0 fewer per 1000 (from 50 fewer to 60 more)

1 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
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 65: Clinical evidence summary: Exercise versus manual therapy

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Exercise versus manual therapy (95% CI)
Pain at ≤3 months (NRS, high is poor outcome, final values, 0-10)	101 (1 study) 11 weeks	⨁⨁⊖⊖ LOW1,2 due to risk of		The mean pain score in the	The mean pain score in the intervention groups was

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Control	Risk difference with Exercise versus manual therapy (95% CI)
		bias, imprecision		control groups was 3.7	1.3 lower (2.11 to 0.49 lower)
Pain at >3 months (NRS, high is poor outcome, final values, 0-10)	101 (1 study) 52 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean pain score in the control groups was 3.9	The mean pain score in the intervention groups was 0.5 lower (1.42 lower to 0.42 higher)
Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50)	94 (1 study) 11 weeks	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 18.7	The mean physical function score in the intervention groups was 5.9 lower (10.6 to 1.2 lower)
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)	94 (1 study) 24 months	⊕⊕⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean physical function score in the control groups was 20.5	The mean physical function score in the intervention groups was 3.9 lower (9.14 lower to 1.34 higher)
Discontinuation at ≤3 months	127 (1 study) 11 weeks	⊕⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision	RR 1.34 (0.74 to 2.43)	222 per 1000	75 more per 1000 (from 58 fewer to 317 more)

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

See appendix F for full GRADE tables.

1.5 Economic evidence

1.5.1 Included studies

Two health economic studies were identified with the relevant comparisons and have been included in this review. This is summarised in the health economic evidence profile below and the health economic evidence tables in appendix H.

1.5.2 Excluded studies

Three additional health economic studies were identified as relevant to this question, but were selectively excluded as the committee judged that other available evidence was of greater applicability and methodological quality. [181],[263],[264] These are listed in appendix I, with reason for exclusion given.

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

1.5.3 Summary of studies included in the economic evidence review

Note that **Table 66** includes only the relevant comparisons for this review, although the evidence table in Appendix H: includes all comparators in the study.

Table 66: Health economic evidence profile: Aerobic exercise therapy vs. psychological therapy or usual care

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental QALYs	Cost effectiveness	Uncertainty
Beasley, 2015[28] [UK]	Directly applicable[(a)]	Potentially serious limitations [(b)]	• Within-trial analysis (same paper) • Cost-utility analysis (QALYs) • Population: > 25 years and over with chronic widespread pain according to the definition in the American College of Rheumatology (ACR) 1990 criteria for fibromyalgia, for which they have consulted their general practitioner in the previous year. • 6 month interventions • Follow-up: 30 months (24 months post treatment) Comparators: 1. Treatment as usual. 2. Telephone-delivered cognitive behaviour therapy (TCBT): initial assessment (45-60 mins) followed by 7 weekly sessions (30-45 mins each), 1 session at three months, and 1 session	**Complete case analysis:**			Used non-parametric bootstrapping.
				(3-1): £1,924	(3-1): 0.025	ICER: £76,960 per QALY gained	
				(3-2) £1,350	(3-2): -0.072	Dominated	
				Multiple imputation analysis:			
				(3-1): £1,256	(3-1): 0.071	ICER: £17,690 per QALY gained	
				(3-2): £702	(3-2): -0.069	Dominated	

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental QALYs	Cost effectiveness	Uncertainty
			at 6 months after randomisation. 3. Exercise therapy: leisure-facility-and-gym-based exercise program consistent with American College of Sport Medicine guidelines for improving cardiorespiratory fitness. (only partly supervised with monthly instructor led appointments and people otherwise used the gym)				

Abbreviations: ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years; RCT: randomised controlled trial

(a) UK NHS study, used EQ-5D. Participation in study based on self-reported symptoms and recruited through primary care, may not necessarily be representative of general population with chronic widespread pain caused by fibromyalgia.

(b) Treatment as usual not defined, usual care provided by GP was not restricted and may not be the same across all participants in that group. Within-study analysis which may not reflect full body of evidence. The imputed results are also quite different to the complete case data results, leading to a change in conclusion on cost effectiveness. It is hard to know which results should be used without knowing the details of the imputations and the nature of the missing data.

Table 67: Health economic evidence profile: Aquatic based aerobic exercise + usual care versus usual care

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental QALYs	Cost effectiveness	Uncertainty
Gusi 2008 [116](Spain)	Partially applicable(a)	Potentially serious limitations(b)	• Within trial analysis[252,253] • Cost-utility analysis (QALYs) • Population: women with Fibromyalgia. • 8 month intervention. • Follow-up: 8 months Comparing: • Exercise + usual care: Exercise programme in a	£475(c)	0.131 QALYs	£3,630 per QALY gained	Probability exercise cost effective: Determined by reading off the graph based on the '2005 adjusted investment ceiling set at €34,729/QALY): approx. 97% Various sensitivity analyses tested such as varying the number of people per group (participation), the salaries

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental QALYs	Cost effectiveness	Uncertainty
			waist high pool of warm water (33°C). A qualified exercise leader instructed and trained the group three times a week for 1 h per session over a period of 8 months. • Treatment as usual				of the staff. And testing worst and best case scenarios based on participation, salaries, and extremes of confidence interval for QALY difference. Only the worst case scenario led to the intervention not being cost effective based on the threshold published in the Spanish literature.

Abbreviations: ICER= incremental cost-effectiveness ratio; QALY= quality-adjusted life years; RCT= randomised controlled trial FM = Fibromyalgia.

(a) Uses EQ-5D. Non-UK study.

(b) Only based on one study. Date and costs may not reflect current NHS context. Costs of staff look very low compared to UK costs which will affect the ICER. Recruitment of participants was through local FM association, perhaps not representative of wider population with FM.

(c) 2005 Spanish Euros converted to UK pounds.[208] Cost components incorporated: Programme cost (based on staff costs, renting the pool, management costs of the programme like insurance). Health care costs (consultations, drug process).

1.5.4 Health economic modelling

This area was prioritised for new economic modelling. The rationale, methods and results are summarised below. Full details are available in the 'Exercise modelling report'.

Methods

The clinical review showed a benefit of exercise compared to usual care in reducing pain and improving quality of life. When comparing types of exercise to each other, there was less evidence and it was difficult to draw conclusions about a hierarchy of types of exercise.

Two economic evaluations were identified for this review comparing exercise to treatment as usual. One was a UK within trial analysis (cost utility analysis) looking at a gym based exercise program (gym membership provided), and 6 fitness instructor-led monthly sessions, for a duration of 6 months. The committee view was that this study was quite different to most of the other studies in the clinical review, which tended to be structured class-based interventions, generally group based, with varying frequency/intensity. The study found exercise was not cost effective in the base case analysis using complete case data, but it was cost effective when using imputed data. The second economic evaluation was a Spanish within trial analysis (cost utility analysis, comparing 8 months of group pool-based exercised to usual care. This found exercise to be cost effective. Pool-based exercises are not considered to be current practice in the UK because they have higher costs. This was an older study than the UK one (2008), and had limitations like the costs of the staff involved seem very low compared to UK costs, which is likely to increase the ICER.

Uncertainty remained about the cost effectiveness of exercise from the included data, therefore, a lifetime cost utility analysis was undertaken, from the NHS perspective, that compared exercise with no exercise (both groups had usual care therefore this was not included in the model). The analysis is based on studies from the clinical review that reported utilities (EQ-5D), or the SF-36 that could be mapped to utilities (12 studies). All exercise types were pooled. All studies except one used supervised exercise, and most were group based (or assumed to be).

For each study, the difference between follow up EQ-5D (whether this was at the end of treatment or later) and the baseline EQ-5D was taken for the intervention and usual care group, to take account of any baseline differences between the two groups. The difference in EQ-5D was then taken between the intervention and usual care group for each study. Therefore, the treatment benefit is the EQ-5D gain from exercise compared to usual care, taking into account baseline differences. Where there were several studies that reported quality of life at the same time point, these were pooled in a meta-analysis. A linear trend line was fitted to the QoL gain points over time, based on weighted least squares regression to attach more weight to time points where there was more certainty about the treatment effect. The available data on the difference in utility between the comparators were combined with assumptions about what is likely to happen to treatment effect beyond the follow-up in the trials (treatment effect was extrapolated), to calculate the average QALY gain with exercise compared to no exercise. Extrapolation assumptions were based on committee opinion, and different assumptions were needed for different scenarios that occurred in probabilistic analyses. Note the treatment effect was extrapolated only until there was no additional quality of life benefit from exercise. Two base cases were analysed; one with a lifetime horizon and one where treatment effect is not extrapolated beyond the trial data.

The key difference in costs was agreed to be those related to delivering an exercise programme. No other costs were incorporated in the analysis. The average resource use from the interventions in each study were identified and costed, and a weighted average cost calculated, weighting by the number of participants in the studies.

Results

The probabilistic and deterministic base case results can be seen in the table below. Results are presented for both base cases. Both analyses show the ICER is below the NICE threshold of £20,000, and therefore exercise would be considered cost effective. The probability of exercise being cost effective is also high.

Table 68: Base case results (discounted)

Base case	Analysis	Incremental cost	Incremental QALYs	Cost per QALY gained	Probability cost effective at £20k
Lifetime	Probabilistic	£380	0.04	£9,121	86%
	Deterministic	£380	0.031	£12,327	NA
No extrapolation beyond last trial observation (36 weeks)	Probabilistic	£380	0.03	£12,683	93%
	Deterministic	£380	0.030	£12,739	NA

Abbreviations: QALYs: quality adjusted life years, £20k: £20,000.

The deterministic results are slightly different to the probabilistic in the lifetime analysis because there is a larger incremental QALY gain in the probabilistic analysis from the QALY gains having a skewed distribution, as there are some simulations with quite flat slopes which lead to a large QALY gain because of the extrapolation assumptions exacerbating the gain, and the point at which there is no longer a difference in treatment effect from exercise being far into the future. This was proven by looking at the distribution of the QALY gains in a probabilistic analysis and plotting them graphically. Additionally, when looking at the analysis where no extrapolation of the data was assumed, then the probabilistic and deterministic results are very close, proving that the extrapolation assumptions and the nature of the data in the probabilistic analysis is creating this discord between the types of results, and both types of results are still well below the NICE threshold.

Various sensitivity analyses were undertaken for both base cases, where long term data points were included that were not included in the base case, and also data points that followed a 'de-training' period were also only used in a sensitivity analysis. Sensitivity analysis also tested using final QoL values in the meta-analysis as opposed to changes from baseline. Assumptions were also made about less staff and lower staff bands, as the most conservative assumptions about resource use were made in the base case. All sensitivity analyses did not change the conclusions.

Limitations of the analysis include that data was pooled from different studies that had different interventions of different intensities. This is likely to affect costs but also treatment effect. There is uncertainty around whether the costs that have been pooled appropriately correspond to/or are leading to the pooled treatment effect. This is because it is unclear what it is about exercise that causes a benefit. The analysis only used a subset of studies from the clinical review. The linear trend line representing treatment effect over time is a simplification of how people's quality of life would fluctuate in reality. The quality of life gain taken from the studies could also be an overestimate because it is likely that people who respond to follow up questionnaires or that have not dropped out of a trial are more engaged with the intervention. Additionally, it is uncertain what was happening after the intervention and whether people were continuing the intervention so assumptions were made. No other costs have been accounted for in the analysis except for intervention costs.

Overall, this analysis has pooled a subset of data from the clinical review that reported quality of life, to estimate the potential cost effectiveness of supervised exercise in general, not being specific to a particular type of exercise. Given the differences between the studies

and how few studies were used compared to the review as a whole, this analysis should be interpreted carefully.

1.6 Evidence statements

1.6.1 Clinical evidence statements

1.6.1.1 Aerobic exercise versus usual care

Pain reduction

Very low quality evidence from 1 study with 40 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 9 studies with 528 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 1 study with 95 participants showed no clinically important difference between exercise and usual care at >3 months.

Health related quality of life

Very low quality evidence from 5 studies with 372 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Very low to low quality evidence from 1 study with 54 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Very low to low quality evidence from 1 study with 95 participants showed usual care to lead to a clinically important benefit compared to exercise at ≤3 months. Very low quality evidence from 2 studies with 259 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Very low quality evidence from 1 study 95 participants showed no clinically important difference between exercise and usual care at ≤3 months or at >3 months.

Physical function

Very low quality evidence from 2 studies with 155 participants showed no clinically important difference between exercise and usual care at ≤3 months and very low quality evidence from 1 study with 95 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 3 studies with 169 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Very low quality evidence from 3 studies with 246 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Psychological distress

Low quality evidence from 1 study with 60 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Low quality evidence from 3 studies with 123 participants showed no clinically important difference between exercise and usual care at >3 months. Low quality evidence from 4 studies with 306 participants showed no clinically important difference between exercise and usual care at >3 months. Low quality evidence from 4 studies with 320 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 1 study with 50 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 1 study with 95 participants showed no clinically important difference between exercise and usual care at >3 months.

Use of healthcare services

Very low to low quality evidence from 1 study with 95 participants was identified but clinical importance could not be determined (unclear if high or low healthcare service use is a clinically important benefit).

Sleep

Very low quality evidence from 5 studies with 414 participants showed no clinically important difference between exercise and usual care at >3 months.

Discontinuation

Very low quality evidence from 9 studies with 607 participants showed more people discontinued from exercise compared to usual care.

1.6.1.2 Strength training versus usual care

Pain reduction

Very low quality evidence from 3 studies with 156 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 3 studies with 251 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 4 studies with 449 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Health related quality of life

Very low quality evidence from 2 studies with 102 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Low quality evidence from 1 study with 42 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 2 studies with 52 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months.

Physical function

Low quality evidence from 3 studies with 146 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 2 studies with 151 participants showed no clinically important difference between exercise and usual care at ≤3 months. Very low quality evidence from 1 study with 20 participants showed no clinically important difference between exercise and usual care at ≤3 months. Low quality evidence from 2 studies with 163 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 3 studies with 105 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Psychological distress

Very low quality evidence from 1 study with 25 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Low quality evidence from 1 study with 21 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Use of healthcare services

Very low to low quality evidence from 1 study with 179 participants was identified but clinical importance could not be determined (unclear if high or low healthcare service use is a clinically important benefit).

Sleep

Low quality evidence from 1 study with 21 participants showed no clinically important difference between exercise and usual care at >3 months.

Discontinuation

Low quality evidence from 4 studies with 252 participants showed no clinically important difference between exercise and usual care at >3 months.

1.6.1.3 Aerobic and strength exercise versus usual care

Pain reduction

Low quality evidence from 2 studes with 129 participants showed no clinically important difference between between exercise and usual care at ≤3 months. Very low quality evidence from 3 studies with 161 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Health related quality of life

Low quality evidence from 1 study with 30 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months and >3 months. Low quality evidence from 2 studies with 54 participants showed no clinically important difference between exercise and usual care at ≤3 months. Very low quality evidence from 4 studies with 171 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Low quality evidence from 1 study with 42 participants showed both a clinically important benefit of exercise compared to usual care and no clinically important difference at >3 months (various subscales).

Physical function

Low quality evidence from 1 study with 32 participants showed no clinically important difference between exercise and usual care at ≤3 months. Low quality evidence from 1 study with 16 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Low quality evidence from 1 study with 37 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Very low quality evidence from 1 study with 30 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Psychological distress

Low quality evidence from 2 studies with 54 participants showed no clinically important difference between exercise and usual care at ≤3 months. Very low quality evidence from 1 study with 58 participants showed no clinically important difference between exercise and usual care at ≤3 months. Low quality evidence from 1 study wih 32 participants showed no clinically important difference between between exercise and usual care at ≤3 months. Low quality evidence from 4 studies with 125 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 2 studies with 83 participants showed a clinically important of exercise compared to usual care at >3 months.

Use of healthcare services

Very low quality evidence from 1 study with 78 participants showed no clinically important difference between exercise and usual care at >3 months.

Sleep

Low quality evidence from 1 study with 58 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Discontinuation

Low quality evidence from 4 studies with 125 participants showed no clinically important difference between exercise and usual care at ≤3 months. Very low quality evidence from 7 studies with 230 participants showed no clinically important difference between exercise and usual care at >3 months.

1.6.1.4 Aerobic, strength and flexibility versus usual care

Low quality evidence from 1 study with 25 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months for quality of life.

No other evidence identified.

1.6.1.5 Strength and flexibility versus usual care

Pain reduction

Low quality evidence from 2 studies with 110 participants showed no clinically important difference between exercise and usual care at ≤3 months. Low quality evidence from 2 studies with 144 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Health related quality of life

Low quality evidence from 1 study with 70 participants showed no clinically important difference between exercise and usual care at ≤3 months. Low quality evidence from 1 study with 144 participants showed no clinically important difference between exercise and usual care at >3 months.

Physical function

Low quality evidence from 1 study with 70 participants showed no clinically important difference between exercise and usual care at ≤3 months. Moderate quality evidence from 2 studies with 144 participants showed no clinically important difference between exercise and usual care at >3 months.

Psychological distress

Low quality evidence from 1 study with 70 participants showed no clinically important difference between exercise and usual care at ≤3 months or >3 months.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 2 studies with 157 participants showed no clinically important difference between exercise and usual care at >3 months.

1.6.1.6 Strength, proprioception and flexibility versus usual care

Pain reduction

Low quality evidence from 1 study with 76 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months and >3 months

Health related quality of life

Low quality evidence from 1 study with 76 participants showed both a clinically important benefit of exercise compared to usual care and no clinically important difference at ≤3 months and >3 months (various subscales).

Physical function

Low quality evidence from 1 study with 76 participants showed no clinically important difference between exercise compared to usual care at ≤3 months and >3 months.

Psychological distress

Low quality evidence from 1 study with 76 participants showed no clinically important difference between exercise compared to usual care at ≤3 months and >3 months.

Use of healthcare services

Very low to low quality evidence from 1 study with 95 participants was identified but clinical importance could not be determined (unclear if high or low healthcare service use is a clinically important benefit).

Sleep

No evidence identified.

Discontinuation

Low quality evidence from 1 study with 76 participants showed more people discontinued from exercise compared to usual care at ≤3 months.

1.6.1.7 Proprioception versus usual care

Low to very low quality evidence from 1 study with 46 participants showed no clinically important difference between exercise and usual care at ≤3 months and >3 months for pain or quality of life. Low quality evidence from the same study showed a clinically important benefit of exercise compared to usual care at ≤3 months and >3 months for psychological distress, and a clinically important benefit at ≤3 months for physical function, but no clinically important difference at >3 months.

No other evidence identified.

1.6.1.8 Mind-body exercise versus usual care

Pain reduction

Very low quality evidence from 8 studies with 393 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Low quality evidence from 1 study with 117 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 1 study with 117 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Low quality evidence from 1 study with 80 participants showed a clinically important benefit of exercise compared to usual care at >3 months. Low quality evidence from 3 studies with 221 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Health related quality of life

Low quality evidence from 1 study with 57 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 3 studies with 106 participants showed no clinically important difference between exercise and usual care at ≤3 months. Moderate quality evidence from 3 studies with 220 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Very low quality evidence from 3 studies with 220 participants showed no clinically important difference between exercise and usual care at ≤3 months. Very low quality evidence from 3 studies with 221 participants showed no clinically important difference between exercise and usual care at >3 months. Very low quality evidence from 1 study with 80 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Physical function

Very low quality evidence from 7 studies with 363 participants showed no clinically important difference between exercise and usual care at >3 months. Low quality evidence from 3 studies with 225 participants showed no clinically important difference between exercise and usual care at >3 months. Low quality evidence from 1 study with 80 participants showed a clinically important benefit of exercise compared to usual care at >3 months.

Psychological distress

Very low quality evidence from 5 studies with 306 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Low quality evidence from 1 study with 57 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months. Low quality evidence from 2 studies with 77 participants showed no clinically important difference between exercise and usual care at ≤3 months. Moderate quality evidence from 3 studies with 223 participants showed no clinically important difference between exercise and usual care at >3 months. Low quality evidence from 1 study with 77 participants showed no clinically important difference between exercise and usual care at >3 months.

Use of healthcare services

No evidence identified.

Sleep

Very low quality evidence from 2 studies with 60 participants showed no clinically important difference between exercise and usual care at ≤3 months.

Discontinuation

Very low quality evidence from 12 studies with 784 participants showed no clinically important difference between exercise and usual care at >3 months.

1.6.1.9 Flexibility versus usual care

Pain reduction

Very low quality evidence from 1 study with 28 participants showed a clinically important benefit of exercise compared to usual care at ≤3 months.

Health related quality of life

No evidence identified.

Physical function

Very low quality evidence from 1 study with 28 participants showed no clinically important difference between exercise and usual care at ≤3 months.

Psychological distress

No evidence identified.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 1 study with 34 participants showed more people discontinued from exercise compared to usual care at ≤3 months.

1.6.1.10 Aerobic versus strength

Pain reduction

Very low quality evidence from 4 studies with 199 participants showed no clinically important difference between aerobic and strength at ≤3 months. Very low quality evidence from 1 study with 60 participants showed no clinically important difference between aerobic and strength at >3 months.

Health related quality of life

Very low quality evidence from 3 studies with 127 participants showed a clinically important benefit of aerobic compared to strength at ≤3 months.

Physical function

Very low quality evidence from 1 study with 26 participants showed no clinically important difference between aerobic and strength at ≤3 months. Moderate quality evidence from 1 study with 75 participants showed no clinically important difference between aerobic and

strength at ≤3 months. Low quality evidence from 2 studies with 86 participants showed no clinically important difference between aerobic and strength at >3 months.

Psychological distress

Very low quality evidence from 2 studies with 52 participants showed no clinically important difference between aerobic and strength at ≤3 months. Very low quality evidence from 1 study with 75 participants showed a clinically important benefit of aerobic compared to strength at ≤3 months.

Use of healthcare services

No evidence identified.

Sleep

Very low quality evidence from 1 study with 26 participants showed no clinically important difference between aerobic and strength at ≤3 months.

Discontinuation

Low quality evidence from 4 studies with 196 participants showed no clinically important difference between aerobic and strength at ≤3 months.

1.6.1.11 Aerobic exercise versus flexibility

Pain reduction

Very low quality evidence from 1 study with 60 participants showed no clinically important difference between aerobic and flexibility at ≤3 months. Very low quality evidence from 1 study with 60 participants showed a clinically important benefit of aerobic compared to flexibility at >3 months.

Health related quality of life

Very low quality evidence from 1 study with 60 participants showed a clinically important benefit of aerobic compared to flexibility at ≤3 months and >3 months.

Physical function

No evidence identified.

Psychological distress

Very low quality evidence from 1 study with 60 participants showed no clinically important difference between aerobic and flexibility at ≤3 months, and both clinically important benefit of aerobic (for depression subscale) and no clinically important difference (for anxiety subscale) at >3 months.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 1 study with 76 participants showed more people discontinued from aerobic compared to flexibility at >3 months.

1.6.1.12 Aerobic exercise versus biomechanical exercise

Moderate to very low quality evidence from 1 study with 42 participants showed a clinically important benefit of aerobic exercise compared with biomechanical exercise for quality of life at ≤3 months, but no clinically important difference between aerobic and biomechanical exercise for pain reduction, psychological distress or sleep. More people discontinued from biomechanical exercise than aerobic exercise.

No other evidence identified.

1.6.1.13 Aerobic and strength versus aerobic exercise

Low to very low quality evidence from 1 study with 43 participants showed no clinically important difference between aerobic and strength and aerobic at >3 months for quality of life, psychological distress or discontinuation.

No other evidence identified.

1.6.1.14 Aerobic and strength versus flexibility

Very low quality evidence from 1 study with 85 participants showed no clinically important difference between aerobic and strength and flexibility at ≤3 months for pain or psychological distress but a benefit or aerobic and strength for quality of life. Very low quality evidence from 1 study with 76 participants showed a clinically important benefit of aerobic and strength compared to flexibility at >3 months for pain and quality and life but not clinically important difference for psychological distress. Very low quality evidence from 2 studies with 103 participants showed more people discontinued from aerobic and strength compared to flexibility at ≤3 months.

No other evidence identified.

1.6.1.15 Aerobic and flexibility versus mind-body exercise

Very low to low quality evidence from 1 study with 111 participants showed no clinically important difference between aerobic and flexibility and mind-body at ≤3 months and >3 months for quality of life, physical function, psychological distress and sleep (other than a benefit of aerobic and flexibility for a mental quality of life subscale at ≤3 months and a physical quality of life subscale at >3 months. Very low quality evidence from the same study showed more people discontinued from aerobic and flexibility compared to mind-body exercise at ≤3 months.

No other evidence identified.

1.6.1.16 Aerobic and flexibility versus aerobic exercise

Moderate quality evidence from 1 study with 64 participants showed a clinically important benefit of aerobic and flexibility exercise compared with aerobic exercise alone for quality of life and sleep at ≤3 months and >3 months, but no clinically important difference between aerobic and flexibility exercise and aerobic exercise alone for pain reduction at either time point, or discontinuation.

No other evidence identified.

1.6.1.17 Aerobic, strength, mind-body and proprioception versus flexibility

Low quality evidence from 1 study with 21 participants showed a clinically important benefit of aerobic, strength, mind-body and proprioception exercise compared with flexibility for quality of life and discontinuation, but no clinically important difference for physical function at ≤3 months.

No other evidence identified.

1.6.1.18 Strength training versus mind-body exercise

Pain reduction

Very low quality evidence from 1 study with 36 participants showed a clinically important benefit of mind-body exercise compared to strength training at ≤3 months.

Health related quality of life

Very low quality evidence from 1 study with 36 participants showed no clinically important difference between strength training and mind-body exercise at ≤3 months.

Physical function

Very low quality evidence from 1 study with 36 participants showed a clinically important benefit of mind-body exercise compared to strength training at ≤3 months.

Psychological distress

Very low quality evidence from 1 study with 36 participants showed a clinically important benefit of mind-body exercise compared to strength training at ≤3 months.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 1 study showed more people discontinued from strength compared to mind-body exercise at ≤3 months.

1.6.1.19 Strength training versus biomechanical exercise

Pain reduction

Very low quality evidence from 1 study with 38 participants showed a clinically important benefit of biomechanical exercise compared to strength training at ≤3 months.

Health related quality of life

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between strength training and biomechanical exercise at ≤3 months.

Physical function

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between strength training and biomechanical exercise at ≤3 months.

Psychological distress

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between strength training and biomechanical exercise at ≤3 months.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

No evidence identified.

1.6.1.20 Strength training versus flexibility

Pain reduction

Moderate quality evidence from 2 studies with 86 participants showed no clinically important difference between strength and flexibility at ≤3 months.

Health related quality of life

Very low quality evidence from 1 study with 60 participants showed both a clinically important benefit and no clinically important difference of/between strength compared to flexibility at >3 months.

Physical function

Very low quality evidence from 1 study with 30 participants showed clinically important benefit of flexibility compared to strength at ≤3 months.

Psychological distress

Low quality evidence from 1 study with 56 participants showed clinically important benefit of flexibility compared to strength (anxiety subscale) and no clinically important difference between strength and flexibility (depression subscale) at ≤3 months.

Use of healthcare services

No evidence identified.

Sleep

Moderate quality evidence from 1 study with 56 participants showed a clinically important benefit of strength compared to flexibility at ≤3 months.

Discontinuation

Very low quality evidence from 3 studies with 157 participants showed a clinically important benefit of strength compared to flexibility at >3 months.

1.6.1.21 Strength and flexibility versus flexibility

Very low quality evidence from 1 study with 86 participants showed both a clinically important benefit of strength and flexibility compared to flexibility and no clinically important difference at >3 months (various subscales). Very low quality evidence from the same study showed a clinically important benefit of strength and flexibility compared to flexibility for discontinuation at >3 months.

No other evidence identified.

1.6.1.22 Strength and flexibility versus mind-body exercise

Pain reduction

Very low quality evidence from 2 studies with 117 participants showed a clinically important benefit of strength and flexibility compared to mind-body at ≤3 months. Moderate quality evidence from 2 studies with 140 participants showed no clinically important difference between strength and flexibility compared to mind-body at >3 months.

Health related quality of life

Moderate quality evidence from 2 studies with 117 participants showed no clinically important difference between strength and flexibility compared to mind-body at ≤3 months. Moderate to low quality evidence from 2 studies with 140 participants showed no clinically important difference between strength and flexibility compared to mind-body at >3 months.

Physical function

Low quality evidence from 2 studies with 117 participants showed no clinically important difference between strength and flexibility compared to mind-body at ≤3 months. Moderate to low quality evidence from 2 studies with 140 participants showed no clinically important difference between strength and flexibility compared to mind-body at >3 months.

Psychological distress

Low quality evidence from 1 study with 66 participants showed no clinically important difference between strength and flexibility compared to mind-body at ≤3 months.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 3 studies with 209 participants showed no clinically important difference between strength and flexibility compared to mind-body at >3 months.

1.6.1.23 Strength, flexibility and proprioception versus mind-body exercise

Very low to moderate quality evidence from 1 study with 75 participants showed no clinically important difference between strength and flexibility and flexibility at ≤3 months and >3 months for pain, quality of life, physical function and psychological distress. High quality evidence from the same study showed clinically important benefit of mind-body compared to strength, flexibility and proprioception at ≤3 months for discontinuation.

No other evidence identified.

1.6.1.24 Strength training versus proprioception

Moderate quality evidence from 1 study with 26 participants showed no clinically important difference between strength and proprioception at ≤3 months for physical function.

No other evidence identified.

1.6.1.25 Mind-body exercise versus flexibility

Very low quality evidence from 1 study with 55 participants showed no clinically important difference between mind-body and flexibility at ≤3 months for pain, but a clinically important benefit of mind-body for quality of life. Very low quality evidence from 1 study with 81 participants showed no clinically important difference between mind-body and flexibility at ≤3 months for sleep. Very low quality evidence from 1 study with 62 participants showed more people discontinued from mind-body at ≤3 months.

No other evidence identified.

1.6.1.26 Mind-body exercise versus biomechanical exericse

Pain reduction

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between mind-body and biomechanical exercise at ≤3 months.

Health related quality of life

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between mind-body and biomechanical exercise at ≤3 months.

Physical function

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between mind-body and biomechanical exercise at ≤3 months.

Psychological distress

Very low quality evidence from 1 study with 38 participants showed no clinically important difference between mind-body and biomechanical exercise at ≤3 months.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

No evidence identified.

1.6.1.27 Flexibility and proprioception versus flexibility

Very low quality evidence from 1 study with 57 participants showed a clinically important benefit of flexibility and proprioception compared to flexibility for quality of life and psychological distress at ≤3 months, but no clinically important difference for discontinuation.

No other evidence identified.

1.6.1.28 Flexibility and relaxation versus aerobic

Very low to moderate quality evidence from 1 study with 136 participants showed no clinically important difference between flexibility and relaxation and aerobic at >3 months for quality of life or discontinuation.

1.6.1.29 Exercise versus psychological therapies

Pain reduction

Very low quality evidence from 4 studies with 251 participants showed no clinically important difference between exercise and psychological therapies at ≤3 months. Low quality evidence from 4 studies with 468 participants showed no clinically important difference between exercise and psychological therapies at >3 months.

Health related quality of life

Moderate quality evidence from 4 studies with 292 participants showed no clinically important difference between exercise and psychological therapies at ≤3 months. Very low quality evidence from 1 study with 60 participants showed a clinically important benefit of exercise compared with psychological therapies at ≤3 months . Low quality evidence from 1 study with 152 participants showed no clinically important difference between exercise and psychological therapies at >3 months.

Physical function

Very low quality evidence from 1 study with 98 participants showed a clinically important benefit of exercise compared to psychological therapies at ≤3 months. Low quality evidence from 3 studies with 199 participants showed no clinically important difference between exercise and psychological therapies at ≤3 months. Low quality evidence from 1 study with 105 participants showed a clinically important benefit of exercise compared to psychological therapies at >3 months.

Psychological distress

Low quality evidence from 1 study with 62 participants showed a clinically important benefit of exercise compared to psychological therapies at ≤3 months. Low quality evidence from 1 study with 105 participants showed no clinically important difference between exercise and psychological therapies at >3 months.

Use of healthcare services

No evidence identified.

Sleep

Moderate quality evidence from 1 study with 190 participants showed no clinically important difference between exercise and psychological therapies at >3 months. Low quality evidence from 1 study with 105 participants showed no clinically important difference between exercise and psychological therapies at >3 months.

Discontinuation

Low quality evidence from 10 studies with 1062 participants showed no clinically important difference between exercise and psychological therapies at >3 months.

1.6.1.30 Manual therapy and exercise versus manual therapy

Low quality evidence from 1 study with 101 participants showed no clinically important difference between manual therapy and exercise versus manual therapy for pain at ≤3 months and >3 months, but a clinically important benefit of manual therapy and exercise compared to manual therapy at ≤3 months and >3 months. Very low quality evidence from the same study with 127 participants showed no clinically important difference between the manual therapy and exercise compared to manual therapy for discontinuation.

1.6.1.31 Manual therapy and exercise versus exercise

Pain reduction

Moderate quality evidence from 6 studies with 542 participants showed a clinically important benefit of manual therapy and exercise compared with exercise alone at ≤3 months. Low quality evidence from 3 studies with 394 participants showed no clinically important difference between manual therapy and exercise versus exercise at >3 months.

Health related quality of life

Very low quality evidence from 1 study with 21 participants showed no clinically important difference between manual therapy and exercise versus exercise at >3 months. Moderate quality evidence from 1 study with 180 participants showed no clinically important difference between manual therapy and exercise versus exercise at ≤3 months and >3 months.

Physical function

Low quality evidence from 2 studies with 86 participants showed a clinically important benefit of manual therapy and exercise compared with exercise alone at ≤3 months. Very low quality evidence from 5 studies with 477 participants showed no clinically important difference between manual therapy and exercise versus exercise at ≤3 months. Moderate quality evidence from 3 studies with 394 participants showed no clinically important difference between manual therapy and exercise versus exercise at ≤3 months.

Psychological distress

No evidence identified.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 6 studies with 542 participants showed no clinically important difference between manual therapy and exercise versus exercise at >3 months.

1.6.1.32 Exercise versus manual therapy

Pain reduction

Low quality evidence from 1 study with 101 participants showed a clinically important benefit of exercise compared to psychological therapies at ≤3 months but no clinically important difference between exercise and manual therapies at >3 months.

Health related quality of life

No evidence identified.

Physical function

Low quality evidence from 1 study with 94 participants showed no clinically important difference between exercise and manual therapies at ≤3 months but a clinically important benefit of exercise compared to manual therapies at >3 months.

Psychological distress

No evidence identified.

Use of healthcare services

No evidence identified.

Sleep

No evidence identified.

Discontinuation

Very low quality evidence from 1 study with 127 participants showed more people discontinued from exercise compared to manual therapies at ≤3 months.

1.6.2 Health economic evidence statements

- One cost–utility analysis found that gym-based aerobic exercise therapy was:
    - not cost effective compared to treatment as usual for treating chronic primary pain when using complete case analysis (ICER: £76,960 per QALY). It also found that telephone-delivered cognitive behavioural therapy (TCBT) was dominant (less costly and more effective) compared to exercise therapy.
    - cost effective compared to treatment as usual for treating chronic primary pain when using multiple imputation analysis (ICER: £17,690 per QALY gained). It also found that telephone-delivered cognitive behavioural therapy (TCBT) was dominant (less costly and more effective) compared to exercise therapy.

  This analysis was assessed as directly applicable with potentially serious limitations.

- One cost-utility analysis found that aquatic exercise therapy was cost effective in addition to usual care, compared to usual care (ICER: £3,630 per QALY gained). This analysis was assessed as partially applicable with potentially serious limitations.

- One original cost-utility analysis found that exercise therapy was cost effective compared to no exercise therapy for treating chronic primary pain (probabilistic ICERs: £9,121 per QALY gained (lifetime analysis), £12,683 per QALY gained (no extrapolation analysis), deterministic ICERS: £12,327 per QALY gained (lifetime analysis), £12,739 per QALY gained (no extrapolation analysis). This analysis was assessed as directly applicable with minor limitations.

1.7 The committee's discussion of the evidence

1.7.1 Interpreting the evidence

1.7.1.1 The outcomes that matter most

The committee considered pain reduction, health-related quality of life, physical function and psychological distress to be critical outcomes for decision-making. Use of healthcare services, sleep and discontinuation were also considered to be important outcomes. The critical and important outcomes agreed by the committee were adapted by consensus from relevant core outcome sets registered under the Core Outcome Measures in Effectiveness Trials (COMET) Initiative. This included the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations.

Evidence was identified for all critical and important outcomes.

1.7.1.2 The quality of the evidence

Evidence from 91 randomised controlled trials was identified for 32 different comparisons in this review. Comparisons against usual care with the most evidence were mind-body, aerobic, aerobic plus strength and strength. There were several comparisons of mixed modality exercise versus usual care. A small amount of evidence for some head-to-head comparisons of different types of exercise was also identified. No evidence was identified for graded motor imagery.

The majority of the evidence was of low to very low quality, mainly due to risk of bias and imprecision. There was a lack of blinding in the studies due to the nature of the interventions; this combined with the mostly subjective outcomes resulted in a high risk of performance bias. The majority of the studies had small sample sizes, which increased the uncertainty around the point estimates. Another factor that could have contributed to imprecision was variation in the interventions within the evidence. There were a broad range of exercise programmes which varied in their duration, frequency, intensity, types of exercises and amount of contact with supervisors. This could have influenced the observed effectiveness of each individual intervention within the evidence, leading to greater uncertainty around the point estimates. The committee took into account the low quality evidence, including the uncertainty in their interpretation of the evidence, particularly when considering the small amount of evidence for comparisons between different types of exercise.

The committee noted that the definition of usual care varied across studies or was not clearly reported, which was a general limitation of the review. Usual care generally included: no additional interventions, participants being asked not change their activity levels or to continue normal activities, waiting list controls, low intensity interventions such as advice to stretch or interventions deemed appropriate by the healthcare professionals involved in the study (not including interventions similar to those in the intervention arm of the study).

1.7.1.3 Benefits and harms

The evidence base in general suggested a benefit of exercise therapies over usual care. Although there was uncertainty around the effect estimates for many of the outcomes, the committee agreed that the direction of effect on the whole was positive. Evidence comparing different types of exercise showed little difference in effectiveness between therapies. The majority of evidence involved supervised group exercise.

Exercise versus usual care

Evidence showed that, compared with usual care, there was generally a benefit of both single-modality and mixed-modality exercise therapies for pain reduction and quality of life.

Single-modality exercises

Most types of exercise showed a benefit in terms of improving critical outcomes for people with chronic primary pain (including quality of life, pain, physical function and psychological distress) both in the short-term (less than 3 months) and long-term (more than 3 months), although there was serious uncertainty around the effect estimates for many of the outcomes and in some cases, very serious uncertainty the direction of effect indicated a benefit. Interventions that were shown to be effective include aerobic exercise, strength exercise and mind-body exercises.

Evidence for flexibility alone (for example stretching) or proprioception alone (for example balance exercise) was more limited. Evidence for flexibility exercise was very low quality and was limited to one small study with a short-term follow up and small sample size. This evidence showed a benefit of flexibility in terms of pain, but no difference for physical function. Evidence for other critical outcomes such as psychological distress and quality of life was not available. Similarly evidence for proprioception versus usual care was very low quality and limited to one study with a small number of participants. This showed no benefit of proprioception in the short or long term for pain reduction, quality of life and physical function, and a benefit for psychological distress. The committee agreed that this evidence was not sufficient to determine the effectiveness of flexibility or proprioception exercises alone.

Mixed-modality exercises

Comparisons of mixed-modality exercises versus usual care included:

- Aerobic and strength versus usual care
- Aerobic, strength and flexibility versus usual care
- Strength and flexibility versus usual care
- Strength, proprioception and flexibility versus usual care

Evidence was available for all critical outcomes and generally showed a benefit of these types of exercise for quality of life and pain, although there was uncertainty around the effect estimates for many of the outcomes and in some cases, very serious uncertainty. Evidence for psychological distress and physical function varied across different types of exercise, with some exercise interventions showing a benefit whilst others showed mixed results, again with some uncertainty. There was less evidence for the outcome of sleep, with the majority showing no difference. Evidence for discontinuation was mixed, with some evidence to suggest that more people dropped out of the exercise interventions compared to usual care. However, the committee found the evidence about discontinuation difficult to interpret because usual care was often poorly defined.

Generally, the evidence showed a benefit of mixed-modality exercises for chronic primary pain. No evidence was available to compare mixed-modality exercises to each other, and the committee agreed that evidence was therefore not sufficient to determine whether one type of exercise was more beneficial than another. The committee instead considered that despite

the uncertainty, the evidence reflected an overall benefit of exercise therapies, particularly for reducing pain and improving quality of life, in combination with the lack of negative effects other than discontinuation from the therapy and decided to make a recommendation for exercise.

Head-to-head comparisons (types of exercise compared to each other)

There were 17 different comparisons of different types of exercise compared to each other. The committee found it difficult to draw any firm conclusions regarding a hierarchical order of effectiveness. This was because the evidence was based on small sample sizes, had a high degree of uncertainty and was generally low to very low quality. This contributed to the committee decision not to make a recommendation for one type of exercise over another. When considered alongside the evidence demonstrating that discontinuation from exercise programmes is often an issue, the committee agreed that the choice of type of exercise should be made on an individualised basis, as people are more likely to adhere to an exercise programme that is suited to their needs and preferences.

Exercise versus psychological therapies

Evidence comparing various exercises to psychological therapies was limited, with only a small number of studies available, all of which had small sample sizes. Evidence was available for all critical outcomes but a consistent benefit of either exercise or psychological therapies was not demonstrated. Some outcomes suggested a benefit of exercise in terms of quality of life, physical function and psychological distress. However, there was serious uncertainty around the effect estimates and results were mixed with some evidence suggesting no difference between the two types of interventions (for pain, quality of life, physical function, psychological distress and sleep). Overall, the committee agreed that the evidence was insufficient to determine whether exercise as a whole is more or less effective than psychological therapies. The committee acknowledged that the effects observed with this comparison could have been affected by the type of exercise or psychological therapy in the individual studies contributing to each outcome.

Exercise versus manual therapies

Evidence that directly compared exercise with manual therapies was very limited and inconclusive. When exercise and manual therapies in combination were compared with manual therapies alone, there was a benefit of the addition of exercise for physical function, but no difference in pain or discontinuation. When exercise and manual therapies in combination were compared with exercise therapies alone, evidence showed no difference for pain, quality of life or discontinuation. Evidence for physical function was conflicting, with one outcome based on one small study showing a benefit of exercise and manual therapies in combination, but no difference in any other outcome measures. Overall, the evidence, suggested no benefit of the addition of manual therapy. No evidence was identified for psychological distress, sleep or use of healthcare services for exercise compared with manual therapies.

Summary across comparisons

The committee discussed the applicability of the evidence to the review population and the generalisability to all people with chronic primary pain as the vast majority of the evidence was based on women with fibromyalgia and people with chronic neck pain. The populations were pooled in the clinical review. Where heterogeneity was observed in the effect estimate, this was not explained by subgroup analysis by type of chronic primary pain and therefore the committee agreed that there was no reason recommendations made based on this evidence should not apply for all types of chronic primary pain conditions. The committee considered that despite the uncertainty around the effect estimates, the evidence base was large and benefits were shown across many of the critical and important outcomes, with very little evidence of negative effects except more people discontinuing from exercise

interventions when compared to usual care. There was a clear indication that exercise is beneficial, but the most appropriate type of exercise may depend on the type of pain condition and it should be tailored to individual needs and preferences. This contributed to the committee decision not to make a recommendation about the type of exercise. The committee also noted that the majority of the evidence was based on supervised exercise interventions. In the absence of evidence on unsupervised exercise, the committee agreed to recommend only supervised exercise therapies.

1.7.2 Cost effectiveness and resource use

Two relevant published economic evaluations were identified that compared exercise with usual care. Original economic modelling was also undertaken.

One study was a UK within-trial analysis, looking at a leisure-facility-and-gym-based exercise programme. The comparators included treatment as usual and telephone-delivered cognitive behavioural therapy (TCBT). [NB. The TCBT comparison with usual care is reviewed in the psychological therapies review]. The exercise programme had an ICER of £76,960 per QALY gained compared to treatment as usual using complete case data (the primary analysis in the study) and would therefore not be considered cost effective. When using imputed outcome data, the study found that exercise versus treatment as usual had an ICER of £17,690 per QALY gained and therefore would be considered cost effective. The committee expressed concern over the disparity between the two ICERs, as it is difficult to tell which is a more accurate reflection of the true cost effectiveness of the programme, without knowing the nature of the missing data from the original study. A large amount of data was missing at the follow up 24 months after the intervention ended. This study was rated as directly applicable as it was a UK study from the NHS perspective using the EQ-5D, but with potentially serious methodological limitations such as the fact that the imputed outcomes led to a different conclusion to the complete case data, and the economic evaluation was based on a single RCT. Participation in the study was also based on self-reported symptoms. The committee noted that the cost-effectiveness analysis in the paper would be specific to the exercise programme as described in that particular trial (6 fitness instructor-led monthly sessions, plus a gym membership), which was not typical of the interventions in the other included studies in the review which were more class-based with higher frequency.

The second economic evaluation was a Spanish within-trial analysis, comparing 8 months of group pool-based exercises to usual care. This found exercise to be cost effective with an ICER of £3,630. Pool-based exercises are not considered to be current practice in the UK because they have higher costs. This study was rated as partially applicable with potentially serious limitations because although it uses the EQ-5D, it is not a UK study, it is more out of date than the UK study, and also the costs of the staff involved seem very low compared to UK costs, which is likely to increase the ICER in a UK setting. It is uncertain if this would increase the ICER to above £20,000 per QALY gained.

As both studies had limitations regarding their generalisability because of the types of interventions analysed, and significant uncertainties around cost effectiveness, this question was identified as being a high priority for an original economic analysis.

A cost-utility analysis using a lifetime horizon was undertaken comparing exercise with no exercise. The clinical review looked at each type of exercise separately (for example aerobics, mind body), however the committee agreed they could not infer if one type of exercise had more benefit than another. Therefore, this rationale was also applied to the economic modelling, meaning all the evidence on different types of exercise could be pooled together to make a general recommendation on exercise interventions as a whole. The interventions between studies also varied by intensity, which impacted resource use, however as the clinical review did not stratify by intensity, this supported the committee's decision to pool all the studies for economic analysis.

Treatment effects were based on trials in the review that reported quality of life data, with the model pooling all available quality of life data that reported outcomes at the same time points, to derive an average treatment effect over time. Twelve studies were identified from the review that reported quality of life, either using EQ-5D or SF-36 that could be mapped to the EQ-5D. Differences in quality of life between the exercise and no exercise group in each study were calculated, taking into account the change from baseline in each arm, to derive the quality of life gain from exercise compared to no exercise for each study. A linear trend line was fitted to the pooled quality of life gain at each time point, and this was used to determine the QALY gain of the area under this line. The average treatment effect was also extrapolated beyond the available trial data, based on committee assumptions. Costs included only the costs of the staff time involved in providing an exercise programme. The total resource use from each study being used for treatment effect was identified and costed up, and a weighted average was taken based on the number of participants analysed in the intervention arm of each trial. All studies were looking at supervised exercise, and the majority were assumed to be group based (either because this was stated, or using their description of the intervention, or committee judgement) except one study known to be individual treatment.

Two base cases were modelled, one using a lifetime horizon and the other assuming no extrapolation beyond the trial data. Both base cases showed that exercise was cost effective compared with no exercise, with probabilistic ICERs of £9,121 (86% probability of exercise being cost effective at a threshold of £20,000 per QALY gained), and £12,683 (93% probability) respectively, and deterministic ICERS of £12,327 and £12,739 respectively. Various sensitivity analyses were undertaken, including varying costs, and including data omitted from the base case. The overall conclusion was robust to all sensitivity analyses tested.

The committee discussed the limitations of the analysis, which included how this was only based on a small proportion of studies from the clinical review as a whole (around 12%). However, they agreed that the studies used in the economic analysis were generally representative of the populations in the review as a whole and the populations that would be seen in practice with chronic primary pain (in other words, a mix of people with fibromyalgia and other chronic pain conditions). There was also a wide heterogeneity in the data being used in the model, as studies had very different populations, interventions, and intensities, and these were pooled together in the model. There is also uncertainty around the relationship between resource use and treatment benefit, and this needs to be considered then interpreting the results. It was not considered appropriate to explore this relationship more formally in the model (such as by modelling each study separately), as the clinical review did not establish which characteristics of exercise interventions improve outcomes.

The committee agreed that they had reservations about the two economic evaluations found in the literature, and that the economic analysis undertaken as part of the guideline pooled more data and was therefore considered more robust. The quality of life data from the identified UK economic evaluation was also included in the original economic analysis. The differences in results between the guideline original analysis and the UK economic evaluation are probably attributable to the fact that treatment effects were larger in the other trials included in the model, and additionally the UK economic evaluation found much higher health service costs in the exercise group at 18-24 months after intervention (i.e. they were using more health services). However it is difficult to know if the longer term health service costs were anything to do with the intervention after such long follow up.

Given that the clinical evidence showed there was some benefit from exercise, and taking that into account alongside the highly likely cost effectiveness of exercise, the committee decided to make a strong recommendation to offer exercise.

1.7.3 Other factors the committee took into account

The committee discussed that this review covered the use of exercise interventions to manage chronic primary pain. The committee's experience was that many people with chronic primary pain find it difficult to be physically active. The UK Chief Medical Officers' 'Physical Activity Guidelines' (2019) highlights that sedentary behaviour is an independent risk factor for poor health outcomes, including cardiovascular and cancer mortality, and obesity-related morbidity. NICE has published a range of guidance on physical activity. NICE also published guidance to ensure that interventions, including staff training, to improve population health and wellbeing meet individual needs: Behaviour change: individual approaches.

The committee therefore wished to highlight that there are important public health benefits to engaging in any physical activity for people with chronic primary pain, particularly if they are inactive or sedentary. The committee agreed that, for the chronic primary pain population, it was important to recommend continuing physical activity beyond the end of a formal exercise programme in a manner that is sustainable for the person. The committee discussed that if costs are incurred by engaging in physical activity after a formal exercise programme for management of chronic primary pain ends, this would be a personal cost, and would not fall to the NHS. Therefore, there were no implementation costs attributable to this recommendation.

References

1. Acar B, Yilmaz OT. Effects of different physiotherapy applications on pain and mobility of connective tissue in patients with myofascial pain syndrome. Journal of Back and Musculoskeletal Rehabilitation. 2012; 25(4):261-267

2. Acosta-Gallego A, Ruiz-Montero PJ, Castillo-Rodriguez A. Land- and pool-based intervention in female fibromyalgia patients: A randomized-controlled trial. Turkish Journal of Physical Medicine and Rehabilitation. 2018; 64(4):337-343

3. ACTRN. My Knee Exercise: a 6 month electronically delivered intervention to support self-management for people with knee osteoarthritis: a pragmatic randomised controlled trial. 2018. Available from: https://anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12618001167257 Last accessed: 29/06/2020.

4. Adamse C, Dekker-Van Weering MG, van Etten-Jamaludin FS, Stuiver MM. The effectiveness of exercise-based telemedicine on pain, physical activity and quality of life in the treatment of chronic pain: A systematic review. Journal of Telemedicine and Telecare. 2018; 24(8):511-526

5. Akhter S, Khan M, Ali SS, Soomro RR. Role of manual therapy with exercise regime versus exercise regime alone in the management of non-specific chronic neck pain. Pakistan Journal of Pharmaceutical Sciences. 2014; 27(6 Suppl):2125-2128

6. Alentorn-Geli E, Moras G, Padilla J, Fernandez-Sola J, Bennett RM, Lazaro-Haro C et al. Effect of acute and chronic whole-body vibration exercise on serum insulin-like growth factor-1 levels in women with fibromyalgia. Journal of Alternative and Complementary Medicine. 2009; 15(5):573-578

7. Alentorn-Geli E, Padilla J, Moras G, Lazaro Haro C, Fernandez-Sola J. Six weeks of whole-body vibration exercise improves pain and fatigue in women with fibromyalgia. Journal of Alternative and Complementary Medicine. 2008; 14(8):975-981

8. Allende S, Anandan A, Lauche R, Cramer H. Effect of yoga on chronic non-specific neck pain: An unconditional growth model. Complementary Therapies in Medicine. 2018; 40:237-242

9. Altan L, Bingol U, Aykac M, Koc Z, Yurtkuran M. Investigation of the effects of pool-based exercise on fibromyalgia syndrome. Rheumatology International. 2004; 24(5):272-277

10. Altan L, Korkmaz N, Bingol U, Gunay B. Effect of pilates training on people with fibromyalgia syndrome: a pilot study. Archives of Physical Medicine and Rehabilitation. 2009; 90(12):1983-1988

11. Amanollahi A, Naghizadeh J, Khatibi A, Hollisaz MT, Shamseddini AR, Saburi A. Comparison of impacts of friction massage, stretching exercises and analgesics on pain relief in primary fibromyalgia syndrome: a randomized clinical trial. Tehran University Medical Journal. 2013; 70(10):616-622

12. Amiri Arimi S, Mohseni Bandpei MA, Javanshir K, Rezasoltani A, Biglarian A. The effect of different exercise programs on size and function of deep cervical flexor muscles in patients with chronic nonspecific neck pain: A systematic review of randomized controlled trials. American Journal of Physical Medicine and Rehabilitation. 2017; 96(8):582-588

13. Amris K, Wæhrens EE, Christensen R, Bliddal H, Danneskiold-Samsøe B. Interdisciplinary rehabilitation of patients with chronic widespread pain: primary

endpoint of the randomized, nonblinded, parallel-group IMPROvE trial. Pain. 2014; 155(7):1356-1364

14. Andersen LL, Andersen CH, Zebis MK, Nielsen PK, Sogaard K, Sjogaard G. Effect of physical training on function of chronically painful muscles: A randomized controlled trial. Journal of Applied Physiology. 2008; 105(6):1796-1801

15. Andrade A, Vilarino GT, Sieczkowska SM, Coimbra DR, Steffens RDAK, Vietta GG. Acute effects of physical exercises on the inflammatory markers of patients with fibromyalgia syndrome: A systematic review. Journal of Neuroimmunology. 2018; 316:40-49

16. Andrade CP, Zamuner AR, Forti M, Franca TF, Tamburus NY, Silva E. Oxygen uptake and body composition after aquatic physical training in women with fibromyalgia: a randomized controlled trial. European Journal of Physical and Rehabilitation Medicine. 2017; 53(5):751-758

17. Andrade CP, Zamuner AR, Forti M, Tamburus NY, Silva E. Effects of aquatic training and detraining on women with fibromyalgia: controlled randomized clinical trial. European Journal of Physical and Rehabilitation Medicine. 2019; 55(1):79-88

18. Arami J, Rezasoltani A, Khalkhali Zaavieh M, Rahnama L. The effect of two exercise therapy programs (proprioceptive and endurance training) to treat patients with chronic non-specific neck pain. Journal of babol university of medical sciences. 2012; 14(1):78-84

19. Arcos-Carmona IM, Castro-Sánchez AM, Matarán-Peñarrocha GA, Gutiérrez-Rubio AB, Ramos-González E, Moreno-Lorenzo C. Effects of aerobic exercise program and relaxation techniques on anxiety, quality of sleep, depression, and quality of life in patients with fibromyalgia: a randomized controlled trial. Medicina Clínica. 2011; 137(9):398-401

20. Asenlof P, Denison E, Lindberg P. Individually tailored treatment targeting activity, motor behavior, and cognition reduces pain-related disability: a randomized controlled trial in patients with musculoskeletal pain. Journal of Pain. 2005; 6(9):588-603

21. Asenlof P, Denison E, Lindberg P. Long-term follow-up of tailored behavioural treatment and exercise based physical therapy in persistent musculoskeletal pain: A randomized controlled trial in primary care. European Journal of Pain. 2009; 13(10):1080-1088

22. Assis MR, Silva LE, Alves AM, Pessanha AP, Valim V, Feldman D et al. A randomized controlled trial of deep water running: clinical effectiveness of aquatic exercise to treat fibromyalgia. Arthritis and Rheumatism. 2006; 55(1):57-65

23. Assumpcao A, Matsutani LA, Yuan SL, Santo AS, Sauer J, Mango P et al. Muscle stretching exercises and resistance training in fibromyalgia: which is better? A three-arm randomized controlled trial. European Journal of Physical and Rehabilitation Medicine. 2018; 54(5):663-670

24. Assuncao Junior JC, de Almeida Silva HJ, da Silva JFC, da Silva Cruz R, de Almeida Lins CA, de Souza MC. Zumba dancing can improve the pain and functional capacity in women with fibromyalgia. Journal of Bodywork and Movement Therapies. 2018; 22(2):455-459

25. Astin JA, Berman BM, Bausell B, Lee WL, Hochberg M, Forys KL. The efficacy of mindfulness meditation plus Qigong movement therapy in the treatment of fibromyalgia: a randomized controlled trial. Journal of Rheumatology. 2003; 30(10):2257-2262

26. Bai Z, Guan Z, Fan Y, Liu C, Yang K, Ma B et al. The effects of qigong for adults with chronic pain: Systematic review and meta-analysis. American Journal of Chinese Medicine. 2015; 43(8):1525-1539

27. Baptista AS, Villela AL, Jones A, Natour J. Effectiveness of dance in patients with fibromyalgia: a randomized, single-blind, controlled study. Clinical and Experimental Rheumatology. 2012; 30(6 Suppl 74):18-23

28. Beasley M, Prescott GJ, Scotland G, McBeth J, Lovell K, Keeley P et al. Patient-reported improvements in health are maintained 2 years after completing a short course of cognitive behaviour therapy, exercise or both treatments for chronic widespread pain: long-term results from the MUSICIAN randomised controlled trial. RMD Open. 2015; 1:e000026

29. Beltran-Alacreu H, Lopez-de-Uralde-Villanueva I, Fernandez-Carnero J, La Touche R. Manual therapy, therapeutic patient education, and therapeutic exercise, an effective multimodal treatment of nonspecific chronic neck pain: A randomized controlled trial. American Journal of Physical Medicine and Rehabilitation. 2015; 94(10 Suppl 1):887-897

30. Bergström C, Jensen I, Hagberg J, Busch H, Bergström G. Effectiveness of different interventions using a psychosocial subgroup assignment in chronic neck and back pain patients: a 10-year follow-up. Disability and Rehabilitation. 2012; 34(2):110-118

31. Bertozzi L, Gardenghi I, Turoni F, Villafane JH, Capra F, Guccione AA et al. Effect of therapeutic exercise on pain and disability in the management of chronic nonspecific neck pain: systematic review and meta-analysis of randomized trials. Physical Therapy. 2013; 93(8):1026-1036

32. Bidonde J, Busch A, Schachter C, Webber S, Musselman K, Overend T et al. Mixed exercise training for adults with fibromyalgia. Cochrane Database of Systematic Reviews 2019, Issue 5. Art. No.: CD013340. DOI: 10.1002/14651858.CD013340.

33. Bidonde J, Busch AJ, Schachter CL, Overend TJ, Kim SY, Góes SM et al. Aerobic exercise training for adults with fibromyalgia. Cochrane Database of Systematic Reviews 2017, Issue 6. Art. No.: CD012700. DOI: 10.1002/14651858.CD012700.

34. Bircan C, Karasel SA, Akgun B, El O, Alper S. Effects of muscle strengthening versus aerobic exercise program in fibromyalgia. Rheumatology International. 2008; 28(6):527-532

35. Bjersing JL, Larsson A, Palstam A, Ernberg M, Bileviciute-Ljungar I, Lofgren M et al. Benefits of resistance exercise in lean women with fibromyalgia: involvement of IGF-1 and leptin. BMC Musculoskeletal Disorders. 2017; 18:106

36. Bland P. Tai chi of benefit in fibromyalgia. Practitioner. 2010; 254(1735):8-10

37. Bobos P, Billis E, Papanikolaou DT, Koutsojannis C, Macdermid JC. Does deep cervical flexor muscle training affect pain pressure thresholds of myofascial trigger points in patients with chronic neck pain? A prospective randomized controlled trial. Rehabilitation Research and Practice. 2016; 2016:6480826

38. Bojner-Horwitz E, Theorell T, Anderberg UM. Dance/movement therapy and changes in stress-related hormones: A study of fibromyalgia patients with video-interpretation. The Arts in Psychotherapy. 2003; 30(5):255-264

39. Borisut S, Vongsirinavarat M, Vachalathiti R, Sakulsriprasert P. Effects of strength and endurance training of superficial and deep neck muscles on muscle activities and pain levels of females with chronic neck pain. Journal of Physical Therapy Science. 2013; 25(9):1157-1162

40. Bowering KJ, O'Connell NE, Tabor A, Catley MJ, Leake HB, Moseley GL et al. The effects of graded motor imagery and its components on chronic pain: a systematic review and meta-analysis. Journal of Pain. 2013; 14(1):3-13

41. Brage K, Ris I, Falla D, Sogaard K, Juul-Kristensen B. Pain education combined with neck- and aerobic training is more effective at relieving chronic neck pain than pain education alone--A preliminary randomized controlled trial. Manual Therapy. 2015; 20(5):686-693

42. Bravo C, Skjaerven LH, Espart A, Guitard Sein-Echaluce L, Catalan-Matamoros D. Basic Body Awareness Therapy in patients suffering from fibromyalgia: a randomized clinical trial. Physiotherapy theory and practice. 2019; 35(10):919-929

43. Bronfort G, Evans R, Nelson B, Aker PD, Goldsmith CH, Vernon H. A randomized clinical trial of exercise and spinal manipulation for patients with chronic neck pain. Spine. 2001; 26(7):788-799

44. Buckelew SP, Conway R, Parker J, Deuser WE, Read J, Witty TE et al. Biofeedback/relaxation training and exercise interventions for fibromyalgia: a prospective trial. Arthritis Care and Research. 1998; 11(3):196-209

45. Burckhardt CS, Mannerkorpi K, Bjelle A. A randomized, controlled rial of education and physical therapy for women with fibromyalgia syndrome (FMS). Scandinavian Journal of Rheumatology - Supplement. 1992; 94:48

46. Burckhardt CS, Mannerkorpi K, Hedenberg L, Bjelle A. A randomized, controlled clinical trial of education and physical training for women with fibromyalgia. Journal of Rheumatology. 1994; 21(4):714-720

47. Busch AJ, Barber KAR, Overend TJ, Peloso PMJ, Schachter CL. Exercise for treating fibromyalgia syndrome. Cochrane Database of Systematic Reviews 2007, Issue 4. Art. No.: CD003786. DOI: 10.1002/14651858.CD003786.pub2.

48. Busch AJ, Schachter CL, Overend TJ, Peloso PM, Barber KA. Exercise for fibromyalgia: a systematic review. Journal of Rheumatology. 2008; 35(6):1130-1144

49. Busch AJ, Webber SC, Richards RS, Bidonde J, Schachter CL, Schafer LA et al. Resistance exercise training for fibromyalgia. Cochrane Database of Systematic Reviews 2013, Issue 12. Art. No.: CD010884. DOI: 10.1002/14651858.CD010884.

50. Calandre EP, Rodriguez-Claro ML, Rico-Villademoros F, Vilchez JS, Hidalgo J, Delgado-Rodriguez A. Effects of pool-based exercise in fibromyalgia symptomatology and sleep quality: a prospective randomised comparison between stretching and Ai Chi. Clinical and Experimental Rheumatology. 2009; 27(5 Suppl 56):S21-28

51. Cantarero-Villanueva I, Fernandez-Lao C, Fernandez-de-Las-Penas C, Lopez-Barajas IB, Del-Moral-Avila R, de la-Llave-Rincon AI et al. Effectiveness of water physical therapy on pain, pressure pain sensitivity, and myofascial trigger points in breast cancer survivors: a randomized, controlled clinical trial. Pain Medicine. 2012; 13(11):1509-1519

52. Carbonell-Baeza A, Ruiz JR, Aparicio VA, Ortega FB, Munguia-Izquierdo D, Alvarez-Gallardo IC et al. Land- and water-based exercise intervention in women with fibromyalgia: the al-Andalus physical activity randomised controlled trial. BMC Musculoskeletal Disorders. 2012; 13:18

53. Carson JW, Carson KM, Jones KD, Bennett RM, Wright CL, Mist SD. A pilot randomized controlled trial of the Yoga of Awareness program in the management of fibromyalgia. Pain. 2010; 151(2):530-539

54. Carson JW, Carson KM, Jones KD, Mist SD, Bennett RM. Follow-up of yoga of awareness for fibromyalgia: results at 3 months and replication in the wait-list group. Clinical Journal of Pain. 2012; 28(9):804-813

55. Carvalho MS, Carvalho LC, Menezes FDS, Frazin A, Gomes EDC, Iunes DH. Effects of Exergames in Women with Fibromyalgia: A Randomized Controlled Study. Games for Health Journal. 2020; 05:05

56. Cerrillo-Urbina AJ, Garcia-Hermoso A, Sanchez-Lopez M, Martinez-Vizcaino V. Effect of exercise programs on symptoms of fibromyalgia in peri-menopausal age women: A systematic review and meta-analysis of randomized controlled trials. Myopain. 2015; 23(1-2):56-70

57. Champagne R, Ronzil Y, Roche-Leboucher G, Begue C, Dubus V, Bontoux L et al. Effectiveness of an outpatient rehabilitation program with multidisciplinary approach on return to work for patients with non-specific chronic lombal pain. Annals of Physical and Rehabilitation Medicine. 2018; 61S:e16

58. Chan CL, Wang CW, Ho RT, Ng SM, Ziea ET, Wong VT. Qigong exercise for the treatment of fibromyalgia: a systematic review of randomized controlled trials. Journal of Alternative and Complementary Medicine. 2012; 18(7):641-646

59. Chiu TT, Lam TH, Hedley AJ. A randomized controlled trial on the efficacy of exercise for patients with chronic neck pain. Spine. 2005; 30(1):E1-7

60. Cho Y, Do J, Jung S, Kwon O, Jeon JY. Effects of a physical therapy program combined with manual lymphatic drainage on shoulder function, quality of life, lymphedema incidence, and pain in breast cancer patients with axillary web syndrome following axillary dissection. Supportive Care in Cancer. 2016; 24(5):2047-2057

61. Chung S, Jeong YG. Effects of the craniocervical flexion and isometric neck exercise compared in patients with chronic neck pain: A randomized controlled trial. Physiotherapy Theory & Practice. 2018; 34(12):916-925

62. Collado-Mateo D, Dominguez-Munoz FJ, Adsuar JC, Garcia-Gordillo MA, Gusi N. Effects of exergames on quality of life, pain, and disease effect in women with fibromyalgia: A randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2017; 98(9):1725-1731

63. Cramer H, Klose P, Brinkhaus B, Michalsen A, Dobos G. Effects of yoga on chronic neck pain: a systematic review and meta-analysis. Clinical Rehabilitation. 2017; 31(11):1457-1465

64. Cramer H, Lauche R, Hohmann C, Langhorst J, Dobos G. Yoga for chronic neck pain: a 12-month follow-up. Pain Medicine. 2013; 14(4):541-548

65. Cramer H, Lauche R, Hohmann C, Ludtke R, Haller H, Michalsen A et al. Randomized-controlled trial comparing yoga and home-based exercise for chronic neck pain. Clinical Journal of Pain. 2013; 29(3):216-223

66. Cramer H, Lauche R, Klose P, Lange S, Langhorst J, Dobos GJ. Yoga for improving health-related quality of life, mental health and cancer-related symptoms in women diagnosed with breast cancer. Cochrane Database of Systematic Reviews 2017, Issue 1. Art. No.: CD010802. DOI: 10.1002/14651858.CD010802.pub2.

67. Da Costa D, Abrahamowicz M, Lowensteyn I, Bernatsky S, Dritsa M, Fitzcharles MA et al. A randomized clinical trial of an individualized home-based exercise programme for women with fibromyalgia. Rheumatology. 2005; 44(11):1422-1427

68. de Araujo Cazotti L, Jones A, Roger-Silva D, Ribeiro LHC, Natour J. Effectiveness of the pilates method in the treatment of chronic mechanical neck pain: A randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2018; 99(9):1740-1746

69. de Medeiros SA, de Almeida Silva HJ, do Nascimento RM, da Silva Maia JB, de Almeida Lins CA, de Souza MC. Mat Pilates is as effective as aquatic aerobic exercise in treating women with fibromyalgia: a clinical, randomized and blind trial. Advances in Rheumatology. 2020; 60(1):21

70. Demir-Gocmen D, Altan L, Korkmaz N, Arabaci R. Effect of supervised exercise program including balance exercises on the balance status and clinical signs in patients with fibromyalgia. Rheumatology International. 2013; 33(3):743-750

71. Dobkin PL, Abrahamowicz M, Fitzcharles MA, Dritsa M, da Costa D. Maintenance of exercise in women with fibromyalgia. Arthritis and Rheumatism. 2005; 53(5):724-731

72. Dunleavy K, Kava K, Goldberg A, Malek MH, Talley SA, Tutag-Lehr V et al. Comparative effectiveness of Pilates and yoga group exercise interventions for chronic mechanical neck pain: quasi-randomised parallel controlled study. Physiotherapy. 2016; 102(3):236-242

73. Duray M, Simsek S, Altug F, Cavlak U. Effect of proprioceptive training on balance in patients with chronic neck pain. Agri Dergisi. 2018; 30(3):130-137

74. Duruturk N, Tuzun EH, Culhaoglu B. Is balance exercise training as effective as aerobic exercise training in fibromyalgia syndrome? Rheumatology International. 2015; 35(5):845-854

75. Dusunceli Y, Ozturk C, Atamaz F, Hepguler S, Durmaz B. The comparison of cervicothoracic stabilization exercises with isometric and streching exercises in patients with chronic neck pain. Journal of rheumatology and medical rehabilitation. 2006; 17(3):185-192

76. Effectiveness of cervical stabilisation exercises on respiratory strength in chronic neck pain patients with forward head posture-A pilot study. Journal of Clinical and Diagnostic Research. 2019; 13(4):YC06-YC09

77. Ekici G, Yakut E, Akbayrak T. [Effects of Pilates exercises and connective tissue manipulation on pain and depression in females with fibromyalgia: a randomized controlled trial]. Fizyoterapi rehabilitasyon. 2008; 19(2):47-54

78. El-Gendy MH, Lasheen YR, Rezkalla WKS. Multimodal approach of electrotherapy versus myofascial release in patients with chronic mechanical neck pain: A randomized controlled trial. Physiotherapy Quarterly. 2019; 27(4):6-12

79. Emilson C, Demmelmaier I, Bergman S, Lindberg P, Denison E, Asenlof P. A 10-year follow-up of tailored behavioural treatment and exercise-based physiotherapy for persistent musculoskeletal pain. Clinical Rehabilitation. 2017; 31(2):186-196

80. Ericsson A, Bremell T, Cider A, Mannerkorpi K. Effects of exercise on fatigue and physical capacity in men with chronic widespread pain - a pilot study. BMC Sports Science, Medicine and Rehabilitation. 2016; 8:29

81. Ericsson A, Palstam A, Larsson A, Lofgren M, Bileviciute-Ljungar I, Bjersing J et al. Resistance exercise improves physical fatigue in women with fibromyalgia: a randomized controlled trial. Arthritis Research & Therapy. 2016; 18:176

82. Ernberg M, Christidis N, Ghafouri B, Bileviciute-Ljungar I, Lofgren M, Larsson A et al. Effects of 15 weeks of resistance exercise on pro-inflammatory cytokine levels in the

vastus lateralis muscle of patients with fibromyalgia. Arthritis Research & Therapy. 2016; 18(1):137

83. Espi-Lopez GV, Ingles M, Ruescas-Nicolau MA, Moreno-Segura N. Effect of low-impact aerobic exercise combined with music therapy on patients with fibromyalgia. A pilot study. Complementary Therapies in Medicine. 2016; 28(October 2016):1-7

84. Etnier JL, Karper WB, Gapin JI, Barella LA, Chang YK, Murphy KJ. Exercise, fibromyalgia, and fibrofog: a pilot study. Journal of Physical Activity & Health. 2009; 6(2):239-246

85. Evans R, Bronfort G, Nelson B, Goldsmith CH. Two-year follow-up of a randomized clinical trial of spinal manipulation and two types of exercise for patients with chronic neck pain. Spine. 2002; 27(21):2383-2389

86. Evans R, Bronfort G, Schulz C, Maiers M, Bracha Y, Svendsen K et al. Supervised exercise with and without spinal manipulation performs similarly and better than home exercise for chronic neck pain: a randomized controlled trial. Spine. 2012; 37(11):903-914

87. Evcik D, Yigit I, Pusak H, Kavuncu V. Effectiveness of aquatic therapy in the treatment of fibromyalgia syndrome: a randomized controlled open study. Rheumatology International. 2008; 28(9):885-890

88. Falla D, Jull G, Hodges P, Vicenzino B. An endurance-strength training regime is effective in reducing myoelectric manifestations of cervical flexor muscle fatigue in females with chronic neck pain. Clinical Neurophysiology. 2006; 117(4):828-837

89. Falla D, Jull G, Russell T, Vicenzino B, Hodges P. Effect of neck exercise on sitting posture in patients with chronic neck pain. Physical Therapy. 2007; 87(4):408-417

90. Falla D, Lindstrom R, Rechter L, Boudreau S, Petzke F. Effectiveness of an 8-week exercise programme on pain and specificity of neck muscle activity in patients with chronic neck pain: a randomized controlled study. European Journal of Pain. 2013; 17(10):1517-1528

91. Fernandes G, Jennings F, Nery Cabral MV, Pirozzi Buosi AL, Natour J. Swimming improves pain and functional capacity of patients with fibromyalgia: A randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2016; 97(8):1269-1275

92. Field T, Delage J, Hernandez-Reif M. Movement and massage therapy reduce fibromyalgia pain. Journal of Bodywork and Movement Therapies. 2003; 7(1):49-52

93. Fontaine KR, Conn L, Clauw DJ. Effects of lifestyle physical activity in adults with fibromyalgia: results at follow-up. JCR: Journal of Clinical Rheumatology. 2011; 17(2):64-68

94. Fontaine KR, Conn L, Clauw DJ. Effects of lifestyle physical activity on perceived symptoms and physical function in adults with fibromyalgia: results of a randomized trial. Arthritis Research & Therapy. 2010; 12(2):R55

95. Fontaine KR, Haaz S. Effects of lifestyle physical activity on health status, pain, and function in adults with fibromyalgia syndrome. Journal of Musculoskeletal Pain. 2007; 15(1):3-9

96. Galindez-Ibarbengoetxea X, Setuain I, Ramirez-Velez R, Andersen LL, Gonzalez-Izal M, Jauregi A et al. Short-term effects of manipulative treatment versus a therapeutic home exercise protocol for chronic cervical pain: A randomized clinical trial. Journal of Back and Musculoskeletal Rehabilitation. 2018; 31(1):133-145

97. Gallego Izquierdo T, Pecos-Martin D, Lluch Girbes E, Plaza-Manzano G, Rodriguez Caldentey R, Mayor Melus R et al. Comparison of cranio-cervical flexion training versus cervical proprioception training in patients with chronic neck pain: A randomized controlled clinical trial. Journal of Rehabilitation Medicine. 2016; 48(1):48-55

98. Garcia-Hermoso A, Saavedra JM, Escalante Y. Effects of exercise on functional aerobic capacity in adults with fibromyalgia syndrome: A systematic review of randomized controlled trials. Journal of Back and Musculoskeletal Rehabilitation. 2015; 28(4):609-619

99. Garcia-Martinez AM, De Paz JA, Marquez S. Effects of an exercise programme on self-esteem, self-concept and quality of life in women with fibromyalgia: a randomized controlled trial. Rheumatology International. 2012; 32(7):1869-1876

100. Gavi MB, Vassalo DV, Amaral FT, Macedo DC, Gava PL, Dantas EM et al. Strengthening exercises improve symptoms and quality of life but do not change autonomic modulation in fibromyalgia: a randomized clinical trial. PloS One. 2014; 9(3):e90767

101. Gavish A, Winocur E, Astandzelov-Nachmias T, Gazit E. Effect of controlled masticatory exercise on pain and muscle performance in myofascial pain patients: A pilot study. Cranio. 2006; 24(3):184-190

102. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database of Systematic Reviews 2017, Issue 4. Art. No.: CD011279. DOI: 10.1002/14651858.CD011279.pub3.

103. Ghaderi F, Jafarabadi MA, Javanshir K. The clinical and EMG assessment of the effects of stabilization exercise on nonspecific chronic neck pain: A randomized controlled trial. Journal of Back and Musculoskeletal Rehabilitation. 2017; 30(2):211-219

104. Ghodrati M, Mosallanezhad Z, Shati M, Noroozi M, Moghadam AN, Rostami M et al. Adding Temporomandibular joint treatments to routine physiotherapy for patients with non-specific chronic neck pain: A randomized clinical study. Journal of Bodywork and Movement Therapies. 2020;

105. Giannotti E, Koutsikos K, Pigatto M, Rampudda ME, Doria A, Masiero S. Medium-/long-term effects of a specific exercise protocol combined with patient education on spine mobility, chronic fatigue, pain, aerobic fitness and level of disability in fibromyalgia. BioMed Research International. 2014; 2014:474029

106. Giubilei G, Mondaini N, Minervini A, Saieva C, Lapini A, Serni S et al. Physical activity of men with chronic prostatitis/chronic pelvic pain syndrome not satisfied with conventional treatments--could it represent a valid option? The physical activity and male pelvic pain trial: a double-blind, randomized study. Journal of Urology. 2007; 177(1):159-165

107. Glasgow A, Stone TM, Kingsley JD. Resistance exercise training on disease impact, pain catastrophizing and autonomic modulation in women with fibromyalgia. International Journal of Exercise Science. 2017; 10(8):1184-1195

108. Gomez-Hernandez M, Gallego-Izquierdo T, Martinez-Merinero P, Pecos-Martin D, Ferragut-Garcias A, Hita-Contreras F et al. Benefits of adding stretching to a moderate-intensity aerobic exercise programme in women with fibromyalgia: a randomized controlled trial. Clinical Rehabilitation. 2020; 34(2):242-251

109. Gowans SE, deHueck A. Pool exercise for individuals with fibromyalgia. Current Opinion in Rheumatology. 2007; 19(2):168-173

110. Gowans SE, DeHueck A, Abbey SE. Measuring exercise-induced mood changes in fibromyalgia: a comparison of several measures. Arthritis and Rheumatism. 2002; 47(6):603-609

111. Gowans SE, deHueck A, Voss S, Richardson M. A randomized, controlled trial of exercise and education for individuals with fibromyalgia. Arthritis Care and Research. 1999; 12(2):120-128

112. Gowans SE, Dehueck A, Voss S, Silaj A, Abbey SE. Six-month and one-year followup of 23 weeks of aerobic exercise for individuals with fibromyalgia. Arthritis and Rheumatism. 2004; 51(6):890-898

113. Gowans SE, deHueck A, Voss S, Silaj A, Abbey SE, Reynolds WJ. Effect of a randomized, controlled trial of exercise on mood and physical function in individuals with fibromyalgia. Arthritis and Rheumatism. 2001; 45(6):519-529

114. Gross A, Langevin P, Burnie SJ, Bédard-Brochu M-S, Empey B, Dugas E et al. Manipulation and mobilisation for neck pain contrasted against an inactive control or another active treatment. Cochrane Database of Systematic Reviews 2015, Issue 9. Art. No.: CD004249. DOI: 10.1002/14651858.CD004249.pub4.

115. GunendiZ, MerayJ, OzdemS. The effect of a 4-week aerobic exercise program on muscle performance in patients with fibromyalgia. Journal of Back and Musculoskeletal Rehabilitation. 2008; 21(3):185-191

116. Gusi N, Tomas-Carus P. Cost-utility of an 8-month aquatic training for women with fibromyalgia: a randomized controlled trial. Arthritis Research & Therapy. 2008; 10(1):R24

117. Gutierrez-Espinoza H, Araya-Quintanilla F, Gutierrez-Monclus R, Rios-Riquelme M, Alvarez-Bueno C, Martinez-Vizcaino V et al. Does pectoralis minor stretching provide additional benefit over an exercise program in participants with subacromial pain syndrome? A randomized controlled trial. Musculoskeletal science and practice. 2019; 44

118. Haak T, Scott B. The effect of Qigong on fibromyalgia (FMS): a controlled randomized study. Disability and Rehabilitation. 2008; 30(8):625-633

119. Hakkinen A, Kautiainen H, Hannonen P, Ylinen J. Strength training and stretching versus stretching only in the treatment of patients with chronic neck pain: a randomized one-year follow-up study. Clinical Rehabilitation. 2008; 22(7):592-600

120. Hakkinen K, Pakarinen A, Hannonen P, Hakkinen A, Airaksinen O, Valkeinen H et al. Effects of strength training on muscle strength, cross-sectional area, maximal electromyographic activity, and serum hormones in premenopausal women with fibromyalgia. Journal of Rheumatology. 2002; 29(6):1287-1295

121. Hammond A, Freeman K. Community patient education and exercise for people with fibromyalgia: a parallel group randomized controlled trial. Clinical Rehabilitation. 2006; 20(10):835-846

122. Har E. Influence of neck exercises, combined with either the Chace technique of dance therapy or aerobic training, on pain perception, mood state and cervical range of motion of adults with chronic mechanical neck pain. New York. New York University. 2000

123. Hoeger Bement MK, Weyer A, Hartley S, Drewek B, Harkins AL, Hunter SK. Pain perception after isometric exercise in women with fibromyalgia. Archives of Physical Medicine and Rehabilitation. 2011; 92(1):89-95

124. Holmer ML, Gevirtz R, Spira JL, Greenberg MA. The effects of yoga on symptoms and psychosocial adjustment in fibromyalgia syndrome patients. Conference abstracts from papers presented at the 35th annual meeting. 2004; 29:302

125. Hooten WM, Qu W, Townsend CO, Judd JW. Effects of strength vs aerobic exercise on pain severity in adults with fibromyalgia: a randomized equivalence trial. Pain. 2012; 153(4):915-923

126. Humphreys BK, Irgens PM. The effect of a rehabilitation exercise program on head repositioning accuracy and reported levels of pain in chronic neck pain subjects. Journal of whiplash & related disorders. 2002; 1(1):99-112

127. Iaroshevskyi OA, Morozova OG, Logvinenko AV, Lypynska YV. Non-pharmacological treatment of chronic neck-shoulder myofascial pain in patients with forward head posture. Wiadomosci lekarskie (Warsaw, Poland : 1960). 2019; 72(1):84-88

128. Ide MR, Laurindo LMM, Rodrigues-Junior AL, Tanaka C. Effect of aquatic respiratory exercise-based program in patients with fibromyalgia. International Journal of Rheumatic Diseases. 2008; 11(2):131-140

129. Im SH, Han EY. Improvement in anxiety and pain after whole body whirlpool hydrotherapy among patients with myofascial pain syndrome. Annals of Rehabilitation Medicine. 2013; 37(4):534-540

130. Isomeri R, Mikkelsson M, Latikka P. Effects of amitriptyline and cardiovascular fitness training on the pain of fibromyalgia patients. Scandinavian Journal of Rheumatology - Supplement. 1992; 94:47

131. Isomeri R, Mikkelsson M, Latikka P, Kammonen K. Effects of amitriptyline and cardiovascular fitness training on pain in patients with primary fibromyalgia. Journal of Musculoskeletal Pain. 1993; 1(3-4):253-260

132. Izquierdo-Alventosa R, Ingles M, Cortes-Amador S, Gimeno-Mallench L, Chirivella-Garrido J, Kropotov J et al. Low-Intensity Physical Exercise Improves Pain Catastrophizing and Other Psychological and Physical Aspects in Women with Fibromyalgia: A Randomized Controlled Trial. International Journal of Environmental Research & Public Health [Electronic Resource]. 2020; 17(10):21

133. Jensen IB, Bergstrom G, Ljungquist T, Bodin L, Nygren AL. A randomized controlled component analysis of a behavioral medicine rehabilitation program for chronic spinal pain: are the effects dependent on gender? Pain. 2001; 91(1-2):65-78

134. Jentoft ES, Kvalvik AG, Mengshoel AM. Effects of pool-based and land-based aerobic exercise on women with fibromyalgia/chronic widespread muscle pain. Arthritis and Rheumatism. 2001; 45(1):42-47

135. Jones KD. Nordic walking in fibromyalgia: a means of promoting fitness that is easy for busy clinicians to recommend. Arthritis Research & Therapy. 2011; 13(1):103

136. Jones KD, Sherman CA, Mist SD, Carson JW, Bennett RM, Li F. A randomized controlled trial of 8-form Tai chi improves symptoms and functional mobility in fibromyalgia patients. Clinical Rheumatology. 2012; 31(8):1205-1214

137. Jordan A, Bendix T, Nielsen H, Hansen FR, Host D, Winkel A. Intensive training, physiotherapy, or manipulation for patients with chronic neck pain. A prospective, single-blinded, randomized clinical trial. Spine. 1998; 23(3):311-319

138. Jull GA, Falla D, Vicenzino B, Hodges PW. The effect of therapeutic exercise on activation of the deep cervical flexor muscles in people with chronic neck pain. Manual Therapy. 2009; 14(6):696-701

139. Kalamir A, Bonello R, Graham P, Vitiello AL, Pollard H. Intraoral myofascial therapy for chronic myogenous temporomandibular disorder: A randomized controlled trial. Journal of Manipulative and Physiological Therapeutics. 2012; 35(1):26-37

140. Kaleth AS, Saha CK, Jensen MP, Slaven JE, Ang DC. Effect of moderate to vigorous physical activity on long-term clinical outcomes and pain severity in fibromyalgia. Arthritis Care and Research. 2013; 65(8):1211-1218

141. Kay JA, Carlson CR. The role of stretch-based relaxation in the treatment of chronic neck tension. Behavior Therapy. 1992; 23(3):423-431

142. Kayo AH, Peccin MS, Sanches CM, Trevisani VF. Effectiveness of physical activity in reducing pain in patients with fibromyalgia: a blinded randomized clinical trial. Rheumatology International. 2012; 32(8):2285-2292

143. Keel PJ, Bodoky C, Gerhard U, Muller W. Comparison of integrated group therapy and group relaxation training for fibromyalgia. Clinical Journal of Pain. 1998; 14(3):232-238

144. Kelley GA, Kelley KS, Hootman JM, Jones DL. Exercise and global well-being in community-dwelling adults with fibromyalgia: a systematic review with meta-analysis. BMC Public Health. 2010; 10:198

145. Khan AA, Srivastava A, Passi D, Devi M, Chandra L, Atri M. Management of myofascial pain dysfunction syndrome with meditation and yoga: healing through natural therapy. National journal of maxillofacial surgery. 2018; 9(2):155-159

146. Khan M, Soomro RR, Ali SS. The effectiveness of isometric exercises as compared to general exercises in the management of chronic non-specific neck pain. Pakistan Journal of Pharmaceutical Sciences. 2014; 27(5 Suppl):1719-1722

147. Kibar S, Yildiz HE, Ay S, Evcik D, Ergin ES. New approach in fibromyalgia exercise program: A preliminary study regarding the effectiveness of balance training. Archives of Physical Medicine and Rehabilitation. 2015; 96(9):1576-1582

148. Kim JY, Kwag KI. Clinical effects of deep cervical flexor muscle activation in patients with chronic neck pain. Journal of Physical Therapy Science. 2016; 28(1):269-273

149. Kim M, Lee M, Kim Y, Oh S, Lee D, Yoon B. Myofascial Pain Syndrome in the Elderly and Self-Exercise: A Single-Blind, Randomized, Controlled Trial. Journal of Alternative and Complementary Medicine. 2016; 22(3):244-251

150. Kim S, Busch A, Overend T, Schachter C, van dSI, Boden C et al. Flexibility exercise training for adults with fibromyalgia. Cochrane Database of Systematic Reviews 2019, Issue 9. Art. No.: CD013419. DOI: 10.1002/14651858.CD013419.

151. Kim SD. Effects of yoga on chronic neck pain: a systematic review of randomized controlled trials. Journal of Physical Therapy Science. 2016; 28(7):2171-2174

152. King SJ, Wessel J, Bhambhani Y, Sholter D, Maksymowych W. The effects of exercise and education, individually or combined, in women with fibromyalgia. Journal of Rheumatology. 2002; 29(12):2620-2627

153. Kingsley JD, Panton LB, Toole T, Sirithienthad P, Mathis R, McMillan V. The effects of a 12-week strength-training program on strength and functionality in women with

fibromyalgia. Archives of Physical Medicine and Rehabilitation. 2005; 86(9):1713-1721

154. Lagueux E, Beaulieu-Boire L, Courtemanche-Harel R, Bourgault P, Tousignant-Laflamme Y. Can TDCS enhance the therapeutic effect of graded motor imagery in patients presenting chronic crps type I? Preliminary results. Pain Research and Management. 2014; 19(3):e83

155. Langhorst J, Klose P, Dobos GJ, Bernardy K, Hauser W. Efficacy and safety of meditative movement therapies in fibromyalgia syndrome: a systematic review and meta-analysis of randomized controlled trials. Rheumatology International. 2013; 33(1):193-207

156. Langhorst J, Musial F, Klose P, Hauser W. Efficacy of hydrotherapy in fibromyalgia syndrome--a meta-analysis of randomized controlled clinical trials. Rheumatology. 2009; 48(9):1155-1159

157. Lansinger B, Carlsson JY, Kreuter M, Taft C. Health-related quality of life in persons with long-term neck pain after treatment with qigong and exercise therapy respectively. European journal of physiotherapy. 2013; 15(3):111-117

158. Latorre PA, Santos MA, Heredia-Jimenez JM, Delgado-Fernandez M, Soto VM, Manas A et al. Effect of a 24-week physical training programme (in water and on land) on pain, functional capacity, body composition and quality of life in women with fibromyalgia. Clinical and Experimental Rheumatology. 2013; 31(6 Suppl 79):S72-80

159. Latorre Roman PA, Santos ECMA, Garcia-Pinillos F. Effects of functional training on pain, leg strength, and balance in women with fibromyalgia. Modern Rheumatology. 2015; 25(6):943-947

160. Lauche R, Stumpe C, Fehr J, Cramer H, Cheng YW, Wayne PM et al. The effects of tai chi and neck exercises in the treatment of chronic nonspecific neck pain: A randomized controlled trial. Journal of Pain. 2016; 17(9):1013-1027

161. Lauche R, Wayne PM, Fehr J, Stumpe C, Dobos G, Cramer H. Does postural awareness contribute to exercise-induced improvements in neck pain intensity? A secondary analysis of a randomized controlled trial evaluating tai chi and neck exercises. Spine. 2017; 42(16):1195-1200

162. Law RY, Harvey LA, Nicholas MK, Tonkin L, De Sousa M, Finniss DG. Stretch exercises increase tolerance to stretch in patients with chronic musculoskeletal pain: a randomized controlled trial. Physical Therapy. 2009; 89(10):1016-1026

163. Lee KW, Kim WH. Effect of thoracic manipulation and deep craniocervical flexor training on pain, mobility, strength, and disability of the neck of patients with chronic nonspecific neck pain: a randomized clinical trial. Journal of Physical Therapy Science. 2016; 28(1):175-180

164. Letafatkar A, Rabiei P, Alamooti G, Bertozzi L, Farivar N, Afshari M. Effect of therapeutic exercise routine on pain, disability, posture, and health status in dentists with chronic neck pain: a randomized controlled trial. International Archives of Occupational and Environmental Health. 2020; 93(3):281-290

165. Lima TB, Dias JM, Mazuquin BF, da Silva CT, Nogueira RM, Marques AP et al. The effectiveness of aquatic physical therapy in the treatment of fibromyalgia: a systematic review with meta-analysis. Clinical Rehabilitation. 2013; 27(10):892-908

166. Liu W, Zahner L, Cornell M, Le T, Ratner J, Wang Y et al. Benefit of Qigong exercise in patients with fibromyalgia: a pilot study. International Journal of Neuroscience. 2012; 122(11):657-664

167. Lopez-de-Uralde-Villanueva I, Beltran-Alacreu H, Fernandez-Carnero J, La Touche R. Pain management using a multimodal physiotherapy program including a biobehavioral approach for chronic nonspecific neck pain: a randomized controlled trial. Physiotherapy Theory & Practice. 2020; 36(1):45-62

168. Lopez-Pousa S, Bassets Pages G, Monserrat-Vila S, de Gracia Blanco M, Hidalgo Colome J, Garre-Olmo J. Sense of Well-Being in Patients with Fibromyalgia: Aerobic Exercise Program in a Mature Forest-A Pilot Study. Evidence-Based Complementary & Alternative Medicine: eCAM. 2015; 2015:614783

169. Lopez-Rodriguez MDM, Castro-Sanchez AM, Fernandez-Martinez M, Mataran-Penarrocha GA, Rodriguez-Ferrer ME. Comparison between aquatic-biodanza and stretching for improving quality of life and pain in pations with fibromyalgia. Atencion Primaria. 2012; 44(11):641-650

170. López-Rodríguez MM, Fernández-Martínez M, Matarán-Peñarrocha GA, Rodríguez-Ferrer ME, Granados Gámez G, Aguilar Ferrándiz E. Effectiveness of aquatic biodance on sleep quality, anxiety and other symptoms in patients with fibromyalgia. Medicina Clínica. 2013; 141(11):471-478

171. Lorena SB, Lima Mdo C, Ranzolin A, Duarte AL. Effects of muscle stretching exercises in the treatment of fibromyalgia: a systematic review. Revista Brasileira de Reumatologia. 2015; 55(2):167-173

172. Lynch M, Sawynok J, Hiew C, Marcon D. A randomized controlled trial of qigong for fibromyalgia. Arthritis Research & Therapy. 2012; 14(4):R178

173. Mannerkorpi K, Ahlmen M, Ekdahl C. Six- and 24-month follow-up of pool exercise therapy and education for patients with fibromyalgia. Scandinavian Journal of Rheumatology. 2002; 31(5):306-310

174. Mannerkorpi K, Arndorw M. Efficacy and feasibility of a combination of body awareness therapy and qigong in patients with fibromyalgia: a pilot study. Journal of Rehabilitation Medicine. 2004; 36(6):279-281

175. Mannerkorpi K, Nordeman L, Cider A, Jonsson G. Does moderate-to-high intensity Nordic walking improve functional capacity and pain in fibromyalgia? A prospective randomized controlled trial. Arthritis Research & Therapy. 2010; 12(5):R189

176. Mannerkorpi K, Nordeman L, Ericsson A, Arndorw M, Group GAUS. Pool exercise for patients with fibromyalgia or chronic widespread pain: a randomized controlled trial and subgroup analyses. Journal of Rehabilitation Medicine. 2009; 41(9):751-760

177. Mannerkorpi K, Nyberg B, Ahlmen M, Ekdahl C. Pool exercise combined with an education program for patients with fibromyalgia syndrome. A prospective, randomized study. Journal of Rheumatology. 2000; 27(10):2473-2481

178. Martin-Martinez JP, Villafaina S, Collado-Mateo D, Perez-Gomez J, Gusi N. Effects of 24-week exergame intervention on physical function under single- and dual-task conditions in fibromyalgia: A randomized controlled trial. Scandinavian Journal of Medicine and Science in Sports. 2019; 29(10):1610-1617

179. Martin L, Nutting A, MacIntosh BR, Edworthy SM, Butterwick D, Cook J. An exercise program in the treatment of fibromyalgia. Journal of Rheumatology. 1996; 23(6):1050-1053

180. Matsutani LA, Marques AP, Ferreira EA, Assumpcao A, Lage LV, Casarotto RA et al. Effectiveness of muscle stretching exercises with and without laser therapy at tender points for patients with fibromyalgia. Clinical and Experimental Rheumatology. 2007; 25(3):410-415

181. McBeth J, Prescott G, Scotland G, Lovell K, Keeley P, Hannaford P et al. Cognitive behavior therapy, exercise, or both for treating chronic widespread pain. Archives of Internal Medicine. 2012; 172(1):48-57

182. McCain GA. Role of physical fitness training in the fibrositis/fibromyalgia syndrome. American Journal of Medicine. 1986; 81(3A):73-77

183. McCain GA, Bell DA, Mai FM, Halliday PD. A controlled study of the effects of a supervised cardiovascular fitness training program on the manifestations of primary fibromyalgia. Arthritis and Rheumatism. 1988; 31(9):1135-1141

184. McDowell CP, Cook DB, Herring MP. The effects of exercise training on anxiety in fibromyalgia patients: A meta-analysis. Medicine and Science in Sports and Exercise. 2017; 49(9):1868-1876

185. McVeigh JG, McGaughey H, Hall M, Kane P. The effectiveness of hydrotherapy in the management of fibromyalgia syndrome: a systematic review. Rheumatology International. 2008; 29(2):119-130

186. Meiworm L, Jakob E, Walker UA, Peter HH, Keul J. Patients with fibromyalgia benefit from aerobic endurance exercise. Clinical Rheumatology. 2000; 19(4):253-257

187. Meiworm L, Strass D, Jakob E, Walker UA, Peter HH, Keul J. The effects of an aerobic exercise training on symptomatic pain of patients with Fibromyalgia. Deutsche zeitschrift fur sportmedizin. 1999; 50(6):188-192

188. Mendez-Rebolledo G, Gatica-Rojas V, Torres-Cueco R, Albornoz-Verdugo M, Guzman-Munoz E. Update on the effects of graded motor imagery and mirror therapy on complex regional pain syndrome type 1: A systematic review. Journal of Back and Musculoskeletal Rehabilitation. 2017; 30(3):441-449

189. Mengshoel AM, Komnaes HB, Forre O. The effects of 20 weeks of physical fitness training in female patients with fibromyalgia. Clinical and Experimental Rheumatology. 1992; 10(4):345-349

190. Mesquita C, Lopes S, Silva D, Neves I, Araujo F, Vitorino A. A7.3 Evaluation the effect of a home-exercise program based on disability caused by back pain in individuals with fibromyalgia. Annals of the Rheumatic Diseases. 2014; 73(Suppl. 1):A75

191. Meyer BB, Lemley KJ. Utilizing exercise to affect the symptomology of fibromyalgia: a pilot study. Medicine and Science in Sports and Exercise. 2000; 32(10):1691-1697

192. Michalsen A, Traitteur H, Ludtke R, Brunnhuber S, Meier L, Jeitler M et al. Yoga for chronic neck pain: a pilot randomized controlled clinical trial. Journal of Pain. 2012; 13(11):1122-1130

193. Miles ALS. The effects of gentle yoga vs. cognitive behavioral therapy on physical and psychological symptoms; neurocognitive functioning; and physiology in women with fibromyalgia [Thesis]. San Diego. Alliant International University, California School of Professional Psychology. 2014

194. Molinari G, Garcia-Palacios A, Enrique A, Roca P, Fernandez-Llanio Comella N, Botella C. The power of visualization: Back to the future for pain management in fibromyalgia syndrome. Pain Medicine. 2018; 19(7):1451-1468

195. Moseley GL. Graded motor imagery for pathologic pain: a randomized controlled trial. Neurology. 2006; 67(12):2129-2134

196. Moseley GL. Graded motor imagery is effective for long-standing complex regional pain syndrome: a randomised controlled trial. Pain. 2004; 108(1-2):192-198

197. Moseley GL. Is successful rehabilitation of complex regional pain syndrome due to sustained attention to the affected limb? A randomised clinical trial. Pain. 2005; 114(1-2):54-61

198. Moustafa IM, Diab AA. The addition of upper cervical manipulative therapy in the treatment of patients with fibromyalgia: a randomized controlled trial. Rheumatology International. 2015; 35(7):1163-1174

199. Munguia-Izquierdo D, Legaz-Arrese A. Assessment of the effects of aquatic therapy on global symptomatology in patients with fibromyalgia syndrome: a randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2008; 89(12):2250-2257

200. Munguia-Izquierdo D, Legaz-Arrese A. Exercise in warm water decreases pain and improves cognitive function in middle-aged women with fibromyalgia. Clinical and Experimental Rheumatology. 2007; 25(6):823-830

201. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

202. NCT. Efficiency of Modified Pilates Exercises in Patients With Chronic Neck Pain. 2018. Available from: Https://clinicaltrials.gov/show/nct03782584 Last accessed: 10/12/2019.

203. NCT. A Functional Exercise Program Improves Pain and Health Related Quality of Life in Patients With Fibromyalgia. 2018. Available from: Https://clinicaltrials.gov/show/nct03682588 Last accessed: 10/12/2019.

204. Nichols DS, Glenn TM. Effects of aerobic exercise on pain perception, affect, and level of disability in individuals with fibromyalgia. Physical Therapy. 1994; 74(4):327-332

205. Nickel MK, Nickel C, Lahmann C, Mitterlehner FO, Tritt K, Leiberich PK et al. Changes in instrumental activities of daily living disability after treatment of depressive symptoms in elderly women with chronic musculoskeletal pain: a double-blind, placebo-controlled trial. Aging-Clinical & Experimental Research. 2005; 17(4):293-296

206. Norouzi E, Hosseini F, Vaezmosavi M, Gerber M, Puhse U, Brand S. Zumba dancing and aerobic exercise can improve working memory, motor function, and depressive symptoms in female patients with Fibromyalgia. European Journal of Sport Science EJSS : Official Journal of the European College of Sport Science. 2019:1-11

207. Norregaard J, Lykkegaard JJ, Mehlsen J, Danneskiold-Samsoe B. Exercise training in treatment of fibromyalgia. Journal of Musculoskeletal Pain. 1997; 5(1):71-79

208. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: http://www.oecd.org/std/ppp Last accessed: 13/12/2019.

209. Ote Karaca S, Demirsoy N, Gunendi Z. Effects of aerobic exercise on pain sensitivity, heart rate recovery, and health-related quality of life in patients with chronic musculoskeletal pain. International Journal of Rehabilitation Research. 2017; 40(2):164-170

210. Panton LB, Figueroa A, Kingsley JD, Hornbuckle L, Wilson J, St John N et al. Effects of resistance training and chiropractic treatment in women with fibromyalgia. Journal of Alternative and Complementary Medicine. 2009; 15(3):321-328

211. Perez-De la Cruz S, Lambeck J. Effects of a programme of aquatic Ai Chi exercise in patients with fibromyalgia. A pilot study. Revista de Neurología. 2015; 60(2):59-65

212. Peters S, Stanley I, Rose M, Kaney S, Salmon P. A randomized controlled trial of group aerobic exercise in primary care patients with persistent, unexplained physical symptoms. Family Practice. 2002; 19(6):665-674

213. Petersen SB, Cook C, Donaldson M, Hassen A, Ellis A, Learman K. The effect of manual therapy with augmentative exercises for neck pain: a randomised clinical trial. Journal of Manual & Manipulative Therapy. 2015; 23(5):264-275

214. Phattharasupharerk S, Purepong N, Eksakulkla S, Siriphorn A. Effects of Qigong practice in office workers with chronic non-specific low back pain: A randomized control trial. Journal of Bodywork and Movement Therapies. 2019; 23(2):375-381

215. Pico-Espinosa OJ, Aboagye E, Cote P, Peterson A, Holm LW, Jensen I et al. Deep tissue massage, strengthening and stretching exercises, and a combination of both compared with advice to stay active for subacute or persistent non-specific neck pain: A cost-effectiveness analysis of the Stockholm Neck trial (STONE). Musculoskeletal Science & Practice. 2020; 46:102109

216. Pike M, Sawynok J, Lynch M, Clark AJ, Marcon D. Observational trial of qigong as a complementary practice in a chronic pain program. Pain Research and Management. 2015; 20(3):e56

217. Plumbe L, Peters S, Bennett S, Vicenzino B, Coppieters MW. Mirror therapy, graded motor imagery and virtual illusion for the management of chronic pain. Cochrane Database of Systematic Reviews 2016, Issue 4. Art. No.: CD010329. DOI: 10.1002/14651858.CD010329.pub2.

218. Rajalaxmi V, Jasim A, Sudhakar S, Mohan Kumar G. To analyse the effectiveness of yoga, pilates and tai chi exercise for chronic mechanical neck pain -a randomized controlled trial. Biomedicine (India). 2018; 38(1):147-151

219. Ramel J, Bannuru R, Griffith M, Wang C. Exercise for fibromyalgia pain: A meta-analysis of randomized controlled trials. Current Rheumatology Reviews. 2009; 5(4):188-193

220. Ramsay C, Moreland J, Ho M, Joyce S, Walker S, Pullar T. An observer-blinded comparison of supervised and unsupervised aerobic exercise regimens in fibromyalgia. Rheumatology. 2000; 39(5):501-505

221. Redondo JR, Justo CM, Moraleda FV, Velayos YG, Puche JJ, Zubero JR et al. Long-term efficacy of therapy in patients with fibromyalgia: a physical exercise-based program and a cognitive-behavioral approach. Arthritis and Rheumatism. 2004; 51(2):184-192

222. Rendant D, Pach D, Ludtke R, Reisshauer A, Mietzner A, Willich SN et al. Qigong versus exercise versus no therapy for patients with chronic neck pain: a randomized controlled trial. Spine. 2011; 36(6):419-427

223. Reynolds B, Puentedura EJ, Kolber MJ, Cleland JA. Effectiveness of Cervical Spine High Velocity Low Amplitude Thrust Added to Behavioral Education, Soft Tissue Mobilization, and Exercise in Individuals With Temporomandibular Disorder (TMD) With Myalgia: A Randomized Clinical Trial. Journal of Orthopaedic and Sports Physical Therapy. 2020:1-40

224. Richards SC, Scott DL. Prescribed exercise in people with fibromyalgia: parallel group randomised controlled trial. BMJ. 2002; 325(7357):185

225. Ris I, Sogaard K, Gram B, Agerbo K, Boyle E, Juul-Kristensen B. Does a combination of physical training, specific exercises and pain education improve health-related quality of life in patients with chronic neck pain? A randomised control trial with a 4-month follow up. Manual Therapy. 2016; 26(December):132-140

226. Rivas Neira S, Pasqual Marques A, Pegito Perez I, Fernandez Cervantes R, Vivas Costa J. Effectiveness of aquatic therapy vs land-based therapy for balance and pain in women with fibromyalgia: A study protocol for a randomised controlled trial. BMC Musculoskeletal Disorders. 2017; 18:22

227. Rolving N, Christiansen DH, Andersen LL, Skotte J, Ylinen J, Jensen OK et al. Effect of strength training in addition to general exercise in the rehabilitation of patients with non-specific neck pain. A randomized clinical trial. European journal of physical & rehabilitation medicine. 2014; 50(6):617-626

228. Ryan JM. Reducing pain and disability for patients with chronic neck pain : results of a double-blind randomised controlled trial comparing strength to endurance training. Canberra. Australian National University. 2002

229. Saadat M, Salehi R, Negahban H, Shaterzadeh MJ, Mehravar M, Hessam M. Traditional physical therapy exercises combined with sensorimotor training: The effects on clinical outcomes for chronic neck pain in a double-blind, randomized controlled trial. Journal of Bodywork and Movement Therapies. 2019; 23(4):901-907

230. Salo P, Ylonen-Kayra N, Hakkinen A, Kautiainen H, Malkia E, Ylinen J. Effects of long-term home-based exercise on health-related quality of life in patients with chronic neck pain: a randomized study with a 1-year follow-up. Disability and Rehabilitation. 2012; 34(23):1971-1977

231. Salo PK, Hakkinen AH, Kautiainen H, Ylinen JJ. Effect of neck strength training on health-related quality of life in females with chronic neck pain: a randomized controlled 1-year follow-up study. Health & Quality of Life Outcomes. 2010; 8:48

232. Sanudo B, Carrasco L, de Hoyo M, Figueroa A, Saxton JM. Vagal modulation and symptomatology following a 6-month aerobic exercise program for women with fibromyalgia. Clinical and Experimental Rheumatology. 2015; 33(1 Suppl 88):S41-45

233. Sanudo B, Carrasco L, de Hoyo M, McVeigh JG. Effects of exercise training and detraining in patients with fibromyalgia syndrome: a 3-yr longitudinal study. American Journal of Physical Medicine and Rehabilitation. 2012; 91(7):561-573

234. Sanudo B, Galiano D, Carrasco L, Blagojevic M, de Hoyo M, Saxton J. Aerobic exercise versus combined exercise therapy in women with fibromyalgia syndrome: a randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2010; 91(12):1838-1843

235. Sanudo B, Galiano D, Carrasco L, de Hoyo M, McVeigh JG. Effects of a prolonged exercise program on key health outcomes in women with fibromyalgia: a randomized controlled trial. Journal of Rehabilitation Medicine. 2011; 43(6):521-526

236. Sarmento CVM, Moon S, Pfeifer T, Smirnova IV, Colgrove Y, Lai SM et al. The therapeutic efficacy of Qigong exercise on the main symptoms of fibromyalgia: A pilot randomized clinical trial. Integrative Medicine Research. 2020; 9(4):100416

237. Sawynok J, Lynch M, Marcon D. Extension trial of qigong for fibromyalgia: a quantitative and qualitative study. Evidence-Based Complementary and Alternative Medicine. 2013; 2013:726062

238. Saxena R, Gupta M, Shankar N, Jain S, Saxena A. Effects of yogic intervention on pain scores and quality of life in females with chronic pelvic pain. International Journal of Yoga. 2017; 10(1):9-15

239. Schachter CL, Busch AJ, Peloso PM, Sheppard MS. Effects of short versus long bouts of aerobic exercise in sedentary women with fibromyalgia: a randomized controlled trial. Physical Therapy. 2003; 83(4):340-358

240. Segura-Jimenez V, Carbonell-Baeza A, Aparicio VA, Samos B, Femia P, Ruiz JR et al. A warm water pool-based exercise program decreases immediate pain in female fibromyalgia patients: uncontrolled clinical trial. International Journal of Sports Medicine. 2013; 34(7):600-605

241. Sencan S, Ak S, Karan A, Muslumanoglu L, Ozcan E, Berker E. A study to compare the therapeutic efficacy of aerobic exercise and paroxetine in fibromyalgia syndrome. Journal of Back and Musculoskeletal Rehabilitation. 2004; 17(2):57-61

242. Sevimli D, Kozanoglu E, Guzel R, Doganay A. The effects of aquatic, isometric strength-stretching and aerobic exercise on physical and psychological parameters of female patients with fibromyalgia syndrome. Journal of Physical Therapy Science. 2015; 27(6):1781-1786

243. Silva HJA, Assuncao Junior JC, de Oliveira FS, Oliveira JMP, Figueiredo Dantas GA, Lins CAA et al. Sophrology versus resistance training for treatment of women with fibromyalgia: A randomized controlled trial. Journal of Bodywork and Movement Therapies. 2019; 23(2):382-389

244. Skillgate E, Bill AS, Cote P, Viklund P, Peterson A, Holm LW. The effect of massage therapy and/or exercise therapy on subacute or long-lasting neck pain--the Stockholm Neck trial (STONE): study protocol for a randomized controlled trial. Trials. 2015; 16:414

245. Skillgate E, Pico-Espinosa OJ, Cote P, Jensen I, Viklund P, Bottai M et al. Effectiveness of deep tissue massage therapy, and supervised strengthening and stretching exercises for subacute or persistent disabling neck pain. The Stockholm Neck (STONE) randomized controlled trial. Musculoskeletal Science & Practice. 2020; 45:102070

246. Song CH, Jang HJLKJ, Lee YW. Active exercise program in patient with nonspecific chronic neck pain: a randomised controlled trial. Pain Practice. 2012; 12(Suppl 1):1-199

247. Suvarnnato T, Puntumetakul R, Uthaikhup S, Boucaut R. Effect of specific deep cervical muscle exercises on functional disability, pain intensity, craniovertebral angle, and neck-muscle strength in chronic mechanical neck pain: a randomized controlled trial. Journal of Pain Research. 2019; 12:915-925

248. Taggart HM, Arslanian CL, Bae S, Singh K. Effects of T'ai Chi exercise on fibromyalgia symptoms and health-related quality of life. Orthopaedic Nursing. 2003; 22(5):353-360

249. Taimela S, Takala EP, Asklöf T, Seppälä K, Parviainen S. Active treatment of chronic neck pain: a prospective randomized intervention. Spine. 2000; 25(8):1021-1027

250. Theadom A, Cropley M, Smith HE, Feigin VL, McPherson K. Mind and body therapy for fibromyalgia. Cochrane Database of Systematic Reviews 2015, Issue 4. Art. No.: CD001980. DOI: 10.1002/14651858.CD001980.pub3.

251. Thompson DP, Oldham JA, Woby SR. Does adding cognitive-behavioural physiotherapy to exercise improve outcome in patients with chronic neck pain? A randomised controlled trial. Physiotherapy. 2016; 102(2):170-177

252. Tomas-Carus P, Gusi N, Hakkinen A, Hakkinen K, Leal A, Ortega-Alonso A. Eight months of physical training in warm water improves physical and mental health in women with fibromyalgia: a randomized controlled trial. Journal of Rehabilitation Medicine. 2008; 40(4):248-252

253. Tomas-Carus P, Gusi N, Hakkinen A, Hakkinen K, Raimundo A, Ortega-Alonso A. Improvements of muscle strength predicted benefits in HRQOL and postural balance in women with fibromyalgia: an 8-month randomized controlled trial. Rheumatology. 2009; 48(9):1147-1151

254. Tomas-Carus P, Hakkinen A, Gusi N, Leal A, Hakkinen K, Ortega-Alonso A. Aquatic training and detraining on fitness and quality of life in fibromyalgia. Medicine and Science in Sports and Exercise. 2007; 39(7):1044-1050

255. Tomas-Carus P, Raimundo A, Timon R, Gusi N. Exercise in warm water decreases pain but no the number of tender points in women with fibromyalgia: a randomized controlled trial. Seleccion. 2007; 16(2):98-102

256. Toprak Celenay S, Anaforoglu Kulunkoglu B, Yasa ME, Sahbaz Pirincci C, Un Yildirim N, Kucuksahin O et al. A comparison of the effects of exercises plus connective tissue massage to exercises alone in women with fibromyalgia syndrome: a randomized controlled trial. Rheumatology International. 2017; 37(11):1799-1806

257. Ulug N, Yilmaz OT, Kara M, Ozcakar L. Effects of Pilates and yoga in patients with chronic neck pain: A sonographic study. Journal of Rehabilitation Medicine. 2018; 50(1):80-85

258. Valencia M, Alonso B, Alvarez MJ, Barrientos MJ, Ayan C, Martin Sanchez V. Effects of 2 physiotherapy programs on pain perception, muscular flexibility, and illness impact in women with fibromyalgia: a pilot study. Journal of Manipulative and Physiological Therapeutics. 2009; 32(1):84-92

259. Valim V, Oliveira L, Suda A, Silva L, de Assis M, Barros Neto T et al. Aerobic fitness effects in fibromyalgia. Journal of Rheumatology. 2003; 30(5):1060-1069

260. Valkeinen H, Alen M, Hakkinen A, Hannonen P, Kukkonen-Harjula K, Hakkinen K. Effects of concurrent strength and endurance training on physical fitness and symptoms in postmenopausal women with fibromyalgia: a randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2008; 89(9):1660-1666

261. Valkeinen H, Hakkinen K, Pakarinen A, Hannonen P, Hakkinen A, Airaksinen O et al. Muscle hypertrophy, strength development, and serum hormones during strength training in elderly women with fibromyalgia. Scandinavian Journal of Rheumatology. 2005; 34(4):309-314

262. van Dessel N, den Boeft M, van der Wouden JC, Kleinstäuber M, Leone SS, Terluin B et al. Non-pharmacological interventions for somatoform disorders and medically unexplained physical symptoms (MUPS) in adults. Cochrane Database of Systematic Reviews 2014, Issue 11. Art. No.: CD011142. DOI: 10.1002/14651858.CD011142.pub2.

263. van Eijk-Hustings Y, Kroese M, Creemers A, Landewe R, Boonen A. Resource utilisation and direct costs in patients with recently diagnosed fibromyalgia who are offered one of three different interventions in a randomised pragmatic trial. Clinical Rheumatology. 2016; 35(5):1307-1315

264. van Eijk-Hustings Y, Kroese M, Tan F, Boonen A, Bessems-Beks M, Landewe R. Challenges in demonstrating the effectiveness of multidisciplinary treatment on quality of life, participation and health care utilisation in patients with fibromyalgia: a randomised controlled trial. Clinical Rheumatology. 2013; 32(2):199-209

265. van Koulil S, van Lankveld W, Kraaimaat FW, van Helmond T, Vedder A, van Hoorn H et al. Tailored cognitive-behavioural therapy and exercise training improves the physical fitness of patients with fibromyalgia. Annals of the Rheumatic Diseases. 2011; 70(12):2131-2133

266. Verstappen FTJ, Van Santen-Hoeuftt HMS, Bolwijn PH, Van Der Linden S, Kuipers H. Effects of a group activity program for fibromyalgia patients on physical fitness and well being. Journal of Musculoskeletal Pain. 1997; 5(4):17-28

267. Viljanen M, Malmivaara A, Uitti J, Rinne M, Palmroos P, Laippala P. Effectiveness of dynamic muscle training, relaxation training, or ordinary activity for chronic neck pain: randomised controlled trial. BMJ. 2003; 327(7413):475

268. Villafaina S, Borrega-Mouquinho Y, Fuentes-Garcia JP, Collado-Mateo D, Gusi N. Effect of Exergame Training and Detraining on Lower-Body Strength, Agility, and Cardiorespiratory Fitness in Women with Fibromyalgia: Single-Blinded Randomized Controlled Trial. International Journal of Environmental Research & Public Health [Electronic Resource]. 2019; 17(1):24

269. Villafaina S, Collado-Mateo D, Dominguez-Munoz FJ, Fuentes-Garcia JP, Gusi N. Benefits of 24-Week Exergame Intervention on Health-Related Quality of Life and Pain in Women with Fibromyalgia: A Single-Blind, Randomized Controlled Trial. Games for health journal. 2019; 28

270. Vitorino DF, Carvalho LB, Prado GF. Hydrotherapy and conventional physiotherapy improve total sleep time and quality of life of fibromyalgia patients: randomized clinical trial. Sleep Medicine. 2006; 7(3):293-296

271. von Trott P, Wiedemann AM, Ludtke R, Reishauer A, Willich SN, Witt CM. Qigong and exercise therapy for elderly patients with chronic neck pain (QIBANE): a randomized controlled study. Journal of Pain. 2009; 10(5):501-508

272. Vonk F, Verhagen AP, Twisk JW, Koke AJ, Luiten MW, Koes BW. Effectiveness of a behaviour graded activity program versus conventional exercise for chronic neck pain patients. European Journal of Pain. 2009; 13(5):533-541

273. Waling K, Jarvholm B, Sundelin G. Effects of training on female trapezius Myalgia: An intervention study with a 3-year follow-up period. Spine. 2002; 27(8):789-796

274. Wang C, Schmid CH, Fielding RA, Harvey WF, Reid KF, Price LL et al. Effect of tai chi versus aerobic exercise for fibromyalgia: comparative effectiveness randomized controlled trial. BMJ. 2018; 360:k851

275. Wang C, Schmid CH, Lee Y, McAlindon T. Does obesity in patients with fibromyalgia modify response to Tai Chi therapy: Analysis of a randomized controlled trial. Arthritis and Rheumatism. 2010; Supplement:S38

276. Wigers SH, Stiles TC, Vogel PA. Effects of aerobic exercise versus stress management treatment in fibromyalgia. A 4.5 year prospective study. Scandinavian Journal of Rheumatology. 1996; 25(2):77-86

277. Wiklund T, Linton SJ, Alfoldi P, Gerdle B. Is sleep disturbance in patients with chronic pain affected by physical exercise or ACT-based stress management? - A randomized controlled study. BMC Musculoskeletal Disorders. 2018; 19:111

278. Wong A, Figueroa A, Sanchez-Gonzalez MA, Son WM, Chernykh O, Park SY. Effectiveness of Tai Chi on Cardiac Autonomic Function and Symptomatology in Women With Fibromyalgia: A Randomized Controlled Trial. Journal of Aging & Physical Activity. 2018; 26(2):214-221

279. Wu WH, Bandilla E, Ciccone DS, Yang J, Cheng SC, Carner N et al. Effects of qigong on late-stage complex regional pain syndrome. Alternative Therapies in Health and Medicine. 1999; 5(1):45-54

280. Yang KH, Kim YH, Lee MS. Efficacy of Qi-therapy (external Qigong) for elderly people with chronic pain. International Journal of Neuroscience. 2005; 115(7):949-963

281. Ylinen J, Hakkinen A, Nykanen M, Kautiainen H, Takala EP. Neck muscle training in the treatment of chronic neck pain: a three-year follow-up study. Europa Medicophysica. 2007; 43(2):161-169

282. Ylinen J, Takala EP, Kautiainen H, Nykanen M, Hakkinen A, Pohjolainen T et al. Effect of long-term neck muscle training on pressure pain threshold: a randomized controlled trial. European Journal of Pain. 2005; 9(6):673-681

283. Ylinen J, Takala EP, Nykänen M, Häkkinen A, Kautiainen H, Mälkiä E et al. Exercise of neck and shoulder muscles as a relief for the chronic neck pain. Duodecim. 2004; 120(16):1958-1967

284. Ylinen J, Takala EP, Nykanen M, Hakkinen A, Malkia E, Pohjolainen T et al. Active neck muscle training in the treatment of chronic neck pain in women: a randomized controlled trial. JAMA. 2003; 289(19):2509-2516

285. Ylinen JJ, Hakkinen AH, Takala EP, Nykanen MJ, Kautiainen HJ, Malkia EA et al. Effects of neck muscle training in women with chronic neck pain: one-year follow-up study. Journal of Strength and Conditioning Research. 2006; 20(1):6-13

286. Ylinen JJ, Takala EP, Nykanen MJ, Kautiainen HJ, Hakkinen AH, Airaksinen OV. Effects of twelve-month strength training subsequent to twelve-month stretching exercise in treatment of chronic neck pain. Journal of Strength and Conditioning Research. 2006; 20(2):304-308

287. Zamuner AR, Andrade CP, Forti M, Marchi A, Milan J, Avila MA et al. Effects of a hydrotherapy programme on symbolic and complexity dynamics of heart rate variability and aerobic capacity in fibromyalgia patients. Clinical and Experimental Rheumatology. 2015; 33(1 Suppl 88):S73-S81

288. Zijlstra TR, van de Laar MA, Bernelot Moens HJ, Taal E, Zakraoui L, Rasker JJ. Spa treatment for primary fibromyalgia syndrome: a combination of thalassotherapy, exercise and patient education improves symptoms and quality of life. Rheumatology. 2005; 44(4):539-546

289. Zonneveld LN, Rood YR, Timman R, Kooiman CG, Van't Spijker A, Busschbach JJ. Effective group training for patients with unexplained physical symptoms: a randomized controlled trial with a non-randomized one-year follow-up. PloS One. 2012; 7(8):e42629

Appendices

Appendix A: Review protocols

Review protocol for exercise

ID	Field	Content
0.	PROSPERO registration number	Not registered.
1.	Review title	What is the clinical and cost effectiveness of exercise interventions for the management of chronic primary pain?
2.	Review question	What is the clinical and cost effectiveness of exercise interventions for the management of chronic primary pain?
3.	Objective	To determine the clinical and cost effectiveness of exercise interventions for the management of chronic primary pain.
4.	Searches	The following databases will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE • CINAHL, Current Nursing and Allied Health Literature 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 committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review.
5.	Condition or domain being studied	Chronic pain in one or more anatomical regions that is characterized by significant emotional distress (anxiety, anger/frustration or depressed mood) and functional disability (interference in daily life activities and reduced participation in social roles). The diagnosis is appropriate independently of identified biological or psychological contributors unless another diagnosis would better account for the presenting symptoms.
6.	Population	Inclusion: People, aged 16 years and over, with chronic primary pain (whose pain management is not addressed by existing NICE guidance) (chronic widespread pain, complex regional pain syndrome, chronic visceral pain, chronic orofacial pain, chronic musculoskeletal pain other than orofacial) Exclusion: Those whose pain management is addressed by existing NICE guidance
7.	Intervention/Exposure/Test	Interventions: • mind-body exercises (e.g. yoga, Tai Chi) • biomechanical (e.g. pilates) • proprioceptive • strength and conditioning • flexibility • aerobics (e.g. swimming, walking programme, aerobic exercise) • graded motor imagery • mixed modality exercise (aerobics and/or mind-body and/or biomechanical).
8.	Comparator/Reference standard/Confounding factors	Comparators: • each other • usual care

		• psychological therapies • other physical therapies (e.g. manual therapy) • manual therapy + exercise.
9.	Types of study to be included	Randomised controlled trials (RCTs) and systematic reviews of RCTs Cross-over RCTs will be considered if no non-cross-over RCT evidence is identified.
10.	Other exclusion criteria	Non-English language studies.
11.	Context	A clear understanding of the evidence for the effectiveness of chronic primary pain treatments: • improves the confidence of healthcare professionals in their conversations about pain, and • helps healthcare professionals and patients to have realistic expectations about outcomes of treatment.
12.	Primary outcomes (critical outcomes)	• Pain reduction (any validated scale) • health related quality of life (including meaningful activity) • physical function (e.g. 5 minute walk, sit to stand, Roland Morris Disability Questionnaire, Oswestry Disability Index, Canadian Occupational Performance Measure) • psychological distress (depression/anxiety) (preferably Hospital Anxiety and Depression Scale).
13.	Secondary outcomes (important outcomes)	• Use of healthcare services • sleep • discontinuation.
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 Cochrane Risk of Bias (2.0) tool. 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.	
17.	Analysis of sub-groups	Proposed sensitivity / subgroup analysis to be explored where there is heterogeneity: • chronic widespread pain • complex regional pain syndrome • chronic visceral pain • chronic orofacial pain • chronic primary musculoskeletal pain • cognitive impairment • learning difficulties • first language not English • sensory impairment • homelessness.	
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	NA – not registered on PROSPERO	
22.	Anticipated completion date	19/08/2020	
23.	Named contact	5a. Named contact National Guideline Centre 5b Named contact e-mail Chronicpain@nice.org.uk 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE) and the National Guideline Centre	
24.	Review team members	From the National Guideline Centre: Serena Carville, Guideline Lead Maria Smyth, Senior Systematic Reviewer Rebecca Boffa, Senior Systematic Reviewer Margaret Constanti, Senior Health Economist Joseph Runicles, Information Specialist Katie Broomfield, Project Manager	
25.	Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.	

26.	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.
27.	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-ng10069
28.	Other registration details	NA
29.	Reference/URL for published protocol	NA
30.	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.
31.	Keywords	-
32.	Details of existing review of same topic by same authors	NA
33.	Additional information	-
34.	Details of final publication	www.nice.org.uk

Table 69: 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 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 2002. Abstract-only studies and studies from non-OECD countries or the USA will also be excluded. 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 2002 or later but that depend on unit costs and resource data entirely or predominantly from before 2002 will be rated as 'Not applicable'. • Studies published before 2002 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

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 Methods Report published as part of the accompanying documents for this guideline.

B.1 Clinical search literature search strategy

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies for interventions as these concepts may not be well described in title, abstract or indexes and therefore difficult to retrieve. Search filters were applied to the search where appropriate.

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 20 May 2020	Exclusions Randomised controlled trials Systematic review studies
Embase (OVID)	1974 – 20 May 2020	Exclusions Randomised controlled trials Systematic review studies
The Cochrane Library (Wiley)	Cochrane Reviews to 2020 Issue 5 of 12 CENTRAL to 2020 Issue 5 of 12	None

Medline (Ovid) search terms

1.	Chronic pain/
2.	((chronic or persist* or idiopathic or atypical or a-typical) adj4 pain).ti,ab.
3.	exp Complex Regional Pain Syndromes/
4.	(complex regional pain syndrome* or CRPS or causalgia).ti,ab.
5.	((reflex or sympathetic) adj2 dystroph*).ti,ab.
6.	fibromyalgia/
7.	(fibromyalgia* or fibrositis or myofascial pain syndrome).ti,ab.
8.	vulvodynia/
9.	(vulvodynia or vestibulodynia or dyspareunia or vulvar vestibulitis or vulvitis).ti,ab.
10.	interstitial cystitis/
11.	(interstitial adj2 cystitis).ti,ab.
12.	algodystrophy/
13.	(algodystroph* or sudek or sudeck*).ti,ab.
14.	exp myofascial pain syndromes/
15.	cystitis, interstitial/
16.	(loin pain adj (haematuria or hematuria) adj syndrome*).ti,ab.
17.	(LPHS or prostatodynia or CPPS or atypic* odontalgia or a-typic* odontalgia or burning mouth syndrome* or phantom tooth pain or neuropathic orofacial pain or "myofascial pain" or MPS).ti,ab.
18.	((pelvic or pelvis) adj pain syndrome*).ti,ab.
19.	((non-cardiac or noncardiac) adj3 chest adj3 pain).ti,ab.
20.	(temporomandibular adj3 joint adj3 pain).ti,ab.
21.	((prostate or vulv* or bladder or perineal) adj3 pain).ti,ab.
22.	(functional pain syndrome* or non-cancer pain or noncancer pain).ti,ab.
23.	((pelvic or pelvis or abdominal) adj3 pain adj3 (unknown or un-known or idiopathic or atypic* or a-typic*)).ti,ab.
24.	or/1-23
25.	letter/
26.	editorial/
27.	news/
28.	exp historical article/
29.	Anecdotes as Topic/
30.	comment/
31.	case report/
32.	(letter or comment*).ti.
33.	or/25-32
34.	randomized controlled trial/ or random*.ti,ab.
35.	33 not 34
36.	animals/ not humans/
37.	exp Animals, Laboratory/
38.	exp Animal Experimentation/
39.	exp Models, Animal/
40.	exp Rodentia/
41.	(rat or rats or mouse or mice).ti.
42.	or/35-41
43.	24 not 42
44.	limit 43 to English language

45.	exp exercise/
46.	exp exercise therapy/
47.	exp Exercise Movement Techniques/
48.	exp "physical education and training"/
49.	(pilates or yoga or feldenkrais or swim* or walk* or run* or jog* or treadmill* or tread mill*).ti,ab.
50.	(stretch* adj3 (active* or passive* or relax* or static* or dynamic* or gentl* or ballistic* or force* or isometric or technique* or exercis* or therap*)).ti,ab.
51.	(aerobic* adj (exercise* or train* or therap*)).ti,ab.
52.	((corrective* or biomechanic* or propiocet* or balance or flexib*) adj2 (exercise* or train* or therap*)).ti,ab.
53.	((biomechanic* or mckenzie) adj (method* or course*)).ti,ab.
54.	((strength* or stabil* or program* or train* or therap* or technique* or treat*) adj3 exercise*).ti,ab.
55.	(physical adj (fitness or conditioning or education or training or mobility or activit$ or exertion or effort)).ti,ab.
56.	danc*.ti,ab.
57.	(fitness* adj3 (program* or train* or therap*)).ti,ab.
58.	(tai ji or tai chi or taichi or taiji or taijiquan).ti,ab.
59.	(qigong or ch'i k#ng or ch'i g#ng or chi k#ng or chi g#ng or qi k#ng or qi g#ng).ti,ab.
60.	core stability.ti,ab.
61.	exp hydrotherapy/
62.	((water* or bath* or pool or pools or shower* or underwater* or spa or spas or aqua*) adj2 (exercise* or train* or therap* or treat*)).ti,ab.
63.	(hydrotherap* or hydro-therap*).ti,ab.
64.	(graded motor imagery or GMI or mirror therapy).ti,ab.
65.	or/45-64
66.	44 and 65
67.	randomized controlled trial.pt.
68.	controlled clinical trial.pt.
69.	randomi#ed.ti,ab.
70.	placebo.ab.
71.	randomly.ti,ab.
72.	Clinical Trials as topic.sh.
73.	trial.ti.
74.	or/67-73
75.	Meta-Analysis/
76.	exp Meta-Analysis as Topic/
77.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
78.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
79.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
80.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
81.	(search* adj4 literature).ab.
82.	(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.
83.	cochrane.jw.
84.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.

85.	or/75-84
86.	66 and (74 or 85)

Embase (Ovid) search terms

1.	Chronic pain/
2.	((chronic or persist* or idiopathic or atypical or a-typical) adj4 pain).ti,ab.
3.	exp Complex regional pain syndrome/
4.	(complex regional pain syndrome* or CRPS or causalgia).ti,ab.
5.	((reflex or sympathetic) adj2 dystroph*).ti,ab.
6.	fibromyalgia/
7.	(fibromyalgia* or fibrositis or myofascial pain syndrome).ti,ab.
8.	vulvodynia/
9.	(vulvodynia or vestibulodynia or dyspareunia or vulvar vestibulitis or vulvitis).ti,ab.
10.	interstitial cystitis/
11.	(interstitial adj2 cystitis).ti,ab.
12.	algodystrophy/
13.	(algodystroph* or sudek or sudeck*).ti,ab.
14.	myofascial pain/
15.	noncardiac chest pain/
16.	cystalgia/
17.	Pelvis pain syndrome/
18.	(loin pain adj (haematuria or hematuria) adj syndrome*).ti,ab.
19.	(LPHS or prostatodynia or CPPS or atypic* odontalgia or a-typic* odontalgia or burning mouth syndrome* or phantom tooth pain or neuropathic orofacial pain or "myofascial pain" or MPS).ti,ab.
20.	((pelvic or pelvis) adj pain syndrome*).ti,ab.
21.	((non-cardiac or noncardiac) adj3 chest adj3 pain).ti,ab.
22.	(temporomandibular adj3 joint adj3 pain).ti,ab.
23.	((prostate or vulv* or bladder or perineal) adj3 pain).ti,ab.
24.	(functional pain syndrome* or non-cancer pain or noncancer pain).ti,ab.
25.	((pelvic or pelvis or abdominal) adj3 pain adj3 (unknown or un-known or idiopathic or atypic* or a-typic*)).ti,ab.
26.	or/1-25
27.	letter.pt. or letter/
28.	note.pt.
29.	editorial.pt.
30.	case report/ or case study/
31.	(letter or comment*).ti.
32.	or/27-31
33.	randomized controlled trial/ or random*.ti,ab.
34.	32 not 33
35.	animal/ not human/
36.	nonhuman/
37.	exp Animal Experiment/
38.	exp Experimental Animal/
39.	animal model/
40.	exp Rodent/
41.	(rat or rats or mouse or mice).ti.

42.	or/34-41
43.	26 not 42
44.	exp exercise/
45.	exp kinesiotherapy/
46.	exp physical education/
47.	(pilates or yoga or feldenkrais or swim* or walk* or run* or jog* or treadmill* or tread mill*).ti,ab.
48.	(stretch* adj3 (active* or passive* or relax* or static* or dynamic* or gentl* or ballistic* or force* or isometric or technique* or exercis* or therap*)).ti,ab.
49.	(aerobic* adj (exercise* or train* or therap*)).ti,ab.
50.	((corrective* or biomechanic* or propiocet* or balance or flexib*) adj2 (exercise* or train* or therap*)).ti,ab.
51.	((biomechanic* or mckenzie) adj (method* or course*)).ti,ab.
52.	((strength* or stabil* or program* or train* or therap* or technique* or treat*) adj3 exercise*).ti,ab.
53.	(physical adj (fitness or conditioning or education or training or mobility or activit$ or exertion or effort)).ti,ab.
54.	danc*.ti,ab.
55.	(fitness* adj3 (program* or train* or therap*)).ti,ab.
56.	(tai ji or tai chi or taichi or taiji or taijiquan).ti,ab.
57.	(qigong or ch'i k#ng or ch'i g#ng or chi k#ng or chi g#ng or qi k#ng or qi g#ng).ti,ab.
58.	core stability.ti,ab.
59.	exp hydrotherapy/
60.	((water* or bath* or pool or pools or shower* or underwater* or spa or spas or aqua*) adj2 (exercise* or train* or therap* or treat*)).ti,ab.
61.	(hydrotherap* or hydro-therap*).ti,ab.
62.	(graded motor imagery or GMI or mirror therapy).ti,ab.
63.	or/44-62
64.	43 and 63
65.	limit 64 to English language
66.	randomized controlled trial.pt.
67.	controlled clinical trial.pt.
68.	randomi#ed.ti,ab.
69.	placebo.ab.
70.	randomly.ti,ab.
71.	Clinical Trials as topic.sh.
72.	trial.ti.
73.	or/66-72
74.	Meta-Analysis/
75.	exp Meta-Analysis as Topic/
76.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
77.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
78.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
79.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
80.	(search* adj4 literature).ab.
81.	(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.

82.	cochrane.jw.
83.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
84.	or/74-83
85.	65 and (73 or 84)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Chronic Pain] explode all trees
#2.	((chronic or persist* or idiopathic or atypical or a-typical) near/4 pain):ti,ab
#3.	MeSH descriptor: [Complex Regional Pain Syndromes] explode all trees
#4.	(complex regional pain syndrome* or CRPS or causalgia):ti,ab
#5.	((reflex or sympathetic) near/2 dystroph*):ti,ab
#6.	MeSH descriptor: [Fibromyalgia] explode all trees
#7.	(fibromyalgia* or fibrositis or myofascial pain syndrome):ti,ab
#8.	MeSH descriptor: [Vulvodynia] explode all trees
#9.	(vulvodynia or vestibulodynia or dyspareunia or vulvar vestibulitis or vulvitis):ti,ab
#10.	MeSH descriptor: [Cystitis, Interstitial] explode all trees
#11.	(interstitial near/2 cystitis):ti,ab
#12.	MeSH descriptor: [Reflex Sympathetic Dystrophy] explode all trees
#13.	(algodystroph* or sudek or sudeck*):ti,ab
#14.	MeSH descriptor: [Myofascial Pain Syndromes] explode all trees
#15.	(loinpain near (haematuria or hematuria) near syndrome*):ti,ab
#16.	(LPHS or prostatodynia or CPPS or atypic* odontalgia or a-typic* odontalgia or burning mouth syndrome* or phantom tooth pain or neuropathic orofacial pain or "myofascial pain" or MPS):ti,ab
#17.	((pelvic or pelvis) near pain syndrome*):ti,ab
#18.	((non-cardiac or noncardiac) near/3 chest near/3 pain):ti,ab
#19.	(temporomandibular near/3 joint near/3 pain):ti,ab
#20.	((prostate or vulv* or bladder or perineal) near/3 pain):ti,ab
#21.	(functional pain syndrome* or non-cancer pain or noncancer pain):ti,ab
#22.	((pelvic or pelvis or abdominal) near/3 pain near/3 (unknown or un-known or idiopathic or atypic* or a-typic*)):ti,ab
#23.	(or #1-#22)
#24.	MeSH descriptor: [Exercise] explode all trees
#25.	MeSH descriptor: [Exercise Therapy] explode all trees
#26.	MeSH descriptor: [Exercise Movement Techniques] explode all trees
#27.	MeSH descriptor: [Physical Education and Training] explode all trees
#28.	(pilates or yoga or feldenkrais or swim* or walk* or run* or jog* or treadmill* or tread mill*):ti,ab
#29.	(stretch* near/3 (active* or passive* or relax* or static* or dynamic* or gentl* or ballistic* or force* or isometric or technique* or exercis* or therap*)):ti,ab
#30.	(aerobic* near (exercise* or train* or therap*)):ti,ab
#31.	((corrective* or biomechanic* or propiocet* or balance or flexib*) near/2 (exercise* or train* or therap*)):ti,ab
#32.	((biomechanic* or mckenzie) near (method* or course*)):ti,ab
#33.	((strength* or stabil* or program* or train* or therap* or technique* or treat*) near/3 exercise*):ti,ab
#34.	(physical near (fitness or conditioning or education or training or mobility or activit$ or exertion or effort)):ti,ab
#35.	danc*:ti,ab

#36.	(fitness* near/3 (program* or train* or therap*)):ti,ab
#37.	(tai ji or tai chi or taichi or taiji or taijiquan):ti,ab
#38.	(qigong or ch'i k?ng or ch'i g?ng or chi k?ng or chi g?ng or qi k?ng or qi g?ng):ti,ab
#39.	core stability:ti,ab
#40.	MeSH descriptor: [Hydrotherapy] explode all trees
#41.	((water* or bath* or pool or pools or shower* or underwater* or spa or spas or aqua*) near/2 (exercise* or train* or therap* or treat*)):ti,ab
#42.	(hydrotherap* or hydro-therap*):ti,ab
#43.	(graded motor imagery or GMI or mirror therapy):ti,ab
#44.	(or #24-#43)
#45.	#23 and #44

B.2 Health Economics literature search strategy

Health economic evidence was identified by conducting a broad search relating to a Chronic Pain population in NHS Economic Evaluation Database (NHS EED – this ceased to be updated after March 2015) and the Health Technology Assessment database (HTA) with no date restrictions. 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 and economic modelling.

Table 70: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	2014 – 30 September 2019	Exclusions Health economics studies Health economics modelling studies
Embase	2014 – 30 September 2019	Exclusions Health economics studies Health economics modelling studies
Centre for Research and Dissemination (CRD)	HTA - Inception – 30 September 2019 NHSEED - Inception to March 2015	None

Medline search terms

1.	chronic pain/ or pain, intractable/
2.	((persist* or intract* or chronic or longstanding or long standing or longterm or long term or refractory or prolong* or long last* or sustain* or linger* or syndrome*) adj3 pain*).ti,ab.
3.	((chronic or persist* or idiopathic or atypical or a-typical) adj4 pain).ti,ab.
4.	exp Complex Regional Pain Syndromes/
5.	(complex regional pain syndrome* or CRPS or causalgia).ti,ab.
6.	fibromyalgia/
7.	((reflex or sympathetic) adj2 dystroph*).ti,ab.
8.	vulvodynia/
9.	(vulvodynia or vestibulodynia or dyspareunia or vulvar vestibulitis or vulvitis).ti,ab.

10.	interstitial cystitis/
11.	(interstitial adj2 cystitis).ti,ab.
12.	algodystrophy/
13.	(algodystroph* or sudek or sudeck*).ti,ab.
14.	exp myofascial pain syndromes/
15.	cystitis, interstitial/
16.	(loin pain adj (haematuria or hematuria) adj syndrome*).ti,ab.
17.	(LPHS or prostatodynia or CPPS or atypic* odontalgia or a-typic* odontalgia or burning mouth syndrome* or phantom tooth pain or neuropathic orofacial pain or "myofascial pain" or MPS).ti,ab.
18.	((pelvic or pelvis) adj pain syndrome*).ti,ab.
19.	((non-cardiac or noncardiac) adj3 chest adj3 pain).ti,ab.
20.	(temporomandibular adj3 joint adj3 pain).ti,ab.
21.	((prostate or vulv* or bladder or perineal) adj3 pain).ti,ab.
22.	(functional pain syndrome* or non-cancer pain or noncancer pain).ti,ab.
23.	((pelvic or pelvis or abdominal) adj3 pain adj3 (unknown or un-known or idiopathic or atypic* or a-typic*)).ti,ab.
24.	(fibromyalgia* or fibrositis or myofascial pain syndrome).ti,ab.
25.	or/1-24
26.	letter/
27.	editorial/
28.	news/
29.	exp historical article/
30.	Anecdotes as Topic/
31.	comment/
32.	case report/
33.	(letter or comment*).ti.
34.	or/26-33
35.	randomized controlled trial/ or random*.ti,ab.
36.	34 not 35
37.	animals/ not humans/
38.	exp Animals, Laboratory/
39.	exp Animal Experimentation/
40.	exp Models, Animal/
41.	exp Rodentia/
42.	(rat or rats or mouse or mice).ti.
43.	or/36-42
44.	25 not 43
45.	Economics/
46.	Value of life/
47.	exp "Costs and Cost Analysis"/
48.	exp Economics, Hospital/
49.	exp Economics, Medical/
50.	Economics, Nursing/
51.	Economics, Pharmaceutical/
52.	exp "Fees and Charges"/
53.	exp Budgets/
54.	budget*.ti,ab.

55.	cost*.ti.
56.	(economic* or pharmaco?economic*).ti.
57.	(price* or pricing*).ti,ab.
58.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
59.	(financ* or fee or fees).ti,ab.
60.	(value adj2 (money or monetary)).ti,ab.
61.	or/45-60
62.	exp models, economic/
63.	*Models, Theoretical/
64.	*Models, Organizational/
65.	markov chains/
66.	monte carlo method/
67.	exp Decision Theory/
68.	(markov* or monte carlo).ti,ab.
69.	econom* model*.ti,ab.
70.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
71.	or/62-70
72.	44 and (61 or 71)

Embase (Ovid) search terms

1.	chronic pain/ or pain, intractable/
2.	((persist* or intract* or chronic or longstanding or long standing or longterm or long term or refractory or prolong* or long last* or sustain* or linger* or syndrome*) adj3 pain*).ti,ab.
3.	((chronic or persist* or idiopathic or atypical or a-typical) adj4 pain).ti,ab.
4.	exp Complex regional pain syndrome/
5.	(complex regional pain syndrome* or CRPS or causalgia).ti,ab.
6.	((reflex or sympathetic) adj2 dystroph*).ti,ab.
7.	fibromyalgia/
8.	(fibromyalgia* or fibrositis or myofascial pain syndrome).ti,ab.
9.	vulvodynia/
10.	(vulvodynia or vestibulodynia or dyspareunia or vulvar vestibulitis or vulvitis).ti,ab.
11.	interstitial cystitis/
12.	(interstitial adj2 cystitis).ti,ab.
13.	algodystrophy/
14.	(algodystroph* or sudek or sudeck*).ti,ab.
15.	myofascial pain/
16.	noncardiac chest pain/
17.	cystalgia/
18.	Pelvis pain syndrome/
19.	(loin pain adj (haematuria or hematuria) adj syndrome*).ti,ab.
20.	(LPHS or prostatodynia or CPPS or atypic* odontalgia or a-typic* odontalgia or burning mouth syndrome* or phantom tooth pain or neuropathic orofacial pain or "myofascial pain" or MPS).ti,ab.
21.	((pelvic or pelvis) adj pain syndrome*).ti,ab.
22.	((non-cardiac or noncardiac) adj3 chest adj3 pain).ti,ab.
23.	(temporomandibular adj3 joint adj3 pain).ti,ab.
24.	((prostate or vulv* or bladder or perineal) adj3 pain).ti,ab.

25.	(functional pain syndrome* or non-cancer pain or noncancer pain).ti,ab.
26.	((pelvic or pelvis or abdominal) adj3 pain adj3 (unknown or un-known or idiopathic or atypic* or a-typic*)).ti,ab.
27.	or/1-26
28.	letter.pt. or letter/
29.	note.pt.
30.	editorial.pt.
31.	case report/ or case study/
32.	(letter or comment*).ti.
33.	or/28-32
34.	randomized controlled trial/ or random*.ti,ab.
35.	33 not 34
36.	animal/ not human/
37.	nonhuman/
38.	exp Animal Experiment/
39.	exp Experimental Animal/
40.	animal model/
41.	exp Rodent/
42.	(rat or rats or mouse or mice).ti.
43.	or/35-42
44.	27 not 43
45.	health economics/
46.	exp economic evaluation/
47.	exp health care cost/
48.	exp fee/
49.	budget/
50.	funding/
51.	budget*.ti,ab.
52.	cost*.ti.
53.	(economic* or pharmaco?economic*).ti.
54.	(price* or pricing*).ti,ab.
55.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
56.	(financ* or fee or fees).ti,ab.
57.	(value adj2 (money or monetary)).ti,ab.
58.	or/45-57
59.	statistical model/
60.	exp economic aspect/
61.	59 and 60
62.	*theoretical model/
63.	*nonbiological model/
64.	stochastic model/
65.	decision theory/
66.	decision tree/
67.	monte carlo method/
68.	(markov* or monte carlo).ti,ab.
69.	econom* model*.ti,ab.
70.	(decision* adj2 (tree* or analy* or model*)).ti,ab.

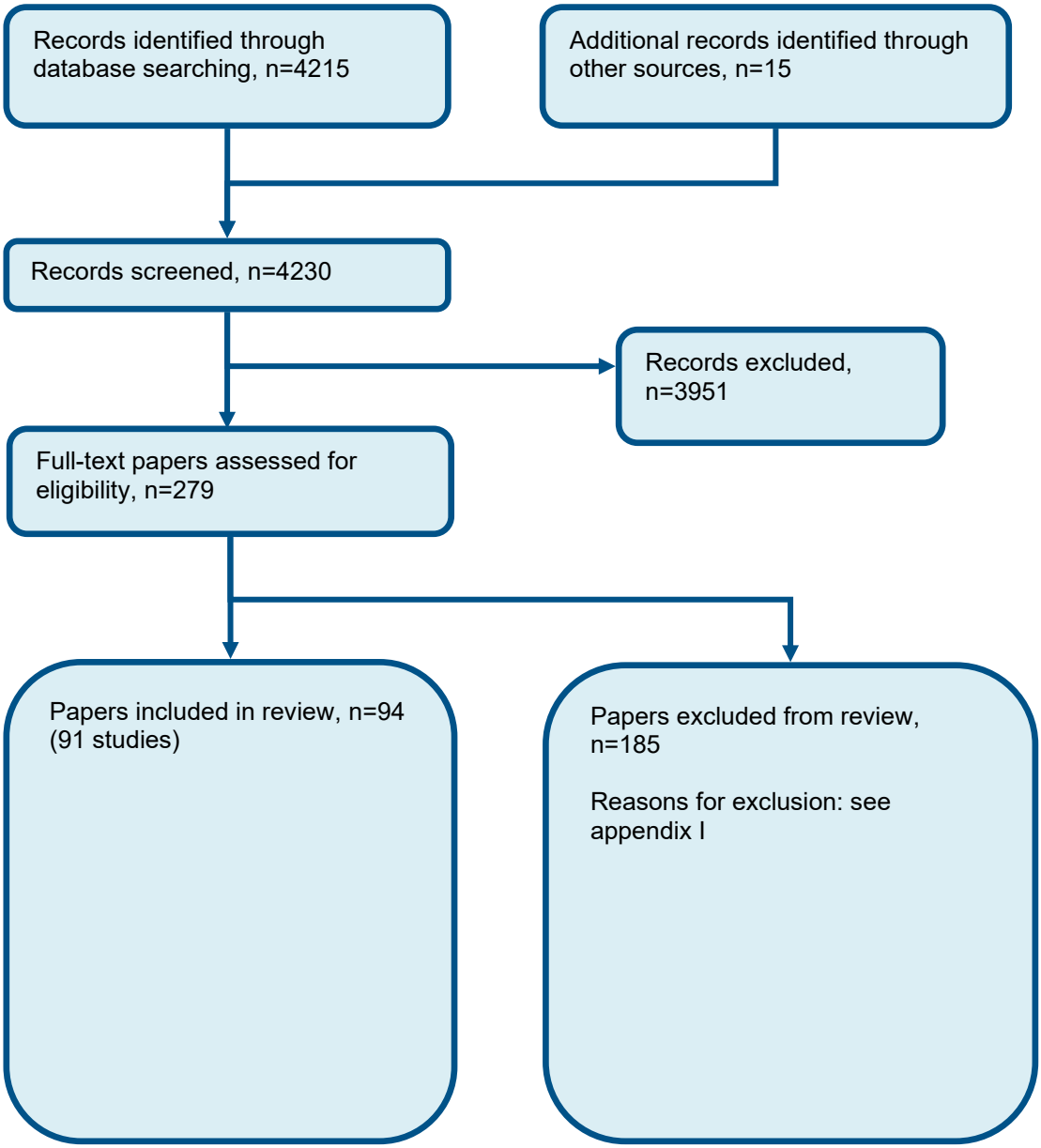
71.	or/61-70
72.	44 and (58 or 71)

NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Chronic Pain EXPLODE ALL TREES
#2.	(((persist* or intract* or chronic or longstanding or long standing or longterm or long term or refractory or prolong* or long last* or sustain* or linger* or syndrome*) adj3 pain*))
#3.	(((chronic or persist* or idiopathic or atypical or a-typical) adj4 pain))
#4.	MeSH DESCRIPTOR Complex Regional Pain Syndromes EXPLODE ALL TREES
#5.	((complex regional pain syndrome* or CRPS or causalgia))
#6.	MeSH DESCRIPTOR Fibromyalgia EXPLODE ALL TREES
#7.	(((reflex or sympathetic) adj2 dystroph*))
#8.	MeSH DESCRIPTOR Vulvodynia EXPLODE ALL TREES
#9.	((vulvodynia or vestibulodynia or dyspareunia or vulvar vestibulitis or vulvitis))
#10.	MeSH DESCRIPTOR Cystitis, Interstitial EXPLODE ALL TREES
#11.	((interstitial adj2 cystitis))
#12.	MeSH DESCRIPTOR Reflex Sympathetic Dystrophy EXPLODE ALL TREES
#13.	((algodystroph* or sudek or sudeck*))
#14.	MeSH DESCRIPTOR Myofascial Pain Syndromes EXPLODE ALL TREES
#15.	((loin pain adj (haematuria or hematuria) adj syndrome*))
#16.	((LPHS or prostatodynia or CPPS or atypic* odontalgia or a-typic* odontalgia or burning mouth syndrome* or phantom tooth pain or neuropathic orofacial pain or "myofascial pain" or MPS))
#17.	(((pelvic or pelvis) adj pain syndrome*))
#18.	(((non-cardiac or noncardiac) adj3 chest adj3 pain))
#19.	((temporomandibular adj3 joint adj3 pain))
#20.	(((prostate or vulv* or bladder or perineal) adj3 pain))
#21.	((functional pain syndrome* or non-cancer pain or noncancer pain))
#22.	(((pelvic or pelvis or abdominal) adj3 pain adj3 (unknown or un-known or idiopathic or atypic* or a-typic*)))
#23.	((fibromyalgia* or fibrositis or myofascial pain syndrome))
#24.	(#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23)

Appendix C: Clinical evidence selection

Figure 1: Flow chart of clinical study selection for the review of exercise



Appendix D: Clinical evidence tables

D.1 Evidence tables

Study	Acar 2012[1]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=60)
Countries and setting	Conducted in Turkey; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 2 weeks
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	(1) Under age of 65 years (2) no problems with cervical region but experiencing pain in the area within the last 6 months (3) not using pain killers.
Exclusion criteria	(1) Other conditions that cause pain
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 38(11.75) years. Gender (M:F): 3:17. Ethnicity: Not specified
Further population details	Chronic primary musculoskeletal pain subgroup
Extra comments	Exercise group duration of pain 43.65(48.17) years, control group 50.4(58.93) months
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Mixed modality exercise - Other mixed modality exercise. Strengthening exercises for multiple muscles and neck stretching exercises. 10 sessions 5 days a week, supervised by physiotherapists. Duration 2 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness

	(n=20) Intervention 2: Other. No treatment; no details. Duration 2 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH AND STRETCHING EXERCISES versus NO TREATMENT Protocol outcome 1: Pain reduction - Actual outcome: McGill Pain Questionnaire at 2 weeks; Group 1: mean 3.72 (SD 2.73); n=20, Group 2: mean 5.07 (SD 2.18); n=20; VAS 0-10 Top=High is poor outcome; Comments: Baseline: Exercise group 4.85(2.36); Control group 6.1(2.9) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference in VAS baseline scores and duration of pain; Group 1 Number missing: Not reported; Group 2 Number missing: not reported	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Altan 2004[9]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=46)
Countries and setting	Conducted in Turkey
Line of therapy	Unclear
Duration of study	Intervention + follow up: Intervention time 12 weeks, plus 12 weeks follow up
Method of assessment of guideline condition	ACR diagnostic criteria for fibromyalgia
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Not specified, although none of the participants had accompanying rheumatoid disease, unstable hypertension, cardiopulmonary problems, heat intolerance or any psychiatric disorder that could affect compliance
Exclusion criteria	Those with abnormal results were excluded (routine blood count and chemistry, ESR and urinalysis)
Age, gender and ethnicity	Age: Mean 43.9 years: . Gender (M:F): All female Ethnicity: Not specified
Further population details	Subgroup: Chronic primary musculoskeletal pain: fibromyalgia
Indirectness of population	No indirectness
Interventions	(n= 24) Intervention 1: Pool-based exercises All patients were given two educational sessions of 1 h each for 2 days by a physiatrist about the description and available diagnosis and treatment methods of FMS. Next, they were assigned randomly into two groups by the researcher other than the one who performed the evaluation throughout the study. In group 1, a pool-based exercise program was given by a physiotherapist to 25 patients in a therapeutic pool at 37°C for 35 min a day three times a week for 12 weeks. The program included warming (walking back and forth in the pool), activity (jumping in the pool and active joint motion range and stretching of the neck and the extremities), relaxation (lying supine on the water and slow swimming), and out-of-pool exercises (bending back and forth, squatting, and relaxing with deep breaths) for a period of 35 min. (n=22) Intervention 2: Control Warm balnefontainotherapy pool.
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PROPRIOCEPTION versus CONTROL

Protocol outcome 1: Pain reduction
- Actual outcome: Pain at end of treatment; Group 1: mean 5.81 (SD 2.7); n=24, Group 2: mean 5.63 (SD 1.62); n=22; VAS 0-10 Top=High is poor outcome;
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness
Baseline 7.91 (SD 1.81)
- Actual outcome: Pain at 24 week follow up; Group 1: mean 5.39 (SD 2.84); n=24, Group 2: mean 6.36 (SD 2.33); n=22; VAS 0-10 Top=High is poor outcome;
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness; Comments: baseline 7.91 (SD 1.81)

Protocol outcome 2: Quality of life
- Actual outcome: Quality of life at end of treatment; Group 1: mean 48.29 (SD 19.4); n=24, Group 2: mean 50.17 (SD 11.95); n=22; FIQ 0-100 Top=High is poor outcome;
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness
Baseline 7.91 (SD 1.81)
- Actual outcome: Quality of life at 24 week follow up; Group 1: mean 49.37 (SD 20.35); n=24, Group 2: mean 52.96 (SD 16.92); n=22; FIQ 0-100 Top=High is poor outcome;
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness; Comments: Baseline 62.58(13.14)

Protocol outcome 3: Physical function
- Actual outcome: Physical function at end of treatment; Group 1: mean 24.21 (SD 3.82); n=24, Group 2: mean 28.59(SD 4.56); n=22; Chair test Top=High is good outcome;
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness
- Actual outcome: Physical function at 24 weeks; Group 1: mean 24.91 (SD 2.87); n=24, Group 2: mean 25.77 (SD 4.82); n=22; Chair test Top=High is good outcome;
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness
Baseline: 24.95(3.19); 27(5.71)

Protocol outcome 4: Psychological Distress

- Actual outcome: Psychological distress at end of treatment; Group 1: mean 9.21 (SD 6.97); n=24, Group 2: mean 13.95 (SD 5.79); n=22; BDI 0-21 Top=High is poor outcome; Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness Baseline 7.91 (SD 1.81) - Actual outcome: Psychological Distress at 24 week follow up; Group 1: mean 10 (SD 7.57); n=24, Group 2: mean 14.86 (SD 9.45); n=22; BDI 0-21 Top=High is poor outcome; Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Comments: Baseline 14.08 (5.2) Protocol outcome 5: Discontinuation - Actual outcome: Discontinuation at end of treatment (12 weeks); Group 1: 1/25, Group 2: 3/25 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness	
Protocol outcomes not reported by the study	Pain interference; pain self-efficacy; Use of healthcare services ; Sleep ;

Study	Akhter 2014[5]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=62)
Countries and setting	Conducted in Pakistan; Setting: not reported
Line of therapy	Unclear
Duration of study	Intervention time: 3 months
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with history of more than 3 months neck pain with no related medical dysfunction
Exclusion criteria	Spinal instability, whiplash injury, osteoporosis, fracture of cervical spine, tumor of spine, unexplained headache, pain post cervical spine surgery, disc herniation, injection therapy application in cervical spine, radiculopathy of cervical spine, stenosis of cervical spine, rheumatoid arthritis, behaviour therapy rehabilitation and VBI symptoms (dizziness, drop attack, double vision), difficulty in swallowing, difficulty in finding words and patients who already had spinal manipulative session.
Recruitment/selection of patients	not reported
Age, gender and ethnicity	Age - Mean (range): exercise + manual therapy 38.1 (23-49); exercise only 39.5 (25-45). Gender (M:F): 23/39. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable 5. complex regional pain syndrome: Not applicable
Extra comments	Duration of symptoms (months): exercise + manual therapy 4.12 (1-6); exercise 4.78 (1-6)
Indirectness of population	No indirectness
Interventions	(n=31) Intervention 1: Manual therapy and exercise. Manual therapy (Maitland's approach Grade V, High velocity thrust, low amplitude application, rotation/lateral flexion technique on painful and stiff cervical spinal segments in supine position, maximum 6 sessions in 3 weeks) with supervised exercise regime for 20 minutes. The exercise regime

	included a set of strengthening exercises consisted of isometric, concentric and eccentric exercises with rest in between and a set of stretching exercises of cervical spine; rotation side to side, lateral flexion side to side, Extension and Sternocleidomastoid stretches 10 repetitions each to the left and right, Levator scapulae and pectolaris muscles stretches10 repetitions each to the left and right. After the end of 3 weeks intervention both groups taught and practiced a home exercise program. A printed exercise sheet was provided with frequency and repetition details: twice a day, 7 days a week, for 3 months. This home exercise program consisted of strengthening exercises for neck/scapular stability, stretching exercises and general range of motion exercises for neck with advice regarding posture awareness and correction . Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=31) Intervention 2: Strength. Participants performed supervised exercise regime same as the other group, and also followed the same home exercise programme. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EXERCISE versus STRENGTH Protocol outcome 1: Pain reduction - Actual outcome: Pain at end of treatment; Group 1: mean 2.4 (SD 1.17); n=31, Group 2: mean 3.1 (SD 1.13); n=31; VAS 0-10 Top=High is poor outcome; Comments: Baseline: manual + exercise 7.3 (1.08); exercise 7.6 (0.85) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness Protocol outcome 2: Physical function - Actual outcome: Neck disability at end of treatment; Group 1: mean 16.83 (SD 2.3); n=31, Group 2: mean 19.13 (SD 2.2); n=31; Neck Disability Index 0-100 Top=High is poor outcome; Comments: Baseline: manual + exercise 24.1 (3.2); exercise 27.1 (3.1) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness	
Protocol outcomes not reported by the study	Quality of life ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Altan 2009[10]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=50)
Countries and setting	Conducted in Turkey; Setting: No details
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 week intervention plus 12 weeks follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	None specified
Exclusion criteria	Routine blood count and chemistry, erythrocyte sedimentation rate, and urinalysis were performed for each patient, and those with abnormal results were excluded. All patients were instructed to discontinue nonsteroidal anti-inflammatory drug medication throughout the study period. The patients who had begun with antidepressive and/or sedative drugs at or prior to 1 month before the start of the study were allowed to continue their medications.
Recruitment/selection of patients	No details
Age, gender and ethnicity	Age - Mean (SD): 49.16(7.51) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup of people with chronic widespread pain
Extra comments	None of the patients had accompanying rheumatoid disease, unstable hypertension, severe cardiopulmonary problems, heat intolerance, or any psychiatric disorder affecting patient compliance
Indirectness of population	No indirectness
Interventions	(n=25) Intervention 1: Biomechanical - Pilates. The Pilates exercise program of 1 hour was given by a certified trainer to 25 participants 3 times a week for 12 weeks. The exercise program follows the basic principles of the Pilates method. Our protocol comprised 9 modules: postural education, search for neutral position, sitting exercise, antalgic exercises, stretching exercises, proprioceptivity improvement exercises, and breathing education. Resistance bands and 26cm Pilates balls were used as supportive equipment. Duration 12 weeks. Concurrent medication/care: Participants were allowed to take acetaminophen when they had severe pain. For a more accurate pain assessment, patients were asked

	to not take acetaminophen on the morning of the assessment day. Indirectness: No indirectness (n=25) Intervention 2: Mixed modality exercise - Other mixed modality exercise. Stretching and relaxation exercises. Participants were given a home exercise relaxation/stretching program, which has previously been routinely used for FMS patients in our clinic. The participants were instructed about this program of 1 hour 3 times a week for 12 weeks. We checked on this group's execution of the exercise program once a month. This exercise program consisted of relaxation techniques based on the published regimen by Ost and dynamic (slow, controlled leg and arm swings), active stretching (i.e., bringing the leg up high and holding it there without anything to keep it in that extended position), and passive stretching(i.e., reaching out to the feet while sitting up). Duration 12 weeks. Concurrent medication/care: Participants were allowed to take acetaminophen when they had severe pain. For a more accurate pain assessment, patients were asked to not take acetaminophen on the morning of the assessment day. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PILATES versus STRETCHING AND RELAXATION EXERCISES

Protocol outcome 1: Pain reduction
- Actual outcome: VAS final scores at 12 weeks (post intervention); Group 1: mean 4.1 (SD 1.7); n=25, Group 2: mean 6 (SD 2.1); n=24; VAS 0-10 Top=High is poor outcome
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 1

Protocol outcome 2: Quality of life
- Actual outcome: Fibromyalgia impact questionnaire final values at 12 weeks (post intervention); Group 1: mean 63.5 (SD 19.6); n=25, Group 2: mean 77.5 (SD 21.4); n=24; FIQ 0-100 Top=High is poor outcome
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 1

Protocol outcome 3: Physical function
- Actual outcome: Chair test at 12 weeks (post intervention); Group 1: mean 23.3 (SD 4.6); n=25, Group 2: mean 20.7 (SD 4.9); n=24; FIQ 0-100 Top=High is poor outcome
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 1
Baseline: 21.4(5.36); 22(5.2)

Protocol outcomes not reported by the study	Psychological distress (depression/anxiety); pain interference; pain self-efficacy; Use of healthcare services ; Sleep; Discontinuation

Study	**Andrade 2019[17]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=54)
Countries and setting	Conducted in Brazil; Setting: Department of Physical Therapy of the Federal University of São Carlos.
Line of therapy	Unclear
Duration of study	Intervention + follow up: 16 week intervention (plus 16 week follow up after detraining)
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR criteria for fibromyalgia
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Participants aged 30-60 years and had low level of physical activity according to the International Physical Activity Questionnaire (iPAQ)
Exclusion criteria	Volunteers with cardiovascular diseases, systemic arterial hypertension, arrhythmias, diabetes mellitus, musculoskeletal and neurological disorders that could directly interfere with assessments (for example, advanced joint diseases), presence of infections and any other rheumatic diseases (e.g., osteoarthritis, connective tissue disease, rheumatoid arthritis) were excluded.
Recruitment/selection of patients	Participants were recruited through posters and leaflets distributed at strategic points in the city (rheumatology, orthopedics and physiotherapy clinics and offices) from December 2013 to December 2014.
Age, gender and ethnicity	Age - Mean (SD): 47.5(8) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with pain conditions other than chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with chronic widespread pain 5. complex regional pain syndrome: people with pain conditions other than complex regional pain syndrome
Extra comments	7.5(9.5) years (NB: study states duration of diagnosis 75 years; assumed error).
Indirectness of population	No indirectness

Interventions	(n=27) Intervention 1: Aerobics - Swimming. The APT program was performed in a heated pool (30±2 °C). The protocol consisted of 32 sessions of 45 min, twice a week (alternating days) for 16 weeks. The sessions were conducted in groups of up to 5 women and were supervised by three physiotherapists. The APT protocol has already been described in a previous study conducted by our research group. 14 The progression of aerobic exercises was adjusted throughout the sessions in order to maintain HR and the subjective perceived exertion (RPE) reached at VAT level identified in the CPET. Duration 16 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=27) Intervention 2: No treatment. No treatment; no further details. Duration 16 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Financial support from Sao Paulo research foundation Support (FAPESP) and from National Council for Scientific and Technological Development

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC versus NO TREATMENT

Protocol outcome 1: Pain reduction
- Actual outcome: VAS pain reduction at 16 weeks; Group 1: mean 5.4 (SD 2.4); n=27, Group 2: mean 6.4 (SD 2.1); n=27; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 5.8(2.7); 5.5(2.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 functional capacity subscale at 16 weeks; Group 1: mean 50.5 (SD 17.6); n=27, Group 2: mean 38 (SD 14.7); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 44.6(17.6) 38.2(13.9)
Risk of bias: All domain – Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3
- Actual outcome: SF-36 physical appearance subscale at 16 weeks; Group 1: mean 29.8 (SD 41); n=27, Group 2: mean 13.8 (SD 27.8); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 10.2(28); 11(25.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3
- Actual outcome: SF-36 pain subscale at 16 weeks; Group 1: mean 36.7 (SD 41); n=27, Group 2: mean 29.2 (SD 12.1); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 31.8(16.3); 25.5(11)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3

- Actual outcome: SF-36 vitality subscale at 16 weeks; Group 1: mean 37.9 (SD 22.4); n=27, Group 2: mean 30.2 (SD 15.1); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 33.5(18.6); 25.4(14.7)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3
- Actual outcome: SF-36 social aspect subscale at 16 weeks; Group 1: mean 54.3 (SD 22.2); n=27, Group 2: mean 45.4 (SD 23); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 48.1(17.9); 44.5(20.2)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3
- Actual outcome: SF-36 emotional aspect subscale at 16 weeks; Group 1: mean 32.1 (SD 40.8); n=27, Group 2: mean 22.4 (SD 35.5); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 24.7 (35.3) / 18.7 (29.4)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3
- Actual outcome: SF-36 mental health subscale at 16 weeks; Group 1: mean 46.8 (SD 23); n=27, Group 2: mean 43.4 (SD 17.3); n=27; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline 48.6(22.1); 53.7(21.2)
Risk of bias: All domain - Very high, Selection –High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3

Protocol outcome 3: Psychological distress (depression/anxiety)
- Actual outcome: Beck depression inventory at 16 weeks; Group 1: mean 15.8 (SD 9); n=27, Group 2: mean 19.6 (SD 8.6); n=27; BDI 0-21 Top=High is poor outcome; Comments: Baseline 18.2(9.6); 20.6(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; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3
- Actual outcome: Beck anxiety inventory at 16 weeks; Group 1: mean 15.3 (SD 9.1); n=27, Group 2: mean 19.5 (SD 9); n=27; BAI 0-21 Top=High is poor outcome; Comments: baseline 16.1(9.1);21.2(9.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3

Protocol outcome 4: Sleep
- Actual outcome: Pittsburgh sleep quality index at 16 weeks; Group 1: mean 8.8 (SD 4.4); n=27, Group 2: mean 11.2 (SD 3.3); n=27; PSQI 0-21 Top=High is poor outcome; Comments: Baseline: 9.4(4.3); 11(3.8)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3

Protocol outcome 5: Discontinuation
- Actual outcome: Discontinuation at 16 weeks; Group 1: 3/27, Group 2: 3/27
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low,

Subgroups - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 3	
Protocol outcomes not reported by the study	Physical function; Use of healthcare services

Study	Assumpcao 2018[23]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=53)
Countries and setting	Conducted in Brazil; Setting: Fibromyalgia outpatient clinic
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR classification (by rheumatologist)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 30 to 55 years
Exclusion criteria	non-controlled systemic disorders (diabetes, hypertension), neurological and musculoskeletal conditions that could compromise assessments, impaired alertness or comprehension, relevant joint disorders (severe arthritis, arthroplasty of the hip or knee, rheumatoid arthritis), recent changes in physical activity, and recent changes in therapy for FM (medication, educational programs, alternative medicine, psychotherapy).
Recruitment/selection of patients	People who were referred to the physical therapy service, fibromyalgia outpatient clinic at Hospital das Clinacas HCFMUSP, Faculdade de Medcina, Universidade de Sao Paulo.
Age, gender and ethnicity	Age - Mean (SD): 47(6.2) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=18) Intervention 1: Flexibility. Patients underwent a 12 week supervised exercise program of 40-minute sessions performed twice a week. Segmental active muscle stretching was conducted without therapist assistance. Large muscles were chosen for their role in the muscular chains of global posture. Patients started with three repetitions up to a maximum of 5 by week 9. The stretch was held until the point of moderate discomfort, for 30 seconds. Duration 12 weeks. Concurrent medication/care: 57% were taking concomitant medication for fibromyalgia (antidepressants, analgesics, anti-inflammatories or psychotropic medications).a. Indirectness: No indirectness (n=19) Intervention 2: Strength. 12 week supervised resistance training programme of 40-minute sessions performed twice a week with progressive overload. Equipment included dumbbells, shin pads. No load was used in the first 2

	sessions, after which time 0.5kg was added each week if the patient identified the effort as slightly intense on the Borg scale. 8 repetitions for: triceps, quadriceps, hip adductors and abductors, hip flexors, elbow flexors and extensors, pectoralis major and rhomboids. Duration 12 weeks. Concurrent medication/care: 62% were taking concomitant medication for fibromyalgia (antidepressants, analgesics, anti-inflammatories or psychotropic medications). Indirectness: No indirectness (n=16) Intervention 3: Other. Control group: usual medical treatment. After 12 weeks patients were reassessed and offered physical therapy based on stretching and resistance training. Duration 12 weeks. Concurrent medication/care: 43% were taking medication for fibromyalgia (antidepressants, analgesics, anti-inflammatories or psychotropic medication). Indirectness: No indirectness
Funding	Academic or government funding (Fundacao de Amparo a)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: FLEXIBILITY (STRETCHING) versus STRENGTH (RESISTANCE TRAINING)

Protocol outcome 1: Pain reduction
- Actual outcome: VAS final values at 12 weeks; Group 1: mean 4.6 (SD 2.6); n=14, Group 2: mean 4.4 (SD 3); n=16; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.6 (1.8); 5.3(2.5)
Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4, Reason: Lost to follow up; Group 2 Number missing: 3, Reason: Lost to follow up

Protocol outcome 2: Physical function
- Actual outcome: FIQ physical function subscale final values at 12 weeks; Group 1: mean 9.5 (SD 5.2); n=14, Group 2: mean 15.5 (SD 5); n=16; FIQ physical function subscale 0-30 Top=High is poor outcome; Comments: Baseline: 6.5(5.5); 10.9(6.3)
Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4, Reason: Lost to follow up; Group 2 Number missing: 3, Reason: Lost to follow up

Protocol outcome 3: Discontinuation
- Actual outcome: Discontinuation at 12 weeks; Group 1: 3/17, Group 2: 2/18
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: Lost to follow up; Group 2 Number missing: 1, Reason: Lost to follow up

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: FLEXIBILITY (STRETCHING) versus USUAL CARE

Protocol outcome 1: Pain reduction

- Actual outcome: VAS final values at 12 weeks; Group 1: mean 4.6 (SD 2.6); n=16, Group 2: mean 6.4 (SD 2.7); n=14; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.6(1.8); 6(2.6)
Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4, Reason: Lost to follow up (1), discontinued intervention (3); Group 2 Number missing: 2, Reason: Lost to follow up

Protocol outcome 2: Physical function
- Actual outcome: FIQ physical function subscale final values at 12 weeks; Group 1: mean 9.5 (SD 5.2); n=14, Group 2: mean 10.5 (SD 5.3); n=14; FIQ physical function subscale 0-30 Top=High is poor outcome; Comments: Baseline: 6.5(5.5); 9.6(3.8)
Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data – Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4, Reason: Lost to follow up (1), discontinued intervention (3); Group 2 Number missing: 2, Reason: Lost to follow up

Protocol outcome 3: Discontinuation
- Actual outcome: Discontinuation at 12 weeks; Group 1: 3/17, Group 2: 0/14
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: Lost to follow up; Group 2 Number missing: 2, Reason: Lost to follow up

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH (RESISTANCE TRAINING) versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: VAS final values at 12 weeks; Group 1: mean 4.4 (SD 3); n=16, Group 2: mean 6.4 (SD 2.7); n=14; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.3(2.5); 6(2.6)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3, Reason: Lost to follow up (1), discontinued (2); Group 2 Number missing: 2, Reason: Lost to follow up

Protocol outcome 2: Physical function
- Actual outcome: FIQ physical function subscale final values at 12 weeks; Group 1: mean 14.5 (SD 5); n=16, Group 2: mean 10.5 (SD 5.3); n=14; FIQ physical function subscale 0-30 Top=High is poor outcome; Comments: Baseline: 10.9(6.3); 9.6(3.8)
Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3, Reason: Lost to follow up (1), discontinued (2); Group 2 Number missing: 2, Reason: Lost to follow up

Protocol outcome 3: Psychological distress (depression/anxiety)
- Actual outcome: **extract median and interquartile range data into report (too many outcomes to extract in here). FIQ anxiety, depression, SF-36 8 subscales at 12

weeks; Risk of bias: All domain - ; Indirectness of outcome: No indirectness Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at 12 weeks; Group 1: 2/18, Group 2: 0/14 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: Lost to follow up; Group 2 Number missing: 2, Reason: Lost to follow up	
Protocol outcomes not reported by the study	Quality of life ; pain interference; pain self-efficacy; Use of healthcare services ; Sleep

Study	Baptista 2012[27]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=80)
Countries and setting	Conducted in Brazil; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention + follow up: 16 week intervention plus 16 weeks follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR criteria
Stratum	Overall
Subgroup analysis within study	Not applicable:
Inclusion criteria	Diagnosis of fibromyalgia based on the criteria of the American College of Rheumatology (1); female gender; age between 18 and 65 years; not having altered treatment in previous four weeks; and having signed an informed consent document.
Exclusion criteria	Patients with other rheumatic diseases, painful joint diseases, uncontrolled cardiopulmonary diseases, diseases of the lower limbs or uncontrolled diabetes were excluded
Recruitment/selection of patients	From rheumatology outpatient clinic
Age, gender and ethnicity	Age - Mean (SD): 49.3 years (SD 11.2) (range 18-65 years). Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=40) Intervention 1: Mind-body exercises - Other. 1 hour belly dance class twice a week for 16 weeks. Each class had a maximum of 8 students and was led by physiotherapists. Classes began with warm up, followed by movements for the day, choreography and a cool-down exercise. Participants also received a disc with music and an exercise book with all movements for the programme. From the 4th week a set sequence of movements in the form of choreography was established for training at home. Duration 16 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=40) Intervention 2: Other. Offered intervention at the end of study. Duration 16 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness

Funding	Academic or government funding (CAPES)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MIND-BODY EXERCISE (BELLY DANCING) versus CONTROL (WAITING LIST CONTROL)

Protocol outcome 1: Pain reduction
- Actual outcome: VAS final values at 32 weeks (follow up, including 16 week intervention); Group 1: mean 4.7 (SD 2.6); n=40, Group 2: mean 7.3 (SD 1.7); n=40; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 7.7(1.7); 7.5(1.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales not balanced at baseline; Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 functional capacity subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 56.3 (SD 19.9); n=40, Group 2: mean 39.1 (SD 22); n=40; sf-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 44.9(1.89); 32.6(18.9)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression
- Actual outcome: SF-36 physical aspects subscale at 32 weeks follow up (including 16 week intervention); Group 1: mean 36.5 (SD 32.4); n=40, Group 2: mean 13.8 (SD 26.5); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 24.7(32.2), 8.8(17.9)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression
- Actual outcome: SF-36 pain subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 46 (SD 19.2); n=40, Group 2: mean 29.1 (SD 21.1); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 29.6(17.5); 25.7(13.4)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression
- Actual outcome: SF-36 general health subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 44.9 (SD 15.6); n=40, Group 2: mean 41.5 (SD 21.4); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: baseline: 46(21.7); 38(16.5)
Risk of bias: All domain - ; Indirectness of outcome: No indirectness
- Actual outcome: SF-36 vitality subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 47.6 (SD 23.8); n=40, Group 2: mean 37.1 (SD 21.8); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 41.3(18.8); 29(18.2)
Risk of bias: All domain - ; Indirectness of outcome: No indirectness
- Actual outcome: SF-36 social subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 57.2 (SD 27); n=40, Group 2: mean 51.3 (SD 25.5); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 52.6(27.7); 47.6(23.1)

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression - Actual outcome: SF-36 emotional subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 51.9 (SD 39.6); n=50, Group 2: mean 31.5 (SD 38.7); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 34.2(36.9); 21.2(33.1) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression - Actual outcome: SF-36 mental health subscale at 32 weeks (follow up, including 16 week intervention); Group 1: mean 52.3 (SD 20.8); n=40, Group 2: mean 46.2 (SD 22.6); n=40; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 46(19.9); 43.4(24) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression Protocol outcome 1: Physical function - Actual outcome: 6 minute walk test at 32 weeks (follow up, including 16 week intervention); Group 1: mean 431 (SD (88.7); n=40, Group 2: mean 343 (SD 77.9); n=40; Metres; Comments: Baseline: 372.8(80.2);332(66.7) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales not balanced at baseline; Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression Protocol outcome 3: Psychological distress (depression/anxiety) - Actual outcome: Beck inventory final values at 32 weeks (follow up, including 16 week intervention); Group 1: mean 23.1 (SD 15.3); n=40, Group 2: mean 23.5 (SD 13.7); n=40; BDI 0-63 Top=High is poor outcome; Comments: Baseline: 23.9(14.7); 21.2(13.0) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at 32 weeks (follow up, including 16 week intervention); Group 1: 2/40, Group 2: 3/40 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 functional capacity and limitation due to physical aspects (8.8 vs 24.7); Group 1 Number missing: 2, Reason: RA diagnosis, family problems; Group 2 Number missing: 3, Reason: Moved to a different city, breast cancer, severe depression	
Protocol outcomes not reported by the study	Physical function ; pain interference; pain self-efficacy; Use of healthcare services ; Sleep

Study	Bircan 2008[34]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=30)
Countries and setting	Conducted in Turkey; Setting: Outpatient clinic, no further details
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	None specified
Exclusion criteria	Presence of serious cardiovascular, pulmonary, endocrine, neurological or renal disease, inflammatory rheumatic disease, or participation in a physical therapy or exercise program in the last 6 months.
Recruitment/selection of patients	Through outpatient clinic. No further details
Age, gender and ethnicity	Age - Mean (SD): 47.2(7.1) years. Gender (M:F): All female. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=15) Intervention 1: Aerobics - Walking. Aerobic exercise program comprised walking on tread- mill, initially for 20 min and increasing up to 30 min as the patient tolerated. Exercise intensity was adjusted to generate heart rates equivalent to 60–70% of age-adjusted maxi- mum heart rates (220 ¡ age in years). Heart rate monitoring was performed by using a pulse oximeter (Nonin Medical, Inc., MN, USA). At the beginning and end of each session mild stretches were included for 5 min. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=15) Intervention 2: Strength. Patients in the SE group received a supervised, progressive physical training program in a group setting with muscle strengthening exercises performed in the standing, sitting, and lying positions. Exercises strengthened the upper and lower limb muscles and trunk muscles, initially with 4–5 repetitions and progressing to 12 repetitions gradually. Free weights and body weight were used for strengthening. Patients began with resistance levels they could do easily, and weight was increased gradually according to patient's tolerance. Exercise sessions began with a low intensity warm up of marching in place and gentle stretching for 5 min, followed by 30 min of muscle

	strengthening, and concluded with 5 min of cool down and stretching. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: WALKING versus STRENGTH

Protocol outcome 1: Pain reduction
- Actual outcome: VAS final values at 8 weeks; Group 1: mean 2.19 (SD 1.88); n=13, Group 2: mean 2.65 (SD 1.41); n=13; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 6.07(1.86); 5.21(2.18)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 2, Reason: Pneumonia, transportation problems; Group 2 Number missing: 2, Reason: Transportation problems

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 physical component summary score at 8 weeks; Group 1: mean 38.92 (SD 6.11); n=13, Group 2: mean 43.01 (SD 7.02); n=13; SF-36 physical component summary score 0-100 Top=High is good outcome; Comments: Baseline: 34.49(6.02); 35.81(8.26)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 2, Reason: Pneumonia, transportation problems; Group 2 Number missing: 2, Reason: Transportation problems
- Actual outcome: SF-36 mental component summary score at 8 weeks; Group 1: mean 41.07 (SD 8.53); n=13, Group 2: mean 45.44 (SD 7.71); n=13; SF-36 mental component summary score 0-100 Top=High is good outcome; Comments: Baseline: 35.51(7.92); 38.66(9.78)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 2, Reason: Pneumonia, transportation problems; Group 2 Number missing: 2, Reason: Transportation problems

Protocol outcome 3: Psychological distress (depression/anxiety)
- Actual outcome: HAD-anxiety score at 8 weeks; Group 1: mean 8.31 (SD 3.79); n=12, Group 2: mean 9.54 (SD 3.62); n=13; Hospital anxiety and depression scale (anxiety subscore) 0-21 Top=High is poor outcome; Comments: Baseline: 9.46(4.45); 10.08(4.59)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 2, Reason: Pneumonia, transportation problems; Group 2 Number missing: 2, Reason: Transportation problems
- Actual outcome: HAD-depression score at 8 weeks; Group 1: mean 6.39 (SD 3.79); n=13, Group 2: mean 5.69 (SD 3.28); n=13; Hospital anxiety and depression scale (depression subscore) 0-21 Top=High is poor outcome; Comments: Baseline: 8.39(3.97); 8.23(4.51)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 2, Reason: Pneumonia, transportation problems; Group 2 Number missing: 2, Reason: Transportation problems Protocol outcome 4: Sleep - Actual outcome: VAS sleep final values at 8 weeks; Group 1: mean 1.25 (SD 1.71); n=13, Group 2: mean 2.58 (SD 2.97); n=13; VAS sleep scale 0-10 Top=High is poor outcome; Comments: Baseline: 4.6(2.01); 4.45(2.98) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 2, Reason: Pneumonia, transportation problems; Group 2 Number missing: 2, Reason: Transportation problems Protocol outcome 5: Discontinuation - Actual outcome: Discontinuation at 8 weeks; Group 1: 2/15, Group 2: 2/15 Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 0, Reason: Pneumonia, transportation problems; Group 2 Number missing: 0, Reason: Transportation problems	
Protocol outcomes not reported by the study	Physical function ; Use of healthcare services

Study	**Borisut 2013[39]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=100)
Countries and setting	Conducted in Thailand; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Females, aged between 20 and 35 with a history of intermittent work-related neck pain lasting for more than 6 months who worked with a computer at least 4 hours each working day. The pain level at the time of examination exceeded 30 mm on a visual analogue scale of

	0–100 mm
Exclusion criteria	Participants were excluded if they had neck or shoulder pain from non-musculoskeletal causes, demonstrated neurological signs, or had a history of malignancy, pregnancy, or menstruation at the time of examination
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): 32.72 (3.11); 30.40 (3.54); 30.16 (2.96); 29.32 (3.11). Gender (M:F): All female. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with pain conditions other than chronic widespread pain 5. complex regional pain syndrome: people with pain conditions other than complex regional pain syndrome
Indirectness of population	No indirectness
Interventions	(n=75) Intervention 1: Strength and conditioning. Within this group, 25 participants underwent strength-endurance training, 25 participants underwent cranio-flexion exercise, and 25 participants underwent a combination of strength-endurance and cranio-flexion. The strength-endurance training consisted of a progressive resistance exercise program forthe neck muscles, especially the superficial neck flexor and extensor muscles (SCM,AS and CE). Neck flexion and extension were performed in the supine and pronepositions, respectively, with the head supported in a comfortable resting position. Subjects slowly moved the head and neck through the total range of motion avoiding discomfort or symptom reproduction. This exercise program included two phases. The first phase of 4 weeks and the second of 8 weeks were recommended for initiating a weight program in untrained individuals). In phase one, each subject performed 12–15 repetitions with a weight that they could lift 12 times on the first training session (12 repetitions maximum) andprogress to 15 repetitions. They were maintained at this level for 4 weeks. In phase two, subjects performed 3 sets of 15 repetitions of the initial 12 repetition sat maximum load with one minute rest interval between sets. The craniocervical flexion exercise consisted of a low load exercise for the cranio-cervical flexor muscles. Subjects lay supine and slowly moved the head to the inner

	range of cranio-cervical flexion, guided by feedback from an air filled pressure sensor placed suboccipitally behind the neck and inflated to a baseline pressure of 20 mmHg. Subjects moved the head toincrease the pressure to between 22 to 30 mmHg; and maintained this position for 10 seconds in 15 repetitions. The subjects maintained the 10-second contraction with no pain. Ten seconds rest was allowed between each contraction. The targets of this exercise are the deep flexors of the uppercervical region, the longus capitis and colli, rather than the superficialflexors, which flex the neck but not the head. The combined exercise group performed both strength endurance and cranio-cervical flexion exercises. First, subjects lay supine and performed the cranio-cervical flexion exercise. A five minute rest was then taken before performing the strength-endurance exercise. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=25) Intervention 2: No treatment. After finishing data collection, participants in the control group were advised to perform both the strength-endurance and cranio-cervical exercises. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH AND CONDITIONING versus NO TREATMENT Protocol outcome 1: Pain reduction - Actual outcome: Pain at 12 weeks; Group 1: mean 32.87 (SD 17.12); n=75, Group 2: mean 61.32 (SD 11.29); n=25; VAS 0-100 Top=High is poor outcome; Comments: Baseline values: 57.51 (17.34); 59.04 (10.49) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 2: Physical function - Actual outcome: Neck disability at 12 weeks; Group 1: mean 14.41 (SD 4.94); n=75, Group 2: mean 33.86 (SD 5.04); n=25; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline values: 29.13 (5.11); 31.56 (5.14) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life at Define; Psychological distress (depression/anxiety) at Define; Use of healthcare services at Define; Sleep at Define; Discontinuation at Define

Study	Bronfort 2001[43]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=191)
Countries and setting	Conducted in USA; Setting: Minneapolis, Minnesota
Line of therapy	Unclear
Duration of study	Intervention + follow up: 11 weeks and 1 year follow up
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 20 to 65 years, neck pain persisting for at least 12 weeks (mechanical neck pain , no specific identifiable etiology).
Exclusion criteria	Referred neck pain, osteopenia, any neurological or vascular conditions that could affect the neck, spine surgery, inability to work because of neck pain, and previous involvement in manipulation therapy or exercise in the last 3 months.
Recruitment/selection of patients	Local newspaper advertisements
Age, gender and ethnicity	Age - Mean (SD): 44.3(10.6) years. Gender (M:F): 78:113. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: 4. chronic widespread pain:
Extra comments	Median duration of pain 5 years (range 0.3 to 34)
Indirectness of population	No indirectness
Interventions	(n=63) Intervention 1: Manual therapy and exercise. Spinal manipulation therapy and exercise. Participants underwent treatment from an experienced chiropractor for 15 minutes, followed by a supervised exercise session for 45 minutes. Manipulation therapy was administered to the cervical and thoracic spine, as well as light soft-tissue massage. The exercise component involved progressive strengthening exercises for the neck and upper body preceded by a short aerobic warm up of the upper body and light stretching. 2 sets of 15-30 repetitions were conducted and resistance was increased gradually over time. Duration 11 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=60) Intervention 2: Mixed modality exercise - Aerobic, Strength exercise. Warm up of stretching and upper body strengthening followed by 15 to 20 minutes of aerobic exercise using a stationary bike. Resistance exercises were

	performed on the MedX cervical extension and rotation machines, and resistance was increased periodically, with patients performing approximately 20 repetitions of each exercise. Duration 11 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated

AEROBIC AND STRENGTH EXERCISE VERSUS STRENGTH AND MANUAL THERAPY

Protocol outcome 1: Pain reduction
- Actual outcome: VAS final values at 11 weeks; Group 1: mean 24.1 (SD 19.7); n=56, Group 2: mean 23.6 (SD 18); n=63; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 57.1(15); 56(15)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness
- Actual outcome: VAS final values at 12 months follow up; Group 1: mean 29.8 (SD 20.4); n=56, Group 2: mean 31.1(SD 22.7); n=63; NDI 0-50 Top=High is poor outcome; Comments: Baseline: 57.1(15); 56(15)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness

Protocol outcome 2: Physical function
- Actual outcome: Neck disability index at 11 weeks; Group 1: mean 17.1 (SD 10.3); n=56, Group 2: mean 18.6 (SD 9.2); n=63; NDI 0-50 Top=High is poor outcome; Comments: Baseline: 57.1(15); 56(15)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness
- Actual outcome: Neck disability index at 12 months follow up; Group 1: mean 15.6 (SD 13.1); n=56, Group 2: mean 16.1(SD 11.2); n=63; NDI 0-50 Top=High is poor outcome; Comments: Baseline: 57.1(15); 56(15)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness

Protocol outcome 3: Discontinuation
- Actual outcome: Discontinuation at 11 weeks; Group 1: 4/60, Group 2: 5/63
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: SF-36 subscales, VAS; Group 1 Number missing: 0, Reason: Pneumonia, transportation problems; Group 2 Number missing: 0, Reason: Transportation problems

Quality of life ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ;	
Protocol outcomes not reported by the study	

Study	Carvalho 2020[55]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=35)
Countries and setting	Conducted in Brazil; Setting: This study was conducted in the Laboratory of Movement Analysis of the Department of Physiotherapy, Federal University of Alfenas
Line of therapy	Unclear
Duration of study	Intervention time: 7 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women with a minimum age of 18 yearsand a diagnosis of fibromyalgia in accordance with the parameters of the American College of Rheumatology (ACR). The diagnosis requires a history of widespread pain (i.e., in >7 regions), at least moderate severity (a score >5) of pain, fatigue, sleep disruption, and cognitive symptoms, duration of symptoms >3 months, and absence of another disorder that could explain the condition. Criteria are also satisfied if only three to six regions are affected by pain, but the symptoms are more severe (a score q9)
Exclusion criteria	Cardiovascular, pulmonary, orthopedic, neurological, or dermatological conditions, which negatively affect muscle strength and physical capabilities and pregnancy. Men were excluded to avoid a heterogeneous sample and due to low prevalence
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Exercise group: 55.64 (9.16); stretch group: 47.70 (15.46). Gender (M:F): All female. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain 5. complex regional pain syndrome:
Indirectness of population	No indirectness
Interventions	(n=16) Intervention 1: Mixed modality exercise - Other mixed modality exercise. The intervention was named exergames. It was performed thrice per week with each session lasting 1 hour. The intervention took place using a Nintendo Wii system. Before beginning the intervention, participants were instructed and trained to play the

	games and handle the game console. Six subgames of Wii Fit Plus were chosen for this group. These included Jogging Plus, an activity in which the subjects perform stationary running. It results in active and constant movement of the lower limb muscles for 15 minutes. The ''Bird's-eye Bull's-eye game'' was performed for 9 minutes. It is a game that requires active movement of the upper limbs in isolation from weight and balance training. The ''Yoga game'' was used for 3 minutes. It stimulates not just control of expiratory and inspiratory movements but also active control of the body's center of gravity. The ''Super Hula Hoop game'' was performed for 9 minutes. It requires the action of the trunkmuscles associated with circular rhythmic movements as well as balance control. A ''Step game'' was used for 15 minutes and consists of active and alternating movements of the lower limb muscles, as well as balance and unipodal discharge. Finally, ''Rhythm Parade'' was performed for 9 minutes. It consists of stationary walking associated with active and rhythmic movements of the lower limb muscles.. Duration 7 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=19) Intervention 2: Flexibility. Chain muscle stretching technique thrice per week with each session lasting 1 hour. The positions were held during four deep and prolonged expirations. Exercises were chosen to include standing, sitting, and lying positions. In addition, they were chosen to engage all muscle groups in a global manner. Duration 7 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Academic or government funding (Grants from the Research Support Foundation of Minas Gerais and Tutorial Education program, and part financed by the Coordination for the Improvement of Higher Education Personel)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: OTHER MIXED MODALITY EXERCISE versus STRETCHING

Protocol outcome 1: Quality of life

- Actual outcome: Fibromyalgia Impact Questionnaire - total score at 7 weeks (After 20 sessions); Group 1: mean 33.4 (SD 6.29); n=11, Group 2: mean 46.44 (SD 13.01); n=10; FIQ - total score Not reported Top=High is poor outcome; Comments: Baseline values: exercise group 64.55 (16.09); stretching group 72.00 (9.10)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - ; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5; Group 2 Number missing: 9

Protocol outcome 2: Physical function
- Actual outcome: Number of steps climbed at 7 weeks (After 20 sessions); Group 1: mean 112.58 steps (SD 12.11); n=11, Group 2: mean 103.39 steps (SD 30.87); n=10; Comments: Basleine values: exercise group 97.55 (16.36); stretching group 93.00 (36.07)

Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - High, Crossover - Low, Comments - ; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5; Group 2 Number missing: 9 Protocol outcome 3: Discontinuation - Actual outcome: Dicontinuation at 7 weeks (After 20 sessions); Group 1: 5/16, Group 2: 9/19 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - ; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Pain reduction; Psychological distress (depression/anxiety); Use of healthcare services ; Sleep

Study	Chiu 2005[59]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=145)
Countries and setting	Conducted in Hong Kong (China); Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Patients with chronic neck pain (of various intensity) that had lasted longer than 3 months , age 20-70 years, and able to read Chinese. Both genders were included
Exclusion criteria	A previous history of injury to the neck or upper back from T1-T6, an inflammation condition e.g. rheumatoid arthritis, previous surgery to the neck, a history of malignancy, congenital abnormality of the spine, been receiving concurrent treatment e.g. chiropractor or bone setting, contraindication for infrared irradiation e.g. loss of skin sensation, neurologic signs and symptoms e.g. muscle weakness or changes in spinal reflex jerks, other musculoskeletal problems at the same time, acute neck pain with no freedom of movement, received physiotherapy manipulation, or training because of neck pain in the 6 months before examination, or work related injuries
Recruitment/selection of patients	Recruited from physiotherapy outpatient departments
Age, gender and ethnicity	Age - Mean (SD): exercise 43.3 (9.7); control 44.3 (9.8). Gender (M:F): 45/100. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not stated / Unclear 4. chronic widespread pain: Not stated / Unclear Not stated / Unclear
Indirectness of population	No indirectness
Interventions	(n=67) Intervention 1: Strength. The exercise program began with one set (10 minutes) of activation of the deep neck muscles to enhance its ability for active stabilisation of the cervical spine. Then the patient was asked to perform 15 repetitions of flexion and extension of the neck using the MCRU as a warming up exercise for the superficial torque producing muscles. The resistance used during the warm up was set at approximately 20% of the PIS. After the warm up, dynamic training started, which consisted of 3 sets of variable resistance load allowing 8-12 repetitions of full flexion and extension within pain tolerance. A 5 minute rest between session was given. For the initial training session, the dynamic weight load used for each subject was calculated from about 30% of the PIS. The weight load was

	increased approximately 5 % when a set of 12 or more repetitions had been achieved. There were 2 training sessions per week for a period of 6 weeks. Duration 6 weeks. Concurrent medication/care: Infrared irradiation was given to both the exercise group and the control group. The irradiation time was 20 minutes. For the exercise group, irradiation was given before the exercise program. Indirectness: No indirectness (n=78) Intervention 2: Other. The control group received infrared irradiation twice a week for 6 weeks. Duration 6 weeks. Concurrent medication/care: Infrared irradiation was given to both the exercise group and the control group. The irradiation time was 20 minutes. Indirectness: No indirectness
Funding	Academic or government funding (Supported by the Area of Strategic Development Fund of the Hong Kong Polytechnic University and the Hong Kong Health Services Research Committee)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus OTHER

Protocol outcome 1: Pain reduction
- Actual outcome: Pain at End of treatment; Group 1: mean 3 (SD 2.3); n=59, Group 2: mean 3.8 (SD 2.3); n=62; Verbak NRS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 4.6 (1.9); control 4.3 (2.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8; Group 2 Number missing: 16
- Actual outcome: Pain at 6 months; Group 1: mean 3.1 (SD 2.4); n=48, Group 2: mean 3.9 (SD 2.4); n=61; Verbal NRS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 4.6 (1.9); control 4.3 (2.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 19; Group 2 Number missing: 17

Protocol outcome 2: Physical function
- Actual outcome: Disability at End of treatment; Group 1: mean 1 (SD 0.5); n=59, Group 2: mean 1.1 (SD 0.6); n=62; Northwick Park Questionnaire 0-4 Top=High is poor outcome; Comments: Baseline: exercise 1.4 (0.6); control 1.4 (0.5)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 8; Group 2 Number missing: 16
- Actual outcome: Disability at 6 months; Group 1: mean 1 (SD 0.5); n=48, Group 2: mean 1.2 (SD 0.7); n=61; NPQ 0-4 Top=High is poor outcome; Comments: Baseline: exercise 1.4 (0.6); control 1.4 (0.5)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 19; Group 2 Number missing: 17

Protocol outcome 3: Discontinuation

- Actual outcome: Discontinuation at 6 months; Group 1: 19/67, Group 2: 17/78 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; - Actual outcome: Discontinuation at End of treatment; Group 1: 8/67, Group 2: 16/78 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Quality of life ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Cramer 2013[65]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=51)
Countries and setting	Conducted in Germany; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 9 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 18-60 years old and had non-specific neck pain for at least the previous 12 weeks at least 5 days a week. The mean neck pain intensity had to be at least 40mm on a 100mm visual analogue scale, with 0mm meaning no pain and 100mm meaning worst pain imaginable
Exclusion criteria	Neck pain due to specific causes (disc protrusion, radicular syndrome, whiplash, congenital deformity of the spine, spinal canal stenosis, and neoplasm), inflammatory rheumatic disease, active oncologic disease, affective disorder, addiction, and psychosis. Patients who were pregnant or who had had invasive treatment of the spine within the previous 4 weeks or spinal surgery within the previous 12 months were not included. Patients who had physical disability precluding yoga practice and those who had practiced yoga or pilates within the previous 12 weeks were excluded. Patients who had started a new treatment for neck pain within the previous month or were planning to start a new treatment within the next 9 weeks were excluded
Recruitment/selection of patients	Local newspaper announcement
Age, gender and ethnicity	Age - Mean (SD): 47.8 (10.4). Gender (M:F): 9/42. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable Not applicable
Extra comments	Duration of pain (years): 8.1 (6.3)
Indirectness of population	No indirectness
Interventions	(n=25) Intervention 1: Mind-body exercises - Yoga. The yoga group participated in weekly 90 minute yoga classes of 10-15 participants over a period of 9 weeks. The intervention was designed for patients with chronic neck pain without previous experience in yoga. Each class consisted of 8 to 11 yoga postures chosen from a pool of 14 standing, sitting and supine postures, starting with relatively simple postures and succeeding to more complex ones. The focus of

	postures was given on lengthening and strengthening muscles of the neck and shoulder region and to improve stability and posture. Each class started with the mountain pose, a basic standing posture, and ended with the corpse pose, lying supine during a 15 minute guided relaxation. Each class was built up on the previous ones. To enhance alignment and stability and to prevent injury, props, including belts, blocks and blankets were used. Patients were required to practice at home for 10 minutes each day. Patients received a manual describing and depicting 3 basic standing and 3 basic sitting postures. Duration 9 weeks. Concurrent medication/care: Patients in both groups were allowed to continue their usual pain medication and physical activity. They were asked not to change their treatment regimen during the course of the study and to daily record pain medications and other treatments for neck pain in their diaries. Indirectness: No indirectness (n=26) Intervention 2: Mixed modality exercise - Other mixed modality exercise. Participants received a self care manual designed by a large statutory German health insurance company to relieve neck pain and stiffness. The manual described and depicted a staged seated exercise program for the neck and shoulder region. The program began with taking a proper upright sitting posture, followed by stretching exercises for the neck and shoulders. Then, strengthening exercises and isometric exercises for the neck-shoulder region were performed. The program ended with combined stretching and strengthening exercises for the neck-shoulder region using a towel as an aid. Patients were required to practice at home for 10 minutes each day and to record their practice in a diary. Duration 9 weeks. Concurrent medication/care: Patients in both groups were allowed to continue their usual pain medication and physical activity. They were asked not to change their treatment regimen during the course of the study and to daily record pain medications and other treatments for neck pain in their diaries. Indirectness: No indirectness
Funding	Other (Supported by a research Grant from the Karl and Veronica Carstens Foundation, Essen, Germany)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: YOGA versus OTHER MIXED MODALITY EXERCISE

Protocol outcome 1: Pain reduction
- Actual outcome: Pain intensity at end of treatment; Group 1: mean 20.7 (SD 13.6); n=25, Group 2: mean 37.2 (SD 24.4); n=26; VAS 0-100 Top=High is poor outcome; Comments: Baseline: yoga 49.3 (19.2); exercise 40.3 (17.6)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (years): 7.7 (6.6), 8.4 (6.1); pain intensity: 49.3 (19.2); 40.3 (17.6);

Protocol outcome 2: Quality of life
- Actual outcome: QoL mental component at end of treatment; Group 1: mean 50.9 (SD 6.6); n=25, Group 2: mean 45.1 (SD 12.4); n=26; SF36 0-100 Top=High is good outcome; Comments: Baseline: yoga 45.1 (8.9); exercise 45.5 (12.5)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (years): 7.7 (6.6), 8.4 (6.1); pain intensity: 49.3 (19.2); 40.3 (17.6); - Actual outcome: QoL physical component at end of treatment; Group 1: mean 47.3 (SD 7.3); n=25, Group 2: mean 44.2 (SD 10.4); n=26; SF36 0-100 Top=High is good outcome; Comments: Baseline: yoga 42.2 (7.7); exercise 43.8 (8.3) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (years): 7.7 (6.6), 8.4 (6.1); pain intensity: 49.3 (19.2); 40.3 (17.6); Protocol outcome 3: Physical function - Actual outcome: Functional disability at end of treatment; Group 1: mean 20 (SD 9.8); n=25, Group 2: mean 26.2 (SD 15); n=26; Neck disability index 0-50 Top=High is poor outcome; Comments: Baseline: yoga 30 (10); exercise 25.8 (9.8) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (years): 7.7 (6.6), 8.4 (6.1); pain intensity: 49.3 (19.2); 40.3 (17.6); Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at end of treatment; Group 1: 3/25, Group 2: 0/26 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (years): 7.7 (6.6), 8.4 (6.1); pain intensity: 49.3 (19.2); 40.3 (17.6);	
Protocol outcomes not reported by the study	Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Da costa 2005[67]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=80)
Countries and setting	Conducted in Canada; Setting: Not specified; conducted from 1999 to 2002
Line of therapy	Unclear
Duration of study	Intervention + follow up: 3 month intervention plus 9 months follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	None specified
Exclusion criteria	Concomitant diseases which precluded exercise, contraindication to exercise, recent change in medication, regular participation in moderate intensity exercise at the time of study entry.
Recruitment/selection of patients	Recruited through hospitals or community rheumatologists through letters of invitation or newspaper advertisements
Age, gender and ethnicity	Age - Mean (SD): 51.2(9.5 years). Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Disease duration 11(8) years
Indirectness of population	No indirectness
Interventions	(n=39) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. 12 week exercise programme meeting four times with an exercise physiologist. Visits were 90 minutes with 30 minute follow ups. Exercises were individualised for each participant and following the American college of sports medicine guidelines. Exercise focused mainly on aerobic fitness with exercises at heart rate intensity of 60-70% initially then to 75-85% depending on progress, and duration of exercise depended on the intensity although the guidelines suggested individuals should perform 60-120minutes per week. Stretching and strength exercises were also prescribed with the amount depending on the needs of each participant. Participants were provided with a heart rate monitor. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=41) Intervention 2: Other. Usual care control group. Duration 12 weeks. Concurrent medication/care: Not specified.

	Indirectness: No indirectness
Funding	Academic or government funding (The Arthritis Society)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus USUAL CARE Protocol outcome 1: Quality of life - Actual outcome: Fibromyalgia impact questionnaire at 12 months follow up (including 3 month intervention); Group 1: mean -10.1 (SD 16.33); n=28, Group 2: mean -0.024 (SD 12.16); n=33; FIQ 0-33 Top=High is poor outcome; Comments: SD calculated from CIs: E: -16.1 to -4 UC: -4.4 to 3.9 Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11, Reason: Not specified ; Group 2 Number missing: 8, Reason: Not specified	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	De medeiros 2020[69]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=42)
Countries and setting	Conducted in Brazil; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women with FM diagnosis were selected according to the 2010 American College of Rheumatology classification criteria, between 18 and 60 years of age and with pain between 3 and 8 on the Visual Analogue Pain Scale (VAS)
Exclusion criteria	Women with uncontrolled hypertension, decompensated cardiorespiratory disease, history of exercise induced syncope or arrhythmias, decompensated diabetes, severe psychiatric illness, history of regular exercise (at least twice a week) in the last 6 months or any another condition that made the patient unable to perform physical exercise
Recruitment/selection of patients	Participants were recruited from the waiting list of patients of the Clinic Physiotherapy School and Basic Health Units of the city
Age, gender and ethnicity	Age - Mean (SD): Aerobic group: 50.7 (9.7); Pilates group: 45.5 (10.6). Gender (M:F): Women only. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain 5. complex regional pain syndrome:
Indirectness of population	No indirectness
Interventions	(n=21) Intervention 1: Aerobics - Other aerobic exercise. Aquatic aerobic exercise group participants performed aquatic aerobic exercises at a swimming pool. Each session lasted about 40min and was directed by a physiotherapist experienced in water exercises. The program consisted of six main exercises lasting 30min with different intensity exercises moderated by the Borg scale. Two warm-up exercises and two cool-down exercises were performed before and after the program.

	Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=21) Intervention 2: Biomechanical - Pilates. Exercises based on the Mat Pilates method were performed in a group of up to 4 women in a large and comfortable room. Each session lasted about 50min and was led by a physiotherapist experienced in the technique. All the recommendations of the Traditional Pilates method were followed in relation to its six principles to carry out the exercise program, namely: centralization, concentration, control, precision, breathing and flow. Nine exercises were performed for the main muscle groups with progressions each month. The exercises were initially performed in 1 series of 8 repetitions in the first month. Then they were performed in 2 sets of 10 repetitions in the second month. Finally, they were performed in 3 sets of 8 repetitions in the last month. Three Swiss ball relaxation exercises were performed in 1 set of 30s each at the end of each session. . Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Academic or government funding (Partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Master's degree scholarship)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PILATES versus AQUATIC AEROBICS

Protocol outcome 1: Pain reduction
- Actual outcome: Pain VAS at 12 weeks (Post intervention); Group 1: mean 6.2 (SD 1.4); n=21, Group 2: mean 5.6 (SD 2.4); n=21; Visual analogue scale 0-10 Top=High is poor outcome; Comments: Baseline: pilates group 7.5 (1.6); aerobics group 7.5 (1.8)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Quality of life

- Actual outcome: SF36 - role social at 12 weeks (Post intervention); Group 1: mean 64.2 (SD 22.1); n=21, Group 2: mean 53.6 (SD 32.3); n=21; Brazilian version of the Short Form-36 Health Survey (SF-36) 0-100 Top=High is good outcome; Comments: Baseline: pilates 54.2 (21.3); aerobics 49.5 (24.5)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: SF36 - General health status at 12 weeks (Post intervention); Group 1: mean 39 (SD 23.6); n=21, Group 2: mean 37 (SD 22.3); n=21; SF36 0-100 Top=High is good outcome; Comments: Baseline scores: pilates 38.2 (19.2); aerobics 29.7 (22.6)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0

- Actual outcome: SF36 - Vitality at 12 weeks (Post intervention); Group 1: mean 43.8 (SD 19.5); n=21, Group 2: mean 42.6 (SD 17.6); n=21; SF36 0--100 Top=High is good outcome; Comments: Baseline scores: pilates 34.6 (17.5); aerobics 36.2 (18.9)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: SF36 - Functional capacity at 12 weeks (Post intervention); Group 1: mean 43.5 (SD 22); n=21, Group 2: mean 33.9 (SD 18); n=21; SF36 0-100 Top=High is good outcome; Comments: Baseline: pilates 34.0 (17.1); aerobics 28.5 (16.6)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: SF36 - Role physical at 12 weeks (Post intervention); Group 1: mean 36.2 (SD 38.6); n=21, Group 2: mean 21.9 (SD 32.4); n=21; SF36 0-100 Top=High is good outcome; Comments: Baseline group: pilates 23.7 (28.8); aerobics 17.8 (30.7)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: SF36 - Emotional aspects at 12 weeks (Post intervention); Group 1: mean 43.6 (SD 43.6); n=21, Group 2: mean 34.6 (SD 41.2); n=21; SF36 0-100 Top=High is good outcome; Comments: Baseline scores: Pilates 44.4 (46.3); aerobics 22.2 (33.9)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: SF36 - Pain at 12 weeks (Post intervention); Group 1: mean 44.9 (SD 18.4); n=21, Group 2: mean 37.9 (SD 20.3); n=21; SF36 0-100 Top=High is good outcome; Comments: Baseline scores: Pilates 33.3 (17.2); aerobics 29.4 (18.0)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: SF36 - Mental health at 12 weeks (Post intervention); Group 1: mean 65.9 (SD 27.8); n=21, Group 2: mean 55 (SD 19.3); n=21; SF36 0-100 Top=High is good outcome; Comments: Baseline: Pilates 57.5 (21.9); aerobics 47.1 (22.7)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Psychological distress
- Actual outcome: Pain catastrophising at Post intervention; Group 1: mean 2.3 (SD 1.5); n=21, Group 2: mean 2.5 (SD 1.4); n=21; Brazilian version of the Catastrophic Thoughts on Pain Scale (PRCTS) 0-5 Top=High is poor outcome; Comments: Baseline scores: Pilates 2.64 (1.2); aerobics 3.04 (1.2)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 4: Sleep - Actual outcome: Sleep quality at 12 weeks (Post intervention); Group 1: mean 9.9 (SD 3.7); n=21, Group 2: mean 9.5 (SD 3.7); n=21; Pittsburgh Sleep Quality Index 0-21 Top=High is poor outcome; Comments: Baseline: Pilates 10.3 (3.8); aerobics 12.3 (4.1) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 5: Discontinuation - Actual outcome: Dicontinuation at 12 weeks (Post intervention); Group 1: 2/21 Group 2: 4/21 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Pilates group were younger (45.5 years vs 50.7 years), and had higher scores on SF36 general health, emotional aspects and mental health (all ~10 points difference); Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Physical function; Psychological distress (depression/anxiety); Use of healthcare services

Study	**El-gendy 2019[78]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=60)
Countries and setting	Conducted in Egypt; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 4 weeks

Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Mechanical neck pain for at least 3 months with or without shoulder girdle and upper limb unilateral or bilateral symptoms and myofascial trigger points
Exclusion criteria	A positive neurological examination result (presence of positive motor, reflex, or sensory abnormalities indicating spinal root compression) or abnormal neurological signs in the upper limbs relating to nerve entrapment, inflammation, infection, or advanced degeneration due to a systemic rheumatologic disease (e.g., rheumatoid arthritis), congenital malformation, trauma, cerebrovascular abnormalities, cervical spine surgery or stenosis, metabolic or systemic disorders, cancer, known photosensitivity or other illnesses unrelated to neck pain which precluded involvement for practical reasons, pregnancy
Recruitment/selection of patients	Recruited from the Orthopedic Outpatient Clinic, Shoubra General Hospital, Cairo, Egypt
Age, gender and ethnicity	Age - Mean (SD): Manual therapy + exercise group: 33.9 ± 5.51; stretching group 33.65 ± 5.7. Gender (M:F): Not reported. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: 4. chronic widespread pain: 5. complex regional pain syndrome:
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Manual therapy and exercise. Myofascial release therapy plus traditional therapeutic exercises in the form of strength and stretch. Myofascial release therapy comprised superficial stroke massage for 2–3 mins followed by myofascial release technique with pressure with the patient's pain tolerance. At the end of the treatment session, about 2–3-minute surface stroke massage was performed again and the treatment was ended. Each treatment session took 20 minutes; there were 3 sessions per week for 4 weeks. Strength and stretch involved gentle stretching of the pectoral muscle, trapezius muscle, scaleni muscles, levator scapulae muscle, suboccipital muscle, and strengthening consisting of cervical flexion and extension, shoulder retraction, seated upright rowing and push ups if tolerated. Duration 4 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=20) Intervention 2: Flexibility. Strength and stretching protocol as described for the exercise component of the manual therapy and exercise group, 3 sessions per week for 4 weeks. Duration 4 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness

Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EXERCISE versus FLEXIBILITY Protocol outcome 1: Pain reduction - Actual outcome: Pain at 4 weeks (End of intervention); Group 1: mean 3.4 (SD 1.87); n=20, Group 2: mean 4.95 (SD 0.99); n=20; Visual analogue scale 0-10 Top=High is poor outcome; Comments: baseline: manual therapy + exercise 6.65 ± 0.87; strength/stretch 6.5 ± 0.82 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age reported; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 2: Physical function - Actual outcome: Neck disability index at 4 weeks (End of intervention); Group 1: mean 15.35 (SD 5.87); n=20, Group 2: mean 21.8 (SD 4.03); n=20; Neck disability index 0-50 Top=High is poor outcome; Comments: Baseline values: manual therapy + exercise 24.85 ± 3.82; exercise 24.7 ± 3.78 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age reported; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 3: Discontinuation - Actual outcome: Dicontinuation at 4 weeks (End of intervention); Group 1: 0/20 Group 2: 0/20 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age reported; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life; Psychological distress (depression/anxiety); Use of healthcare services; Sleep

Study	Ericsson 2016[80]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=34)
Countries and setting	Conducted in Sweden; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Met the ACR criteria for chronic widespread pain, having experienced pain for at least 3 months
Exclusion criteria	Inability to understand Swedish, severe psychiatric or somatic disorders, or having participated in resistance exercise or pool exercise at a physical therapy clinic during the preceding six months.
Recruitment/selection of patients	5 primary health care centres in western Sweden
Age, gender and ethnicity	Age - Mean (SD): 59(8.1) years. Gender (M:F): All male. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Duration of pain 5.3(2.3) years
Indirectness of population	No indirectness
Interventions	(n=17) Intervention 1: Aerobics - Swimming. Pool exercise programme. 50 minute sessions in groups of 6-8 participants twice a week for 12 weeks, supervised by a physiotherapist. Sessions included aerobic exercise with endurance, strength, flexibility, coordination and relaxation. patients were instructed to exercise at their own rhythm and modify exercises with respect to thresholds of pain and fatigue. They were encouraged to increase intensity and resistance with or without water equipment, based on the rate of perceived exertion on the Borg scale. Duration 12 weeks. Concurrent medication/care: 41% were taking analgesics/NSAIDs, 59% were taking psychotropic. Indirectness: No indirectness (n=17) Intervention 2: Strength. Twice a week sessions for 12 weeks with free weights and resistance machines in groups of 8-10 patients, supervised by a physiotherapist. The sessions lasted approximately 1 hour and include exercises for multiple main muscle groups. Load was increased from 40% to 80% of one repetition maximum

	established at baseline. Participants performed 3 sets with 15-20 repetitions of each exercise, when the load increased they performed 2 sets but fewer repetitions. All sessions started with 10 minute warm up on an ergometer bicycle. Duration 12 weeks. Concurrent medication/care: 71% were taking analgesics/NSAIDs, 24% were taking psychotropics. Indirectness: No indirectness
Funding	Academic or government funding (Fyrbodal research development council and the health care committee of the regional executive board, Vastra Gotaland, Sweden.)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: SWIMMING versus STRENGTH

Protocol outcome 1: Pain reduction
- Actual outcome: FIQ pain score at 12 weeks; Group 1: mean -2.5 (SD 25.3); n=14, Group 2: mean -3.3 (SD 13.4); n=12; FIQ pain scale 0-100 Top=High is poor outcome; Comments: Baseline: 53.4(28.3); 69.5(17.7)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Different number on pharmacological treatment at baseline; Group 1 Number missing: 3, Reason: Time restriction, increased pain, unknown reason ; Group 2 Number missing: 5, Reason: Surgery, cardiac infarction, time restriction, infection, car accident

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 physical component summary score at 12 weeks; Group 1: mean 4.9 (SD 6.2); n=14, Group 2: mean 2.2 (SD 5.8); n=12; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 33.8(9.8); 36.7(6.9)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Different number on pharmacological treatment at baseline; Group 1 Number missing: 3, Reason: Time restriction, increased pain, unknown reason ; Group 2 Number missing: 5, Reason: Surgery, cardiac infarction, time restriction, infection, car accident
- Actual outcome: SF-36 mental component summary score at 12 weeks; Group 1: mean 1.9 (SD 8.1); n=14, Group 2: mean 0.5 (SD 9.1); n=12; SF-36 subscale 0-100 Top=High is poor outcome; Comments: Baseline: 46(14.1); 35.6(13.5)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Different number on pharmacological treatment at baseline; Group 1 Number missing: 3, Reason: Time restriction, increased pain, unknown reason ; Group 2 Number missing: 5, Reason: Surgery, cardiac infarction, time restriction, infection, car accident

Protocol outcome 3: Physical function
- Actual outcome: Multidimensional fatigue inventory-20 reduced activity subscale at 12 weeks; Group 1: mean -0.3 (SD 3.5); n=14, Group 2: mean -1.3 (SD 2.1); n=12; MFI subscale 4-20 Top=High is poor outcome; Comments: Baseline: 11.8(4); 13.6(5.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Different number on pharmacological treatment at baseline; Group 1 Number missing: 3, Reason: Time restriction, increased pain, unknown reason ; Group 2 Number missing: 5, Reason: Surgery, cardiac infarction, time restriction, infection, car accident

Protocol outcome 4: Psychological distress (depression/anxiety) - Actual outcome: Hospital anxiety and depression scale anxiety subscale at 12 weeks; Group 1: mean -1.6 (SD 2.2); n=14, Group 2: mean -0.8 (SD 2.5); n=12; HADS:A 0-21 Top=High is poor outcome; Comments: Baseline: 8.4(5.7); 8.3(5.3) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Different number on pharmacological treatment at baseline; Group 1 Number missing: 3, Reason: Time restriction, increased pain, unknown reason ; Group 2 Number missing: 5, Reason: Surgery, cardiac infarction, time restriction, infection, car accident - Actual outcome: Hospital anxiety and depression scale depression subscale at 12 weeks; Group 1: mean -0.1 (SD 2.2); n=14, Group 2: mean 0.1 (SD 2.1); n=12; HADS:D 0-21 Top=High is poor outcome; Comments: Baseline: 5.4(5.4); 7.1(4) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Different number on pharmacological treatment at baseline; Group 1 Number missing: 3, Reason: Time restriction, increased pain, unknown reason ; Group 2 Number missing: 5, Reason: Surgery, cardiac infarction, time restriction, infection, car accident Protocol outcome 5: Discontinuation - Actual outcome: Discontinuation at 12 weeks; Group 1: 3/17, Group 2: 5/17; Comments: Due to time restrictions, increased pain, surgery, cardiac infarction, infection and car accident. Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness; Baseline details: Different number on pharmacological treatment at baseline;	
Protocol outcomes not reported by the study	Use of healthcare services ; Sleep

Study	Ericsson 2016[81]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=130)
Countries and setting	Conducted in Sweden; Setting: Multiple centres across Sweden
Line of therapy	Unclear
Duration of study	15 week intervention
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 20–65 years, meeting the American College of Rheumatology (ACR) 1990 classification criteria for FM
Exclusion criteria	Other severe somatic or psychiatric disorders, participation in a rehabilitation program within the past year, or inability to understand Swedish.
Recruitment/selection of patients	Recruited by newspaper advertisement in the local newspapers of three cities in Sweden
Age, gender and ethnicity	Age - Range: 22 to 64 years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Duration of pain not specified
Indirectness of population	No indirectness
Interventions	(n=67) Intervention 1: Strength. Exercise sessions were twice a week for 15 weeks at physiotherapy premises and at a local gym and were supervised by experienced physiotherapists. The exercise program was standardized and performed in groups of five to seven participants but the load was adjusted individually. The exercise session started with 10 minutes of warm up followed by 50 minutes of resistance exercises focused on large muscle groups in all four extremities and trunk. The resistance exercise was initiated at 40 % of 1 repetition maximum (RM) and progressed up to 80 % of 1 RM during the 15 weeks. Possibilities for progression of loads were evaluated every 3–4 weeks. Forty-two participants (62.7 %) in the resistance exercise group reached exercise loads of 80 % of 1 RM while seven participants (10.4 %) reached exercise loads of 60 % of 1 RMv. Duration 15 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=63) Intervention 2: Mixed modality exercise - Other mixed modality exercise. Relaxation therapy, which was

	performed twice a week for 15 weeks, guided by experienced physiotherapists and conducted at physiotherapy premises in groups of five to eight participants. It was performed as autogenic training. which refers to a series of mental exercises including autosuggestion and relaxation. The relaxation therapy lasted for approximately 25 minutes, followed by stretching exercises. Duration 15 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Swedish Rheumatism Association, the Swedish Research Council, the Health and Medical Care Executive Board of Västra Götaland Region, ALF-LUA at Sahlgrenska University Hospital, Stockholm and Östergötland County Councils (ALF), and AFA Insurance and Gothenburg Center for Person Centered Care (GPCC))

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus RELAXATION AND STRETCHING COMBINATION

Protocol outcome 1: Pain reduction
- Actual outcome: Pain catastrophising scale total scores at 15 weeks; Group 1: mean -2.7 (SD 7.6); n=56, Group 2: mean -2.8 (SD 7.9); n=49; PCS 0-54 Top=High is poor outcome; Comments: Baseline: 19.4(10); 20.3(11.9)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11; Group 2 Number missing: 14
- Actual outcome: VAS at 15 weeks; Group 1: mean 38.6 (SD 25.2); n=56, Group 2: mean 53.4 (SD 20); n=49; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 49.3(23.9); 52.4(18.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11; Group 2 Number missing: 14

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 physical component summary score at 15 weeks; Group 1: mean 34.5 (SD 9.1); n=56, Group 2: mean 30.7 (SD 8.3); n=49; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 31.2(7.0); 29.9(8.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11; Group 2 Number missing: 14
- Actual outcome: SF-36 mental component summary score at 15 weeks; Group 1: mean 42 (SD 12.6); n=56, Group 2: mean 38.8 (SD 12.9); n=49; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 37.7(12.2); 39.6(12.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11, Reason: Due to increased pain, personal reason, no contact; Group 2 Number missing: 14, Reason: Due to personal reasons, no contact

Protocol outcome 3: Physical function
- Actual outcome: 6 minute walking test (metres) at 15 weeks; Group 1: mean 579.7 (SD 73.7); n=5656, Group 2: mean 533.9 (SD 73.1); n=49; Comments: Baseline:

556.6(75.1); 540.7(64.5) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11; Group 2 Number missing: 14 Protocol outcome 4: Psychological distress (depression/anxiety) - Actual outcome: Hospital anxiety and depression scale depression subscale at 15 weeks; Group 1: mean -0.7 (SD 3.7); n=56, Group 2: mean 0.3 (SD 2.8); n=48; HADS subscale 0-21 Top=High is poor outcome; Comments: Baseline: 7.0 (3.9); 6.7(3.5) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11; Group 2 Number missing: 14 - Actual outcome: Hospital anxiety and depression scale anxiety subscale at 15 weeks; Group 1: mean -0.3 (SD 3.6); n=56, Group 2: mean 0.5 (SD 2.7); n=49; HADS subscale 0-23 Top=High is poor outcome; Comments: Baseline: 7.9 (4.7); 8(4.5) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11; Group 2 Number missing: 14 Protocol outcome 5: Sleep - Actual outcome: Pittsburgh Sleep Quality Index, total score at 15 weeks; Group 1: mean -0.6 (SD 3.4); n=56, Group 2: mean 0.5 (SD 3); n=49; PSQI total scores 0-21 Top=High is poor outcome; Comments: Baseline: 10.9 (4.3); 10.8(4) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 11, Reason: Due to increased pain, personal reason, no contact; Group 2 Number missing: 14, Reason: Due to personal reasons, no contact Protocol outcome 6: Discontinuation - Actual outcome: Discontinuation at 15 weeks; Group 1: 11/67, Group 2: 14/63; Comments: Due to increased pain, personal reasons and no contact Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness, Comments: Unclear if dropped out of study or intervention; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Use of healthcare services

Study	Espi-lopez 2016[83]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=22)
Countries and setting	Conducted in Spain; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	The inclusion criteria were being aged between 30 and 80 years and meeting the ACR 2010 criteria for FMS. Additional inclusion criteria from the clinical trial registry: Mett some or several of the following characteristics: depression, anxiety, muscle pain, fatigue, sleep disturbance. May have limited mobility as long as it is caused by fibromyalgia.
Exclusion criteria	The exclusion criteria included medical contraindication for physical activity, deafness or limited hearing, vestibular disorders that compromise balance, very low vision or blind people, psychotic disorder, cognitive disabilities, decompensation or changes in medication.
Recruitment/selection of patients	Patients were belonged to the 'Association of People Affected by Fibromyalgis of Valencia'
Age, gender and ethnicity	Age - Mean (SD): 53.6(8.1) years. Gender (M:F): 1:21. Ethnicity: Not stated
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=13) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. Low-impact aerobic exercise with low impact strengthening exercises. Two sessions per week. Each session consisted of 60min and was divided into three parts: warm up (15 min); games, group dynamics and aerobics (30 min); and cool down with stretching for 15 min. The warm up consisted of combined low impact aerobic exercises, free range of motion exercises of limbs and spine, and coordination exercises plus stretching. This was followed by active low load resistance exercises involving arms and legs, followed by a circuit of coordination and agility exercises and then low-impact strengthening exercises of the trunk. This was followed by a cool down with stretches. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness

	(n=9) Intervention 2: Other. Control group: no intervention. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus NO TREATMENT Protocol outcome 1: Quality of life - Actual outcome: Fibromyalgia impact questionnaire at 8 weeks; Group 1: mean 59 (SD 15.55); n=13, Group 2: mean 58.72 (SD 19.42); n=9; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: 63.48(14.3); 59.53(20.96) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Health problems and personal problems; Group 2 Number missing: 1, Reason: Inability to attend assessment sessions Protocol outcome 2: Psychological distress (depression/anxiety) - Actual outcome: Beck depression scale at 8 weeks; Group 1: mean 17.69 (SD 11.62); n=13, Group 2: mean 14.11 (SD 10.15); n=9; BDI 0-30 Top=High is poor outcome; Comments: Baseline (downgraded for difference at baseline): 22.23(11.25); 17.89(9.29) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Health problems and personal problems; Group 2 Number missing: 1, Reason: Inability to attend assessment sessions Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at 8 weeks; Group 1: 5/13, Group 2: 1/9 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness, Comments: Unclear if participants dropped out of intervention or study; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Use of healthcare services ; Sleep

Study	Etnier 2009[84]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=16)
Countries and setting	Conducted in USA; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 18 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Over 18 years of age, currently inactive (defined as participating in exercise one day or less per week), and must satisfy the American College of Sport Medicine criteria for the safe conduct of exercise. Must also be willing to be assigned to either treatment condition
Exclusion criteria	Not reported
Recruitment/selection of patients	Referred by local rheumatologists
Age, gender and ethnicity	Age - Mean (SD): not reported. Gender (M:F): Women only. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Extra comments	Duration of pain not reported, but most participants reported having symptoms as teenagers and received a medical diagnosis within the last 1-10 years
Indirectness of population	No indirectness
Interventions	(n=8) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. The exercise sessions were 60 minutes in duration 3 days a week. During the sessions, participants walked, performed light resistance exercises, and performed static bridging and stretching exercises. All sessions were conducted and directly supervised by one of the authors. In terms of the walking portion, participants were encouraged to walk a comfortable/brisk pace (55-65% of maximal heart rate reserve) for 15 minutes. Over the course of the intervention, they were encouraged to try to walk a greater distance in the 15 minute period and used this as a self-measure of aerobic fitness. In terms of the light resistance exercises, participants moved through an 8 station light resistance exercise circuit. When subjects were able to easily complete the required number of repetitions for a certain exercise, resistance was increased by 1 pound. Often, this caused participants to reduce the number of repetitions for a short time followed by slowly working back to the

	required number. Static-bridging exercises require that the exerciser support her body (holding the body very still) in various positions to increase core (abdominal, back and pelvic), muscle strength/endurance. Usually 10 repetitions of approximately 3 seconds were completed in each session. Duration 18 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=8) Intervention 2: No treatment. No treatment control condition. Duration 18 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Other (Funding was provided by the University of North Carolina Greensboro Office of Research and Public/Private Sector Partnerships)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus NO TREATMENT

Protocol outcome 1: Quality of life
- Actual outcome: FMS symptoms at end of treatment; Group 1: mean 41.4 (SD 18.19); n=8, Group 2: mean 66.58 (SD 18.19); n=8; FIQ 0-100 Top=High is poor outcome; Comments: Baseline not reported
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 2: Physical function
- Actual outcome: Quarter mile walk test at end of treatment; Group 1: mean 282.85 seconds (SD 26.42); n=8, Group 2: mean 320.15 seconds (SD 26.42); n=8
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 3: Psychological distress (depression/anxiety)
- Actual outcome: depression at end of treatment; Group 1: mean 19.97 (SD 8.91); n=8, Group 2: mean 28.91 (SD 8.91); n=8; The Centre for Epidemiological Scale - Depression 0-60 Top=High is poor outcome; Comments: Baseline not reported
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 4: Discontinuation
- Actual outcome: discontinuation at end of treatment; Group 1: 0/8, Group 2: 0/8
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Pain reduction ; Use of healthcare services ; Sleep	
Protocol outcomes not reported by the study	

Study	Evans 2002[85]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=191)
Countries and setting	Conducted in USA; Setting: University and Neck and Back Clinic
Line of therapy	Unclear
Duration of study	Intervention + follow up: 11 weeks + 24 months
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Between 20-65 years of age, primary complaint of mechanical neck pain that had lasted for 12 weeks or more
Exclusion criteria	Neck pain referred from peripheral joints of viscera, severe osteopenia, progressive neurologic deficits, vascular disease of the neck or upper extremity, significant infectious disease or other severe disability health conditions, previous cervical spine surgery, current or pending mitigation, inability to work because of neck pain, spinal manipulative therapy or exercise in the 3 months before study entry, or concurrent treatment for neck pain by other health care providers
Recruitment/selection of patients	Newspaper advertisements
Age, gender and ethnicity	Age - Mean (SD): combined group 45 (10.5); manual therapy group 44.3 (11). Gender (M:F): 53/75. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable 5. complex regional pain syndrome: Not applicable
Extra comments	Duration of pain (median years, range): combined 6.5 (0.3-29); manual therapy 5.5 (0.4-4.34)
Indirectness of population	No indirectness
Interventions	(n=64) Intervention 1: Manual therapy and exercise. Spinal manipulation combined with rehabilitative exercise. Spinal manipulation treatment included manual spinal manipulation with light soft tissue massage as facilitate the spinal manipulative therapy. Rehabilitative exercise began each session with a warm up on a stationary bike with arm levers and light stretching, followed by upper body strengthening exercises including push-ups and dumbbell shoulder exercises. Dynamic neck extension, flexion, and rotation exercises were performed with the patient lying on a therapy table wearing headgear with variable weight attachments (1.25 to 10 lbs.) guided by a simple pulley system attached to

	a physical therapy table. Beginning weights were determined by baseline strength performance and were increased gradually during the treatment phase. Each session was 1 hour and there were 20 sessions. Duration 11 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=63) Intervention 2: Strength. Each appointment began with a warm up of stretching and aerobic exercise using a dual action stationary bike, followed by strengthening exercises of the shoulders and upper back using variable resistance equipment. Neck strengthening exercises were performed on the MedX variable resistance, cervical extension, and rotation machines. Patients were stabilized with torso restraints to isolate and specifically exercise the cervical musculature. They were encouraged to perform repetitions to volitional muscle fatigue (maximum 20 reps) even if the pain was exacerbated, and resistance was increased periodically. Duration 11 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=64) Intervention 3: Physical therapy - Manual therapy. Patients received the same spinal manipulation treatment as in the combined treatment group. Duration 11 weeks. Concurrent medication/care: Patients were also given 45 minutes of micronutrient therapy (sham) to minimize the effects of attention bias. Indirectness: No indirectness
Funding	Other (Foundation funds were received)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EXERCISE versus STRENGTH

Protocol outcome 1: Pain reduction
- Actual outcome: Neck pain over the past week at 3 months; Group 1: mean 2.9 (SD 2.1); n=51, Group 2: mean 2.4 (SD 1.8); n=44; NRS 0-10 Top=High is poor outcome; Comments: Baseline: combined 5.6 (1.5); exercise 5.6 (1.5)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, exercise 5, 0.3-24); Group 1 Number missing: 13; Group 2 Number missing: 19
- Actual outcome: Neck pain over the past week at 24 months; Group 1: mean 3.4 (SD 2.4); n=51, Group 2: mean 3.4 (SD 2.4); n=44; NRS 0-10 Top=High is poor outcome; Comments: Baseline: combined 5.6 (1.5); exercise 5.6 (1.5)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, exercise 5, 0.3-24); Group 1 Number missing: 13; Group 2 Number missing: 19

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at 3 months; Group 1: mean 13.6 (SD 10.2); n=51, Group 2: mean 12.8 (SD 10.2); n=44; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: combined 26.3 (8.4); exercise 26.4 (10.2)

Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, exercise 5, 0.3-24); Group 1 Number missing: 13; Group 2 Number missing: 19
- Actual outcome: Neck disability at 24 months; Group 1: mean 15.6 (SD 11.8); n=51, Group 2: mean 16.6 (SD 12.4); n=44; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: combined 26.3 (8.4); exercise 26.4 (10.2)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, exercise 5, 0.3-24); Group 1 Number missing: 13; Group 2 Number missing: 19

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at 24 months; Group 1: 13/64, Group 2: 19/63
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, exercise 5, 0.3-24);

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EXERCISE versus MANUAL THERAPY

Protocol outcome 1: Pain reduction
- Actual outcome: Neck pain over the past week at 3 months; Group 1: mean 2.9 (SD 2.1); n=51, Group 2: mean 3.7 (SD 2.3); n=50; NRS 0-10 Top=High is poor outcome; Comments: Baseline: combined 5.6 (1.5); manual therapy 5.6 (1.4)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, manual 5.5 0.4-34); Group 1 Number missing: 13; Group 2 Number missing: 14
- Actual outcome: Neck pain over the past week at 24 months; Group 1: mean 3.4 (SD 2.4); n=51, Group 2: mean 3.9 (SD 2.3); n=50; NRS 0-10 Top=High is poor outcome; Comments: Baseline: combined 5.6 (1.5); manual therapy 5.6 (1.4)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, manual 5.5 0.4-34); Group 1 Number missing: 13; Group 2 Number missing: 14

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at 3 months; Group 1: mean 13.6 (SD 10.2); n=51, Group 2: mean 18.7 (SD 13); n=50; Neck disability index 0-100 Top=High is good outcome; Comments: Baseline: combined 26.3 (8.4); manual therapy 27.9 (10.2)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, manual 5.5 0.4-34); Group 1 Number missing: 13; Group 2 Number missing: 14
- Actual outcome: Neck disability at 24 months; Group 1: mean 15.6 (SD 11.8); n=51, Group 2: mean 20.5 (SD 13.5); n=50; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: combined 26.3 (8.4); manual therapy 27.9 (10.2)

Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (combined 6.5 0.3-29, manual 5.5 0.4-34); Group 1 Number missing: 13; Group 2 Number missing: 14

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at 24 months; Group 1: 13/64, Group 2: 14/64
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus MANUAL THERAPY

Protocol outcome 1: Pain reduction
- Actual outcome: Neck pain over the past week at 3 months; Group 1: mean 2.4 (SD 1.8); n=44, Group 2: mean 3.7 (SD 2.3); n=50; NRS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 5.6 (1.5); manual therapy 5.6 (1.4)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 19; Group 2 Number missing: 14
- Actual outcome: Neck pain over the past week at 24 months; Group 1: mean 3.4 (SD 2.4); n=44, Group 2: mean 3.9 (SD 2.3); n=50; NRS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 5.6 (1.5); manual therapy 5.6 (1.4)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 19; Group 2 Number missing: 14

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at 3 months; Group 1: mean 12.8 (SD 10.2); n=44, Group 2: mean 18.7 (SD 13); n=50; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: exercise 26.4 (10.2); manual therapy 27.9 (10.2)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 19; Group 2 Number missing: 14
- Actual outcome: Neck disability at 24 months; Group 1: mean 16.6 (SD 12.4); n=44, Group 2: mean 20.5 (SD 13.5); n=50; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: exercise 26.4 (10.2); manual therapy 27.9 (10.2)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 19; Group 2 Number missing: 14

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at 24 months; Group 1: 19/63, Group 2: 14/64
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep	
Protocol outcomes not reported by the study	

Study	Evans 2012[86]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=180)
Countries and setting	Conducted in USA; Setting: Wolfe-Harris center for clinical studies, Minnesota
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks plus 52 weeks follow up
Method of assessment of guideline condition	Partially adequate method of assessment/diagnosis: Grade I or II classification according to the Neck Pain Task Force
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	18 to 65 years old, primary complaint of chronic nonspecific neck pain for at least 12 weeks, with a neck pain score greater than 3 (on 0-10 scale)
Exclusion criteria	Previous cervical spine conditions or surgery, neck pain referred from other joints of viscera, any neurological, musculoskeletal conditions or cardiac disease that require medical treatment or could cause pain, pregnancy, substance abuse, or those with ongoing treatment of neck pain by other health care providers.
Recruitment/selection of patients	Newspaper adverts, posters, mass mailings.
Age, gender and ethnicity	Age - Mean (SD): Mean age 46.3(10.7). Gender (M:F): 75:195. Ethnicity: Not specified
Further population details	Subgroup: people with chronic primary musculoskeletal pain (Chronic cervical pain)
Extra comments	Duration of pain 9.4(9.1) years
Indirectness of population	No indirectness
Interventions	(n=89) Intervention 1: Strength. Predominantly upper body and neck exercises that were partially individualised in terms of intensity, according to the participants' abilities. One-on-one supervision in 20 1 hour sessions. The main focus was cervical strengthening exercises using low-tech methods performed with the patient lying on a therapy table, wearing headgear with variable weight attachments. 3 sets of 15-25 repetitions were conducted. There was also light aerobic warm up (5 minutes) and stretching before and after strengthening. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=91) Intervention 2: Manual therapy and exercise. Identical exercises as strength intervention (as described) which was preceded by a 15-20 minute session with a licensed chiropractor who administered spinal manipulation therapy. Sessions focused mainly on manual manipulation to the cervical and thoracic spines using high velocity, low amplitude

	pressure applied to the joints. Up to 5 minutes of light soft tissue massage was also used. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Federal funs)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND STRENGTH versus STRENGTH

Protocol outcome 1: Pain reduction
- Actual outcome: VAS pain scores at 12 weeks; Group 1: mean 2.3 (SD 1.8); n=91, Group 2: mean 2.6 (SD 1.9); n=89; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.6(1.4); 5.7(1.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns
- Actual outcome: VAS pain scores at 52 weeks; Group 1: mean 3.4 (SD 2.3); n=91, Group 2: mean 3.1 (SD 2.2); n=89; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.6(1.4); 5.7(1.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 physical component summary score at 52 weeks; Group 1: mean 50 (SD 6.4); n=91, Group 2: mean 49.8 (SD 7.2); n=89; SF-36 summary score 0-100 Top=High is good outcome; Comments: Baseline:45.7(6.6); 46.6(6.8)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns
- Actual outcome: SF-36 physical component summary score at 12 weeks; Group 1: mean 50.7 (SD 6.7); n=91, Group 2: mean 50.1 (SD 6.6); n=89; SF-36 summary score 0-100 Top=High is good outcome; Comments: Baseline: 45.7(6.6); 46.6(6.8)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns
- Actual outcome: SF-36 mental component summary score at 12 weeks; Group 1: mean 53.9 (SD 9.8); n=91, Group 2: mean 54.6 (SD 9.7); n=89; SF-36 summary score 0-100 Top=High is good outcome; Comments: Baseline: 51.5(9.9); 53.7(9.2)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns

- Actual outcome: SF-36 mental component summary score at 52 weeks; Group 1: mean 53 (SD 8.9); n=91, Group 2: mean 54.8 (SD 8.5); n=89; SF-36 summary score 0-100 Top=High is good outcome; Comments: Baseline: 51.5(9.9); 53.7(9.2) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns Protocol outcome 3: Physical function - Actual outcome: Neck disability index at 52 weeks; Group 1: mean 18 (SD 11.3); n=91, Group 2: mean 17.5 (SD 13.3); n=89; NDI 0-50? Top=High is poor outcome; Comments: Baseline: 27.8(9); 26.1(9.8) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 13, Reason: Didn't want to participate, time commitments, other health concerns - Actual outcome: Neck disability index at 12 weeks; Group 1: mean 14.5 (SD 9.5); n=91, Group 2: mean 16 (SD 11.3); n=89; NDI 0-50? Top=High is poor outcome; Comments: Baseline: 27.8(9); 26.1(9.8) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns; Group 2 Number missing: 5, Reason: Didn't want to participate, time commitments, other health concerns Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation of intervention at 12 weeks; Group 1: 9/91, Group 2: 5/89 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Psychological distress (depression/anxiety); Use of healthcare services ; Sleep

Study	**Falla 2013[90]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=46)
Countries and setting	Conducted in Denmark; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks

Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women, between 18-50 years of age, suffering from persistent neck pain and disability limiting their daily physical activity for at least 1 year
Exclusion criteria	Trauma induced neck pain, neck pain attributed to an inflammatory or infectious condition, neurological signs, previous cervical spine surgery, exercise therapy within 3 months prior to entry into the study, current treatment for neck pain from health care providers or pregnancy
Recruitment/selection of patients	Referral from a Pain Management Centre, general practitioners or through general advertising in the popular press
Age, gender and ethnicity	Age - Mean (SD): exercise 39.1 (8.7); control 38.6 (9). Gender (M:F): Women only. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable 5. complex regional pain syndrome: Not applicable
Extra comments	Duration of pain (years): exercise 10 (7.4); control 8.4 (5.1)
Indirectness of population	No indirectness
Interventions	(n=23) Intervention 1: Strength. An 8 week progressive exercise programme for the neck flexors and extensor muscles. Participants received personal instruction and supervision by a physiotherapist for 30 minutes once per week for 8 weeks. The therapist examined the exercises and progressed the participant if appropriate. The programme consisted of 2 stages. The first stage was 6 weeks duration. The principal exercise task during this period was incremental cranio-cervical flexion in a relaxed supine lying position. The exercise targets the deep flexors of the upper cervical region, the longus capitis and colli, rather than the superficial flexors, sternocleidomastoid and anterior scalene muscles. The patients were instructed to perform and hold progressively inner range positions of cranio-cervical flexion. Patients were guided by a pressure unit. Patients also performed cranio cervical extension, flexion and rotation in a prove on elbows position while maintaining the cervical spine in a neutral position, to target the cranio-cervical extensors of the cervical spine. The second stage was 2 weeks and involved higher load exercise with head weight as the load. During this stage, participants performed up to 15 repetitions of a head lift for flexors, which was performed in supine, and neck extension for the extensor group, which was performed in 4 point kneeling. For the head lift, the patients were instructed to perform cranio-cervical flexion followed by cervical flexion to just lift the head from the bed. For the neck extension exercise, the patients were instructed to keep their cranio-cervical region in a mid-position while they extended the cervical region. For the higher load exercises, all repetitions were performed over a 3 second period with no rests in between repetitions. Participants practiced twice per day, and the programme was 10-20 minutes/day. Duration 8 weeks. Concurrent medication/care: Not reported

	(n=23) Intervention 2: Usual care. The control group did not receive any intervention, however they patients were not asked to refrain from seeking treatment. Duration 8 weeks. Concurrent medication/care: Not reported
Funding	-- (Supported by the Danish Medical Research Council and Gigforeningen Denmark)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: Average pain intensity over the last 4 weeks at end of treatment; Group 1: mean -1.7 (SD 2.2); n=22, Group 2: mean -0.3 (SD 2.1); n=20; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 5.3 (2.8); control 5.1 (2)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (10 vs 8.4 years); quality of life (52.3 vs 68.6); Group 1 Number missing: 1; Group 2 Number missing: 3

Protocol outcome 2: Quality of life
- Actual outcome: SF36 total at end of treatment; Group 1: mean 8.3 (SD 15.2); n=22, Group 2: mean 2.6 (SD 11.5); n=20; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 52.3 (17.8); control 68.6 (17.0)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (10 vs 8.4 years); quality of life (52.3 vs 68.6); Group 1 Number missing: 1; Group 2 Number missing: 3
- Actual outcome: SF36 physical component at end of treatment; Group 1: mean 9.6 (SD 15); n=22, Group 2: mean 2 (SD 10.8); n=20; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 46.8 (16.5); control 63.7 (18.5)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (10 vs 8.4 years); quality of life (52.3 vs 68.6); Group 1 Number missing: 1; Group 2 Number missing: 3
- Actual outcome: SF36 mental component at end of treatment; Group 1: mean 6.7 (SD 16.4); n=22, Group 2: mean 2.5 (SD 14.2); n=20; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 55.7 (20.6); control 70.3 (15.5)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (10 vs 8.4 years); quality of life (52.3 vs 68.6); Group 1 Number missing: 1; Group 2 Number missing: 3

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at end of treatment; Group 1: mean -4.1 (SD 4.8); n=22, Group 2: mean -1 (SD 4.4); n=20; Neck Disability Index 0-50 Top=High is poor outcome; Comments: Baseline: exercise 18.2 (7.4); control 17.5 (6.3)

Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (10 vs 8.4 years); quality of life (52.3 vs 68.6); Group 1 Number missing: 1; Group 2 Number missing: 3 Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at end of treatment; Group 1: 1/23, Group 2: 3/23 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain (10 vs 8.4 years); quality of life (52.3 vs 68.6);	
Protocol outcomes not reported by the study	Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Gallego Izquierdo 2016 [97]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=28)
Countries and setting	Conducted in Spain; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	History of non-specific neck pain for greater than 3 months, Inclusion criteria were: age between 18 and 55 years, score ≤ 15/50 on the Neck Disability Index (NDI), showing signs of cervical movement control dysfunction and manual physical examination revealing muscle tenderness. A cervical movement control dysfunction was defined as the presence of aberrant or uncontrolled movements of the cervical spine observed during prescribed active movements of the neck and/or upper limb.
Exclusion criteria	Subjects were excluded if they had vascular, neoplastic or vestibular disease, a diagnosis of fibromyalgia or rheumatoid arthritis, or any medical condition that prevented exercise.
Recruitment/selection of patients	Via advertisements in 2014
Age, gender and ethnicity	Age - Mean (SD): 29.2(7.2) years. Gender (M:F): 10:18. Ethnicity: Not specified
Further population details	Chronic primary musculoskeletal pain: chronic primary cervical pain
Extra comments	Duration of pain not specified (more than 3 months)
Indirectness of population	No indirectness
Interventions	(n=14) Intervention 1: Strength. Cranio-cervical flexion training. Low load training of flexor muscles to target deep flexors and aiming to minimize the activation of the superficial flexor muscles. Initially, patients were taught to perform the CCF movement slowly and in a controlled manner in a supine position, with the head and neck in a neutral position. Once the correct CCF motion was achieved, subjects began to hold progressively increasing ranges of CCF using feedback from an air-filled pressure sensor (StabilizerTM, Chattanooga Group Inc., Tennessee, USA) placed behind the neck. The patient initially performed CCF to sequentially reach 5 pressure targets in 2 mmHg increments from a baseline of 20 mmHg to the final level of 30 mmHg. The physiotherapist identified the target level that the patient could hold steadily for 5 s without resorting to retraction, without dominant use of the superficial neck flexor muscles,

	and without a quick, jerky cranio-cervical flexion movement. Training commenced at this target level. For each target level, the contraction duration was increased to 10 s, and the subject trained to perform 10 repetitions with brief rest periods between each contraction (~3–5 s). Once one set of 10 repetitions of 10 s was achieved at one target level, the exercise was progressed to train at the next target level up to the final target of 10 repetitions of 10 s at 30 mmHg. The exercise load prescribed to each patient was based on their assessment performance. Participants were taught to do exercises at home without biofeedback. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=14) Intervention 2: Proprioceptive - Proprioceptive exercise. Patients trained cervical proprioception following the protocol described by Revel et al. This regime consisted of exercises of head relocation, eye-follow, gaze stability and eye-head coordination. For head relocation exercises, subjects started in a sitting position, with a laser attached to a helmet at the apex of their head, and a target located at eye level on a wall 90 cm away. This was established as the natural head posture. Subjects then practiced relocating their head to the natural head posture after active neck movements, first with eyes open using feedback from the laser attached to their head, then with pupillary glasses preventing pupillary excursion, and finally with their eyes closed. All active movements of the cervical spine (flexion, extension, rotation, lateral flexion) were performed. Oculomotor exercises were progressed through several stages. First, eye movement following a target located at a comfortable distance was practiced with the head stationary, progressing to movements of the head with visual fixation on a target (i.e. gaze stability). Pupillary glasses were used in the clinic to ensure a steady gaze during this exercise. Eye-head coordination exercises started with rotation of the eyes and head to the same side, both left and right. After that, patients practiced following a target with the eyes first, followed by the head, ensuring that they maintained focus on the target. As a further progression, the eyes moved first, and then the head, to look between 2 targets positioned horizontally or vertically, and finally, the eyes and head rotated in opposite directions, both left and right. All these exercises were progressed by increasing the speed and range of motion of the target and with patients in a standing position. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH (CRANIO-CERVICAL FLEXION) versus PROPRIOCEPTIVE EXERCISE

Protocol outcome 2: Physical function
- Actual outcome: Neck disability index total scores at 8 weeks; Group 1: mean 4.46 (SD 2.02); n=12, Group 2: mean 4.14 (SD 2.62); n=14; NDI Not specified Top=High is poor outcome; Comments: Baseline: 7.71(2.78); 7.42(2.87)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Garcia-martinez 2012[99]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=28)
Countries and setting	Conducted in Spain
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	None stated
Exclusion criteria	Exclusion criteria were the presence of serious cardiovascular, pulmonary, endocrine, neurological or renal disease, inflammatory rheumatic disease or participation in a physical therapy or exercise programme in the last 6 months.
Recruitment/selection of patients	Recruited from the Leon FM and chronic fatigue syndrome association.
Age, gender and ethnicity	Age - Mean (SD): 58.9(6.2). Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: chronic widespread pain: fibromyalgia
Extra comments	Mean duration of symptoms 10.3(4) years
Indirectness of population	No indirectness
Interventions	(n=14) Intervention 1: Mixed modality exercise - Aerobic, strength and stretching exercise. Exercised 3 times a week for 12 weeks. The exercise protocol was individualized and followed the guidelines from the ACSM for developing and maintaining cardio-respiratory fitness. Each session was 60 min long and included 10 min of warming-up with slow walks and easy movements of progressive intensity, 20 min of aerobic exercise that began at 60–70% of maximal heart rate and was gradually increased to as high as 75–85% maximum, depending on the subjects' adaptation, 20 min of stretching and strength exercise and 10 min of cooling down with low-intensity exercises. Duration 12 weeks. Concurrent medication/care: Not specified (n=14) Intervention 2: Other. Subjects continued their daily activities which did not include any physical exercise. Duration 8 weeks. Concurrent medication/care: Not stated. Indirectness: No indirectness

Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH AND STRETCHING EXERCISE versus USUAL CARE Protocol outcome 1: Quality of life - Actual outcome: SF-36 mental component at 12 weeks; Group 1: mean 45 (SD 12.7); n=12, Group 2: mean 32.9 (SD 12.7); n=13; Comments: Baseline: 37.9(9.9); 36.9(13.2) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2; Group 2 Number missing: 1 - Actual outcome: SF-36 physical component at 12 weeks; Group 1: mean 36.4 (SD 12.9); n=12, Group 2: mean 31.3 (SD 7.2); n=13; SF-36 PCS 0-100 Top=High is good outcome; Comments: Baseline: 30(8); 32.1(4.6) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2; Group 2 Number missing: 1	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Gavi 2014[100]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=80)
Countries and setting	Conducted in Brazil; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 16 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women, between 18 and 65 years old, who met the criteria according to the American College of Rheumatology.
Exclusion criteria	Any diseases or conditions that could limit exercise, autonomic dysfunctioning, the use of medication such as beta blockers or CCBs or other medications that could interfere with cardiovascular or autonomic responses, taking part in exercise in the last 3 months, receipt of social security benefits.
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 46.71(8.82) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: Chronic widespread pain: fibromyalgia
Extra comments	Duration of pain not specified
Indirectness of population	No indirectness
Interventions	(n=40) Intervention 1: Strength. 45 minute sessions 2 times a week for 16 weeks. Supervised progressive training in standing and sitting positions using weight machines. Moderate intensity with load of 45% the estimated maximum. Multiple muscle groups were trained in 12 different exercises, with 3 sets of 12 repetitions. Duration 16 weeks. Concurrent medication/care: 7% were using low doses of cyclobenzaprine or amitriptyline. Indirectness: No indirectness (n=40) Intervention 2: Flexibility. 45 minute sessions 2 times a week for 16 weeks. Stretching of the major muscles. No further details. Duration 16 weeks. Concurrent medication/care: 7% taking amitriptyline of benzodiazepines. Indirectness: No indirectness

Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus FLEXIBILITY (STRETCHING) Protocol outcome 1: Quality of life - Actual outcome: SF-36 physical component at 16 weeks (post intervention); Group 1: mean 35.65 (SD 7.8); n=35, Group 2: mean 34.15 (SD 9.2); n=31; SF-36 PCS 0-100 Top=High is good outcome; Comments: Baseline: 27.01(7.61); 24.37(7.58) Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Employment, death in family, arthritis; Group 2 Number missing: 9, Reason: Employment, childcare, moved, illness in the family, lost to follow up, arthrosis - Actual outcome: SF-36 mental component at 16 weeks (post intervention); Group 1: mean 39.16 (SD 12.64); n=35, Group 2: mean 44.55 (SD 13.6); n=31; sf-36 MCS 0-100 Top=High is good outcome; Comments: Baseline: 33.47(12.33); 36.98(12.73) Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Employment, death in family, arthritis; Group 2 Number missing: 9, Reason: Employment, childcare, moved, illness in the family, lost to follow up, arthrosis Protocol outcome 2: Discontinuation - Actual outcome: Discontinuation at 16 weeks (post intervention); Group 1: 5/35, Group 2: 9/31 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness, Comments: Dropped out of study; not defined as discontinuation of intervention;	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	**Gavish 2006[101]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=20)
Countries and setting	Conducted in Israel; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: According to Research Diagnostic Criteria for TMD (RDC/TMD)

Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Females aged 20-45 years old with a dolichocephalism face configuration, masticatory muscle pain for at least 6 months before the study, sensitivity to palpation of the masseter muscle at moderate to severe level at the pain side, masseter muscle that did not significantly increase in volume in maximal clench, natural definition with no more than one missing tooth per quadrant, no evidence of carious lesions or periodontal disease, and an increased pain level during a chewing test of at least 15.100 mm on the VAS
Exclusion criteria	Patients with temporomandibular joint disease or disorder diagnosed clinically or radiographically, systemic chronic disease or continuous use of medication, history of trauma to the facial or cervical regions, and previous treatment related to the myofascial pain within the last 6 months
Recruitment/selection of patients	Recruited from the patients transferred for treatment at the TMD clinic
Age, gender and ethnicity	Age - Mean (SD): exercise 27.1 (10.1); control 27.3 (5.9). Gender (M:F): Women only. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: people with chronic orofacial pain 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable 5. complex regional pain syndrome: Not applicable
Extra comments	Duration of pain not reported
Indirectness of population	No indirectness
Interventions	(n=10) Intervention 1: Strength. Chewing exercise. Two units of sugarless chewing gum were chewed three times daily for 10 minutes (weeks 1 and 2), increasing to 15 minutes three times daily (weeks 5 and 6), and 30 minutes 3 times daily (weeks 7 and 8). Patients were instructed to chew at their own rate. All patients received a detailed explanation of their disorder, its cyclic nature and possible etiology at the initial examination. They then received a detailed description of the chewing exercise protocol (at session 1). Sessions 2, 3, and 4 were to report the patient's condition, reassurance, support, and encouragement. They also reported their performance. Duration 8 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=10) Intervention 2: Psychological intervention - Pain education. All patients received a detailed explanation of their disorder, its cyclic nature and possible etiology at the initial examination. Sessions 2, 3, and 4 were to report the patient's condition, reassurance, support, and encouragement. Duration 8 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus PAIN EDUCATION Protocol outcome 1: Pain reduction - Actual outcome: Pain relief at post intervention; Group 1: mean 47 (SD 27); n=10, Group 2: mean 19 (SD 22); n=10; VAS 0-100 Top=High is good outcome Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Protocol outcome 2: Discontinuation - Actual outcome: Discontinuation at post intervention; Group 1: 0/10, Group 2: 0/10 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Giubilei 2007[106]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=103)
Countries and setting	Conducted in Afghanistan, Italy; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 18 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis: Men with NIH type III CP
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Men with chronic prostatitis/chronic pelvic pain syndrome. No medical or psychological contraindications for moderate intensity exercise. Experienced pain for at least 3 month
Exclusion criteria	People older than 50 years, Any concurrent condition that could cause the pain or concurrent treatment such as chemotherapy or thermotherapy that could influence the results of the study.
Recruitment/selection of patients	From outpatient clinics
Age, gender and ethnicity	Age - Mean (SD): 36.7(8.1)years. Gender (M:F): All men. Ethnicity: Not specified
Further population details	Subgroup: chronic visceral pain
Extra comments	Mean symptom duration 5.72(4.1) years.
Indirectness of population	No indirectness
Interventions	(n=52) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. 18 week walking program, 3 times per week. Each exercise session included a warm up and cool down regimen of slow paced walking, specific postural muscle and isometric strengthening exercises, and 40 minutes of fast paced walking on in-outdoor track, at 70-80% of maximum heart rate. Duration 18 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=51) Intervention 2: Flexibility. Participants participated in a flexibility and motion exercise program for the same period of time and frequency as the aerobic group. Patients were instructed about the correct exercise execution and were advised to maintain their heart rate under 110bpm. Exercises were simply stretches with some motion exercises such as leg lifts. Duration 18 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus FLEXIBILITY

Protocol outcome 1: Pain reduction
- Actual outcome: VAS at 6 weeks; Group 1: mean 4.3 (SD 1.4); n=41, Group 2: mean 4.7 (SD 1.4); n=44; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.1(1.6); 5.1(1.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 11; Group 2 Number missing: 7
- Actual outcome: VAS at 18 weeks; Group 1: mean 3.4 (SD 1.4); n=36, Group 2: mean 4.2 (SD 1.2); n=40; VAS 0-10 Top=High is poor outcome; Comments: Baseline: 5.1(1.6); 5.1(1.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 11; Group 2 Number missing: 7

Protocol outcome 2: Quality of life
- Actual outcome: NIH CPSI quality of life subscale at 18 weeks; Group 1: mean 4.4 (SD 1.8); n=36, Group 2: mean 6.2 (SD 2.1); n=40; NIH CPSI quality of life subscale 0-12 Top=High is poor outcome; Comments: Baseline: 6.5(2.8); 8(2.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 11; Group 2 Number missing: 7
- Actual outcome: NIH CPSI quality of life subscale at 6 weeks; Group 1: mean 5.1 (SD 2.1); n=41, Group 2: mean 6.9 (SD 2.1); n=44; nih-cpsi 0-12 Top=High is poor outcome; Comments: Baseline: 6.5(2.8); 6.9(2.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 11; Group 2 Number missing: 7

Protocol outcome 3: Psychological distress
- Actual outcome: Beck depression inventory at 6 weeks; Group 1: mean 9.8 (SD 4.3); n=41, Group 2: mean 9.3 (SD 4.3); n=44; BDI 0-21 Top=High is poor outcome; Comments: Baseline: 12.1(6.4); 11.2(5.6)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 11; Group 2 Number missing: 7
- Actual outcome: Beck depression inventory at 18 weeks; Group 1: mean 8.3 (SD 3.5); n=36, Group 2: mean 7.8 (SD 3); n=40; BDI 0-21 Top=High is poor outcome; Comments: Baseline: 12.1(6.4);11.2(5.6)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 11; Group 2 Number missing: 7

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at 18 weeks; Group 1: 10/52, Group 2: 5/51

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Quality of life different at baseline; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Physical function ; Use of healthcare services ; Sleep

Study	Glasgow 2017[107]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=26)
Countries and setting	Conducted in USA; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Met ACR criteria for fibromyalgia
Exclusion criteria	Exclusion criteria included having engaged in any form of exercise within the past year, smoking within the past year, history of cardiovascular, pulmonary or metabolic diseases and using any medications that may affect heart rate or blood pressure.
Recruitment/selection of patients	Fliers and newspaper advertisements in local community
Age, gender and ethnicity	Age - Mean (SD): 51(10.5) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=14) Intervention 1: Strength. Supervised resistance exercises twice a week for 8 weeks, each lasting 30 minutes. 3 sets of 8-12 repetitions followed by 90 second rest periods between each set. Exercises were chest presses, leg extensions, leg curls and seated rows, initially at a training intensity of 50-60% of maximum. Resistance was increased when participants could complete 12 repetitions on all 3 sets over 2 consecutive training days. Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=12) Intervention 2: Other. Control group (non-exercising, no further details). Duration 8 weeks. Concurrent medication/care: Not specified. Indirectness: Serious indirectness; Indirectness comment: Control treatment unclear
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus NO TREATMENT	

Protocol outcome 1: Psychological distress - Actual outcome: Pain catastrophising scale at 8 weeks; Group 1: mean 11 (SD 12); n=13, Group 2: mean 20 (SD 15); n=12; Comments: Baseline 18(13); 28(14) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference of over 16 at baseline ; Group 1 Number missing: 1; Group 2 Number missing: 0 Protocol outcome 2: Quality of life - Actual outcome: Fibromyalgia impact questionnaire at 8 weeks; Group 1: mean 41 (SD 24); n=13, Group 2: mean 71.8 (SD 8); n=12; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: 59(12); 72.7(7) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference of over 12 at baseline (out of 100); Group 1 Number missing: 1; Group 2 Number missing: 0 Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at 8 weeks; Group 1: 1/14, Group 2: 0/12 Risk of bias: All domain - ; Indirectness of outcome: Serious indirectness, Comments: Unclear definition of discontinuation	
Protocol outcomes not reported by the study	Physical function ; Use of healthcare services ; Sleep

Study	Gomez-hernandez 2020[108]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=64)
Countries and setting	Conducted in Spain; Setting: the clinical laboratory of the Physiotherapy Department at Universidad Cardenal Herrera-CEU
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: diagnosis of fibromyalgia syndrome according to the American College of Rheumatology criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	women with fibromyalgia syndrome according to the American College of Rheumatology criteria
Exclusion criteria	any health condition for which physical exercise was contraindicated, a history of regular physical exercise (three times a week) in the previous three months, severe cardiopulmonary problems, a serious psychiatric disorder, inflammatory rheumatoid disease, or unstable hypertension
Recruitment/selection of patients	participants were recruited through the local fibromyalgia association
Age, gender and ethnicity	Age - Mean (SD): Intervention group: 53.97 (5.00); control group: 54.58 (8.52). Gender (M:F): All female. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable 5. complex regional pain syndrome: Not applicable
Indirectness of population	No indirectness
Interventions	(n=32) Intervention 1: Aerobics - Other aerobic exercise. A supervised stationary cycling programme consisting of three 12-minute sessions per week for 12 weeks. Each session consisted of a 2-minute cycling warm-up and 10 minutes of moderate intensity cycling (50%–70% of the age-predicted maximum heart rate). Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness

	(n=32) Intervention 2: Mixed modality exercise - Aerobic and flexibility exercise. The same exercise programme as the control group, plus an additional 45 minutes stretching session per week for 12 weeks. Each session consisted of three repetitions of 10 seconds for each trunk muscle and two repetitions of 10 seconds for each extremity muscle. After each repetition, there was a 10-second pause.. Duration 12 weeks. Concurrent medication/care: No information. Indirectness: No indirectness
Funding	No funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: CYCLING + STRETCHING versus CYCLING

Protocol outcome 1: Pain reduction
- Actual outcome: Pain perception at 4 weeks; Group 1: mean 6.68 (SD 0.48); n=32, Group 2: mean 7.33 (SD 0.38); n=32; VAS 0-10 Top=High is poor outcome; Comments: Baseline values: experimental group - 7.79 ± 0.39; control group - 7.92 ± 0.31
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; G0roup 2 Number missing: 0
- Actual outcome: Pain perception at 12 weeks; Group 1: mean 5.77 (SD 0.4); n=32, Group 2: mean 6.71 (SD 0.42); n=32; VAS 0-10 Top=High is poor outcome; Comments: Baseline values: experimental group - 7.79 ± 0.39; control group - 7.92 ± 0.31
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Quality of life
- Actual outcome: Impact on QoL at 4 weeks; Group 1: mean 64.32 (SD 3.99); n=32, Group 2: mean 69.81 (SD 4.07); n=32; Fibromyalgia Impact Questionnaire 0-100 Top=High is poor outcome; Comments: Baseline: experimental - 84.10 ± 4.12; control - 83.65 ± 3.36
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Impact on QoL at 12 weeks; Group 1: mean 55.48 (SD 2.63); n=32, Group 2: mean 66.1 (SD 4.21); n=32; Fibromyalgia Impact Questionnaire 0-100 Top=High is poor outcome; Comments: Baseline: experimental - 84.10 ± 4.12; control - 83.65 ± 3.36
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Sleep
- Actual outcome: Sleep quality at 4 weeks; Group 1: mean 8.45 (SD 1.33); n=32, Group 2: mean 12.39 (SD 1.45); n=32; Pittsburgh Sleep Quality Index 0–21 Top=High is poor outcome; Comments: Baseline: Experimental - 15.42 ± 2.09; control - 14.68 ± 1.64
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Sleep quality at 12 weeks; Group 1: mean 5.42 (SD 0.98); n=32, Group 2: mean 10.45 (SD 0.99); n=32; Pittsburgh Sleep Quality Index 0-26 Top=High is poor outcome; Comments: Baseline: Experimental - 15.42 ± 2.09; control - 14.68 ± 1.64

Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0 Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at 12 weeks; Group 1: 0/32, Group 2: 0/32 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Physical function; Psychological distress (depression/anxiety); Use of healthcare services

Study	Haak 2008[118]
Study type	RCT (randomised; Parallel)
Number of studies (number of participants)	(n=57)
Countries and setting	Conducted in Sweden; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention + follow up: 4 week intervention plus 16 week follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	At least 18 years old, diagnosis for at least 6 months
Exclusion criteria	Severe depression, psychosis, other severe diseases, suicidal risk, drug or alcohol dependency
Recruitment/selection of patients	Local press, Patient's association for fibromyalgia, care centres and the Swedish National Insurance Scheme
Age, gender and ethnicity	Age - Mean (range): 53 years (range 27 - 73). Gender (M:F): All female. Ethnicity: Not specified
Further population details	Subgroup: chronic widespread pain
Extra comments	Mean duration of symptoms 15 years
Indirectness of population	No indirectness
Interventions	(n=29) Intervention 1: Mind-body exercises - Qigong. Total Qigong time 711.5 hours. Participants were instructed to practice Qigong at home with the support of a free instruction tape, twice a day for 20 minutes. Supervisors of the intervention were experienced Qigong masters. The sessions included internal and external methods of Qigong (influenced by oneself and influenced by the Qigong master). Duration 7 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=28) Intervention 2: No treatment. Waiting list control . Duration 7 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus NO TREATMENT	

Protocol outcome 1: Pain reduction - Actual outcome: Visual numerological scale (pain) at 7 weeks; Group 1: mean 3.31 (SD 0.81); n=29, Group 2: mean 4.2 (SD 0.85); n=28; VNS 0-10 Top=High is poor outcome; Comments: Baseline: 3.87(0.77); 4.33(0.95) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Protocol outcome 2: Quality of life - Actual outcome: WHOQOL-BREF at 7 weeks; Group 1: mean 3.37 (SD 0.68); n=29, Group 2: mean 2.79 (SD 0.92); n=28; World health organisation quality of life scale 0-5 Top=High is good outcome; Comments: Baseline: 2.89(0.92); 2.78(0.96) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Protocol outcome 3: Psychological distress (depression/anxiety) - Actual outcome: Beck depression inventory at 7 weeks; Group 1: mean 12.88 (SD 7.54); n=29, Group 2: mean 17.1 (SD 8); n=28; BDI 0-21 Top=High is poor outcome; Comments: 15.28(8.79);15.1(5.49) Risk of bias: All domain - ; Indirectness of outcome: No indirectness - Actual outcome: State trace anxiety inventory at 7 weeks; Group 1: mean 41.77 (SD 11.03); n=29, Group 2: mean 51.68 (SD 10.84); n=28; STAI-S 0-100 Top=High is poor outcome; Comments: Baseline: 44.51(11.12); 49.51(8.69) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Physical function ; Use of healthcare services ; Sleep ; Discontinuation

Study	Hooten 2012[125]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=72)
Countries and setting	Conducted in USA; Setting: Mayo Comprehensive pain rehabilitation centre, USA
Line of therapy	Unclear
Duration of study	Intervention time: 3 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Established diagnosis of fibromyalgia according to the ACR criteria, aged over 18 years
Exclusion criteria	Cardiovascular, pulmonary, orthopedic, or other systematic disease that could limit strength training or aerobic conditioning. Other exclusion criteria included pregnancy, schizophrenia, dementia.
Recruitment/selection of patients	From the Mayo pain clinic between 2006 and 2008
Age, gender and ethnicity	Age - Mean (SD): 46.5(10.8) years. Gender (M:F): 7:65 Ethnicity: 97% White, 1% African American, 1% Hispanic, 1% Arabic
Further population details	Subgroup: people with chronic widespread pain
Extra comments	Mean pain duration12.5(12.9) years
Indirectness of population	No indirectness
Interventions	(n=36) Intervention 1: Strength. Upper and lower body strengthening exercises were performed daily using resistive techniques, all supervised by a physical therapist with experience in treating patients with fibromyalgia. Each daily strength training session was 25-30 minutes in duration and also involved a warm up and cool down period. Participants were encouraged to train at the maximal amount of load tolerated, using one set of 10 repetitions. Duration 3 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=36) Intervention 2: Aerobics - Other aerobic exercise. Stationary bicycle exercises supervised by a physical therapist. Sessions also had a warm up and cool down and intensity of exercises was gradually increased to achieve 70-75% of maximal heart rate based on age. Exercise started at 10 minutes daily during week 1 (5 times a week), 15 minutes in week 2 and up to 20 to 30 minutes daily during week 3. Duration 3 weeks. Concurrent medication/care: Not specified.

	Indirectness: No indirectness
Funding	Academic or government funding (Mayo Foundation)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus AEROBIC (CYCLING) Protocol outcome 1: Pain reduction - Actual outcome: Multidimensional pain inventory at 3 weeks; Group 1: mean 34.4 (SD 11.5); n=36, Group 2: mean 37.6 (SD 11.9); n=36; MDPI 0-100 Top=High is poor outcome; Comments: baseline: 46.4(9.8); 48.6(6.7) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Current opioid use difference of 11%; Group 1 Number missing: 4, Reason: Lost to follow up, lack of efficacy, other conditions; Group 2 Number missing: 6, Reason: Lost to follow up, lack of efficacy, other conditions Protocol outcome 2: Discontinuation - Actual outcome: Discontinuation at 3 weeks; Group 1: 3/36, Group 2: 6/36 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Current opioid use difference of 11%; Group 1 Number missing: 4, Reason: Lost to follow up, lack of efficacy, other conditions; Group 2 Number missing: 6, Reason: Lost to follow up, lack of efficacy, other conditions	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Izquierdo-alventosa 2020[132]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=32)
Countries and setting	Conducted in Spain; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 8 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women between 30–70 years old, an age range in which FM becomes more prevalent, diagnoses according to the 2016 American College of Rheumatology criteria for FM, and having received pharmacological treatment for more than three months with no clinical improvement
Exclusion criteria	pPegnancy or breast-feeding, any known advanced-stage pathology associated with the locomotor system that contraindicates physical activity (arthritis,osteoarthritis, uric acid), epilepsy, in take of drugs that reduce the seizure threshold, history of intense headaches, neurological disorder, peripheral neuropathy, known serious cardiovascular disease (i.e., endocranial hypertension, uncontrolled arterial hypertension, heart failure, cardiac pacemaker), pneumothorax, neoplasia, surgery in the last four months, diagnosis of alcohol addiction, and use of psychoactive drugs or narcotics. Moreover, patients should not have been enrolled in any PE program in the two months before the study began.
Recruitment/selection of patients	Recruited from several Fibromyalgia Associations
Age, gender and ethnicity	Age - Mean (SD): Exercise group: 53.06 (8.4); control group: 55.13 (7.35) . Gender (M:F): Female only. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain 5. complex regional pain syndrome: Not applicable
Indirectness of population	No indirectness
Interventions	(n=16) Intervention 1: Mixed modality exercise - Aerobic, strength and conditioning exercise. A low-intensity PE program combining endurance training (i.e., aerobic and low-load resistance exercises aimed at improving endurance) and coordination. There were 16 sessions

	performed twice a week, each lasting 1 hour. Each session was divided into three parts: warm-up (walking at a slow pace and moving the main joint structures), training, and cool-down (walking at a slow pace, trunk stretching, deep breathing). Training included exercises conducted using 1-kg dumbbells and weights at a velocity determined by a metronome set at 60 beats per minute. Exercises included preacher curl, leg extension, dumbbell front raise, hip abduction, pull ups, shoulder rotation, sitting down/standing up, throwing and catching a ball, calf raise, step ups. Duration 8 weeks. Concurrent medication/care: Continued to take their usual medication. Indirectness: No indirectness (n=16) Intervention 2: No treatment. No intervention, participants were asked to perform their daily routines. Duration 8 weeks. Concurrent medication/care: Continued to take their usual medication. Indirectness: No indirectness
Funding	No funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH AND CONDITIONING EXERCISE versus NO TREATMENT

Protocol outcome 1: Quality of life
- Actual outcome: Quality of lifeat Post-treatment (8 weeks); Group 1: mean 61.49 (SD 17.65); n=16, Group 2: mean 67.07 (SD 15.87); n=16; Spanish validated version of the Revised Fibromyalgia Impact Questionnaire (FIQR) 0-100 Top=High is poor outcome; Comments: Baseline: exercise group 71.47 (14.21); control group 62.44 (17.33)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Physical function
- Actual outcome: Endurance and functional capacity - 6 minute walk test at Post-treatment (8 weeks); Group 1: mean 513 distance in meters (SD 64.84); n=16, Group 2: mean 497.31 distance in meters (SD 76.29); n=16; Comments: Baseline: exercise group 481.00 (71.23); control group 493.19 (68.48)
Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Psychological distress (depression/anxiety)
- Actual outcome: Depression at Post-treatment (8 weeks); Group 1: mean 23.81 (SD 7.93); n=16, Group 2: mean 27.94 (SD 11.14); n=16; validated Spanish version of the Beck Depression Inventory-Second Edition (BDI-II) 0-63 Top=High is poor outcome; Comments: Baseline: exercise group 31.13 (9.06); control group 29.31 (11.55)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Anxiety at Post-treatment (8 weeks); Group 1: mean 9.94 (SD 3.57); n=16, Group 2: mean 11.19 (SD 3.69); n=16; Comments: Baseline: exercise group 11.81 (3.54); control group 12.19 (4.07)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at Post-treatment (8 weeks); Group 1: 0/16, Group 2: 0/16 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Pain reduction; Use of healthcare services; Sleep

Study	Kibar 2015[147]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=68)
Countries and setting	Conducted in Turkey; Setting: not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Based on the 2010 American College of Rheumatology diagnostic criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 18-65 years with fibromyalgia syndrome
Exclusion criteria	People with vitamin B12, 25OH vitamin D, and folate deficiencies; diabetes mellitus; neurologic diseases; rheumatoid diseases; eye and internal ear pathologies; advanced cardiovascular or lung pathologies; and uncontrolled hypertension or hypotension were excluded. Patients who previously underwent surgery, who had injuries in their lower extremities (knees, hips, ankles, feet), and who were admitted to a physical therapy and/or exercise programme for their pain within the last year were also not included
Recruitment/selection of patients	not reported
Age, gender and ethnicity	Age - Mean (SD): Flexibility + balance: 48.11 (13.42); flexibility: 48.17 (12.68). Gender (M:F): 3/54. Ethnicity: not reported
Further population details	Subgroup: people with chronic widespread pain
Extra comments	Duration of pain not reported
Indirectness of population	No indirectness
Interventions	(n=35) Intervention 1: Flexibility and proprioception. Balance exercises included postures that gradually reduced the base of support (2-legged stand, semi-tandem stand, tandem stand, 1-legged stand), dynamic movements that disturbed the centre of gravity (tandem walk, circle turns), exercises that stressed the postural muscle groups (heel or toe stands), and exercises that reduced sensory input (standing with eyes closed). Training was provided by an experienced physiotherapist for 20 sessions over a 4 week period (20 minutes for each session, 5 days/week). The group also received 5 minutes of static and 5 minutes of dynamic balance training with a KAT device 3 days/week. This device has a movable platform and a tilt sensor that is connected to a computer. Participants maintained their balance by tilting the platform in all directions without moving their feet. They could only change their centre of gravity via

	trunk movements. During static balance training, the patients were asked to maintain their equilibrium while standing as motionless as possible on the platform and were told to keep the red X symbol in the centre of the computer screen. In the dynamic balance training, they were asked to superimpose the X symbol onto the moving cursor while it made a 360 degree circle on the screen. For flexibility, active static exercises were performed in order to enable compliance to exercise and its maintenance without being forced. Exercises were performed in 8 large muscle groups (neck, back, lower back, biceps, triceps, gluteus, iliopsoas, quadriceps femoris, hamstring, gastrosoleus) in three 60 second static stretching repetitions. Because in older persons holding a stretch for 30-60 seconds may confer greater benefit for each muscle, to the extent that patients was capable, 30-60 second static stretching was carried out. Ten minutes of walking in place was also recommended as warm up. Duration 6 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=33) Intervention 2: Flexibility. Active static exercises were performed in order to enable compliance to exercise and its maintenance without being forced. Exercises were performed in 8 large muscle groups (neck, back, lower back, biceps, triceps, gluteus, iliopsoas, quadriceps femoris, hamstring, gastrosoleus) in three 60 second static stretching repetitions. Because in older persons holding a stretch for 30-60 seconds may confer greater benefit for each muscle, to the extent that patients was capable, 30-60 second static stretching was carried out. Ten minutes of walking in place was also recommended as warm up. These were performed for 2 sessions and participants were informed of the necessity of exercising 5 days a week. Duration 6 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: FLEXIBILITY AND PROPRIOCEPTION versus FLEXIBILITY

Protocol outcome 1: Quality of life
- Actual outcome: Fibromyalgia impact questionnaire at end of treatment; Group 1: mean 52.85 (SD 15.24); n=28, Group 2: mean 65.55 (SD 17.7); n=29; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: mixed exercise 65.78 (14.73); flexibility 65.89 (18.05)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 4

Protocol outcome 2: Psychological distress (depression/anxiety)
- Actual outcome: depression at end of treatment; Group 1: mean 17.67 (SD 9.37); n=28, Group 2: mean 13.79 (SD 7.18); n=29; BDI 0-63 Top=High is poor outcome; Comments: Baseline: mixed exercise 19.46 (9.33); flexibility 13.89 (7.89)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 4

Protocol outcome 3: Discontinuation - Actual outcome: discontinuation at end of treatment; Group 1: 7/35, Group 2: 4/33 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Use of healthcare services ; Sleep

Study	Kingsley 2005[153]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=29)
Countries and setting	Conducted in USA; Setting: Laboratory and strength training facility
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women diagnosed with fibromyalgia
Exclusion criteria	Uncontrolled hypertension, controlled diabetes, active heart disease, and/or already participating in a strength training programme
Recruitment/selection of patients	Newspaper advertisement
Age, gender and ethnicity	Age - Mean (SD): Intervention group: 45±0; control group 47±4. Gender (M:F): Females only. Ethnicity: Not reported
Further population details	Subgroup: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=15) Intervention 1: Strength. A progressive full body strength training regime twice a week. Sessions consisted of 11 exercises. Six exercises were performed on Nautilus resistance machines, 3 on the Nautilus cable machine and the remaining 2 were performed using the subject's body weight as resistance. Resistance machine exercises included chest press, leg extension, standing leg curl, shoulder press, lumbar extension and abdominal crunch. The cable exercises included low pulley biceps curl, high pulley triceps extension, and the mid pulley standing row. Body weight was used for the standing calf raises and body weight Swiss ball squats. Before and after workouts, participants performed 5 minutes of warm up and cool down that included stretching and walking. Participants began training at 40% of their 1-RM. Once 12 repetitions were performed in proper form, weight was increased by 2.3 to 4.5kg (5-10lb). The duration of each session was 30 minutes. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=14) Intervention 2: No treatment. Participants were asked not to change their activity levels during the 12 week

	intervention period. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus NO TREATMENT Protocol outcome 1: Quality of life - Actual outcome: Fibromyalgia impact questionnaire at Post intervention; Group 1: mean 54.6 (SD 19.9); n=15, Group 2: mean 53.9 (SD 13.2); n=14; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: exercise 60.8 ± 19.9; no treatment 57.1±12.2 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain: exercise group 9±10 years; control group 7±5 years; Group 1 Number missing: 7; Group 2 Number missing: 2 Protocol outcome 2: Physical function - Actual outcome: 6 minute walk test at Post intervention; Group 1: mean 529.9 meters (SD 85.2); n=8, Group 2: mean 538.3 meters (SD 98.5); n=12; Comments: Baseline: exercise 484.2±83.2; no treatment 505.1±99.2 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain: exercise group 9±10 years; control group 7±5 years; Group 1 Number missing: 7; Group 2 Number missing: 2 Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at Post intervention; Group 1: 7/8, Group 2: 2/14 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Duration of pain: exercise group 9±10 years; control group 7±5 years;	
Protocol outcomes not reported by the study	Pain reduction ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Lansinger 2013[157]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=122)
Countries and setting	Conducted in Sweden; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention + follow up: 3 months + 12 months
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age 18-35 years, non-specific neck pain for at least 3 months and an average self-rated neck pain of at least 20mm on a 0-100mm visual analogue scale during the week before screening/baseline
Exclusion criteria	Chronic tension-type headache, migraine, traumatic neck injuries, neurological signs or symptoms, rheumatic diseases, fibromyalgia, or other severe physiological or physical diseases, treatment with anti-depressive and/or anti-inflammatory drugs, and difficulties in understanding the Swedish language
Recruitment/selection of patients	Newspaper advertisement
Age, gender and ethnicity	Age - Mean (SD): 43.8±12.9. Gender (M:F): 86/36. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain (neck pain). 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable Not applicable
Indirectness of population	No indirectness
Interventions	(n=60) Intervention 1: Mind-body exercises - Qigong. 10-12 1 hours sessions conducted on a weekly or biweekly basis over 3 months. Qigong was performed according to medical qigong which is a modality of traditional Chinese medicine and is a way of affecting and directing qi (energy) for medical benefit. Each qigong exercise includes body posture and gentle movement, meditation (concentration) and purposeful relaxation, breathing regulation practice and self-administered massage. Qigong was conducted in groups of 10-15 participants. Duration 12 sessions in 3 months. Concurrent medication/care: Both groups received verbal ergonomic advice for both work and free time, as well as an information pamphlet on neck pain. Indirectness: No indirectness (n=62) Intervention 2: Strength. Exercise therapy was performed individually and the training programme was adjusted for each participant. A physiotherapist instructed the participants throughout the training programme, which focused

	mainly on the cervical and shoulder/thoracic region. Each training session started with a warm up on a stationary bicycle for about 10 minutes, followed by 40 minutes of dynamic exercises. These exercises consisted of active movements aimed to increase range of motion in all neck directions and muscle exercises aimed to maintain/increase circulation, endurance and strength. The amount of load was individualised and was maintained within pain tolerance (aimed not to increase pain). The load at the muscle exercises was to achieve between 30% and 70% of maximum muscle capacity and was gradually increased as endurance and strength were gained. The exercises were performed with low resistance, allowing 20-30 repetitions of maximal voluntary contractions in three sets. Duration 12 sessions in 3 months. Concurrent medication/care: Both groups received verbal ergonomic advice for both work and free time, as well as an information pamphlet on neck pain. Indirectness: No indirectness
Funding	Academic or government funding (Grants from the Vardal Institute, the Ekhaga Foundation, the Herbet and Karin Jacobsson Foundation, the Martina Lundgren Foundation and the Swedish Association of Registered Physiotherapists)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus STRENGTH Protocol outcome 1: Discontinuation - Actual outcome: Discontinuation at After treatment; Group 1: 12/60, Group 2: 8/62 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Pain reduction ; Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Latorre roman 2015[159]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=39)
Countries and setting	Conducted in Spain; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 18 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People who met the Criteria for the Classification of Fibromyalgia established by the American College of Rheumatology, not suffering any other serious somatic disease (i.e. enthesitis or spondyloarthritis) or psychiatric or medical disorder that required immediate treatment or that be incompatible with physical activity (exercise in swimming pools included)
Exclusion criteria	Not reported
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Intervention group: 51.70±9.5; control group 50.25±8.83. Gender (M:F): All women. Ethnicity: Not reported
Further population details	Subgroup: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. Sixty minute sessions of functional training 3 times a week. Of those 3 weekly sessions, 2 consistent of exercise in water and 1 of exercise on land. Both were instructed by a specialist in physical activity. Each session included a warm up (5 minutes) and exercises of muscular strengthening and balance (40 minutes), and a cool down (5 minutes). Exercise intensity was increased during the whole programme by modifying the number of reps per set, by introducing weights (in on land exercises, 0.5-2kg per exercise) and materials that raised the resistance offered by water. Strength training consisted in 1-3 sets of 8-12 reps per exercise and circuit training. The intensity of the exercises was self administered by participants, but they were asked to perform 8-12 repetitions. In the land, the following functional exercises were performed individually and on a circuit, for example, climbing stairs using weights as the external load (medicine ball), pulling used rubber bands at different resistances as external load, picking things up from the floor, carrying heavy objects (medicine ball), sit-to-

	stand from a chair, hurdles, slalom challenges, walking forward, walking backward, and tossing a ball. In the pool with water level at participants' chest height, all exercises were conducted for example, flutter kick with kick board, sit-to-stand from the pool wall, walking forward, walking simulating steps up, lateral walking with large steps, sinking the floats, rowing, and throwing and catching ball with partner. The physical exercise to improve balance includes standing on one leg, reducing base of support, shifting weight from foot to foot, stepping over objects, and sitting on a stability ball and turning and changing its direction in the land; and standing, kneeling and sitting balance in pool noodle in the water. Duration 18 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=19) Intervention 2: Usual care. Participants continued with their daily activities that did not include any kind of physical exercise similar to that of the study group. Duration 18 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus USUAL CARE Protocol outcome 1: Pain reduction - Actual outcome: Pain (VAS in rest) at Post treatment; Group 1: mean 6.47 (SD 3.2); n=20, Group 2: mean 8.75 (SD 1.73); n=16; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: exercise 9.4±1.66; control 9.18±0.75 Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: % employed: exercise 45%; control 25%; Group 1 Number missing: 0; Group 2 Number missing: 3 Protocol outcome 2: Quality of life - Actual outcome: Fibromyalgia impact questionnaire at Post treatment; Group 1: mean 54.72 (SD 14.75); n=20, Group 2: mean 63.86 (SD 15.41); n=16; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 62.26±12.65; control 65.72±15.57 Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: % employed: exercise 45%; control 25%; Group 1 Number missing: 0; Group 2 Number missing: 3 Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at Post treatment; Group 1: 0/20, Group 2: 3/19 Risk of bias: All domain - High, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: % employed: exercise 45%; control 25%;	
Protocol outcomes not reported by the study	Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Lauche 2016[160]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=114)
Countries and setting	Conducted in Germany; Setting: Department of Complementary and Integrative Medicine in Essen
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 weeks + 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	At least 18 years of age and to have chronic nonspecific neck pain for at least 3 consecutive months for at least 5 days a week. They also had to report moderate pain of 45 mm or higher on a visual analogue scale (VAS) ranging from 0 to 100 mm, with 100 mm described as 'worst neck pain imaginable.' Patients with other musculoskeletal pain, such as arm pain or lower back pain, in addition to neck pain as defined previously were eligible
Exclusion criteria	Neck pain caused by trauma, disc protrusion, whiplash, congenital deformity of the spine, spinal stenosis, neoplasm, inflammatory rheumatic disease, neurological disorder, active oncologic disease, severe affective disorder, addiction, and psychosis. In addition, subjects who were pregnant or who had had invasive treatment of the spine within the previous 4 weeks (e.g., acupuncture, injections), or spinal surgery within the previous year, or had initiated or modified their drug regimen recently or were taking opiates were excluded. Finally, subjects with regular practice of Tai Chi, Qigong, or Yoga in the past 6months, or those with any disability precluding exercise practice, were also excluded
Recruitment/selection of patients	recruited via local newspaper advertisements
Age, gender and ethnicity	Age - Mean (SD): tai chi: 52.0 (10.9); neck exercises 47.0 (12.3); waiting list 49.2 (11.7) . Gender (M:F): 23/91. Ethnicity: not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable
Indirectness of population	No indirectness
Interventions	(n=38) Intervention 1: Mind-body exercises - Tai Chi. Participants in the Tai Chi group met once weekly for a 75- to 90-minute session for 12 weeks in total. The Tai Chi intervention was on the basis of a popular and internationally recognized Yang style (13 forms from Mantak Chia). Each session included a warm-up of 5 to 10 minutes, the Tai Chi form practice, and 5 to 10 minutes of relaxation at the end. Tai Chi forms followed explicit protocols outlined in a

	training manual, as required during teacher training certification. Sessions also included educational units and breathing exercises, and they were accompanied by relaxation music. Participants received illustrated written information that covered movement sequences learned in the previous session. They were asked to practice Tai Chi outside of classes for at least 15 minutes each day. This length of home practice was chosen to increase compliance with, and memorization and reinforcement of the exercises taught in class. Fifteen minutes of home practice is also a common recommendation for beginner Tai Chi students. Duration 12 weeks. Concurrent medication/care: "Participants received approximately 2 concomitant therapies per week, with no differences between the groups. Concomitant therapies mainly included massages and the application of heat without differences between the groups". Indirectness: No indirectness (n=37) Intervention 2: Mixed modality exercise - Other mixed modality exercise. Participants in the neck exercise group met once weekly for a 60- to 75-minute session for 12 weeks in total. This group was instructed in neck exercises, which were similar to those taught in rehabilitation programs containing exercises and education for a healthy back. Classes contained basic training of ergonomic principles (bodily alignment while standing), proprioceptive exercises, and isometric and dynamic mobilization, stretching, and strengthening neck and core exercises. Similar to Tai Chi, the sessions opened with 5 to 10 minutes of warm-up exercises and ended with relaxation exercises. Participants also received illustrated and written information that covered the most important exercises, and they were asked to execute the exercises for at least 15 minutes each day. This intervention was to control for effects due to increased levels of physical activity and the group setting in the Tai Chi group. Duration 12 weeks. Concurrent medication/care: "Participants received approximately 2 concomitant therapies per week, with no differences between the groups. Concomitant therapies mainly included massages and the application of heat without differences between the groups". Indirectness: No indirectness (n=39) Intervention 3: No treatment. Participants in this group were advised to continue their usual activities and therapies, but not to initiate any new therapeutic regimen for symptom management. At the trial's end, participants in the wait list group were offered as a courtesy the option to participate in a Tai Chi and neck exercise group. Duration 12 weeks. Concurrent medication/care: "Participants received approximately 2 concomitant therapies per week, with no differences between the groups. Concomitant therapies mainly included massages and the application of heat without differences between the groups". Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TAI CHI versus STRENGTH, PROPRIOCEPTION AND FLEXIBILITY Protocol outcome 1: Pain reduction	

- Actual outcome: Pain at 12 weeks; Group 1: mean 32.4 (SD 23.5); n=38, Group 2: mean 25.2 (SD 18.3); n=37; VAS 0-100 Top=High is poor outcome; Comments: Baseline: Tai chi 54.2 (20.4); exercise 46.2 (19.2)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: tai chi 54.2 (20.5); exercises 46.2 (19.2);
-Actual outcome: Pain at 24 weeks; Group 1: mean 35 (SD 27.7); n=38, Group 2: mean 33.1 (SD 20.9); n=37; VAS 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 54.2 (20.4); exercise 46.2 (19.2)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: tai chi 54.2 (20.5); exercises 46.2 (19.2); Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 2: Quality of life
- Actual outcome: SF36 physical at 12 weeks; Group 1: mean 47.3 (SD 9.1); n=38, Group 2: mean 45.2 (SD 5.4); n=37; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 44.13 (7); exercise 41.8 (7.4)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 mental at 12 weeks; Group 1: mean 46.8(SD 11.9); n=38, Group 2: mean 47.7(SD 8.5); n=37; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 46.3 (10.3); exercise 46.9 (8.3)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:
- Actual outcome: SF36 physical at 24 weeks; Group 1: mean 46.5 (SD 8.9); n=38, Group 2: mean 44 (SD 7.5); n=37; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 44.13 (7); exercise 41.8 (7.4)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 mental at 24 weeks; Group 1: mean 47 (SD 12.2); n=38, Group 2: mean 46.9 (SD 9.1); n=37; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 46.3 (10.3); exercise 46.9 (8.3)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at 12 weeks; Group 1: mean 21.5 (SD 12.2); n=38, Group 2: mean 22.7 (SD 9.3); n=37; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 30.8 (8); exercise 30.1 (9.8)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Neck disability at 24 weeks; Group 1: mean 24.3 (SD 14.1); n=38, Group 2: mean 25.1 (SD 12.9); n=37; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 30.8 (8); exercise 30.1 (9.8)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: Anxiety at 12 weeks; Group 1: mean 6.5 (SD 4.7); n=38, Group 2: mean 5.5 (SD 3.1); n=37; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 6.9 (3.8); exercise 6 (3)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Depression at 12 weeks; Group 1: mean 3.9 (SD 3.8); n=38, Group 2: mean 3.8 (SD 2.3); n=37; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 3.8 (2.9); exercise 3.8 (2.4)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Anxiety at 24 weeks; Group 1: mean 6.1 (SD 4.5); n=38, Group 2: mean 5.5 (SD 3.1); n=37; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 6.9 (3.8); exercise 6 (3)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Depression at 24 weeks; Group 1: mean 4.1 (SD 3.8); n=38, Group 2: mean 4.1 (SD 2.8); n=37; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 3.8 (2.9); exercise 3.8 (2.4)
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing

Protocol outcome 5: Discontinuation
- Actual outcome: Discontinuation at 12 weeks; Group 1: 3/38, Group 2: 13/37
Risk of bias: All domain - Low, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TAI CHI versus NO TREATMENT

Protocol outcome 1: Pain reduction
- Actual outcome: Pain at 12 weeks; Group 1: mean 32.4 (SD 23.5); n=38, Group 2: mean 41.8 (SD 22.5); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 54.2 (20.4); no treatment 51.5 (21.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Pain at 24 weeks; Group 1: mean 35 (SD 27.7); n=38, Group 2: mean 44.6 (SD 20); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 54.2 (20.4); no treatment 51.5 (21.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 2: Quality of life
- Actual outcome: SF36 physical at 12 weeks; Group 1: mean 47.3 (SD 9.1); n=38, Group 2: mean 42.9(SD 5.4); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 44.13 (7); no treatment 43.6 (7.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 mental at 12 weeks; Group 1: mean 46.8 (SD 11.9); n=38, Group 2: mean 46.2(SD 10.7); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 46.3 (10.3); no treatment 46.9 (10.5)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 physical at 24 weeks; Group 1: mean 46.5 (SD 8.9); n=38, Group 2: mean 42 (SD 8); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 44.13 (7); no treatment 43.6 (7.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 mental at 24 weeks; Group 1: mean 47 (SD 12.2); n=38, Group 2: mean 46.4 (SD 10.13); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: tai chi 46.3 (10.3); no treatment 46.9 (10.5)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at 12 weeks; Group 1: mean 21.5(SD 12.2); n=38, Group 2: mean 27.5 (SD 11.4); n=39; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 30.8 (8); no treatment 29.3 (8.2)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Neck disability at 24 weeks; Group 1: mean 24.3 (SD 14.1); n=38, Group 2: mean 29.4 (SD 12.7); n=39; Neck disability index 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 30.8 (8); no treatment 29.3 (8.2)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: Anxiety at 12 weeks; Group 1: mean 6.5 (SD 4.7); n=38, Group 2: mean 6.7 (SD 3.2); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 6.9 (3.8); no treatment 6.7 (3.7)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Depression at 12 weeks; Group 1: mean 3.9 (SD 3.8); n=38, Group 2: mean 4.9 (SD 3.4); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 3.8 (2.9); no treatment 4.5 (3)

Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Anxiety at 24 weeks; Group 1: mean 6.1 (SD 4.5); n=38, Group 2: mean 6.7 (SD 3.4); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 6.9 (3.8); no treatment 6.7 (3.7)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Depression at 24 weeks; Group 1: mean 4.1 (SD 3.8); n=38, Group 2: mean 5.4 (SD 4); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: tai chi 3.8 (2.9); no treatment 4.5 (3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 5: Discontinuation
- Actual outcome: Discontinuation at 12 weeks; Group 1: 3/38, Group 2: 10/39
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH, PROPRIOCEPTION AND FLEXIBILITY versus NO TREATMENT

Protocol outcome 1: Pain reduction
- Actual outcome: Pain at 12 weeks; Group 1: mean 25.2 (SD 18.3); n=37, Group 2: mean 41.8 (SD 22.5); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 46.2 (19.2); control 51.5 (21.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Pain at 24 weeks; Group 1: mean 33.1 (SD 20.9); n=37, Group 2: mean 44.6 (SD 20); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 46.2 (19.2); control 51.5 (21.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 2: Quality of life
- Actual outcome: SF36 physical at 12 weeks; Group 1: mean 45.2 (SD 5.4); n=37, Group 2: mean 42.9 (SD 5.4); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 41.8 (7.4); 43.6 (7.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 mental at 12 weeks; Group 1: mean 47.7 (SD 8.5); n=37, Group 2: mean 46.1 (SD 10.7); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 46.9 (8.3); no treatment 46.9 (10.5)

Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 physical at 24 weeks; Group 1: mean 44 (SD 7.5); n=37, Group 2: mean 42 (SD 8); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 41.8 (7.4); 43.6 (7.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF36 mental at 24 weeks; Group 1: mean 46.9 (SD 9.1); n=37, Group 2: mean 46.4 (SD 10.13); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 46.9 (8.3); no treatment 46.9 (10.5)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 3: Physical function
- Actual outcome: Neck disability at 12 weeks; Group 1: mean 22.7 (SD 9.3); n=37, Group 2: mean 27.5 (SD 11.4); n=39; Neck Disability Index 0-100 Top=High is poor outcome; Comments: Baseline: exercise 30.1 (9.8); no treatment 29.3 (8.2)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Neck disability at 24 weeks; Group 1: mean 25.1 (SD 12.9); n=37, Group 2: mean 29.4 (SD 12.7); n=39; Neck Disability Index 0-100 Top=High is poor outcome; Comments: Baseline: exercise 30.1 (9.8); no treatment 29.3 (8.2)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: Anxiety at 12 weeks; Group 1: mean 5.5 (SD 3.1); n=37, Group 2: mean 6.7 (SD 3.2); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: exercise 6 (3); no treatment 6.7 (3.7)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Depression at 12 weeks; Group 1: mean 3.8 (SD 2.3); n=37, Group 2: mean 4.9 (SD 3.4); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: exercise 3.8 (2.4); no treatment 4.5 (3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Anxiety at 24 weeks; Group 1: mean 5.5 (SD 3.1); n=37, Group 2: mean 6.7 (SD 3.4); n=39; HADS 0-21 Top=High is poor outcome; Comments: Baseline: exercise 6 (3); no treatment 6.7 (3.7)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Depression at 24 weeks; Group 1: mean 4.1 (SD 2.8); n=37, Group 2: mean 5.4 (SD 4); n=39; HADS 0-21 Top=High is poor outcome; Comments:

Baseline: exercise 3.8 (2.4); no treatment 4.5 (3) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 5: Discontinuation - Actual outcome: Discontinuation at 12 weeks; Group 1: 13/37, Group 2: 10/39 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Use of healthcare services ; Sleep

Study	**Lee 2016[163]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=46)
Countries and setting	Conducted in South Korea; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 10 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Diagnosis is of chronic mechanical neck pain, and between the ages of 18 and 60 years; a neck disability index (NDI) score >20%14); and limited craniocervical and thoracic flexion and extension ROM
Exclusion criteria	Pain of vascular or neurological system origin; neurological deficits, including nerve root signs; spinal stenosis; previous craniocervicalor thoracic spine surgery; or receipt of spinal manipulation therapy within 2 months before the study

Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Not reported. Gender (M:F): Not reported. Ethnicity: Not erported
Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with pain conditions other than chronic widespread pain 5. complex regional pain syndrome: people with pain conditions other than complex regional pain syndrome
Indirectness of population	No indirectness
Interventions	(n=16) Intervention 1: Manual therapy and exercise. All patients received treatment for 35 minutes a day, 3 days a week for 10 weeks. Group A received thoracic manipulation (TM) for 10minutes, deep craniocervical flexor training for 15 minutes, and self-stretching of the levator scapulae and upper trapezius muscles as cool-down exercises for 10 minutes. Before TM, a trained therapist confirmed which joints showed hypomobility using a joint play test and a Spinal Mouse device. Patients lay in the supine position, with flexed knee and hip joints, with their hands clasped on the chest. TM was conducted according to the procedures of Krauss et al., with a high-velocity thrust at low amplitude for 10 minutes. A therapist provided instructions and demonstrations on how to exercise the DCF muscles. The exercise intensity was determined by thepatient's status and was increased progressively. Patients were positioned supine, with the knees bent and with a pressure biofeedback unit placed suboccipitally, to detect increases in pressure elicited by the gentle nodding action of craniocervical flexion. Visual feedback of the pressure level was provided. Patients were instructed how to perform craniocervical flexion and practiced progressive targeting at five incremental levels (increments of 2mmHg) between 22 and 30 mmHg). Isometric contraction was performed for 10 seconds, followed by 5 seconds rest in 10 repetitions. Duration 10 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=30) Intervention 2: Strength and conditioning. Half of the participants received only DCF training for 25 minutes, with self-stretching of the levator scapulae and upper trapezius muscle as a cool-down exercise for 10 minutes. Half of participants performed active ROMself-exercise (neck flexion, extension, lateral flexion, and rotation without provocation of pain) for 35 minutes. Duration 10 weeks. Concurrent medication/care: Not reported.

	Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EXERCISE versus STRENGTH AND CONDITIONING Protocol outcome 1: Pain reduction - Actual outcome: Pain at 10 weeks; Group 1: mean 1.4 (SD 0.5); n=16, Group 2: mean 3.15 (SD 0.8); n=30; VAS 0-10 Top=High is poor outcome; Comments: Baseline values: 5.2 (0.6); 5.2 (0.6) Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing: Protocol outcome 2: Physical function - Actual outcome: Neck disability at 10 weeks; Group 1: mean 6.6 (SD 2.1); n=16, Group 2: mean 15.56 (SD 5.38); n=30; Neck disability index 0-50 Top=High is poor outcome; Comments: Baseline values: 27.6 (4.5); 27.15 (3.6) Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: ; Group 2 Number missing:	
Protocol outcomes not reported by the study	Quality of life at Define; Psychological distress (depression/anxiety) at Define; Use of healthcare services at Define; Sleep at Define; Discontinuation at Define

Study	Mannerkorpi 2009[176]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=42)
Countries and setting	Conducted in Canada; Setting: Medex Medical Exercise Clinics, Ontario, Canada
Line of therapy	Unclear
Duration of study	Intervention time: 20 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Smythe criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	The criteria used for the diagnosis of fibromyalgia were those proposed by Smythe, and included each of the following: 1) widespread aching of more than 3 months duration in more than 3 anatomic sites, 2) local tenderness at 12 of 14 specified fibrositic tender points, 3) disturbed sleep with morning fatigue and stiffness, 4) absence of traumatic, neurologic, muscular, infectious, osseous. endocrine, or other rheumatic conditions, and 5) normal Wintrobe erythrocyte sedimentation rate, creatinine phosphokinase level, latex fixation test results, antinuclear antibody factor, and thyroid-stimulating hormone level.
Exclusion criteria	Nonsteroidal anti-inflammatory drugs, hypnotic drugs, and antidepressant agents were discontinued for a minimum of 3 weeks before entry into the trial. Patients treated with amitriptyline within the previous 3 months were excluded from this study. Only acetaminophen was permitted during the study, and each dose was recorded
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 42(9.6) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Duration of pain not specified
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Aerobics - Other aerobic exercise. 60 minutes 3 times weekly. After a 10-minute preliminary warm-up exercise, patients were subjected to sustained heart rate elevation training through the use of a bicycle ergometer (Tunturi, Turku, Finland). Heart rates were maintained in excess of 150 beats per minute for gradually increasing time periods, and were monitored with a Sanyo HRM-97E digital pulse meter. Duration 20 weeks. Concurrent medication/care: All patients were instructed to refrain from additional exercise beyond the supervised

	program. Indirectness: No indirectness (n=20) Intervention 2: Flexibility. Participants met at similar intervals but at different times over the same 20-week observation period. FLEX instruction was administered in a group setting by the same instructors as for CVR training, but consisted only of flexibility maneuvers, such that sustained heart rate responses greater than 115 beats per minute were not attained. Duration 20 weeks. Concurrent medication/care: All patients were instructed to refrain from additional exercise beyond the supervised program. Indirectness: No indirectness
Funding	Funding not stated (Not specified)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC EXERCISE (STAIONARY CYCLING) versus FLEXIBILITY Protocol outcome 1: Pain reduction - Actual outcome: VAS at 20 weeks; Group 1: mean 46.9 (SD 30.6); n=18, Group 2: mean 47.4 (SD 17); n=20; VAS 0-100 Top=High is poor outcome; Comments: Baseline: difference 70.1(15.8); 56.3(19.2) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: VAS difference of over 10;	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Martin 1996[179]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=60)
Countries and setting	Conducted in Canada; Setting: Sports medicine clinic at the university of Calgary
Line of therapy	Unclear
Duration of study	Intervention time: 6 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Diagnosis of FMS according to the ACR criteria
Exclusion criteria	Ant conditions that precluded involvement in an exercise program or if they were taking any medication that would significantly affect their normal physiological response to exercise
Recruitment/selection of patients	Referred by rheumatologists at the University of Calgary, by family practitioners and through the Calgary FM support group
Age, gender and ethnicity	Age - Mean (SD): 44.8(9.8) years. Gender (M:F): 1:37. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Duration of pain 9.2(7.2) years
Indirectness of population	No indirectness
Interventions	(n=30) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. Participants met 3 times a week for 6 weeks and participated in 1 h supervised exercise program. The program included 20 minutes walking at a pace sufficient to raise heart rate to 60-80% of maximum, 20 minutes of flexibility and strength training for multiple muscles. Duration 6 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=30) Intervention 2: Psychological intervention - Relaxation. 3 times per week for 6 week, supervised relaxation program for 1 hour in a quiet room. Patients were taught visualization, yoga and autogenic relaxation by experienced instructors. Duration 6 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Study funded by industry (The Canadian Fitness and Lifestyle Research Institute)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH AND FLEXIBILITY EXERCISE versus RELAXATION

Protocol outcome 1: Quality of life
- Actual outcome: Fibromyalgia impact questionnaire at 6 weeks; Group 1: mean 388.06 (SD 149.68); n=18, Group 2: mean 433.11 (SD 115.55); n=20; FIQ 0-1000 Top=High is poor outcome; Comments: Baseline: 418.63(184.58); 407.44(124.38)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Illness, lack of efficacy, lack of time; Group 2 Number missing: 10, Reason: Illness, lack of efficacy, lack of time

Protocol outcome 2: Discontinuation
- Actual outcome: Discontinuation at 6 weeks; Group 1: 12/30, Group 2: 10/30; Comments: Multiple reasons (illness, lack of efficacy, lack of time)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study (subsidiary papers)	**Mcbeth 2012[181] (Beasley 2015[28])**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=442 (4 arms, only 3 arms (330 participants) relevant to this review))
Countries and setting	Conducted in United Kingdom; Setting: Research nurse led clinic
Line of therapy	Unclear
Duration of study	Intervention time: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	(1) chronic widespread pain for which they had consulted their physician within the last year
Exclusion criteria	Severe psychiatric disorder, contraindications for exercise such as chest pain, syncope or uncontrolled epilepsy, or a condition for which the interventions were not indicated, e.g., metastatic cancer.
Recruitment/selection of patients	From 8 general practices in Aberdeen, Scotland and Macclesfield, Northwest England

Age, gender and ethnicity	Age - Mean (SD): 55.7(12.5) years. Gender (M:F): 70:148. Ethnicity: Not specified
Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with pain conditions other than chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with chronic widespread pain 5. complex regional pain syndrome: people with pain conditions other than complex regional pain syndrome
Indirectness of population	No indirectness
Interventions	(n=109) Intervention 1: Aerobics - Other aerobic exercise. Gym based programme. Induction session followed by 6 (monthly) instructor led appointments for program reassessment. Exercise intensity was increased until exercise levels were sufficient to achieve 40-85% of heart rate, and this was individualised for each participant so actual intensity of treatment varied. Recommended session length 20 to 60 minutes. Duration 6 months. Concurrent medication/care: Participants free to engage in additional exercises (e.g. strength and flexibility) in addition to intervention. Indirectness: No indirectness (n=112) Intervention 2: Psychological intervention - Cognitive behavioural therapy. Telephone-delivered cognitive behaviour therapy (TCBT): initial assessment (45-60mins) followed by 7 weekly sessions (30-45mins each), 1 session at three months, and 1 session at 6 months. Intervention delivered by 4 therapists accredited by the British Association for Behaviour and Cognitive Psychotherapies. Therapists conducted a patient-centred assessment, developed shared understanding and formulation of the participants' problem(s) and identified two to three patient-defined goals. Patients also received a self-management CBT manual that included: behavioural activation, cognitive restructuring, unhelpful thinking and lifestyle changes. Duration 6 months. Concurrent medication/care: Participants free to engage in additional exercises (e.g. strength and flexibility) in addition to intervention. Indirectness: No indirectness (n=109) Intervention 3: Usual care. Usual care from family physician, although precise care delivered, if any, was not recorded. Duration 6 months. Concurrent medication/care: Participants free to engage in additional exercises (e.g. strength and flexibility) in addition to intervention. Indirectness: No indirectness
Funding	Academic or government funding (Arthritis Research UK)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: OTHER AEROBIC EXERCISE versus COGNITIVE BEHAVIOURAL THERAPY

Protocol outcome 1: Quality of life
- Actual outcome: EQ-5D at 9 months (including 6 month intervention); Group 1: mean 0.705 (SD 0.238); n=81, Group 2: mean 0.645 (0.262); n=83; EQ-5D, Top=High is good outcome; Comments: Baseline: 0.649(0.216); 0.686(0.209); difference of over 0.03 at baseline which is the established MID for EQ-5D

Risk of bias: All domain - high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 10, Reason: NR; Group 2 Number missing: 11, Reason: NR Protocol outcome 2: Sleep - Actual outcome: Sleep scale at 9 months (including 6 month intervention); Group 1: mean 12.7 (SD 4.9); n=99, Group 2: mean 12.4 (SD 5.7); n=91; The Sleep Scale 0-20 Top=High is poor outcome; Comments: 13.7(5.9); 13.3(5.5) Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 10, Reason: NR; Group 2 Number missing: 11, Reason: NR Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at 6 months (post-intervention); Group 1: 10/109, Group 2: 21/112 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0, Reason: NA; Group 2 Number missing: 0, Reason: NA RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: OTHER AEROBIC EXERCISE versus USUAL CARE Protocol outcome 1: Quality of life - Actual outcome: EQ-5D at 9 months (including 6 month intervention); Group 1: mean 0.705 (SD 0.238); n=81, Group 2: mean 0.754(0.214); n=71; EQ-5D, Top=High is good outcome; Comments: Baseline: 0.649(0.216); 0.730(0.151); difference of over 0.03 at baseline which is the established MID for EQ-5D Risk of bias: All domain - high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 10, Reason: NR; Group 2 Number missing: 11, Reason: NR Protocol outcome 2: Sleep - Actual outcome: Sleep scale at 9 months (including 6 month intervention); Group 1: mean 12.7 (SD 4.9); n=99, Group 2: mean 13.1 (SD 5.4); n=98; Sleep scale 0-20 Top=High is poor outcome; Comments: 13.7(5.9); 13.8(5.5) Risk of bias: All domain - high, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 10, Reason: NR; Group 2 Number missing: 11, Reason: NR Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at 6 months (post-intervention); Group 1: 10/109, Group 2: 11/109 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing:0 , Reason: NA; Group 2 Number missing:0 , Reason: NA
Protocol outcomes not reported by the study	Pain reduction; Physical function; Psychological distress (depression/anxiety); Use of healthcare services

Study	Mccain 1986[182]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=34)
Countries and setting	Conducted in Canada; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 20 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Smythe's criteria
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	People with fibrositis/fibromyalgia
Exclusion criteria	Not reported
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Flexibility group 46±8; cardiovascular group 39±10. Gender (M:F): 6/28. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Indirectness of population	No indirectness
Interventions	(n=18) Intervention 1: Aerobics - Other aerobic exercise. Three times a week programme. Participants had sustained heart rate elevated training via a bicycle ergometer. Heart rates were maintained in excess of 150 beats per minute for gradually incremental durations. Duration 20 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=16) Intervention 2: Flexibility. Participants met at similar intervals to the aerobic group. Exercise consisted of flexibility maneuvers such that sustained heart rate responses were over 115 beats per minute were not attained. Duration 20 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: OTHER AEROBIC EXERCISE versus FLEXIBILITY	

Protocol outcome 1: Pain reduction - Actual outcome: Pain at End of treatment; Group 1: mean -23.2 (SD 30.6); n=18, Group 2: mean -8.7 (SD 21); n=16; VAS 0-100 Top=High is poor outcome; Comments: Baseline: aerobic 68.6±15; flexibility 58.5±15 Risk of bias: All domain - High, Selection - Very high, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Sex: flexibility 0 males; aerobic 6 males. Duration of pain (month): flexibility 41±41; aerobic 34±54;	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Michalsen 2012[192]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=77)
Countries and setting	Conducted in Germany; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 10 weeks
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged between 18 and 60 years, suffering from a minimum score of 4 out of 10 on the VAS scale, painful restriction of cervical mobility for at least 3 months.
Exclusion criteria	Invasive surgery within the last 6 weeks or treatments planned in the next 10 weeks. Excluded those whose neck pain was complicated or attributable to specific underlying disease. Also excluded those with a coexisting serious comorbidity or those participating in another study or any previous experience with yoga
Recruitment/selection of patients	Press release offering participation in the study
Age, gender and ethnicity	Age - Mean (SD): 47.9(7.9) years. Gender (M:F): 10:67. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain (Chronic primary cervical pain). 3. chronic visceral pain: 4. chronic widespread pain:
Extra comments	Mean duration of pain 6.55(5.3) years
Indirectness of population	No indirectness
Interventions	(n=38) Intervention 1: Mind-body exercises - Yoga. Weekly 90 minute yoga classes using a wide range of postures to enhance flexibility, alignment, stability and mobility in muscles joints and tendons, run by a certified yoga instructor and physician. The exercises specifically addressed neck pain complaints and each class built up on the previous one. Subjects were requested to practice at home for 10-15 minutes, 2 to 3 times a week. Duration 9 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=39) Intervention 2: No treatment. Waiting list control. A standard self care manual about exercise and education for chronic neck pain was given. The manual described exercises that could be carried out to aid chronic neck pain and participants were asked to practice at home for 10-15 minutes at least 3 times a week. Duration 9 weeks. Concurrent

	medication/care: Not specified. Indirectness: No indirectness
Funding	Study funded by industry (Carl and Veronica Cartsens Foundation)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: YOGA versus WAITING LIST CONTROL

Protocol outcome 1: Pain reduction
- Actual outcome: VAS pain scores at 10 weeks; Group 1: mean 13 (SD 11.6); n=38, Group 2: mean 34.4 (SD 21.2); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 44.3(20.1); 41.9(21.9)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Adverse event (n=5), noncompliance (n=5), other reasons ; Group 2 Number missing: 11, Reason: Adverse event (n=1), study noncompliance (n=10)

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 physical component summary score at 10 weeks; Group 1: mean 46.5 (SD 7.3); n=38, Group 2: mean 41.3 (SD 6.4); n=39; SF-36 summary score 0-100 Top=High is good outcome; Comments: Baseline: 38.5(7.1); 40.7(6)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Adverse event (n=5), noncompliance (n=5), other reasons ; Group 2 Number missing: 11, Reason: Adverse event (n=1), study noncompliance (n=10)
- Actual outcome: SF-36 mental component summary score at 10 weeks; Group 1: mean 47.6 (SD 10.4); n=38, Group 2: mean 40.6 (SD 10.7); n=39; Comments: Baseline: 44.3(11.7); 43(10.4)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Adverse event (n=5), noncompliance (n=5), other reasons ; Group 2 Number missing: 11, Reason: Adverse event (n=1), study noncompliance (n=10)

Protocol outcome 3: Physical function
- Actual outcome: Neck disability index score at 10 weeks; Group 1: mean 18.4 (SD 4); n=38, Group 2: mean 24.5 (SD 6); n=39; NDI 0-50 Top=High is poor outcome; Comments: Baseline: 25.4(5.2); 25.8(5.5)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Adverse event (n=5), noncompliance (n=5), other reasons ; Group 2 Number missing: 11, Reason: Adverse event (n=1), study noncompliance (n=10)

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: CES-D depression score at 10 weeks; Group 1: mean 8.4 (SD 5.6); n=38, Group 2: mean 18 (SD 10.4); n=39; CES-D ? Top=High is poor outcome; Comments: Baseline: 17.1(10.3); 17.1(8.2)

Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 12, Reason: Adverse event (n=5), noncompliance (n=5), other reasons ; Group 2 Number missing: 11, Reason: Adverse event (n=1), study noncompliance (n=10) Protocol outcome 5: Discontinuation - Actual outcome: Discontinuation at 10 weeks; Group 1: 12/38, Group 2: 11/39 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness, Comments: Unclear if dropped out of study or intervention; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Use of healthcare services ; Sleep

Study (subsidiary papers)	Munguia-izquierdo 2007[200] (Munguia-izquierdo 2008[199])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=60)
Countries and setting	Conducted in Spain; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 16 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 18 to 60 years
Exclusion criteria	The exclusion criteria included the presence of subjects with a history of morbid obesity, known cardiopulmonary diseases, endocrine or allergic disturbances uncontrolled, severe trauma, frequent migraines, inflammatory rheumatic diseases, and severe psychiatric illness. In addition, subjects with other diseases that prevent physical loading and those who were pregnant were also omitted. Finally, those FM women who attended another type of physical or psychologic therapy were excluded to avoid possible interactions with the present trial. Patients with a history of regular physical activity more strenuous than slow-paced walking a maximum of 2 times a week over 4 months before study entry were excluded from the final analysis according to the criteria of Schachter et al.
Recruitment/selection of patients	From a local FMS association in Spain
Age, gender and ethnicity	Age - Mean (SD): 48 (7.5) year. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Mean duration of symptoms 14(9) years
Indirectness of population	No indirectness
Interventions	(n=35) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. The exercise group trained in a chest-high warm pool (32°C)3 times a week for 16 weeks. Each session included 10 minutes of warming up with slow walks and mobility exercises, 10 to 20 minutes of strength exercises developed at a slow pace using water and aquatic materials as a means of resistance including a stepped progression during the program, 20 to 30 minutes of aerobic exercises developed progressively at intensity sufficient to achieve 50% to 80% of the age predicted maximum heart rate equation (220 – age), and 10 minutes of cooling down with low-intensity and relaxation exercises. Heart rate was

	monitored with a pulse meter. The intervention program met the minimum training standards of the American College of Sports Medicine. Duration 16 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=25) Intervention 2: Usual care. The control group was instructed not to change their habits regarding physical activities during the period. Usual activities and medication allowed. . Duration 16 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (European Social Funds and regional government of Aragon)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus USUAL CARE

Protocol outcome 1: Quality of life
- Actual outcome: Fibromyalgia impact questionnaire at 16 weeks; Group 1: mean -4.8 (SD 9.67); n=34, Group 2: mean -0.9 (SD 9.62); n=24; Comments: Baseline: 68.1(12.4); 63.6(16.7)
SDs calculated from CIs. For change scores: -8.1 to -1.6; -4.8 to 2.9
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: PSQI difference at baseline of 3 (0-21 scale), difference in endurance strength tests, FIQ difference of 5; Group 1 Number missing: 6, Reason: Dropped out, no further details; Group 2 Number missing: 1, Reason: Dropped out, no further details

Protocol outcome 2: Psychological distress (depression/anxiety)
- Actual outcome: State anxiety inventory at 16 weeks; Group 1: mean -0.3 (SD 9.22); n=34, Group 2: mean -0.4 (SD 10.5); n=24; Comments: Baseline: 52.2(10.8); 47.6(11)
SDs calculated from CIs: -3.4 to 2.8, -4.6 to 3.8
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: PSQI difference at baseline of 3 (0-21 scale), difference in endurance strength tests, FIQ difference of 5; Group 1 Number missing: 6, Reason: Dropped out, no further details; Group 2 Number missing: 1, Reason: Dropped out, no further details
- Actual outcome: Pittsburg sleep quality index at 16 weeks; Group 1: mean -1.7 (SD 2.5); n=34, Group 2: mean 0.5 (SD 2.12); n=24; PSQI 0-21 Top=High is poor outcome; Comments: Baseline: 13.4(4.4); 10.4(5)
SDs calculated from CIs (-2.6 to -0.9, -0.4 to 1.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: PSQI difference at baseline of 3 (0-21 scale), difference in endurance strength tests, FIQ difference of 5; Group 1 Number missing: 6, Reason: Dropped out, no further details; Group 2 Number missing: 1, Reason: Dropped out, no further details

Protocol outcome 3: Discontinuation
- Actual outcome: Discontinuation at 16 weeks; Group 1: 6/35, Group 2: 1/24; Comments: Drop out during trial, not attending trial or assessments.

Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: Unclear if discontinued intervention or study; Baseline details: PSQI difference at baseline of 3 (0-21 scale), difference in endurance strength tests, FIQ difference of 5;	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Use of healthcare services ; Sleep

Study	Norouzi 2019[206]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=60)
Countries and setting	Conducted in Iran; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Female, aged between 30 and 40 years, meeting the 1990 American College of Rheumatology criteria for FM (Bigatti & Cronan, 2002), willing to participate in the study and to provide informed consent, willing and able to comply with the study procedures, and having a score on the SCL-90R (Symptom Check List-90-revised) equal or higher than 1 as mean score
Exclusion criteria	The presence of metabolic abnormalities, neurological disorders, drug abuse, uncontrolled blood pressure, uncontrollable blood glucose, regular exercise history (≥ twice per week) during the last six months and severe somatic (e.g., cancer) or psychiatric (e.g., psychotic) diseases
Recruitment/selection of patients	patients were recruited from the FM Association of Urmia (Iran)
Age, gender and ethnicity	Age - Mean (SD): Dancing group: 35.5 (2.42); aerobic group: 35.5 (2.42); control group: 35.4 (2.80) . Gender (M:F): Females only. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain 5. complex regional pain syndrome:
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Aerobics - Dancing. The Zumba dancing program consisted of three weekly 60 minute training sessions. Zumba dancing was taught by a professional coach in a large room with air conditioning and was performed based on Xbox 360 Kinect software. Each session consisted of five minutes of warming up, followed by active upper and lower body movements. This was followed by

	approximately 50 minutes of Zumba dancing, which included movements up to the maximum angle of the upper and lower limbs with a distinction between the pelvic and shoulder movements (shoulder belt). At the end, a 5-min cooling down was performed; this included stretching large muscles and holding them for approximately 30 seconds. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=20) Intervention 2: Aerobics - Walking. Participants practiced on a walking treadmill (RodbyTM , RL 1600E, Enhorna, Sweden) three times per week for 60 minutes. Each training session consisted of 60 minutes of walking with an intensity of 60-75% of estimated maximum heart rate (220 minus age formula). Participants' heart rates were measured by an electric pulse meter. In addition, perceive exertion was measured with the Borg scale of perceived exertion (Borg, 1998). It is used to modulate or refine a prescribed exercise intensity. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=20) Intervention 3: Other. Participants assigned to the control group gathered at the clinic 3 time per 2eek for group meetings. During this time, they could talk with each other and medical staff members. Additionally, they were asked to maintain their current daily physical activity levels, and to refrain from additional exercise or sport activities. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Academic or government funding (Urnia University)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: DANCING versus WALKING

Protocol outcome 1: Psychological distress (depression/anxiety) at 12 weeks
- Actual outcome: Depression at Post intervention; Group 1: mean 13.42 (SD 1.15); n=20, Group 2: mean 21.33 (SD 2.01); n=20; Persian version of the Beck Depression Inventory 0-63 Top=High is poor outcome; Comments: Baseline scores: dancing group 31.99 (3.42); walking group 30.21 (2.98)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Physical function at 12 weeks
- Actual outcome: Physical function at Post intervention; Group 1: mean 9.23 (SD 1.24); n=20, Group 2: mean 9.51 (SD 1.33); n=20; Timed up and go Top=High is good outcome; Comments: Baseline scores: dancing group 9.99 (1.32); walking group 9.92 (1.21)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Discontinuation at 12 weeks
- Actual outcome: Discontinuation at 12 weeks (Post intervention); Group 1: 0/20, Group 2: 0/20
Risk of bias: All domain – Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: DANCING versus ATTENTION CONTROL

Protocol outcome 1: Psychological distress (depression/anxiety) at 12 weeks
- Actual outcome: Depression at 12 weeks (Post intervention); Group 1: mean 13.42 (SD 1.15); n=20, Group 2: mean 30.14 (SD 3.02); n=20; Persian version of the Beck Depression Inventory 0-63 Top=High is poor outcome; Comments: Baseline scores: dancing group 31.99 (3.42); control group 30.98 (3.16)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Physical function at 12 weeks
- Actual outcome: Physical function at Post intervention; Group 1: mean 9.23 (SD 1.24); n=20, Group 2: mean 9.99 (SD 1.52); n=20; Timed up and go Top=High is good outcome; Comments: Baseline scores: dancing group 9.99 (1.32); control group 9.98 (1.26)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Discontinuation at 12 weeks
- Actual outcome: Discontinuation at 12 weeks (Post intervention); Group 1: 0/20, Group 2: 0/20
Risk of bias: All domain – Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: WALKING versus ATTENTION CONTROL

Protocol outcome 1: Psychological distress (depression/anxiety) at 12 weeks
- Actual outcome: Depression at 12 weeks (Post intervention); Group 1: mean 21.33 (SD 2.01); n=20, Group 2: mean 30.14 (SD 3.02); n=20; Persian version of the Beck Depression Inventory 0-63 Top=High is poor outcome; Comments: Baseline scores: walking group 30.21 (2.98); control group 30.98 (3.16)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Physical function at 12 weeks
- Actual outcome: Physical function at Post intervention; Group 1: mean 9.51 (SD 1.33); n=20, Group 2: mean 9.99 (SD 1.52); n=20; Timed up and go Top=High is good outcome; Comments: Baseline scores: walking group 9.92 (1.21); control group 9.98 (1.26)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Discontinuation at 12 weeks - Actual outcome: Discontinuation at 12 weeks (Post intervention); Group 1: 0/20, Group 2: 0/20 Risk of bias: All domain – Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Pain reduction; Quality of life; Physical function; Use of healthcare services; Sleep

Study	Panton 2009[210]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=27)
Countries and setting	Conducted in USA; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 16 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women with fibromyalgia
Exclusion criteria	Uncontrolled hypertension, uncontrolled diabetes, active heart disease, osteoporosis, spinal trauma, spinal instability involving neurologic deficit, known history of cancer, long-term corticosteroid use, endocrine disease, anticoagulant therapy, bleeding disorders, history of stroke, physical examination or radiologic findings that would contraindicate chiropractic manual treatment procedures, currently participating in an exercise programme and/or currently under the care of a chiropractic physician
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Exercise only: 50±7; exercise + manual therapy 47±12. Gender (M:F): Define. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Indirectness of population	No indirectness
Interventions	(n=15) Intervention 1: Strength. Resistance training. Participants met twice a week. Resistance training was chosen to maximise strength gains. Participants performed one set of 8-12 repetitions twice a week on 10 exercises, using 9 resistance machines that included the chest press, leg extension, leg curl, leg press, arm curl, seated dip, overhead press, seated row, abdominal crunch, and one body weight exercise for the lower back extension. Participants began training at approximately 50% of their initial 1-RM measurement and were slowly progressed to approximately 100% of their initial 1RM by the end of the 16 weeks. Once 12 repetitions were completed on 2 consecutive workouts, weights were increased by 5-10 pounds for upper and lower body respectively. Duration 16 weeks. Concurrent medication/care: Participants met once, 4 weeks into the study, with a health educator to re-emphasize the goals or the programme and to address impediments to adherence. Indirectness: No indirectness

	(n=12) Intervention 2: Manual therapy and exercise. Exercise as in the Strength group, plus manual therapy. Participants met twice a week for exercise, and twice a week for chiropractic treatment. Chiropractic treatment consisted of standardised ischemic compression and diversified chiropractic spinal adjustments. Treatments began with 5 minutes of ischemic compression to tender points on the back of the neck and spine. The technique developed by Travell and Simons was followed. Briefly pressure was applied with thumbs over tender points until the patient reacted to the pressure. The pressure was sustained for 10 seconds. This technique was continued throughout the 16 weeks with increasing pressure until an application of 4kg of digital pressure was reached. This 4kg of pressure was continued until the completion of the study. The next 5 minutes consisted of diversified chiropractic spinal adjustments. These adjustments consisted of short lever, low amplitude, high velocity thrusts. Cervical adjustments were performed with the participant in a supine position utilising an index finger proximal or distal interphalangeal joint contact point and a laminar segmental contact point. The thoracic adjustments were performed with the participant in a prone position utilising a double thenar contact point and a double transverse process segmental contact point. The lumbar adjustments were performed with the participant in a lateral decubitus position utilising a pisiform contact point and a mammillary segmental contact point. Target joints were determined at each visit through static and motion palpitation. Duration 16 weeks. Concurrent medication/care: Participants met once, 4 weeks into the study, with a health educator to re-emphasize the goals or the programme and to address impediments to adherence. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus MANUAL THERAPY AND EXERCISE

Protocol outcome 1: Quality of life
- Actual outcome: Fibromyalgic Impact Questionnaire at End of treatment; Group 1: mean 45.9 (SD 14.2); n=10, Group 2: mean 46.9 (SD 15.9); n=11; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: exercise 60.3±8.3; exercise + manual therapy 60.2±10.8
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - reasons for discontinuation: lack of time (n=3); not wanting to continue with massage therapy (n=1); family related issues (n=2); Indirectness of outcome: No indirectness ; Baseline details: FM duration (years): exercise 4±4; exercise + manual 7±5; Group 1 Number missing: 5; Group 2 Number missing: 1

Protocol outcome 2: Physical function
- Actual outcome: Physical function at End of treatment; Group 1: mean 61 (SD 14); n=10, Group 2: mean 67 (SD 9); n=11; Comments: Baseline: exercise 55±11; exercise + manual therapy 55±6
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low, Comments - reasons for discontinuation: lack of time (n=3); not wanting to continue with massage therapy (n=1); family related issues (n=2); Indirectness of outcome: No indirectness ; Baseline details: FM duration (years): exercise 4±4; exercise + manual 7±5; Group 1 Number missing: 5; Group 2 Number missing: 1

Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at End of treatment; Group 1: 5/15, Group 2: 1/12 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: FM duration (years): exercise 4±4; exercise + manual 7±5;	
Protocol outcomes not reported by the study	Pain reduction ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Rendant 2011[222]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=123)
Countries and setting	Conducted in Germany; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 months
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged between 20-60 years of age. The minimum duration of neck pain had to be between 6 months and 5 years and the intensity of the average neck pain over the last 7 days had to be more than 40mm on a 100mm VAS. Patients had to have normal cervical spine flexibility, and predominantly neck pain. If additional back pain was reported, neck pain had to be predominant.
Exclusion criteria	Acute or chronic disorders (physical and mental) that disqualified study participation, pregnancy, participation in qigong or exercise therapy during the last 6 months, whiplash-associated or cancer causing neck flame, inflammatory arthritis column surgery or prolapsed vertebral disc, regular intake of analgesics, planned start of physiotherapy, taking up activities which have a positive influence on the neck pain during the study participation, or participation in another study during the last 6 months
Recruitment/selection of patients	Participants were recruited in Berlin using information material, intranet platforms of the university and other companies (reaching more around 20,000 employees). Also a newspaper advertisement was placed.
Age, gender and ethnicity	Age - Mean (SD): Qigong 44.7±10.8; exercise 44.4±10.9; waiting list 47.8±10.8. Gender (M:F): 15/107. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable Not applicable
Indirectness of population	No indirectness
Interventions	(n=42) Intervention 1: Mind-body exercises - Qigong. Qigong was performed by three qualified teachers certified by the German Qigong Society. Each session of qigong took 90 minutes. Neiyanggong, a special silent and slow form of qigong was chosen by the therapist in a consensus process. The lessons started with up to 12 neck exercises followed by 9 exercises for the shoulder and finished with breathing and moving exercises. There were 18 sessions over a period of 6 months (1 session per week in the first 3 months, and biweekly sessions in the following 3 months). Duration 6 months.

	Concurrent medication/care: Not reported. Indirectness: No indirectness (n=39) Intervention 2: Strength and flexibility - Other mixed modality exercise. Exercise therapy was carried out by 6 qualified therapists. The exercises was based on a standard programme for chronic pain. Each lesson started with a warm up using a softball and was followed by repeated active cervical rotations and strengthening and flexibility exercises. The individual's pain level was not exceeded. There were 18 sessions over a period of 6 months (1 session per week in the first 3 months, and biweekly sessions in the following 3 months). Duration 6 months. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=41) Intervention 3: No treatment. Waiting list control participants received no intervention. Duration 6 months. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus STRENGTH AND FLEXIBILITY

Protocol outcome 1: Pain reduction
- Actual outcome: Average neck pain at End of treatment; Group 1: mean 26.7 (SD 19.6); n=39, Group 2: mean 27.4 (SD 17.05); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 57.7±13.5; exercise 57.5±15.5
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 2

Protocol outcome 2: Quality of life
- Actual outcome: Quality of life - physical component at End of treatment; Group 1: mean 47 (SD 7.65); n=39, Group 2: mean 44.7 (SD 7.55); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 43.1±7.5; exercise 43.7±6.9
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 2
- Actual outcome: Quality of life - mental component at End of treatment; Group 1: mean 47.4 (SD 10.2); n=39, Group 2: mean 47.8 (SD 8.75); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 46±9.6; exercise 45.5±11.8
Risk of bias: All domain - ; Indirectness of outcome: No indirectness

Protocol outcome 3: Physical function
- Actual outcome: Neck pain/disability at End of treatment; Group 1: mean 30 (SD 10.36); n=39, Group 2: mean 31.5 (SD 14.49); n=35; NPDS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 44±12.7; exercise 39.5±15.4
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 2

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at End of treatment; Group 1: 3/42, Group 2: 4/39
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus NO TREATMENT

Protocol outcome 1: Pain reduction
- Actual outcome: Average neck pain at End of treatment; Group 1: mean 26.7 (SD 19.59); n=39, Group 2: mean 41 (SD 20.23); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 57.7±13.5; wait list: 53.4±13.2
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 2

Protocol outcome 2: Quality of life
- Actual outcome: Quality of life - physical component at End of treatment; Group 1: mean 47 (SD 7.65); n=39, Group 2: mean 43.1 (SD 7.17); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 43.1±7.5; waiting list 43.3±7.8
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 2
- Actual outcome: Quality of life - mental component at End of treatment; Group 1: mean 47.4 (SD 10.2); n=39, Group 2: mean 45.4 (SD 8.76); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 46±9.6; waiting list 48.6±9.8
Risk of bias: All domain - ; Indirectness of outcome: No indirectness

Protocol outcome 3: Physical function
- Actual outcome: Neck pain/disability at End of treatment; Group 1: mean 30 (SD 10.36); n=39, Group 2: mean 38.1 (SD 13.7); n=39; NPDS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 44±12.7; waiting list 53.4±13.2
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 2

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at End of treatment; Group 1: 3/42, Group 2: 2/41
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH AND FLEXIBILITY versus NO TREATMENT

Protocol outcome 1: Pain reduction - Actual outcome: Average neck pain at End of treatment; Group 1: mean 27.4 (SD 17.05); n=35, Group 2: mean 41 (SD 20.23); n=39; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 57.5±15.5; waiting list 53.4±13.2 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 2 Protocol outcome 2: Quality of life - Actual outcome: Quality of life - physical component at End of treatment; Group 1: mean 44.7 (SD 7.55); n=35, Group 2: mean 43.1 (SD 7.17); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 43.7±6.9; waiting list 43.3±7.8 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 2 - Actual outcome: Quality of life - mental component at End of treatment; Group 1: mean 47.8 (SD 8.75); n=35, Group 2: mean 45.4 (SD 8.76); n=39; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 45.5±11.8; waiting list 48.6±9.8 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 2 Protocol outcome 3: Physical function - Actual outcome: Neck pain/disability at End of treatment; Group 1: mean 31.5 (SD 14.49); n=35, Group 2: mean 38.1 (SD 13.7); n=39; NPDS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 39.5±15.4; waiting list 43.2±16.1 Risk of bias: All domain - ; Indirectness of outcome: No indirectness Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at End of treatment; Group 1: 4/39, Group 2: 2/41 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Richards 2002[224]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=136)
Countries and setting	Conducted in United Kingdom; Setting: Health living centre
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 week intervention + 40 weeks follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR 1990
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Men and women aged 18-70 years who had fibromyalgia according to the criteria of the American College of Rheumatology 1990
Exclusion criteria	Of those eligible people with alternative diagnoses that explained symptoms or were unable to attend classes (lived to far away, too busy, other reasons) were excluded. Other exclusion criteria were severe pulmonary, cardiovascular, renal or neurological disease precluding involvement in aerobic exercise and inability to cooperate, but no participants were excluded for these reasons.
Recruitment/selection of patients	From rheumatology clinics in a teaching hospital between 1997 to 1998
Age, gender and ethnicity	Age - Median (range): 46.5 years. Gender (M:F): 10:126. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Median duration of disease 5 years
Indirectness of population	No indirectness
Interventions	(n=69) Intervention 1: Aerobics - Other aerobic exercise. Both groups met in hour long classes of up to 18 individuals twice weekly for 12 weeks. Participants continued their medication at entry. They received standardised advice including an explanation of fibromyalgia and encouragement and were told that the exercise offered through prescription would improve their condition. Each week at the classes all individuals received an information leaflet covering an aspect of their condition. The interventions were carried out by personal trainers blinded to the hypothesis of the trial. Exercise therapy comprised an individualised aerobic exercise programme, mostly walking on treadmills and cycling on exercise bicycles. Each individual was encouraged to increase the amount of exercise steadily as tolerated. When

	people first started classes they usually did two periods of exercise per class lasting six minutes. By 12 weeks they were doing two periods of 25 minutes at an intensity that made them sweat slightly while being able to talk comfortably in complete sentences. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=67) Intervention 2: Mixed modality exercise - Other mixed modality exercise. Relaxation and flexibility comprised upper and lower limb stretches and relaxation techniques based on the published regimen by Ost. As the classes continued more techniques were introduced progressing through progressive muscle relaxation, release only relaxation and visualisation, cue controlled relaxation, and differential relaxation. This occupied the whole one hour class, twice weekly. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Research training fellowship (NHS))
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC EXERCISE versus STRETCHING AND RELAXATION Protocol outcome 1: Quality of life - Actual outcome: Fibromyalgia impact questionnaire at 12 months (including 12 week intervention and 40 week follow up); Group 1: mean 55.6 (SD 15.8); n=68, Group 2: mean 56 (SD 13.8); n=65; FIQ 0-100 Top=High is poor outcome; Comments: Baseline: 59.6(56.6 to 62.5); 56.6(53.6 to 59.5) SDs calculated from CIs (52.4 to 59.9; 52.8 to 59.5) Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: Not specified; Group 2 Number missing: 2, Reason: Not specified Protocol outcome 2: Discontinuation - Actual outcome: Discontinuation at 12 months (including 12 week intervention and 40 week follow up); Group 1: 12/69, Group 2: 12/67 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: Not specified; Group 2 Number missing: 2, Reason: Not specified	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Salo 2012[230]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=101)
Countries and setting	Conducted in Finland; Setting: not reported
Line of therapy	Unclear
Duration of study	Intervention time: 12 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age 25-53 years, presence of a non-specific neck pain for more than 6 months and perceived neck pain greater than 30mm on a VAS
Exclusion criteria	Specific disorders of the cervical spine, such as disk prolapse, spinal stenosis, postoperative conditions, severe trauma and hypermobility; spasmodic tortcollis; frequent migraine; peripheral nerve entrapment; fibromyalgia; shoulder disease; inflammatory rheumatic disease; severe psychiatric illness or other difficult mental conditions; and pregnancy
Recruitment/selection of patients	not reported
Age, gender and ethnicity	Age - Mean (SD): stretching: 40 (10); stretching + strength: 41 (9). Gender (M:F): 10/91. Ethnicity: not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable Not applicable
Extra comments	Duration of neck pain (months): stretching 60 (17); stretching + strength 64 (17)
Indirectness of population	No indirectness
Interventions	(n=49) Intervention 1: Mixed modality exercise – Strength and flexibility. Combined strength training and stretching. Participants used elastic rubber bands attached to a leather strap running around the head for the seated isometric neck strength exercises. During each session they performed a series of 15 repetitions directly forward, obliquely toward the right and left and directly backwards. The movement was from the hips with the spine held erect. The aim was to reach the level of resistance that was 80% of the patient's maximum isometric neck strength. The strain was checked for each participant using a handheld digital scale during the supervised group training sessions. In each exercise session, the patients also performed a single series of 15 repetitions of dynamic exercises for the shoulders and upper extremities with an individually adjusted highest load. These exercises involved shrugs, presses, curls, bent over rows, flyers and pullovers using dumbbells. The training programme also involved a single series of squats, sit ups

	and back extension exercises that used only the patient's own body weight; these exercises were performed until muscle tiredness. The training session included stretching exercises for the neck, shoulder, and upper limb muscles with the exercise for each muscle lasting 30 seconds and repeated 3 times. The patients then recording the workout in their training diaries. Supervised meetings were conducted once a week for 6 weeks, then one session was conducted every second month for a total of 10 sessions over the 12 month period. Each group had 6-8 participants. Duration 12 months. Concurrent medication/care: Both groups were instructed to perform their exercises at home regularly three times a week and to keep a weekly exercise diary throughout the year. Both groups received written information about the exercises. Indirectness: No indirectness (n=52) Intervention 2: Flexibility. Those in the stretching group performed the same stretching exercises to the other group. They received training instructions and a lecture about the same topics as the other group in a single group session. Duration 12 months. Concurrent medication/care: Both groups were instructed to perform their exercises at home regularly three times a week and to keep a weekly exercise diary throughout the year. Both groups received written information about the exercises. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH AND FLEXIBILITY versus FLEXIBILITY

Protocol outcome 1: Quality of life
- Actual outcome: QoL physical functioning at End of treatment; Group 1: mean 92 (SD 11.5); n=43, Group 2: mean 92.4 (SD 9.8); n=43; RAND-36 0-100 Top=High is good outcome; Comments: Baseline: combined 86.3 (14.7); stretching 87.5 (11)
Change score (mean, CI): combined 5.7 (1.9-9.8); stretching 4.9 (2.1-8.1)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9
- Actual outcome: QoL role physical at End of treatment; Group 1: mean 78.3 (SD 36.1); n=43, Group 2: mean 79.4 (SD 33.9); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 61.6 (39.1); stretching 70 (34.1)
Change score (mean, CI); combined 16.7 (3.9-29.2); stretching 9.4 (-3.4 to 22.3)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9
- Actual outcome: QoL role emotional at End of treatment; Group 1: mean 89.1 (SD 23.8); n=43, Group 2: mean 87 (SD 31.5); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 86.8 (27.4); stretching 75.6 (37.3)
Change score (mean, CI); combined 2.3 (-7.1, 11.1); stretching 11.4 (1.9, 22.7)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9

- Actual outcome: QoL energy at End of treatment; Group 1: mean 68.6 (SD 16.7); n=43, Group 2: mean 63.4 (SD 21.6); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 65.1 (15.4); stretching 60.7 (22.5) Change score (mean, CI): combined 3.5 (-2, 9.1); stretching 2.7 (-4.2, 10.5) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9 - Actual outcome: QoL emotional well being at End of treatment; Group 1: mean 79.5 (SD 14); n=43, Group 2: mean 75.9 (SD 18.9); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 77.6 (12.8); stretching 73.8 (18.7) Change score (mean, CI): combined 2 (-3, 6.3); stretching 2.1 (-2.7, 7.2) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9 - Actual outcome: QoL social functioning at End of treatment; Group 1: mean 90.4 (SD 17); n=43, Group 2: mean 88.7 (SD 16); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 82 (20.8); stretching 81.7 (17.7) Change score (mean, CI): combined 8.4 (2.8, 14.4); stretching 7 (1.2, 12.5) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9 - Actual outcome: QoL bodily pain at End of treatment; Group 1: mean 69.2 (SD 20.5); n=43, Group 2: mean 70.9 (SD 19.4); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 55.2 (13.1); stretching 54.1 (14.1) Change score (mean, CI): combined 14 (8.1, 19.4); stretching 16.9 (10.5, 23.5) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9 - Actual outcome: QoL general health at End of treatment; Group 1: mean 72.1 (SD 15.2); n=43, Group 2: mean 71.4 (SD 18.3); n=43; RAND36 0-100 Top=High is good outcome; Comments: Baseline: combined 65.9 (16.7); stretching 70 (17.1) Change score (mean, CI): combined 6.2 (1.9, 11); stretching 1.4 (-3.6, 6.8) Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 6; Group 2 Number missing: 9 Protocol outcome 2: Discontinuation - Actual outcome: Discontinuation at End of treatment; Group 1: 6/49, Group 2: 9/52 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Sanudo 2011[235]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=42)
Countries and setting	Conducted in United Kingdom; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 24 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Inclusion criteria for study participants were: women, aged 18 to 65 years, diagnosed with FM based on the America College of Rheumatology
Exclusion criteria	Any significant concomitant illness such as inflammatory rheumatic diseases, respiratory or cardiovascular diseases that would prevent physical exercise, or severe psychiatric illness, or those that had attended physical therapy or psychological therapy in the previous 3 months
Recruitment/selection of patients	From 3 local patient support groups in Spain
Age, gender and ethnicity	Age - Mean (SD): 55.87 (7.8) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Indirectness of population	No indirectness
Interventions	(n=21) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. Twice weekly sessions of combined aerobic and muscle strength training for 24 weeks. 10 minute warm up followed by 10-15 minutes of aerobic exercises at 65-70% of maximum heart rate. Participants were in small groups and performed continuous walking with arm movements and jogging. This was followed by 15-20 minutes of muscle strengthening exercises with a circuit of 8 exercises using multiple muscles. Participants carried out 1 set of 8-10 repetitions and resistance was increased according to the patient's tolerance. This was followed by a cool-down of 10 minutes which consisted of flexibility exercises. Duration 24 weeks. Concurrent medication/care: 81.25% were taking medication for FMS (analgesic or NSAID, antidepressant or other combination). Indirectness: No indirectness

	(n=21) Intervention 2: Usual care. Participants continued their usual treatment and daily activities which did not include any structured exercise. Duration 24 weeks. Concurrent medication/care: 84.2% were taking medication for FMS (analgesics, NSAIDs, antidepressants or other combinations). Indirectness: No indirectness
Funding	Academic or government funding (National institute of health/NHS grants)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH AND FLEXIBILITY EXERCISE versus USUAL CARE

Protocol outcome 1: Quality of life
- Actual outcome: SF-36 physical function subscale at 24 weeks; Group 1: mean 56.8 (SD 17.4); n=21, Group 2: mean 45.2 (SD 14.1); n=21; SF-36 subscale 0-100 Top=High is good outcome; Comments: baseline: 50(22.7); 44.6(15.9)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up
- Actual outcome: SF-36 physical role subscale at 24 weeks; Group 1: mean 21.3 (SD 26.5); n=21, Group 2: mean 19.4 (SD 29.1); n=21; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 13.5(17.4); 19.8(27.6)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up
- Actual outcome: SF-36 bodily pain subscale at 24 weeks; Group 1: mean 29.9 (SD 16.8); n=21, Group 2: mean 19.5 (SD 18.1); n=21; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 23.2(17.4); 23.6(17.7)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up
- Actual outcome: SF-36 general health subscale at 24 weeks; Group 1: mean 43.1 (SD 11); n=21, Group 2: mean 33.5 (SD 11.4); n=21; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline:39.8(16.1); 33.4(12.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up
- Actual outcome: SF-36 vitality subscale at 24 weeks; Group 1: mean 41.3 (SD 13.8); n=21, Group 2: mean 28.6 (SD 18.8); n=21; SF-36 subscale 0-100 Top=High is good outcome
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up
- Actual outcome: SF-36 social function subscale at 24 weeks; Group 1: mean 63.9 (SD 23.8); n=21, Group 2: mean 52.2 (SD 21.1); n=21; SF-36 subscale 0-100 Top=High

is good outcome; Comments: Baseline:55.2(22.9); 48.6(16.5) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up - Actual outcome: SF-36 role emotional subscale at 24 weeks; Group 1: mean 71.1 (SD 41.5); n=21, Group 2: mean 52.1 (SD 44.3); n=21; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 53.3(45.3); 45.6(40.4) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up - Actual outcome: SF-36 mental health subscale at 24 weeks; Group 1: mean 60 (SD 14.9); n=21, Group 2: mean 44.2 (SD 23.9); n=21; SF-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 51.3(18.9); 44(20.7) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up Protocol outcome 2: Psychological distress (depression/anxiety) - Actual outcome: Beck depression inventory at 24 weeks; Group 1: mean 28.9 (SD 12.6); n=21, Group 2: mean 31.5 (SD 11.2); n=21; BDI 0-63 Top=High is poor outcome; Comments: Baseline: 35.1(14.1); 31.4(12.8) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at 24 weeks; Group 1: 3/21, Group 2: 1/21; Comments: 3: concomitant illness, personal reasons 1: lost to follow up Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference for multiple SF-36 subscales at baseline; Group 1 Number missing: 3, Reason: concomitant illness, personal reasons; Group 2 Number missing: 1, Reason: lost to follow up	
Protocol outcomes not reported by the study	Pain reduction ; Physical function ; Use of healthcare services ; Sleep

Study	Sanudo 2012[233]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=41)
Countries and setting	Conducted in Spain; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women who met the American College of Rheumatology criteria for the classification of fibromyalgia
Exclusion criteria	Presence of concomitant conditions such as inflammatory rheumatic diseases, respiratory or cardiovascular diseases, respiratory or cardiovascular diseases and severe psychiatric illness
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Not reported. Gender (M:F): Women only. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Indirectness of population	No indirectness
Interventions	(n=21) Intervention 1: Mixed modality exercise – Strength and aerobic. Exercise was twice weekly for 45-60 minutes. Each session included 10 minutes of warm up activities (slow walking and gently movements of progressive intensity e.g. arm swinging); 10-15 minutes of aerobic exercise at 65% to 70% of maximal heart rate, 15-20 minutes of muscle strengthening exercises (one set of 8-10 repetitions for 8 different muscle groups, with a load of 1-3kg), and 10 minutes of flexibility exercises (1 set of 3 repetitions for 8-9 different exercises, maintaining the stretched position for 30 seconds). Strengthening and flexibility exercises focused on the main areas of pain in patients with FM (deltoids, biceps, neck, hips, back and chest). Duration 6 months. Concurrent medication/care: Not reported . Indirectness: No indirectness (n=20) Intervention 2: Usual care. Usual medical treatment of fibromyalgia and continued normal daily activities which did not include structured exercise. Duration 6 months. Concurrent medication/care: Not reported. Indirectness: No

	indirectness
Funding	Academic or government funding (Supported by the University of Seville)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH AND AEROBIC versus USUAL CARE Protocol outcome 2: Physical function - Actual outcome: Physical function at End of treatment; Group 1: mean 513.87 metres (SD 98.83); n=18, Group 2: mean 459.07 metres (SD 69.54); n=19; 6 minute walk test - Top=High is good outcome; Comments: Baseline exercise 493.25±88.6; control 454.17±69.54 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 1 Protocol outcome 3: Psychological distress (depression/anxiety) - Actual outcome: Depression at End of treatment; Group 1: mean 14.67 (SD 7.4); n=18, Group 2: mean 16.64 (SD 6.37); n=19; BDI 0-63 Top=High is poor outcome; Comments: Baseline: exercise 19.87±7.57; control 20.43±7.73 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 3; Group 2 Number missing: 1 Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at End of treatment; Group 1: 3/21, Group 2: 1/20 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Pain reduction ; Use of healthcare services ; Sleep

Study	Sanudo 2015[232]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=32)
Countries and setting	Conducted in Spain; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 months
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women with fibromyalgia
Exclusion criteria	Pulmonary, cardiovascular, severe psychiatric or inflammatory rheumatic diseases. Those who attended psychological or physical therapy, or received exercise training in the last year were also excluded
Recruitment/selection of patients	Recruited from fibromyalgia support groups
Age, gender and ethnicity	Age - Mean (SD): Exercise 55±2; control 58±2. Gender (M:F): Women only. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Indirectness of population	No indirectness
Interventions	(n=16) Intervention 1: Aerobics - Walking. Two sessions per week of 45-60 minutes duration. Each session included 10 minutes of warm up activities (easy movements and slow walking), 15-20 minutes of steady state exercise at 60-65% of predicted maximum heart rate (including continuous walking with arm movements and jogging) and 15 minutes of interval training at 75-80% (six repetitions of 1.5 minutes with 1 minute interpolated rest intervals), and 5-10 minutes of cool-down activities (slow walks, easy movements, relaxation training). Exercise intensity was monitored by a heart rate telemetric system. The intensity progressively increased as participants improved their exercise capacity to maintain the heart rate in the prescribed range. Duration 24 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=16) Intervention 2: Usual care. Participants continued their normal daily activities which did not include structured exercise. Duration 24 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness

Funding	Academic or government funding (The University of Seville)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: WALKING versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: Pain at End of treatment; Group 1: mean 6.7 (SD 2.2); n=16, Group 2: mean 7 (SD 1.7); n=12; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 7.4±2.2; control 7.2±1.8
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 4

Protocol outcome 2: Psychological distress (depression/anxiety)
- Actual outcome: Depression at End of treatment; Group 1: mean 5.6 (SD 3.4); n=16, Group 2: mean 6.7 (SD 2.2); n=12; VAS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 6.5±3.7; control 7.1±2.7
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 4
- Actual outcome: Anxiety at End of treatment; Group 1: mean 5.7 (SD 3.3); n=16, Group 2: mean 7.5 (SD 2.5); n=12; VAS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 6.9±3.3; control 6.4±3
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 4

Protocol outcome 3: Sleep
- Actual outcome: Sleep disturbances at End of treatment; Group 1: mean 7.2 (SD 2.8); n=16, Group 2: mean 8.6 (SD 1.9); n=12; VAS 0-10 Top=High is poor outcome; Comments: Baseline: exercise 7.5±3.2; control 8.4±2.2
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0; Group 2 Number missing: 4

Protocol outcome 4: Discontinuation
- Actual outcome: Discontinuation at End of treatment; Group 1: 0/16, Group 2: 4/16
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcomes not reported by the study	Quality of life ; Physical function ; Use of healthcare services

Study	Sevimli 2015[242]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=75)
Countries and setting	Conducted in Turkey; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Met the ACR criteria for fibromyalgia and were aged 18 to 50 years
Exclusion criteria	Not specified. Participants were excluded due to other conditions (Cushing syndrome, cardiovascular problems) and for being postmenopausal.
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 35(8.8) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain (Fibromyalgia).
Extra comments	Not specified
Indirectness of population	No indirectness
Interventions	(n=25) Intervention 1: Strength. Isometric strength and stretching exercise program lasting 15 minutes per day. Three minute loadings with 30 seconds rest between 3 sets of low to moderate intensity were repeated in the first month of the exercise programme, and in the second month this was increased to high intensity loadings of 4 sets, and in the third month rest intervals were reduced to 10 seconds with 5 sets of 3 minute loadings. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=25) Intervention 2: Aerobics - Swimming. Pool based aquatic aerobic exercise programme with group therapy 2 times a week. Duration was 40 minutes in the first month, 45 in the second month and 50 minutes in the final month. Duration 12 weeks. Concurrent medication/care: Not specified . Indirectness: No indirectness (n=25) Intervention 3: Aerobics - Other aerobic exercise. Gymnastic-based aerobic exercise programme with group

	therapy 2 times a week. Duration was 40 minutes in the first month, 45 in the second month and 50 minutes in the final month. No further details. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Scientific Research Unit of Cukurova)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus SWIMMING

Protocol outcome 1: Pain reduction
- Actual outcome: VAS total scores at 12 weeks; Group 1: mean 70.4 (SD 12.5); n=25, Group 2: mean 48 (SD 9.3); n=25; VAS 0-100 Top=High is poor outcome; Comments: To note: results in the analysis for gym based and aquatic based exercises were pooled.
Baseline: 68.2(11.8); 71.5(13.1)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 mental component summary score at 12 weeks; Group 1: mean 32.02 (SD 9.4); n=25, Group 2: mean 49.4 (SD 8.3); n=25; SF-36 0-100 Top=High is good outcome; Comments: To note: results in the analysis for gym based and aquatic based exercises were pooled.
Baseline: 31.6(9); 35.2(7.9)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;
- Actual outcome: SF-36 physical component summary score at 12 weeks; Group 1: mean 36.8 (SD 8.4); n=25, Group 2: mean 50.3 (SD 7.4); n=25; SF-36 subscale 0-100 Top=High is good outcome; Comments: To note: results in the analysis for gym based and aquatic based exercises were pooled.
Baseline: 37.3(7.6); 36.4(8.5)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 3: Physical function
- Actual outcome: 6 minute walking test (metres) at 12 weeks; Group 1: mean 540.4 (SD 53.8); n=25, Group 2: mean 619.4 (SD 61.8); n=25; Comments: baseline: 541.4(53.3); 543.3(56.4)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: Beck depression inventory at 12 weeks; Group 1: mean 22.6 (SD 10); n=25, Group 2: mean 6.1 (SD 7.8); n=25; BDI 0-30 Top=High is poor outcome; Comments: To note: results in the analysis for gym based and aquatic based exercises were pooled.

Baseline: 19.4(10.1); 15.7(9)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus GYMNASTIC-BASED AEROBIC EXERCISE

Protocol outcome 1: Pain reduction
- Actual outcome: VAS total scores at 12 weeks; Group 1: mean 70.4 (SD 12.5); n=25, Group 2: mean 48.2 (SD 8.8); n=25; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 68.2(11.8); 70(12.9)
To note: results in the analysis for gym based and aquatic based exercises were pooled.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 mental component summary score at 12 weeks; Group 1: mean 32.02 (SD 9.4); n=25, Group 2: mean 45.2 (SD 7); n=25; SF-36 0-100 Top=High is good outcome; Comments: Baseline: 31.6(9); 23.5(9.7)
To note: results in the analysis for gym based and aquatic based exercises were pooled.
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF-36 physical component summary score at 12 weeks; Group 1: mean 36.8 (SD 8.4); n=25, Group 2: mean 53.6 (SD 5.4); n=25; SF-36 0-100 Top=High is good outcome; Comments: To note: results in the analysis for gym based and aquatic based exercises were pooled.
Baseline: 37.3(7.6); 41.8(8.4)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 3: Physical function
- Actual outcome: 6 minute walking test (metres) at 12 weeks; Group 1: mean 540.4 (SD 52.8); n=25, Group 2: mean 628.8 (SD 55.5); n=25; Comments: Baseline: 541.4(53.3); 569.5(48.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: Beck depression inventory at 12 weeks; Group 1: mean 22.6 (SD 10); n=25, Group 2: mean 9.9 (SD 6.2); n=25; BDI 0-30 Top=High is poor outcome; Comments: Baseline: 19.4(10.1); 20.5(12.3)
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Use of healthcare services ; Sleep ; Discontinuation

Study	**Silva 2019[243]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=60)
Countries and setting	Conducted in Brazil; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: diagnosed according to the Classification Criteria of the American College of Rheumatology
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women with a clinical diagnosis of fibromyalgia with medical referral were included according to the Classification Criteria of the American College of Rheumatology, aged between 18 and 60 years
Exclusion criteria	Patients with arterial insufficiency, decompensated systemic arterial hypertension, decompensated cardiorespiratory disease, history of syncope or arrhythmias induced by physical exercise, decompensated diabetes, severe psychiatric illness, history of regular physical exercise (at least 2 times per week) in the last 6 months, or any other condition that made it impossible for the patient to perform physical exercises
Recruitment/selection of patients	The sample was selected by convenience through the waiting list of the FACISA/UFRN Physiotherapy School Clinic
Age, gender and ethnicity	Age - Mean (SD): resistance trainig group: 44.93±10.30; relaxation group: 49.40±8.30 . Gender (M:F): All female. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable 5. complex regional pain syndrome: Not applicable

Indirectness of population	No indirectness
Interventions	(n=30) Intervention 1: Strength and conditioning. a resistance training program using weight training for calculating one repetition maximum (1 RM), twice a week for 40min for a period of 12 weeks. The exercise program consisted of 3 sets of 12 repetitions, with an interval of 1-2 min for recovery between one set to another, alternating lower limbs. Loads with 60% of 1RM in the first month, 70% of a new 1RM test in the second month, and 80% of a new 1 RM test in the third month. The following muscles were trained: biceps brachial, triceps, pectoralis, trapezius, knee extensors, knee flexors and hip abductors.. Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=30) Intervention 2: Pychological intervention - Relaxation. Performed 2 body relaxation sessions per week based on the sophrology technique. Each session lasted 40 min for a period of 12 weeks. The patients remained lying on comfortable mats with relaxing music playing in the background in a room with pleasant temperature, and were invited to think about their illness, their life, imagining positive and negative points and to analyze everything; the physiotherapist asked them to focus on the negative aspects and concentrate on these negative points ,and they were asked to try to see good aspects of each point . Duration 12 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	No funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH AND CONDITIONING versus RELAXATION

Protocol outcome 1: Pain reduction
- Actual outcome: Pain at 8 weeks; Group 1: mean 5.23 (SD 2.16); n=30, Group 2: mean 4.90 (SD 1.72); n=30; VAS 0-10 Top=High is poor outcome; Comments: Baseline: strength group 6.67 (1.47); relaxation group 6.27 (1.36)
Risk of bias: All domain – Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Pain at 12 weeks; Group 1: mean 4.06 (SD 2.58); n=30, Group 2: mean 5.1 (SD 1.62); n=30; VAS 0-10 Top=High is poor outcome; Comments: Baseline: strength group 6.67 (1.47); relaxation group 6.27 (1.36)
Risk of bias: All domain – Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Quality of life

- Actual outcome: Social Aspects - SF36 at 12 weeks; Group 1: mean 67.3 (SD 28.2); n=30, Group 2: mean 63.9 (SD 21.4); n=30; SF36 0-100 Top=High is good outcome; Comments: Baseline: strength group 52 (29.7); relaxation group 53.5 (21.8)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0

- Actual outcome: General health status - SF36 at 12 weeks; Group 1: mean 47.2 (SD 21); n=30, Group 2: mean 44.6 (SD 21.2); n=30; SF36 0-100 Top=High is good outcome; Comments: Baseline: strength group 35.5 (23.3); relaxation group 38.6 (16)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Functional capacity - SF36 at 12 weeks; Group 1: mean 53.1 (SD 21); n=30, Group 2: mean 40 (SD 20); n=30; SF36 0-100 Top=High is good outcome; Comments: Baseline: strength group 36.6 (20); relaxation group 33.3 (16)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Limitation due to physical aspects - SF36 at 12 weeks; Group 1: mean 45.8 (SD 41); n=30, Group 2: mean 28.6 (SD 38.1); n=30; SF36 0-100 Top=High is good outcome; Comments: Baseline: strength group 15.8 (28.9); relaxation group 18.3 (35.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Limitations due to Emotional Aspects - SF36 at 12 weeks; Group 1: mean 49.4 (SD 38); n=30, Group 2: mean 37.5 (SD 43.4); n=30; SF36 0-100 Top=High is good outcome; Comments: Baseline: strength group 32.4 (39.6); relaxation group 32.1 (40.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Pain - SF36 at 12 weeks; Group 1: mean 34.9 (SD 23.4); n=30, Group 2: mean 29.9 (SD 17.2); n=30; SF36 0-100 Top=High is good outcome; Comments: Baseline: strength group 22.4 (18.3); relaxation group 23.1 (17.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0
- Actual outcome: Mental Health - SF36 at 12 weeks; Group 1: mean 59.5 (SD 23.6); n=30, Group 2: mean 58.6 (SD 23.6); n=30; Comments: Baseline: strength group 50.9 (30); relaxation group 53.3 (22.6)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 3: Physical function
- Actual outcome: Six-minute walk test at 12 weeks; Group 1: mean 472 Minutes (SD 91); n=30, Group 2: mean 415 Minutes (SD 80); n=30; Comments: Baseline: resistance group 429 (92); relaxation group 404 (69)
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 4: Discontinuation
-Actual outcome: Discontinuation at End of treatment; Group 1: 7/30, Group 2: 6/30
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Only age and BMI reported; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcomes not reported by the study	Psychological distress (depression/anxiety); Use of healthcare services; Sleep; Discontinuation

Study	**Suvarnnato 2019[247]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=54)
Countries and setting	Conducted in Australia; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention + follow up: 6 week intervention plus 12 week follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: Neck pain without known cause (see inclusion criteria)
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Mechanical neck pain denied as pain in the area of the neck and/or neck-shoulder with neck pain that could be provoked by mechanical characteristics, including sustained neck postures, cervical movement, ormanual palpation of the cervical musculature. Specifically, the pain had to be localized to the dorsal part of the neck in an area limited by a horizontal line through the inferior portion of the occipital region and a horizontal line through the spinous process of the first thoracic vertebra.29 To be eligible for the study, participants had to meet three criteria: have neck-pain symptoms of at least 3 months' duration, a score ≥10/100 on the Thai Version of the Neck Disability Index (NDI-TH) questionnaire,30 and be aged 18–60 years, to capture adults of working age.
Exclusion criteria	Participants were excluded if they reported any of the following:1) diagnosis of cervical radiculopathy or myelopathy(at least two of myotomal strength, sensation, or reflexes had to be diminished for nerve-root or spinal cord involvement to be considered); 2) history of cervical and thoracic spine fracture and/or dislocation; 3) history of surgery of the cervical and/or thoracic spine; 4) history of spinal osteoporosis, spinal infection, or fibromyalgia syndrome, and 5) history of whiplash injury and/or head/neck injuries. Exclusion criteria included positive neurological signs (n=2) and severe neck pain from spinal infection (n=1).
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 42.94(10.05) years. Gender (M:F): 6:48. Ethnicity: Not specified
Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with pain conditions other than

	chronic widespread pain 5. complex regional pain syndrome: people with pain conditions other than complex regional pain syndrome
Extra comments	Mean duration of pain=12.86(17.6) months
Indirectness of population	No indirectness
Interventions	(n=18) Intervention 1: Strength and conditioning. Semispinalis cervicis-training group. Participants received semispinalis cervicis isometric exercise as described by Schomacher et al in their intramuscular electromyography(EMG) study. In that study, the semispinaliscervicis was selectively activated relative to the splenius capitis by applying manual static resistance to the vertebral arch of C2 and asking the upright-sitting patient to push backward.32The aim of the exercise was to stimulate semispinalis cervicis activation selectively. In the current study, the exercise was performed by subjects while sitting on a stool without a backrestwith hips and knees flexed 90° and feet placed on the floor.The researcher stood on the left of the subject, facing them. Next, the researcher placed the thumb and index finger of the right hand approximately on the posterior vertebral arches ofthe subject's second cervical vertebra (C2) and pushed firmly/gently (slowly to increase resistance) into flexion (anteriorly),while the left hand stabilized the participant's left shoulderto monitor the compensatory body movement. Subjects were asked to resist maximal voluntary contraction in the direction of extension without provocation of neck pain (Figure 2A).The exercise program was performed to hold resistance for 10seconds, ten times per set, with three sets per day. A 30-second rest was allowed between sets. Each subject performedthis exercise twice per week over a 6-week period with th ephysical therapist. The exercise was performed as tolerated without provocation of neck pain. Duration 6 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=18) Intervention 2: Strength and conditioning. Deep cervical flexor-training group. Deep cervical flexor exercise is a low-load exercise focused on deep cervical flexor muscles, as described by Jull et al. This exercise targets the deep flexor muscles of the cervical region, rather than the superficial flexor muscles. In the current study, deep cervical flexor training was conducted in the supine position on the experimental table. Each participant was asked to move their head slowly to the inner range asif to say, "Yes". To correct individual exercise technique,participants were guided in their movements by feedbackfrom an air-filled pressure sensor, which was placed in thesuboccipital region, ie, the posterior neck. The baseline of thepressure sensor was set to 20 mmHg inflation. Subjects wereguided by the researcher to familiarize them with the deep cervical flexor exercise. The deep cervical flexor-exerciseprocedure was correct when performed without contraction ofthe superficial neck-flexor muscles. The action of superficialneck muscles was monitored by researcher palpation. Next, participants were assessed individually for their ability to perform the deep cervical flexor exercises correctly without provocation of neck pain. This assessment was performed at the highest incremental level of pressure appropriate for each individual (22, 24, 26, 28, or 30 mmHg; Figure 2B).The participants were instructed to perform the exercise ten times per set, with a short rest. A 30-second rest was allowed between sets. The exercise program was performed under supervision of the researcher twice per week. Participants were trained to perform deep cervical flexor exercises at the same range of motion as the exercise

	protocol without the air-filled pressure sensor, and each participant was instructed to train with this exercise twice per day at home. Duration 6 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=18) Intervention 3: Usual care. In this study, usual care was treatment deemed appropriate by the physical therapists using any general exercise, including stretching and upper-limb-strengthening exercises, modalities, manual therapy, or electrotherapy within the hospital. Participants randomized to usual care were not eligible to perform the exercises performed in the semispinalis cervicis training and deep cervical flexor-training groups. Participants received usual care over 10–12 treatment appointments within 6 weeks. In the usual-care group, subjects received 20–30minutes for each physiotherapy appointment. Duration 6 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Khon Kean University grant)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH EXERCISE (SCT GROUP) versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: Numeric pain scale at 6 weeks; Group 1: mean 2.3 (SD 3.72); n=18, Group 2: mean 3.49 (SD 3.72); n=18; NPS 0-10 Top=High is poor outcome; Comments: Baseline 4.77(1.89); 4.05(0.87)
Standard deviation estimated from p-value of the mean difference
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Numeric pain scale at 18 week follow up (including 6 week intervention); Group 1: mean 2.79 (SD 4.97); n=18, Group 2: mean 3.37 (SD 4.97); n=18; NPS 0-10 Top=High is poor outcome; Comments: Standard deviation estimated from p-value of the mean difference
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

Protocol outcome 2: Physical function
- Actual outcome: Neck disability index at 18 week follow up (including 6 week intervention); Group 1: mean 12.97 (SD 22.7); n=18, Group 2: mean 21.69 (SD 22.7); n=18; NDI 0-100 Top=High is poor outcome; Comments: Standard deviation estimated from the p-value of the mean difference
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: Neck disability index at 6 weeks; Group 1: mean 13.29 (SD 24.4); n=18, Group 2: mean 20.24 (SD 24.4); n=18; NDI 0-100 Top=High is poor outcome; Comments: Baseline: 30(10.82); 23.11(8.54)
Standard deviation estimated from p-value of the mean difference
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH EXERCISE (DCF GROUP) versus USUAL CARE Protocol outcome 1: Pain reduction - Actual outcome: Numeric pain scale at 6 weeks; Group 1: mean 2.86 (SD 3.5); n=18, Group 2: mean 3.49 (SD 3.5); n=18; NPS 0-10 Top=High is poor outcome; Comments: Baseline Standard deviation estimated from p-value of the mean difference Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; - Actual outcome: Numeric pain scale at 18 week follow up (including 6 week intervention); Group 1: mean 3.27 (SD 10); n=18, Group 2: mean 3.37 (SD 10); n=18; NPS 0-10 Top=High is poor outcome; Comments: Standard deviation estimated from p-value of the mean difference Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Protocol outcome 2: Physical function - Actual outcome: Neck disability index at 6 weeks; Group 1: mean 14.99 (SD 20.77); n=18, Group 2: mean 20.24 (SD 20.77); n=18; NDI 0-100 Top=High is poor outcome; Comments: Baseline 48.22(4.65); 47.55(4.03) Standard deviation estimated from p-value of the mean difference Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; - Actual outcome: Neck disability index at 18 week follow up (including 6 week intervention); Group 1: mean 16.62 (SD 20.1); n=18, Group 2: mean 21.69 (SD 20.1); n=18; NDI 0-100 Top=High is poor outcome; Comments: Standard deviation estimated from p-value of mean difference Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - High, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; *Note: DCF and SCT data pooled in the analysis (compared against usual care)*	
Protocol outcomes not reported by the study	Quality of life ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study (subsidiary papers)	Tomas-carus 2008[252] (Tomas-carus 2007[254], Tomas-carus 2009[253], [116])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=34)
Countries and setting	Conducted in Spain; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 8 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Met ACR diagnostic criteria for fibromyalgia
Exclusion criteria	history of severe trauma; frequent migraines; peripheral nerve entrapment; inflammatory rheumatic diseases; severe psychiatric illness; other diseases that prevent physical loading and pregnancy; attendance at another psychological or physical therapy or regular physical exercise with more than one exercise session of 30 min per week during a 2-week period in the last 5years
Recruitment/selection of patients	Advertisements placed in newsletters of a local FM association in Spain
Age, gender and ethnicity	Age - Mean (SD): 50.8(8.6) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Duration of pain 19.8 (7.5) years.
Indirectness of population	No indirectness
Interventions	(n=17) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. Supervised training in waist high pool of warm water 3 times per week during an 8 month period. Each session 1 hour, 10 minutes warming up with slow walks and easy movements of progressive intensity, 10 minutes of aerobic exercises (60-65% maximal heart rate), 20 minutes of strength exercises using water resistance (4 sets of 10 repetitions), 10 minutes of cooling down with low intensity exercises. Duration 8 months. Concurrent medication/care: Not specified (mean (SD) number of drugs taken 1.3(0.8)). Indirectness: No indirectness (n=16) Intervention 2: Usual care. Control group continuing daily activities which did not include any form of physical exercise similar to those in the therapy . Duration 8 months. Concurrent medication/care: Not specified. Indirectness:

	No indirectness
Funding	Academic or government funding (Regional government of extremadura, Spain)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC, STRENGTH EXERCISE versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: FIQ pain subscale at 8 months; Group 1: mean 5.3 (SD 1.4); n=15, Group 2: mean 6.6 (SD 1.8); n=15; FIQ pain subscale 0-10 Top=High is poor outcome; Comments: Baseline: 5.6(1.9); 6.4(2.3)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference on multiple SF-36 subscales; Group 1 Number missing: 2, Reason: Dropped out ; Group 2 Number missing: 1, Reason: Dropped out
- Actual outcome: VAS at 12 weeks; Group 1: mean -18.4 (SD 27.6); n=17, Group 2: mean 1 (SD 17.4); n=17; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 63.1(26); 63.9(25)
SDs calculated from CIs: -31.5 to -5.3; -7.2 to 9.3
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1, Reason: Lost to follow up ; Group 2 Number missing: 0

Protocol outcome 2: Quality of life
- Actual outcome: EQ-5D at 3 months; Group 1: mean 0.582 (CI 0.434 to 0.729); n=15, Group 2: mean 0.334 (Cis 0.175 to 0.494) ; n=15; EQ-5D, 0-1 Top=High is good outcome; Comments: Baseline: 0.316(0.162 to 0.470); 0.331 (0.15 to 0.511)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference on multiple SF-36 subscales; Group 1 Number missing: 2, Reason: Dropped out ; Group 2 Number missing: 1, Reason: Dropped out
- Actual outcome: EQ-5D at 8 months; Group 1: mean 0.528 (CI 0.380 to 0.675); n=15, Group 2: mean 0.334 (Cis 0.175 to 0.493) ; n=15; EQ-5D, 0-1 Top=High is good outcome; Comments: Baseline: 0.316(0.162 to 0.470); 0.331 (0.15 to 0.511)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference on multiple SF-36 subscales; Group 1 Number missing: 2, Reason: Dropped out ; Group 2 Number missing: 1, Reason: Dropped out

Protocol outcome 3: Physical function
- Actual outcome: FIQ physical function subscale at 8 months; Group 1: mean 2.4 (SD 1.7); n=15, Group 2: mean 3.7 (SD 2); n=15; FIQ PF subscale 0-10 Top=High is poor outcome; Comments: 3(1.5); 3.7(1.5)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference on multiple SF-36 subscales; Group 1 Number missing: 2, Reason: Dropped out ; Group 2 Number

missing: 1, Reason: Dropped out Protocol outcome 4: Psychological distress (depression/anxiety) - Actual outcome: FIQ depression subscale at 8 months; Group 1: mean 4 (SD 3.3); n=15, Group 2: mean 6.1 (SD 1.7); n=15; FIQ depression subscale 0-10 Top=High is poor outcome; Comments: Baseline: 5.4(2.6); 6(2.1) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference on multiple SF-36 subscales; Group 1 Number missing: 2, Reason: Dropped out ; Group 2 Number missing: 1, Reason: Dropped out - Actual outcome: State trait anxiety inventory at 8 months; Group 1: mean 37.5 (SD 8); n=15, Group 2: mean 44.4 (SD 8.9); n=15; STAI 20-80 Top=High is poor outcome; Comments: Baseline: 45.1(9.9); 41.9(8) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Difference on multiple SF-36 subscales; Group 1 Number missing: 2, Reason: Dropped out ; Group 2 Number missing: 1, Reason: Dropped out Protocol outcome 5: Discontinuation - Actual outcome: Discontinuation at 8 months; Group 1: 2/17, Group 2: 1/16; Comments: Discontinued exercise, lost to follow up Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness, Comments: Unclear if discontinued intervention or study; Baseline details: Difference on multiple SF-36 subscales;	
Protocol outcomes not reported by the study	Use of healthcare services; Sleep

Study	Toprak celenay 2017[256]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=49)
Countries and setting	Conducted in Turkey; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women, having fibromyalgia syndrome, 18-65 years of age, and being a volunteer
Exclusion criteria	Neurologic, infectious, endocrine, and other inflammatory rheumatic diseases, severe psychological disorders, any condition interfering with exercise (Advances cardiac respiratory or orthopedic problems), malignancy, being pregnant, and intervention including exercise programme or physical therapy in the last 6 months
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Exercise alone: 39.9±9.5; exercise + manual therapy: 42.5±8.3. Gender (M:F): Women only. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Indirectness of population	No indirectness
Interventions	(n=24) Intervention 1: Mixed modality exercise - Aerobic, Strength exercise. Sessions began with postural education by placing participants in standing position to find a neutral balanced position of the spine curvatures. The participants were asked to maintain neutral spine during the programme. The combined exercise programme was carried out 2 days a week for 6 weeks and took 1 hour. It was composed of 10 minute warm up exercises, 40 minutes aerobic and strengthening exercises including neck, trunk, upper and lower limb muscles. The aerobic exercise consisted of 20 minutes walking on a treadmill. The target heart rate was initially adjusted to 65-70% of the maximal heart rate and to 75-80% of the maximal heart rate in the advanced programme. Muscle strengthening exercises were then performed with elastic resistive bands for 20 minutes, where deep neck muscles, deltoid, latissimus dorsi, serratus anterior, scapular retractor muscles, pectoralis major, shoulder external rotator muscles, erector spine, abdominalis, gluteus, and quadriceps muscles were strengthened. The participants began exercising with yellow or red Thera-Bands with

	mild or medium tension. When they performed 15 repetitions without serious pain or fatigue, they progressed to the next colour resistance band. They had 10 repetitions with a holding period of 10 seconds. Duration 6 weeks. Concurrent medication/care: Using drugs recommended in the clinic was not changed for standardisation. Indirectness: No indirectness (n=25) Intervention 2: Manual therapy and exercise. Connective tissue massage was applied 2 days per week for a total of 12 sessions. While patients were in a sitting position, starting from the lumbosacral region, the lower thoracic, scapular, interscapular, and cervical regions were included in the treatment, respectively. For creating traction between cutaneous tissues, the middle fingers of both hands were used during the application. Each session lasted around 5-20 minutes. Duration 6 weeks. Concurrent medication/care: Using drugs recommended in the clinic was not changed for standardisation. Indirectness: No indirectness
Funding	No funding
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: MANUAL THERAPY AND EXERCISE versus AEROBIC, STRENGTH EXERCISE Protocol outcome 1: Discontinuation - Actual outcome: Discontinuation at End of treatment; Group 1: 5/25, Group 2: 4/20 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Pain reduction ; Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Ulug 2018[257]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=60)
Countries and setting	Conducted in Turkey; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 6 weeks
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Subjects aged 18–50 years and who had chronic neck pain (> 3 months of duration)
Exclusion criteria	Those with a history of cervical spine surgery, cervical trauma, central nervous system diseases, cervical radiculopathy, acute inflammation and malignancy were excluded
Recruitment/selection of patients	Not reproted
Age, gender and ethnicity	Age - Mean (SD): Pilates 38.7 (7.9); yoga 35.9 (9.8); strength 44.6 (4.3). Gender (M:F): 9/47. Ethnicity: Not reported
Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with pain conditions other than chronic widespread pain 5. complex regional pain syndrome: people with pain conditions other than complex regional pain syndrome
Indirectness of population	No indirectness
Interventions	(n=20) Intervention 1: Biomechanical - Pilates. After the initial assessment of the patients, all exercise groups received their exercise programme from a single physiotherapist (NU),using a written and photographic description. Patients were also supervised for the first 3 weeks (home-based thereafter). In the first teaching session, patients were taught how to activate their deep abdominal muscles (transversus abdominis and multifidus). Some visual imagery, verbal cueing or demonstrations were used as facilitation methods. Five key elements of Pilates: lateral costal breathing, centering (pelvic placement), ribcage placement, shoulder blade placement, head and neck placement, were taught. Four Pilates beginner mat

	exercises, including double-leg stretch level, shoulder bridge level, arm openings level and breaststroke level, were taught and patients were encouraged to perform these exercises in 2 sets of 10 repetitions per day. They were also told to pay attention and protect the neutral spine alignment and perform breathing control during all the exercises. Duration 6 weeks. Concurrent medication/care: In addition to the exercises, each group received physical therapy (5 days in a week, a total of 15 sessions over a period of 3 weeks) for neck pain, including hot pack, ultrasound and conventional transcutaneous electrical nerve stimulation (TENS). Indirectness: No indirectness (n=20) Intervention 2: Mind-body exercises - Yoga. Four exercises from Iyengar Yoga asanas: Adho MukhaVirasana, Tadasana, Virabhadrasana and Chair Bharadvajasana (10, 21), were taught to the patients. They were told to maintain each yoga posture starting from at least 10–20 s in the following days. They were encouraged to do these exercises in 2 sets of 10 repetitions per day. Duration 6 weeks. Concurrent medication/care: In addition to the exercises, each group received physical therapy (5 days in a week, a total of 15 sessions over a period of 3 weeks) for neck pain, including hot pack, ultrasound and conventional transcutaneous electrical nerve stimulation (TENS). Indirectness: No indirectness (n=20) Intervention 3: Strength and conditioning. Isometric exercises. In the sitting position, the patients were instructed to place their hands firstly on the front (then the other sides) of their heads and push forward, but resist any movement of the head while maintaining the head and neck in the neutral position for 5 s. They were encouraged to do these exercises in 2 sets of 30 repetitions per day. Duration 6 weeks. Concurrent medication/care: In addition to the exercises, each group received physical therapy (5 days in a week, a total of 15 sessions over a period of 3 weeks) for neck pain, including hot pack, ultrasound and conventional transcutaneous electrical nerve stimulation (TENS). Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PILATES versus YOGA

Protocol outcome 1: Pain reduction at Define
- Actual outcome: Pain at 6 weeks; Group 1: mean 1.7 (SD 1.8); n=20, Group 2: mean 1.4 (SD 2); n=18; VAS 0-10 Top=High is poor outcome; Comments: Baseline value: 6.9 (1.3); 7.0 (0.9)

Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

Protocol outcome 2: Quality of life at Define
- Actual outcome: Quality of life at 6 weeks; Group 1: mean 118.2 (SD 93.1); n=20, Group 2: mean 89.8 (SD 78.6); n=18; Nottingham health profile 0-600 Top=High is poor outcome; Comments: Baseline values: 206.9 (97.9); 189.5 (118.1)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

Protocol outcome 3: Physical function at Define
- Actual outcome: Neck disability at 6 weeks; Group 1: mean 10 (SD 4.8); n=20, Group 2: mean 8.2 (SD 4.8); n=18; Neck disability index Unclear Top=High is poor outcome; Comments: Baseline values: 19.1 (6.6); 15.5 (5.3)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

Protocol outcome 4: Psychological distress (depression/anxiety) at Define
- Actual outcome: Depression at 6 weeks; Group 1: mean 8.5 (SD 6.5); n=20, Group 2: mean 6.4 (SD 6.1); n=18; Beck depression inventory 0-63 Top=High is poor outcome; Comments: Baseline values: 12.9 (7.6); 10.8 (6.2)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: PILATES versus STRENGTH AND CONDITIONING

Protocol outcome 1: Pain reduction at Define
- Actual outcome: Pain at 6 weeks; Group 1: mean 1.7 (SD 1.8); n=20, Group 2: mean 2.5 (SD 2.3); n=18; VAS 0-6 Top=High is poor outcome; Comments: Baseline values: 6.9 (1.3); 6.7 (1.8)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

Protocol outcome 2: Quality of life at Define
- Actual outcome: Quality of life at 6 weeks; Group 1: mean 118.2 (SD 93.1); n=20, Group 2: mean 145.9 (SD 127.8); n=18; Nottingham health profile 0-600 Top=High is poor outcome; Comments: Baseline: 206.9 (97.9); 187.8 (137.4)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

Protocol outcome 3: Physical function at Define

- Actual outcome: Neck disability at 6 weeks; Group 1: mean 10 (SD 4.8); n=20, Group 2: mean 11.3 (SD 6.3); n=18; Neck disability index Unclear Top=High is poor outcome; Comments: Baseline values: 19.1 (6.6); 17.5 (7.1)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

Protocol outcome 4: Psychological distress (depression/anxiety) at Define
- Actual outcome: Depression at 6 weeks; Group 1: mean 8.5 (SD 6.5); n=20, Group 2: mean 9.7 (SD 7.7); n=18; Beck depression inventory 0-63 Top=High is poor outcome; Comments: Baseline values: 12.9 (7.6); 12.4 (9.4)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 0; Group 2 Number missing: 2

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: YOGA versus STRENGTH AND CONDITIONING

Protocol outcome 1: Pain reduction at Define
- Actual outcome: Pain at 6 weeks; Group 1: mean 1.4 (SD 2); n=18, Group 2: mean 2.5 (SD 2.3); n=18; VAS 0-10 Top=High is poor outcome; Comments: Baseline values: 7.0 (0.9); 6.7 (1.8)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 2; Group 2 Number missing: 2

Protocol outcome 2: Quality of life at Define
- Actual outcome: Quality of life at 6 weeks; Group 1: mean 89.8 (SD 78.6); n=18, Group 2: mean 145.9 (SD 127.8); n=18; Nottingham health profile 0-600 Top=High is poor outcome; Comments: Baseline values: 189.5 (118.1); 187.8 (137.4)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 2; Group 2 Number missing: 2

Protocol outcome 3: Physical function at Define
- Actual outcome: Neck disability at 6 weeks; Group 1: mean 8.2 (SD 4.8); n=18, Group 2: mean 11.3 (SD 6.3); n=18; Neck disability index Unclear Top=High is poor outcome; Comments: Baseline values: 15.5 (5.3); 17.5 (7.1)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 2; Group 2 Number missing: 2

Protocol outcome 4: Psychological distress (depression/anxiety) at Define
- Actual outcome: Depression at 6 weeks; Group 1: mean 6.4 (SD 6.1); n=18, Group 2: mean 9.7 (SD 7.7); n=18; Beck Depression Inventory 0-63 Top=High is poor outcome; Comments: Baseline values: 10.8 (6.2); 12.4 (9.4)
Risk of bias: All domain - Very high, Selection - Very high, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover -

Low; Indirectness of outcome: No indirectness ; Baseline details: difference in age between groups; Group 1 Number missing: 2; Group 2 Number missing: 2	
Protocol outcomes not reported by the study	Use of healthcare services at Define; Sleep at Define; Discontinuation at Define

Study	Valim 2003[259]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=76)
Countries and setting	Conducted in Brazil; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention time: 20 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Met ACR criteria for FMS
Exclusion criteria	Cardiorespiratory diseases, neurological disorders, high BMI, hypothyroidism or other rheumatic diseases.
Recruitment/selection of patients	Outpatient clinic
Age, gender and ethnicity	Age - Mean (SD): 46.8(11) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Symptom duration not specified. All patients newly diagnosed and had no previous treatment
Indirectness of population	No indirectness
Interventions	(n=38) Intervention 1: Aerobics - Walking. Walking programme monitored and supervised by a physiotherapist 3 times a week, with 45 minute duration for 20 weeks. Speed was determined by the training heart rate Patients cool down after each session consisted of making rhythmic movements to promote cooling off for 5 minutes. Duration 20 weeks. Concurrent medication/care: Acetaminophen allowed as rescue treatment. Indirectness: No indirectness (n=38) Intervention 2: Flexibility. 3 sessions a week of 45 minute duration including 17 stretching exercises using both muscles and joints. Each position sustained for maximum 30 seconds (supervised by physiotherapist). Duration 20 weeks. Concurrent medication/care: Acetaminophen allowed as rescue treatment. Indirectness: No indirectness
Funding	Academic or government funding (State of Sao Paulo funding)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: WALKING versus FLEXIBILITY

Protocol outcome 1: Pain reduction
- Actual outcome: VAS at 20 weeks; Group 1: mean 3.42 (SD 2.5); n=32, Group 2: mean 4.6 (SD 2.18); n=28; VAS 0-10 Top=High is poor outcome; Comments: 6.19(1.64); 6(2.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation
- Actual outcome: VAS at 10 weeks; Group 1: mean 5 (SD 2.71); n=32, Group 2: mean 4.7 (SD 2.5); n=28; VAS 0-10 Top=High is poor outcome; Comments: 6.19(1.64); 6(2.1)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation

Protocol outcome 2: Quality of life
- Actual outcome: SF-36 physical component summary score at 10 weeks; Group 1: mean 45.37 (SD 8.73); n=32, Group 2: mean 42.55 (SD 7.53); n=28; sf-36 subscale 0-100 Top=High is poor outcome; Comments: Baseline: 37.86(9.53); 34.73(7.32)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation
- Actual outcome: SF-36 physical component summary score at 20 weeks; Group 1: mean 45.37 (SD 8.73); n=32, Group 2: mean 42.82 (SD 9.48); n=28; sf-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 37.86(9.53); 34.73(7.32)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation
- Actual outcome: SF-36 mental component summary score at 10 weeks; Group 1: mean 44.13 (SD 12.1); n=32, Group 2: mean 39.87 (SD 11.4); n=28; sf-36 subscale 0-100 Top=High is good outcome; Comments: Baseline: 34.18(11.36); 37.2(9.51)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation
- Actual outcome: SF-36 mental component summary score at 20 weeks; Group 1: mean 48 (SD 10.23); n=32, Group 2: mean 40.09 (SD 11.28); n=28; sf-36 0-100 Top=High is good outcome; Comments: Baseline: 34.18(11.36); 37.2(9.51)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation

Protocol outcome 3: Psychological distress (depression/anxiety) - Actual outcome: Beck depression inventory at 10 weeks; Group 1: mean 14 (SD 7.892); n=32, Group 2: mean 13.56 (SD 10.26); n=28; BDI 0-21 Top=High is poor outcome; Comments: Baseline: 19.9(7.88); 13.89(7.89) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation - Actual outcome: Beck depression inventory at 20 weeks; Group 1: mean 11.41 (SD 6.24); n=32, Group 2: mean 12.15 (SD 8.4); n=28; BDI 0-21 Top=High is poor outcome; Comments: Baseline: 19.9(7.88); 13.89(7.89) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation - Actual outcome: State trace anxiety inventory at 10 weeks; Group 1: mean 45.57 (SD 9.17); n=32, Group 2: mean 47.4 (SD 8.61); n=28; STAI-state 0-100 Top=High is poor outcome; Comments: Baseline: 46.52(8.34);50.07(8.93) Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation - Actual outcome: State trace anxiety inventory at 20 weeks; Group 1: mean 40.21 (SD 9); n=32, Group 2: mean 45.04 (SD 8.34); n=28; STAI-trace 0-100 Top=High is poor outcome; Comments: Baseline: 46.52(8.34);50.07(8.93) Risk of bias: All domain – Ver9.48y28 high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 10, Reason: Returned to jobs, dropped out without explanation, did not attend all sessions; Group 2 Number missing: 6, Reason: Lack of adherence, no explanation, vacation Protocol outcome 4: Discontinuation - Actual outcome: Discontinuation at 20 weeks; Group 1: 10/38, Group 2: 6/38 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness ; Baseline details: Multiple outcomes including BDI; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Physical function ; Use of healthcare services ; Sleep

Study	Van eijk-hustings 2013 [264]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=203); Note: 3-arm RCT; only 2 arms extracted (third arm included pain management programme evidence review)
Countries and setting	Conducted in Netherlands; Setting: outpatient rheumatology clinics of three medical centres
Line of therapy	Not applicable
Duration of study	Intervention + follow up: 21-24 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: diagnosed FM patients according to the American College of Rheumatology criteria
Stratum	Overall: NA
Subgroup analysis within study	Not applicable: NA
Inclusion criteria	recently (<3 months) diagnosed FM patients according to the American College of Rheumatology criteria, literate and between 18 and 65 years old
Exclusion criteria	pregnancy, involvement in litigation concerning work disability procedures, use of other non-pharmacological treatments such as psychological or physical treatment, interfering with the intervention, alcohol or drugs abuse and use of walking devices
Recruitment/selection of patients	consecutive patients meeting the inclusion criteria during the recruitment period
Age, gender and ethnicity	Age - Range of means: intervention 41 years, control 43 years. Gender (M:F): intervention 148/7. Ethnicity: not reported
Further population details	1. Age 16-18 years: Over 18 years 2. Cognitive impairment: Not stated / Unclear 3. First language not English: Not applicable 4. Homeless: Not stated / Unclear 5. Learning difficulties: Not stated / Unclear 6. Sensory impairment : Not stated / Unclear
Indirectness of population	No indirectness: NA
Interventions	(n=47) Intervention 1: Aerobic exercise. a 12-week group course which was given twice a week by a trained physiotherapist in a community gym, on the floor. Every session started with a 10-min warm up, comprising AE and stretching, followed by an aerobic part during 30 min. The low- intensity aerobic part aimed to reach 55–64 % of the

	predicted maximum heart rate. Patients were instructed to check heart rate by self-control after the warm up and after the aerobic part a few times during the course. They were asked to communicate this with the trainer to check if the intensity of their aerobic training was sufficient. Then, resistance training was applied during 15 min to strengthen major muscle groups. During the course, the intensity of the resistance training increased in weights, frequency and tempo. Finally, every session was finished with a 5-min cool down. Participants received a digital video disc presenting exercises to do at home, and they were advised to perform these once a week. These home exercises were not monitored. The AE group should also consist of nine to ten persons and started when enough participants for the intervention were available. (n=48) Intervention 2: Standard care (a few GP appointments)/waiting list . At least individualised education about FM and lifestyle advice by a rheumatologist or a specialised rheumatology nurse within one or two consultations, but could also include a diversity of other treatments such as physiotherapy or social support from the rheumatology nurse. Duration 1 year. Concurrent medication/care: not reported. Indirectness: No indirectness; Indirectness comment: NA
Funding	Other (supported by Maastricht University Medical Centre and by Care Renewal Grants of medical insurance companies in the region)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC EXERCISE versus STANDARD CARE (A FEW GP APPOINTMENTS)/WAITING LIST

Protocol outcome 1: Quality of life
- Actual outcome: EQ-5D at 12 weeks; Group 1: mean 0.47; n=47, Group 2: mean 0.5; n=48; EQ-5D -0.59-1 Top=High is good outcome; Comments: intervention SE=0.05, control SE=0.04, baseline values: intervention 0.36 (SE 0.03), control 0.51 (SE 0.04),
Risk of bias: All domain – Very high, Selection - Low, Blinding - Low, Incomplete outcome data – Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: EQ-5D at 18 months (after 12 week programme); Group 1: mean 0.54; n=47, Group 2: mean 0.51; n=48; EQ-5D -0.59-1 Top=High is good outcome; Comments: intervention SE=0.05, control SE=0.05,
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: EQVAS at 12 weeks; Group 1: mean 53.9; n=47, Group 2: mean 48.3; n=48; EQ-5D Visual Analogue Scale 0-100 Top=High is good outcome; Comments: intervention SE=3.2, control SE=2.9, baseline values: intervention 48.1 (SE 1.7), control 54 (SE 2.6),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: EQVAS at 18 months (after 12 week programme); Group 1: mean 53.3; n=47, Group 2: mean 51.9; n=48; EQ-5D Visual Analogue Scale 0-100

Top=High is good outcome; Comments: intervention SE=3.6, control SE=3.3, baseline values: intervention 48.1 (SE 1.7), control 54 (SE 2.6)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA

Protocol outcome 2: Physical function
- Actual outcome: FIQ physical function subscale at 12 weeks; Group 1: mean 3.7; n=47, Group 2: mean 4; n=48; FIQ physical function subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.3, control SE=0.3, baseline values: intervention 4.2 (SE 0.2), control 3.4 (SE 0.3)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: FIQ physical function subscale at 18 months (after 12 week programme); Group 1: mean 3.6; n=47, Group 2: mean 3.9; n=48; FIQ physical function subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.6, control SE=0.3, baseline values: intervention 4.2 (SE 0.2)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA

Protocol outcome 3: Psychological distress (depression/anxiety)
- Actual outcome: FIQ anxiety subscale at 12 weeks; Group 1: mean 4.6; n=47, Group 2: mean 5.2; n=48; FIQ anxiety subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.4, control SE=0.4, baseline values: intervention 5.9 (SE 0.3), control 4.8 (SE 0.4)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: FIQ anxiety subscale at 18 months (after 12 week programme); Group 1: mean 5; n=47, Group 2: mean 4.8; n=48; FIQ anxiety subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.5, control SE=0.4, baseline values: intervention 5.9 (SE 0.3)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: FIQ depression subscale at 12 weeks; Group 1: mean 4.6; n=47, Group 2: mean 4.5; n=48; FIQ depression subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.4, control SE=0.4, baseline values: intervention 5.2 (SE 0.3), control 4.2 (SE 0.4),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: FIQ depression subscale at 18 months (after 12 week programme); Group 1: mean 5; n=47, Group 2: mean 4.2; n=48; FIQ depression subscale 0-10 Top=High is poor outco1.5me; Comments: intervention SE=0.5, control SE=0.4, baseline values: intervention 5.2 (SE 0.3), control 4.2 (SE 0.4)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA

Protocol outcome 4: Use of healthcare services
- Actual outcome: GP contacts (2 monthly cost questionnaire) at 12 weeks; Group 1: mean 1.5; n=47, Group 2: mean 0.5; n=48; number of contacts; Comments: intervention SE=0.4, control SE=0.4, baseline values: intervention 2.3 (SE 0.3), control 1.4 (SE 0.3)
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover -

Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: GP contacts (2 monthly cost questionnaire) at 18 months (after 12 week programme); Group 1: mean 1; n=47, Group 2: mean 0.7; n=48; number of contacts; Comments: intervention SE=0.4, control SE=0.3, baseline values: intervention 2.3 (SE 0.3), control 1.4 (SE 0.3),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: medical specialist contacts (2 monthly cost questionnaire) at 12 weeks; Group 1: mean 0.3; n=47, Group 2: mean 0.2; n=48; number of contacts; Comments: intervention SE=0.1, control SE=0.1, baseline values: intervention 1.9 (SE 0.1), control 1.6 (SE 0.1),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: medical specialist contacts (2 monthly cost questionnaire) at 18 months (after 12 week programme); Group 1: mean 0.4; n=47, Group 2: mean 0.2; n=48; number of contacts; Comments: intervention SE=0.1, control SE=0.1, baseline values: intervention 1.9 (SE 0.1), control 1.6 (SE 0.1),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: physiotherapist contacts (2 monthly cost questionnaire) at 12 weeks; Group 1: mean 0.3; n=47, Group 2: mean 3.4; n=48; number of contacts; Comments: intervention SE=0.1, control SE=0.7, baseline values: intervention 2.7 (SE 0.5), control 1 (SE 0.5),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: physiotherapist contacts (2 monthly cost questionnaire) at 18 months (after 12 week programme); Group 1: mean 0.4; n=47, Group 2: mean 2.8; n=48; number of contacts; Comments: intervention SE=0.1, control SE=0.7, baseline values: intervention 2.7 (SE 0.5), control 1 (SE 0.5),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA

Protocol outcome 5: Sleep
- Actual outcome: FIQ unrefreshed sleep subscale at 12 weeks; Group 1: mean 7; n=47, Group 2: mean 7.2; n=48; FIQ unrefreshed sleep subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.33, control SE=0.3, baseline values: intervention 8.2 (SE 0.2), control 7.6 (SE 0.3),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA
- Actual outcome: FIQ unrefreshed sleep subscale at 18 months (after 12 week programme); Group 1: mean 7.2; n=47, Group 2: mean 7.6; n=48; FIQ unrefreshed sleep subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.4, control SE=0.4, baseline values: intervention 8.2 (SE 0.2), control 7.6 (SE 0.3),
Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA

Protocol outcome 6: Discontinuation
- Actual outcome: discontinuation at 12 weeks; Group 1: 28/47, Group 2: 0/48;

Risk of bias: All domain - Low, Selection - Low, Blinding - Low, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 0, Reason: NA; Group 2 Number missing: 0, Reason: NA Protocol outcome 7: Pain reduction - Actual outcome: FIQ pain subscale at 12 weeks; Group 1: mean 5.3; n=47, Group 2: mean 5.7; n=48; FIQ pain subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.31, control SE=0.3, baseline values: intervention 6.3 (SE 0.2), control 5.5 (SE 0.3), Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA - Actual outcome: FIQ pain subscale at 18 months (after 12 week programme); Group 1: mean 5.2; n=47, Group 2: mean 5.3; n=48; FIQ pain subscale 0-10 Top=High is poor outcome; Comments: intervention SE=0.37, control SE=0.3, baseline values: intervention 6.3 (SE 0.2), control 5.5 (SE 0.3), Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: NA; Group 1 Number missing: 28, Reason: NA; Group 2 Number missing: 0, Reason: NA	
Protocol outcomes not reported by the study	Pain interference; Pain self-efficacy

Study	Viljanen 2003[267]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=393)
Countries and setting	Conducted in Finland; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention + follow up: 12 week intervention, 1 year follow up
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women aged 30 to 60 years old
Exclusion criteria	Cancer, major trauma, other causes of neck pain or major rehabilitation in the previous 3 months.
Recruitment/selection of patients	From occupational health physicians
Age, gender and ethnicity	Age - Mean (SD): 44(7) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: 4. chronic widespread pain:
Extra comments	Chronic non-specific neck pain for at least 12 weeks (mean pain duration 10.8(6.3) years
Indirectness of population	No indirectness
Interventions	(n=135) Intervention 1: Strength. Led by trained physiotherapist 3 times a week sessions for 30 minutes each, followed by on week of reinforcement training 6 months after randomisation. Dumbbells were used for dynamic muscle training (weight 1-3kg each according to maximum repetitions with a test weight of 7.5 kg). The Exercises, conducted in the same order in each session, were chosen to activate large muscle groups in the neck and shoulder region. After the 5thweek participants were taught 3 exercises from the program with stretches, after the 9th week they were asked to perform the full training program by themselves. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness (n=128) Intervention 2: Psychological intervention - Relaxation. Led by trained physiotherapist 3 times a week sessions for 30 minutes each, followed by on week of reinforcement training 6 months after randomisation. Exercises aimed to teach participants to activate only those muscles needed for different daily activities and to relax other muscles. Participants were taught to perform the exercises alone from the 5th week. Duration 12 weeks. Concurrent

	medication/care: Not specified. Indirectness: No indirectness (n=130) Intervention 3: Usual care. Usual care, no change to physical activity or means of relaxation during the 12 months of follow up. Duration 12 weeks. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Finnish work environment fund)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus RELAXATION

Protocol outcome 1: Pain reduction
- Actual outcome: Numeric rating scale at 12 months follow up (including 12 week intervention); Group 1: mean 3.1 (SD 2.5); n=135, Group 2: mean 3.3 (SD 2.6); n=128; NRS 0-10 Top=High is poor outcome; Comments: Baseline: 4.8(2.3); 4.8(2.3)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 24; Group 2 Number missing: 18

Protocol outcome 2: Discontinuation
- Actual outcome: Discontinuation at 12 months follow up (including 12 week intervention); Group 1: 24/135, Group 2: 18/128
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: Unclear if dropped out of study or intervention; Group 1 Number missing: 0; Group 2 Number missing: 0

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: Numeric rating scale at 12 months follow up (including 12 week intervention); Group 1: mean 3.1 (SD 2.5); n=135, Group 2: mean 3.2 (SD 2.5); n=130; NRS 0-10 Top=High is poor outcome; Comments: Baseline: 4.8(2.3); 4.1(2.2)
Risk of bias: All domain - Very high, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 24; Group 2 Number missing: 11

Protocol outcome 2: Discontinuation
- Actual outcome: Discontinuation at 12 months follow up (including 12 week intervention); Group 1: 24/135, Group 2: 11/130
Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness, Comments: Unclear if dropped out of study or intervention; Group 1 Number missing: 0; Group 2 Number missing: 0

Protocol outcome 2: Discontinuation

- Actual outcome: Discontinuation at 12 months follow up (including 12 week intervention); Group 1: 18/128, Group 2: 11/130 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: Serious indirectness, Comments: Unclear if dropped out of study or intervention; Group 1 Number missing: 0; Group 2 Number missing: 0	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep

Study	Von trott 2009[271]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=121)
Countries and setting	Conducted in Germany; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 3 months (and 6 months follow up)
Method of assessment of guideline condition	Unclear method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Age 55 or older, had recurrent neck pain for at least 6 months, had an average pain intensity of more than 30 on the 100mm visual analogue scale in the 7 days before baseline assessment, and gave written informed consent
Exclusion criteria	One or more of the following: serious acute or chronic organic illness or mental disorder that disallowed participation in the study, planned start of a physiotherapeutic treatment for neck pain during study participation, or participation in another study during the last 6 months before study entry
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): Qigong: 75.9 (7.6); exercise: 76.0 (7.2); waiting list: 75.7 (7.6). Gender (M:F): 10/111. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain 3. chronic visceral pain: Not applicable 4. chronic widespread pain: Not applicable Not applicable
Indirectness of population	No indirectness
Interventions	(n=38) Intervention 1: Mind-body exercises - Qigong. Twenty-four sessions (each 45 minutes), held over a period of 3 months, in groups of 6-12 participants. Qigong lessons started with about 10 minutes of typical qigong 'opening' exercises, continued with up to 4 exercises of Dantian Qigong, and finished with about 10 minutes of 'closing' exercises. Duration 3 months. Concurrent medication/care: All participants were free to treat their neck pain with the treatment or therapies they were using prior to randomisation . Indirectness: No indirectness (n=39) Intervention 2: Strength/conditioning and flexibility. A standardised programme for computer and workplace related neck pain. It included repeated active cervical rotations as well and strength and flexibility exercises. Special intention as paid so that the patients' individual pain limits were not exceeded. About 90% of the exercises were repeated in each session; some 10% was exchanged regularly. Duration 3 months. Concurrent medication/care: All

	participants were free to treat their neck pain with the treatment or therapies they were using prior to randomisation . Indirectness: No indirectness (n=40) Intervention 3: Usual care. Waiting list control participants did not receive Qigong or exercise therapy. Duration 3 months. Concurrent medication/care: All participants were free to treat their neck pain with the treatment or therapies they were using prior to randomisation . Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus OTHER MIXED MODALITY EXERCISE

Protocol outcome 1: Pain reduction
- Actual outcome: average neck pain at end of treatment; Group 1: mean 47.4 (SD 30.8); n=31, Group 2: mean 44.5 (SD 25.7); n=35; VAS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 56.4±19.7; exercise 47.1±19.6
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 4

Protocol outcome 2: Quality of life
- Actual outcome: QoL (physical) at end of treatment; Group 1: mean 30.4 (SD 7.4); n=31, Group 2: mean 30.3 (SD 7.8); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 30.4±7.9; exercise 28.7±7.2
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 5
- Actual outcome: QoL (mental) at end of treatment; Group 1: mean 48.8 (SD 9.8); n=31, Group 2: mean 49.2 (SD 10.9); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 46.8±9.1; exercise 49.6±10.9
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 4

Protocol outcome 3: Physical function
- Actual outcome: neck pain/disability at end of treatment; Group 1: mean 34.3 (SD 23.6); n=31, Group 2: mean 33.6 (SD 25.5); n=35; NPDS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 38.5±19.2; exercise 41.8±24.9
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 5

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: depression at end of treatment; Group 1: mean 19.7 (SD 7.4); n=31, Group 2: mean 20.2 (SD 9.8); n=35; depression scale 0-60 Top=High is poor

outcome; Comments: Baseline: qigong 18.7±9.1; exercise 18.4±9.4
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 5

Protocol outcome 5: Discontinuation
- Actual outcome: discontinuation at end of treatment; Group 1: 7/38, Group 2: 4/39
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: average neck pain at end of treatment; Group 1: mean 47.4 (SD 30.8); n=31, Group 2: mean 54.9 (SD 28.5); n=35; VAS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 56.4±19.7; usual care 49.9±20.3
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 5

Protocol outcome 2: Quality of life
- Actual outcome: QoL (mental) at end of treatment; Group 1: mean 48.8 (SD 9.8); n=31, Group 2: mean 39.8 (SD 12.6); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 46.8±9.1; usual care 49.9±9.1
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 5
- Actual outcome: QoL (physical) at end of treatment; Group 1: mean 30.4 (SD 7.4); n=31, Group 2: mean 28.6 (SD 9.7); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: qigong 30.4±7.9; usual care 30.6±9.3
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 5

Protocol outcome 3: Physical function
- Actual outcome: neck pain/disability at end of treatment; Group 1: mean 34.3 (SD 23.6); n=31, Group 2: mean 39.1 (SD 21.7); n=35; NPDS 0-100 Top=High is poor outcome; Comments: Baseline: qigong 38.5±19.2; usual care 36.1±20.8
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 5

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: depression at end of treatment; Group 1: mean 19.7 (SD 7.4); n=31, Group 2: mean 18.6 (SD 8); n=35; depression scale 0-60 Top=High is poor outcome; Comments: Baseline: qigong 18.7±9.1; usual care 15.7±7.7

Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 7; Group 2 Number missing: 5

Protocol outcome 5: Discontinuation
- Actual outcome: discontinuation at end of treatment; Group 1: 7/38, Group 2: 5/40
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: OTHER MIXED MODALITY EXERCISE versus USUAL CARE

Protocol outcome 1: Pain reduction
- Actual outcome: Average neck pain at end of treatment; Group 1: mean 44.5 (SD 25.7); n=35, Group 2: mean 54.9 (SD 28.5); n=35; VAS 0-100 Top=High is poor outcome; Comments: Baseline: exercise 47.1±19.6; usual care 49.9±20.3
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 4; Group 2 Number missing: 5

Protocol outcome 2: Quality of life
- Actual outcome: QoL (physical) at end of treatment; Group 1: mean 30.3 (SD 7.8); n=35, Group 2: mean 28.6 (SD 9.7); n=35; SF36 physical component summary 0-100 Top=High is good outcome; Comments: Baseline: exercise 28.7±7.2; usual care 30.6±9.3
Risk of bias: All domain - ; Indirectness of outcome: No indirectness
- Actual outcome: QoL (mental) at end of treatment; Group 1: mean 49.2 (SD 10.9); n=35, Group 2: mean 49.8 (SD 12.6); n=35; SF36 0-100 Top=High is good outcome; Comments: Baseline: exercise 49.6±10.9; usual care 49.9±9.1
Risk of bias: All domain - ; Indirectness of outcome: No indirectness

Protocol outcome 3: Physical function
- Actual outcome: neck pain/disability at end of treatment; Group 1: mean 33.6 (SD 25.5); n=35, Group 2: mean 39.1 (SD 21.7); n=35; Neck pain and disability scale 0-100 Top=High is poor outcome; Comments: Baseline: exercise 41.8±24.9; control 36.1±20.8
Risk of bias: All domain - ; Indirectness of outcome: No indirectness

Protocol outcome 4: Psychological distress (depression/anxiety)
- Actual outcome: depression at end of treatment; Group 1: mean 20.2 (SD 9.8); n=35, Group 2: mean 18.6 (SD 8); n=35; Allgemeine Depressionsskala (depression scale) 0-60 Top=High is poor outcome; Comments: Baseline: exercise 18.4±9.4; usual care 15.7±7.7
Risk of bias: All domain - ; Indirectness of outcome: No indirectness

Protocol outcome 5: Discontinuation
- Actual outcome: discontinuation at end of treatment; Group 1: 4/39, Group 2: 5/40

Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Use of healthcare services ; Sleep

Study	**Waling 2002[273]**
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=126)
Countries and setting	Conducted in Sweden; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention + follow up: 10 weeks + 3 years
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women younger than 45 years who reported work-related trapezius myalgia. The diagnosis of trapezius myalgia was based on the presentation of symptoms such as pain in the descending part of the trapezius muscle, tenderness at palpation, and a limited range of motion in the cervical spine, as well as the exclusion of diseases with other origins. To be defined as work related, the pain and discomfort had to be related to the work situation and assume such intensity that working required extra effort. At least a 1-year history of neck and shoulder problems was required, but sick leave during the last year could not exceed 1 month.
Exclusion criteria	Not reported
Recruitment/selection of patients	Recruited through advertising at workplaces
Age, gender and ethnicity	Age - Mean (SD): 37.9 (5.8). Gender (M:F): Women only. Ethnicity: Not reported

Further population details	1. chronic orofacial pain: people with pain conditions other than chronic orofacial pain 2. chronic primary musculoskeletal pain: people with pain conditions other than chronic primary musculoskeletal pain 3. chronic visceral pain: people with pain conditions other than chronic visceral pain 4. chronic widespread pain: people with pain conditions other than chronic widespread pain 5. complex regional pain syndrome: people with complex regional pain syndrome
Extra comments	Duration of pain: 6.7 (4.2) years
Indirectness of population	No indirectness
Interventions	(n=68) Intervention 1: Mixed modality exercise - Other mixed modality exercise. Half of participants underwent strength training and half underwent aerobic (endurance) training. A physiotherapist supervised the training that was conducted 3 times weekly, 1 hour at a time, over a 10-week period. Strength training consisted of neck and shoulder exercises with individualized loads of 10 to 12maximal voluntary contractions in three sets. Endurance training of the shoulder muscles consisted of arm-cycling and arm exercises with rubber band resistance on the endurance level (30 RM = repetition maximum). Duration 10 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness (n=27) Intervention 2: Other. Participants, led by an occupational nurse, studied stress management once a week, 2 hours at a time, for 10 weeks. No exercises were performed in this group.. Duration 10 weeks. Concurrent medication/care: Not reported. Indirectness: No indirectness
Funding	Funding not stated

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: OTHER MIXED MODALITY EXERCISE versus OTHER

Protocol outcome 1: Pain reduction at Define
- Actual outcome: Pain in general at 3 years; Group 1: mean 30.5 (SD 20.46); n=68, Group 2: mean 20 (SD 18); n=27; VAS 0-100 Top=High is poor outcome; Comments: Baseline values: strength 39 (18); endurance 40 (21); control 43 (19)
Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0, Reason: 0 ; Group 2 Number missing: 0
- Actual outcome: Pain in general at 10 weeks; Group 1: mean 13 (SD 23.05); n=68, Group 2: mean 0 (SD 12.64); n=27; VAS 0-100 Top=High is poor outcome; Comments: Baseline values: strength 18 (95% CI 8-28); endurance 8 (95% CI 3-13); control 0 (95% CI -5-5)

Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0, Reason: 0 ; Group 2 Number missing: 0 Protocol outcome 2: Use of healthcare services at Define - Actual outcome: Health care utilisation at 3 years; Group 1: 23/57, Group 2: 10/21 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 13, Reason: 0 ; Group 2 Number missing: 4	
Protocol outcomes not reported by the study	Quality of life at Define; Physical function at Define; Psychological distress (depression/anxiety) at Define; Sleep at Define; Discontinuation at Define

Study	Wang 2018[274]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=226 (3 arms not extracted))
Countries and setting	Conducted in USA; Setting: Tufts medical center, Boston
Line of therapy	Unclear
Duration of study	Intervention time: 24 weeks plus 1 year follow up
Method of assessment of guideline condition	Adequate method of assessment/diagnosis: ACR
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	21 years or older, fulfilled the ACR 1990 criteria for fibromyalgia and 2010 preliminary diagnostic criteria for fibromyalgia (history of bilateral musculoskeletal pain both above and below the waist for minimum of 3 months and pain in at least 11 of 18 specific tender points, with moderate or greater tenderness on palpation)
Exclusion criteria	Those who had already participated in tai chi or other similar types of complementary and alternative medicine within the last 6 months, those with serious medical conditions that could limit their participation, those with other causes of pain such as inflammation, connective tissue diseases or women who were pregnant or planning a pregnancy.
Recruitment/selection of patients	Advertisements/enrollment through clinics in the Boston area
Age, gender and ethnicity	Age - Mean (SD): 51(13) years. Gender (M:F): 98:3 Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with chronic widespread pain
Extra comments	Mean pain duration 12.5(9.8) years
Indirectness of population	No indirectness
Interventions	(n=36) Intervention 1: Mind-body exercises - Tai Chi. Each session lasted 60 minutes and ran twice a week for 24 weeks. Participants were encouraged to integrate at least 30 minutes of tai chi into their daily routine during the intervention, and to continue this throughout the 52 week follow up. Sessions were run by experienced instructors and sessions were recorded to monitor quality and provide feedback to instructors. Participants also received printed materials on tai chi principles and fibromyalgia. The sessions included warm up, meditative movements, breathing techniques and various relaxation methods. Duration 24 weeks. Concurrent medication/care: Participants were allowed to continue their medication throughout the study. Indirectness: No indirectness

	(n=75) Intervention 2: Aerobic and flexibility. Each session lasted 60 minutes and ran twice a week for 24 weeks. Participants were encouraged to integrate at least 30 minutes of aerobic exercise into their daily routine during the intervention, and to continue this throughout the 52 week follow up. Sessions were closely supervised in a group format and were moderate intensity. Each session consisted of an active warm-up, choreographed aerobic training that progressed gradually from low to moderate intensity and a cool down involving low intensity movements and dynamic and static stretching. During the first week there was a 15 minute warm up, 20 minutes of aerobic training and 25 minutes of cool-down, which increased to 40 minutes of aerobic training by week 10 to (at 60-70% of estimated maximum heart rate). Duration 24 weeks. Concurrent medication/care: Participants were allowed to continue their drugs throughout the duration of the study. Indirectness: No indirectness
Funding	Academic or government funding (National centre for complementary and integrative health of the NIH)

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: AEROBIC AND FLEXIBILITY versus MIND-BODY (TAI-CHI); SDs calculated from CIs

Protocol outcome 1: Quality of life
- Actual outcome: SF-36 physical summary score at 12 weeks; Group 1: mean 1.8 (CIs -0.1-3.6, SD 5.66); n=36, Group 2: mean 3.3 (CIs 0.7-5.8 SD 11.27); n=75; 0-100 Top=High is poor outcome; baseline:30.3(7.5); 28.5(6.5)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
- Actual outcome: SF-36 physical summary score at follow up; Group 1: mean 2.6 (CI 0.4-4.7, SD 6.58); n=36, Group 2: mean 5.4 (CI 2.2-8.6, SD 14.14); n=75; 0-100, Top=High is poor outcome;
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
Baseline:30.3(7.5); 28.5(6.5)
- Actual outcome: SF-36 mental summary score at 12 weeks; Group 1: 0.6 (CI -2.1 to 3.3, SD 8.27);n=36, Group 2: mean 3.8 (CI 0 to 7.6); n=75; 0-100, Top=High is poor outcome;
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
Baseline: 39.4(11.1); 39.1(9.8)
- Actual outcome: SF-36 mental summary score at follow up; Group 1: mean 3 (CI -0.1 to 6, SD 9.34); n=36, Group 2: mean 5.4 (CI 0.8 to 9.9, SD 20.1); n=75; 0-100, Top=High is poor outcome;
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;
Baseline: 39.4(11.1); 39.1(9.8)

Protocol outcome 2: Physical function

- Actual outcome: 6 minute walking test at 12 weeks; Group 1: mean 9.3 (CI -6.1 to 24.8, SD 47.3); n=36, Group 2: mean 7.4 (CI -14.8 to 29.6, SD 98.1); n=75; Top=High is poor outcome; Comments:
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ; Group 1 Number missing: 1; Group 2 Number missing: 5
- Actual outcome: 6 minute walking test at follow up; Group 1: mean 8 (CI -13.3 to 29.4, SD 65.36); n=36, Group 2: mean 30.2 (CI -1.6 to 61.9, SD 140.28); n=75; Top=High is poor outcome; Comments:
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 3: Psychological distress
- Actual outcome: HADS anxiety at 12 weeks; Group 1: mean 0.2 (CI -0.6 to 1, SD 2.45); n=36, Group 2: mean -1.6 (CI -2.7 to -0.4, SD 5.08); n=75; 0-21, Top=High is poor outcome; Comments: 8.8(3.8); 9.5(4.6) SDs:
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;
- Actual outcome: HADS anxiety at follow up; Group 1: mean -0.4 (CI -1.4to 0.6); n=36, Group 2: mean -2.1 (CI -3.6 to -0.7); n=75; 0-21, Top=High is poor outcome; Comments: 8.8(3.8); 9.5(4.6)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;
- Actual outcome: HADS depression at 12 weeks; Group 1: mean -0.5 (CI-1.3 to 0.3, SD 2.45); n=36, Group 2: mean -1.7 (CI -2.8 to 0.6, SD 7.51); n=75; 0-21, Top=High is poor outcome; Comments: Baseline: 8.5(4.2); 7.6(4.4)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;
- Actual outcome: HADS depression at follow up; Group 1: mean -0.6 (CI -1.6 to 0.4, SD 3.06); n=36, Group 2: mean -2.2 (CI -3.7 to 0.8, SD 9.94); n=75; 0-21, Top=High is poor outcome; Comments: Baseline: 8.5(4.2); 7.6(4.4)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;

Protocol outcome 4: Sleep
- Actual outcome: Sleep at 12 weeks; Group 1: mean -0.9 (CI -1.7 to -0.1, SD 2.45); n=36, Group 2: mean -1.6 (CI -2.8 to -0.4, SD 5.3) n=75; Pittsburgh sleep quality index score, 0-21, Top=High is poor outcome; Baseline 8.8(3.8); 9.5(4.6)
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;
Indirectness of outcome: No indirectness ;
- Actual outcome: Sleep at follow up; Group 1: mean -1.2 (CI -2.3 to -0.1, SD 3.37); n=36, Group 2: mean -2 (CI -3.6 to -0.4, SD 7.07) n=75; Pittsburgh sleep quality index score, 0-21, Top=High is poor outcome; Baseline
Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low;

Indirectness of outcome: No indirectness ; Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at end of treatment; Group 1: 11/36, Group 2: 17/75 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Pain reduction ; Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study	Wong 2018[278]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=37)
Countries and setting	Conducted in USA; Setting: Not reported
Line of therapy	Unclear
Duration of study	Intervention time: 12 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Women with fibromyalgia
Exclusion criteria	Pulmonary, cardiovascular, renal, adrenal, pituitary, sever psychiatric, thyroid diseases, and the use of hormone replacement therapy during the 6 months prior to the study. Participants were also excluded if they had any medication changes in the previous year
Recruitment/selection of patients	Not reported
Age, gender and ethnicity	Age - Mean (SD): exercise 51 (2); control 51 (2). Gender (M:F): Women only. Ethnicity: Not reported
Further population details	Subgroup: Not applicable 2. chronic primary musculoskeletal pain: Not applicable 3. chronic visceral pain: Not applicable 4. chronic widespread pain: people with chronic widespread pain Not applicable
Indirectness of population	No indirectness
Interventions	(n=18) Intervention 1: Mind-body exercises - Tai Chi. Supervised sessions 3 times a week for 12 weeks. In the first session, the instructor explained the theory behind tai chi and its procedures providing participants with printed materials on its principles and techniques. In subsequent sessions, participants practiced 10 forms from the classic Yang style of tai chi. The sessions lasted approximately 55 minutes and included a 10 minute warm up, 40 minutes of practice and exercise finalising with a final 5 minute cool down period. During the sessions, the participants heart rate was 40-50% of the HR reserve as they imitated the instructors motion at the same speed. HR during training sessions was monitored using a polar device. Duration 12 weeks. Concurrent medication/care: not reported. Indirectness: No indirectness (n=19) Intervention 2: Usual care. Participants did not participate in any supervised or unsupervised exercise protocol and were asked to maintain their regular lifestyle habits for the duration of the study. Duration 12 weeks. Concurrent

	medication/care: not reported. Indirectness: No indirectness
Funding	Funding not stated
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: TAI CHI versus USUAL CARE Protocol outcome 1: Pain reduction - Actual outcome: Pain at end of treatment; Group 1: mean 5.3 (SD 1.24); n=17, Group 2: mean 7 (SD 1.87); n=14; VAS 0-10 Top=High is poor outcome; Comments: Baseline: tai chi 7.5±1.7; usual care 7.3±1.74 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1; Group 2 Number missing: 5 Protocol outcome 2: Sleep - Actual outcome: Sleep at end of treatment; Group 1: mean 7.8 (SD 1.24); n=17, Group 2: mean 7.6 (SD 1.5); n=14; VAS 0-100 Top=High is poor outcome; Comments: Baseline: tai chi 7.9±1.27; usual care 7.8±2.62 Risk of bias: All domain - Very high, Selection - High, Blinding - High, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 1; Group 2 Number missing: 5 Protocol outcome 3: Discontinuation - Actual outcome: Discontinuation at end of treatment; Group 1: 1/18, Group 2: 5/19 Risk of bias: All domain - High, Selection - High, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ;	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services

Study	Wu 1999[279]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	(n=26)
Countries and setting	Conducted in USA; Setting: New York, no further details
Line of therapy	Unclear
Duration of study	Intervention time: 10 weeks
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Aged 18 to 65 years and met the diagnostic criteria of late-stage CPRS-I (to have at least 5 of the following criteria): Positive 3 phase bone scan, burning pain, aollodynia, swelling, mottling of the skin, dystrophy of skin and/or muscle, negative diagnostic sympathetic blockade. Participants were also required to have failed to achieve 50% pain reduction through drug therapy or palliative physical or chiropractic e therapy (including TENS, hot and cold therapy).
Exclusion criteria	None specified
Recruitment/selection of patients	Not specified
Age, gender and ethnicity	Age - Mean (SD): 38.5(12.4) years. Gender (M:F): 3:19. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: 3. chronic visceral pain: 4. chronic widespread pain: people with complex regional pain syndrome
Indirectness of population	No indirectness
Interventions	(n=13) Intervention 1: Mind-body exercises - Qigong. 6 sessions of qigong training with 2 recognised qigong masters. Sessions included musical compositions and visual images which were coded to represent specific organ systems which qi is believed to believed to stimulate. Each session lasted 40 minutes twice a week for 3 weeks, followed by 7 weeks of home exercises on a daily basis. Duration 10 weeks. Concurrent medication/care: Not specified . Indirectness: No indirectness (n=13) Intervention 2: Other. 6 sessions of simulated qigong training led by a simulated qigong master, in order to maximise nonspecific treatment effects. Participants were shown visual images and listened to recorded music similar to that in the qigong group. After this time a simulated qi adjustment was performed by the facilitator. Each session lasted for 40 minutes. This was followed by 7 weeks of home exercises. Duration 10 weeks. Concurrent

	medication/care: Not specified . Indirectness: No indirectness
Funding	Academic or government funding (NIH grant)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: QIGONG versus CONTROL GROUP (SHAM QIGONG) Protocol outcome 1: Pain reduction - Actual outcome: VAS at 10 weeks; Group 1: mean 53.8 (SD 28.5); n=8, Group 2: mean 58.7 (SD 26.3); n=10; VAS 0-100 Top=High is poor outcome; Comments: Baseline: 66.7(25.5); 64.5(23.7) Risk of bias: All domain - Very high, Selection - Low, Blinding - Low, Incomplete outcome data - Very high, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 5, Reason: Not specified; Group 2 Number missing: 3, Reason: Not specified	
Protocol outcomes not reported by the study	Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Use of healthcare services ; Sleep ; Discontinuation

Study (subsidiary papers)	Ylinen 2003[284] (Ylinen 2007[281], Ylinen 2006[285])
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=180)
Countries and setting	Conducted in Finland; Setting: Not specified
Line of therapy	Unclear
Duration of study	Intervention + follow up: 2 weeks plus 1 year/3 year follow up
Method of assessment of guideline condition	Method of assessment /diagnosis not stated
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	(1) aged 25-53 years (2) office worker, permanently employed (3) constant of frequently occurring neck pain for more than 6 months
Exclusion criteria	(1) Causes of neck pain such as cervical disorders, conditions affecting the neck and shoulder area, sever trauma, instability, migraine, fibromyalgia, shoulder diseases, nerve entrapment, rheumatic diseases or any other psychiatric illness or disease that could prevent physical loading (2) pregnancy
Recruitment/selection of patients	From various workplaces through occupational health care systems.
Age, gender and ethnicity	Age - Mean (SD): 46(6) years. Gender (M:F): All women. Ethnicity: Not specified
Further population details	Subgroup: 2. chronic primary musculoskeletal pain: people with chronic primary musculoskeletal pain (Chronic cervical pain). 3. chronic visceral pain: 4. chronic widespread pain:
Extra comments	All participants were office workers, duration of pain not stated (minimum duration 6 months)
Indirectness of population	No indirectness
Interventions	(n=60) Intervention 1: Strength. 10 patients in each group, 12 day program with 5 sessions per week, each lasting 45 minutes. Exercises aimed to strengthen neck flexor muscles by using an elastic rubber band to train the muscles at a resistance of 80% of maximum (15 repetitions in each direction). Following this the group performed dynamic exercises for the shoulders and upper extremities, with an individually adjusted single dumbbell, performing only 1 set for each exercise with the highest load possible to perform 15 repetitions. This was followed by exercises for the trunk and leg muscles in the same format, which was then concluded by stretching exercises for 20 minutes. Duration 12 days. Concurrent medication/care: Advised to perform aerobic exercise 3 times a week for half an hour and participants were encouraged to practice exercises at home. Indirectness: No indirectness

	(n=60) Intervention 2: Strength. 10 patients in each group, 12 day program with 5 sessions per week, each lasting 45 minutes. Exercises aimed to strengthen neck flexor muscles by lifting head up from the supine position in 3 series of 20 repetitions. Following this the group performed dynamic exercises for the shoulders and upper extremities, at 3 sets of 20 repetitions for each exercise with a pair of dumbbells each weighing 2 kg. This was followed by exercises for the trunk and leg muscles in the same format, which was then concluded by stretching exercises for 20 minutes. Duration 12 days. Concurrent medication/care: Advised to perform aerobic exercise 3 times a week for half an hour and participants were encouraged to practice exercises at home. Indirectness: No indirectness (n=60) Intervention 3: Flexibility. Control group. Performed recreational activities on assessment days. Received written information about the same stretching exercises and were advised to practice these 20 minutes 3 times a week. They were also advised to perform aerobic exercise 3 times a week. Duration 12 days. Concurrent medication/care: Not specified. Indirectness: No indirectness
Funding	Academic or government funding (Social Insurance Institution, Helsinki)
RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: STRENGTH TRAINING versus STRETCHING Protocol outcome 1: Use of healthcare services - Actual outcome: Visits to physician due to neck pain at 12 month follow up; Group 1: 12/60, Group 2: 20/60 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 0, Reason: Personal reasons, other diagnosis; Group 2 Number missing: 1, Reason: Pregnancy RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: ENDURANCE TRAINING versus STRETCHING Protocol outcome 1: Use of healthcare services - Actual outcome: Visits to physician due to neck pain at 12 month follow up; Group 1: 15/59, Group 2: 20/60 Risk of bias: All domain - High, Selection - Low, Blinding - High, Incomplete outcome data - Low, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: 2, Reason: Personal reasons, other diagnosis; Group 2 Number missing: 1, Reason: Pregnancy	
Protocol outcomes not reported by the study	Pain reduction ; Quality of life ; Physical function ; Psychological distress (depression/anxiety) ; Sleep ; Discontinuation

D.2 Cochrane evidence tables

D.2.1 Bidonde 2017

Author and year	Fontaine 2010
Methods	2 groups: lifestyle physical activity (AE); education (control) Length: 12 weeks; follow-up: 26 weeks and 52 weeks Study design: randomized clinical trial with parallel group
Participants	Female:Male: 73:0 Age (years (SD)): 46.4 (11.6); 49 (10.2) Inclusion: diagnosis of fibromyalgia (ACR 1990), patient at Johns Hopkins Arthritis Center, affiliated Johns Hopkins Rheumatology clinics Exclusion: meeting US Surgeon General's 1996 recommendation for physical activity for previous 6 months (ie, not engaging in moderate-intensity physical activity for 30 minutes on 5 days per week or in vigorous physical activity 3 times per week for 20 minutes each time during the previous month), acute or chronic medical condition that could preclude active participation (cancer, coronary artery disease), intent to change medications that might affect mood, intent to seek professional treatment for anxiety or depression during the study period, not unwilling to make the required time commitment Duration of illness (years (SD)): 5.9 (5.1); 9.6 (6.8)
Interventions	Lifestyle physical activity (n = 43): Increase moderate-intensity physical activity by helping participants find ways to accumulate short bouts of physical activity throughout the day. Frequency: 5-7 times/wk; Duration: 60';Intensity: moderate; Mode: walking (the most common form of LPA) and other forms (eg, gardening/mowing the lawn) of household activity (eg,vacuuming); and sports activity (eg, cycling, swimming, field hockey) Education (n = 33): Provide education and control for effects of being enrolled in a clinical trial and receiving increased attention and social support; Frequency: 1/mo; Duration: 90-120'; Intensity: not applicable; Mode: education, question and answer, and social support
Outcomes	Health-related quality of life (FIQ Total), pain (VAS for pain), fatigue (Fatigue Severity Scale - FSS), CR submax (6-minute walk test) Others: depression (Center for Epidemiological Studies Depression Scale - CES-D), tenderness (tender point count), physical activity level (pedometer); perceived improvement ("Since the start of the study, how much change has there been in your fibromyalgia?") Measurements taken at 0 and 12 weeks

Author and year	Fontaine 2010	
Adherence to exercise protocols	Monitoring methods: intensity monitored by pedometer once a week and diaries used to track mode; adherence criteria: not specified; adherence: unknown	
Congruence with ACSM guidelines for aerobic training	Yes	
Notes	Country: United States Language: English Study author contacted: yes, study author confirmed that participants from the 2 studies (Fontaine 2007 and Fontaine 2010) were different Funding source/declaration of interest: Work was supported by NIH/NIAMS (National Institutes of Health/National Institute of Arthritis and Musculoskeletal Skin Diseases)	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Participants were randomized via a coin flip at a 1:1 allocation ratio to each of the two groups" (page 5)
Allocation concealment (selection bias)	Unclear risk	Insufficient information to permit evaluation of risk
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: health-related quality of life (FIQ Total), pain intensity (VAS for pain), fatigue (Fatigue Severity Scale - FSS)
Blinding of objective outcome assessment (detection bias) All outcomes	Unclear risk	CR submax (6-minute walk test): no information on blinding assessors
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Insufficient information on blinding of participants and personnel to permit judgment of risk
Incomplete outcome data (attrition bias) All outcomes	Low risk	Reasons for missing outcome data unlikely to be related to true outcomes; missing outcome data were balanced in numbers across intervention groups
Selective reporting (reporting bias)	Low risk	Study protocol is available (clinicaltrials.gov NCT00383084) and all of the study's prespecified outcomes of interest in the review have been reported in the prespecified way

Author and year	Fontaine 2010	
Other bias	Low risk	Study appears to be free of other sources of bias

Author and year	Gowans 2001
Methods	2 groups: exercise (AE); control Length: 23 weeks; follow-up: none Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 44:6 Age (years (SD)): 44.6 (8.7); 49.8 (7.3) Inclusion: diagnosis of fibromyalgia (ACR 1990), willingness to comply with experimental protocol Exclusion: diagnosis of high blood pressure or symptomatic cardiac disease, other serious systemic diseases (eg, cancer, diabetes), intention of changing medications for anxiety or depression or seeking professional treatment for anxiety or depression during the study period, enrolled in or intended to begin an aerobic exercise program Duration of illness (years (SD)): symptoms: 9.6 (8.6); 8.4 (7.6); diagnosis: 2.8 (2.6); 4.2 (4.4)
Interventions	Exercise (n = 27): Classes for the first 6 weeks were conducted in a warm therapeutic pool; starting at 7 weeks, participants progressed to 2 walking classes in a gym and 1 pool class. Frequency: 3 hospital-based classes/wk; Duration: 30' (5' stretching first, 20' aerobic, 5' stretching after);Intensity: low to moderate (60% to 75% age-adjusted HRmax); Mode: water (warm) walking/running progressing to land walking/running Control (n = 23): "continue ad libitum activity" (page 520)
Outcomes	Health-related quality of life (FIQ Total), CR submax (6-minute walk test) Other: depression (Beck Depression Index), anxiety (state anxiety inventory), self-efficacy (ASES), tenderness (tender point count), muscle function (isokinetic knee extension strength at 60 degrees) Measurements taken at 0 and 23 weeks
Adherence to exercise protocols	Monitoring methods: HR and attendance were monitored; adherence criteria for efficacy analysis: must attend > 45% of exercise classes; adherence: mean attendance at exercise classes 67% (range 46%–84%)
Congruence with ACSM guidelines for aerobic training	No for healthy adults, based on duration (only 20 minutes per session); met ACSM criteria for individuals who are sedentary/have no habitual activity/are extremely deconditioned
Notes	Country: Canada Language: English Study author contacted: no

Author and year	Gowans 2001	
	Funding sources/declaration of interest: Work was supported by a grant from the Toronto Hospital Auxiliary Women's Health Project on Women and Arthritis (page 528)	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Subjects were stratified by sex and randomly assigned to..." (page 520)
Allocation concealment (selection bias)	Unclear risk	No description of the method used for allocation concealment
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instrument: health-related quality of life (FIQ Total)
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	CR submax (6-minute walk test): "Their distance was recorded to the nearest meter by an assessor blinded to subjects' group assignments" (page 520)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants in the intervention group had no contact with those in the control group; control group did not meet
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data were analyzed by ITT
Selective reporting (reporting bias)	Low risk	Published reports include all expected outcomes
Other bias	Low risk	Study appears to be free of other sources of bias

Author and year	Kayo 2011
Methods	3 groups: walking program (AE); strengthening exercise; control Length: 16 weeks; follow-up: 28 weeks Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 90:0

Author and year	Kayo 2011
	Age (years (SD)): 47.7 (5.3); 46.7 (6.3); 46.1 (6.4) Inclusion: women 30-55 years of age who agreed to participate in an exercise program 3/wk for 16 weeks and to discontinue medications for fibromyalgia 4 weeks before the start of the study; individuals who had at least 4 years of schooling Exclusion: women with contraindications to exercise based on clinical rheumatological examination, those involved in cases of medical litigation Duration of illness (years (SD)): 4.0 (3.1); 4.7 (5.7); 5.4 (3.5)
Interventions	Walking program (n = 30): 48 sessions in total. Frequency: 3/wk; Duration: ˜ 60' (warm-up with 5-10' stretching, conditioning stimulus, cool-down 5'); Intensity: moderate at week 1 to vigorous by week 16 (40%-50% to 60%-70% heart rate reserve by week 16); Mode: supervised indoor or outdoor walking monitored by a heart rate monitor Resistance exercise training (n = 30): 48 sessions in total. Frequency: 3/wk; Duration: ˜ 60'; Intensity: high intensity (4 on 10-point Borg scale), exercise load and intensity increased every 2 weeks (reps - weeks 1 + 2: 3 sets of 10 reps with rest intervals of 1' between sets, weeks 3-16; load - weeks 1-4, no load, weeks 5-16, load included). The training load was individually and systematically adjusted every time the participant performed more than 15 repetitions successfully; Mode: supervised exercise protocol consisting of 11 free active exercises for upper and lower limbs and trunk muscles, with free weights and body weight performed in the standing, sitting, and lying positions Control group (n = 30): control conditions not specified, except study authors stated that participants in all 3 groups were asked to discontinue tricyclic antidepressants but were allowed to use acetaminophen (paracetamol) for pain Co-interventions: Exercise was administered in this study as a single modality; the timing of restarting medication was monitored *For this review: only walking program and control group were considered
Outcomes	Health-related quality of life (FIQ Total), pain (VAS), fatigue (SF-36 Vitality Scale), physical function (SF-36 Physical Function Scale) Other: tenderness (tender point count), mental health (SF-36 mental health) as provided by study author on request Measurements taken at 0, 8, 16, and 28 weeks
Adherence to exercise protocols	Monitoring methods: HR monitored; adherence criteria: drop-outs were those who missed more than 20% of sessions or 3 consecutive sessions; adherence: attendance rate 80%
Congruence with ACSM guidelines for aerobic training	Yes
Notes	Country: Brazil Language: English Study author contacted: yes, study authors provided data on outcomes (fatigue and physical function) Funding source/declaration of interest: none reported
Risk of bias	

Author and year	Kayo 2011	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The allocation sequence was based on a random number list (GraphPad Statmate version 1.0), which was organized by an investigator (MSP)" (online page 2)
Allocation concealment (selection bias)	Low risk	Opaque sealed envelopes were used
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: health-related quality of life (FIQ Total), pain intensity (VAS), fatigue (SF-36 - Vitality Scale), physical function (SF-36 Physical Function Scale)
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	"All patients were clinically examined by the same rheumatologist (CSM), who was blinded to group assignment throughout the study" (pages 2-8)
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Unclear blinding of participants and personnel delivering the intervention
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data were analyzed by ITT
Selective reporting (reporting bias)	High risk	Outcome data for important variables (eg, tender points, SF-36 Physical Functioning, SF-36 Vitality, SF-36 Mental Health) were not provided in the published report, but study authors provided these on request. RCT protocol is available (ClinicalTrials.gov ID NCT00498264)
Other bias	Low risk	No other serious sources of bias is evident

Author and year	King 2002
Methods	4 groups: exercise only (AE); education only; education and exercise; control (wait list) Length: 12 weeks; follow-up: 24 weeks Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 170:0 Age (years (SD)): 45.2 (9.4); 44.9 (10); 47.4 (9); 47.3 (7.3)

Author and year	King 2002
	Inclusion: diagnosis of fibromyalgia (ACR 1990), women 18 to 65 years of age, willing to meet 3 weeks × 12 weeks, persons involved in medico-legal cases were not excluded Exclusion: conditions precluding ability to exercise (severe cardiac arrhythmia, dizziness, severe shortness of breath), inflammatory arthritis, systemic lupus erythematosus, rheumatoid arthritis Duration of illness (years (SD)): 7.8; 10.9; 8.9; 9.6
Interventions	Exercise only (AE) (n = 42): Frequency: 3/wk; Duration: starting duration 10' to 15' progressing to 20' to 40', Intensity: light to moderate (60%-75% predicted HRmax/age); Mode: walking, aquacise (deep and shallow water), or low-impact aerobics Education only (n = 41): based upon principles of self-management. Frequency: 1/wk; Duration: 1 1/2 to 2 hour educational session provided by a multidisciplinary team. Topics focused on potential causes of fibromyalgia, principles of self-management (goal setting, maximizing energy for household chores or personal activities, pain or fatigue coping strategies, benefits of exercise, evaluating alternative therapies, and barriers to behaviour change) Exercise + Education (n = 35): exercise same as for exercise only, and education same as for education only. Frequency: 3/wk (2 exercise sessions/wk and 1 combined educational and exercise session per week) Wait list control (n = 34): a page of written instructions for basic stretches and 5 items related to general coping strategies provided on entry to the study For a, b, c, and d: Participants were instructed not to change their present treatment (ie, medications) for the duration of the study *For this review: only exercise only, education only, and wait list control groups were considered
Outcomes	Health-related quality of life (FIQ Total), CR submax (6-minute walk test) Other: pain (Chronic Pain Self-Efficacy Scale), function (Chronic Pain Self-Efficacy Scale), coping with symptoms (Chronic Pain Self-Efficacy Scale), tenderness (tender point count), and total survey site score Measurements taken at 0, 12, and 24 weeks
Adherence to exercise protocols	Monitoring methods: HR and logbooks; adherence criteria: missed 3 consecutive sessions or 12 of the 36 total; adherence: attendance 75% (21%)
Congruence with ACSM guidelines for aerobic training	No, based on frequency and duration (only 3/wk, light to moderate)
Notes	Country: Canada Language: English Stud author contacted: no Funding sources: Work was supported by grants from the Medical Services Incorporated Foundation and from the Health Services Research and Innovation Fund, Alberta Health, administered by Alberta Heritage Foundation for Medical Research

Author and year	King 2002	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Random assignment of subjects to groups was done in blocks of 4 to 16. A list was prepared prior to start of study using a table of random numbers and subject ID number (order of admission to study" (page 2621)
Allocation concealment (selection bias)	Unclear risk	Insufficient information on allocation concealment to permit judgment of risk
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: health-related quality of life (FIQ Total)
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	CR submax (6-minute walk test): "Baseline testing occurred before randomization" and "both assessors were blinded to the subject's group randomization on subsequent visits" (page 2621)
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants were not blinded (pages 2623 and 2626). It is unlikely that care providers were blinded
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data were analyzed by ITT for post-intervention status; follow-up data were reported and analyzed with completer data
Selective reporting (reporting bias)	Low risk	Study protocol is not available but it is clear that the published report includes all expected outcomes
Other bias	Unclear risk	Insufficient information for assessment of whether an important risk of bias exists

Author and year	Mengshoel 1992
Methods	2 groups: low-impact aerobic dance; control Length: 20 weeks; follow-up: none Study design: randomized clinical trial with parallel groups (age)
Participants	Female:Male: 25:0 Age (years (min to max)): 33.5 (21 to 42); 34 (25 to 38)

Author and year	Mengshoel 1992	
	Inclusion: females with fibromyalgia according to 1990 ACR, normal lab test (haemoglobin, liver enzymes, serum creatinine, ESR, ANA, latex, and thyroxine) Exclusion: none stated Duration of illness (years (min to max)): 8.5 (3 to 20), 8 (3 to 23)	
Interventions	Low-impact aerobic dance (n = 11): Frequency: 2/wk; Duration: 60'; Intensity: moderate to vigorous (HR 120 to 150 bpm); Mode: modified low-impact aerobic dance; exercise for upper extremities performed at intervals between periods of rest; exercises modified to prevent pain, fatigue, and static muscle work Control (n = 14): instructed to not change their habits regarding physical activities	
Outcomes	Pain intensity over past 7 days (VAS - 100 mm), fatigue (VAS - 100 mm) - baseline data only, CR submax (Astrand test, RPE) Other: muscle endurance (grip strength at 1st and 20th rep, duration of shoulder hold in seconds, duration in minutes for stair climbing at a constant velocity), sleep (VAS - 100 mm), pain coping (Vanderbilt Pain Management Inventory), fatigue during exercise (Borg's Rating Scale) Measurements taken at 0, 10, and 20 weeks	
Adherence to exercise protocols	Monitoring methods: HR controlled periodically by pulse watch recorder; adherence criteria: not specified; adherence: attendance not specified	
Congruence with ACSM guidelines for aerobic training	Exercise protocol did not meet the frequency requirement; only 2 times/wk	
Notes	Country: Norway Language: English Study author contact: no Funding sources: Financial support was received from the Norwegian Fund for Postgraduate Training in Physiotherapy, the Olga Immerslund Legacy for Rheumatological Research, the Grethe Harbitz Legacy and Hafslund-Nycomed	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Insufficient information to permit judgment of ‘Yes’ or ‘No’
Allocation concealment (selection bias)	Unclear risk	Insufficient information to permit judgment of ‘Yes’ or ‘No’

Author and year	Mengshoel 1992	
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: pain intensity over past 7 days (VAS - 100 mm), fatigue (VAS - 100 mm) - baseline data only
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	Measure: CR submax (Astrand test). "The testing was undertaken by a physical therapist who was blinded to the patients' classification. At the time of re-test neither the patients nor the physiotherapist had access to the results of the baseline tests" (page 346)
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Insufficient information on blinding of participants and personnel to permit judgment of risk
Incomplete outcome data (attrition bias) All outcomes	High risk	Missing outcome data likely led to an imbalance in results across groups
Selective reporting (reporting bias)	High risk	Insufficient information to permit judgment
Other bias	Low risk	Study appears to be free of other sources of bias

Author and year	Nichols 1994
Methods	2 groups: aerobic exercise (AE); control (daily activities not involving physical activity) Length: 8 weeks; follow-up: none Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 17:2 Age (years (SD)): 47.8 (11.1); 50.8 (11.8) Inclusion: diagnosis of fibromyalgia (ACR 1990) Exclusion: history of heart disease, lung disease, uncontrolled hypertension, or orthopaedic disorders that would preclude aerobic activity; participation in any regular aerobic exercise program within 6 months before the study Duration of illness (years (SD)): > 10; > 10 except for person who had 4 (years)
Interventions	Aerobic exercise (n = 10): "Each session included a warm up and cool down regimen of stretching exercises, 1 warm up and cool down lap of slow paced walking" (page 329). Frequency: 3/wk; Duration: unclear; Intensity: light to moderate (60%-70% predicted HRmax/age); Mode: fast-paced walking on an indoor track Control Group (n = 9): daily activities as usual not involving physical activity

Author and year	**Nichols 1994**	
Outcomes	Discontinuation Outcomes not useable: physical function (Sickness Impact Profile), pain (McGill Pain Questionnaire, Brief Symptom Inventory) Measurements taken at 0 and 8 weeks	
Adherence to exercise protocols	Monitoring methods: HR and cadence monitored at midsession; Adherence criteria: not stated; adherence: all participants were able to achieve 60% to 70% of HRmax	
Congruence with ACSM guidelines for aerobic training	No, based on frequency and duration (only twice a week)	
Notes	Country: United States Language: English Study author contacted: no Funding sources: none stated	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Insufficient information on the method used to generate the allocation sequence to permit judgment of risk (page 329)
Allocation concealment (selection bias)	Unclear risk	No description of the method used for allocation concealment
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: physical function (Sickness Impact Profile)
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	Not applicable: Objective outcomes were not assessed
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Conflicting information regarding whether participants in the exercise and control groups interacted (pages 329 and 331)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Missing outcome data were balanced in numbers across exercise and control groups with similar reasons for missing data across groups

Author and year	Nichols 1994	
Selective reporting (reporting bias)	Low risk	Study protocol is not available but the published report includes all expected outcomes
Other bias	Low risk	Study appears to be free of other sources of bias

Author and year	Sanudo 2010
Methods	3 groups: aerobic exercise (AE); mixed exercise (aerobic + resistance + flexibility); control Length: 24 weeks; follow-up: none Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 64:0 Age (years (SE)): 55.9 (1.6); 55.9 (1.7); 56.6 (1.9) Inclusion: women with diagnosis of fibromyalgia (ACR 1990) Exclusion: presence of inflammatory rheumatic disease and severe psychiatric illness, respiratory or cardiovascular disease that prevented physical exertion, women with fibromyalgia receiving psychological or physical therapy to avoid possible interactions with the present trial Duration of illness (years (SD)): not specified for either group
Interventions	Aerobic exercise (n = 22): supervised aerobic exercise intervention. Frequency: 2/wk; Duration: 45-60' (10' warm-up and 5-10' cool-down, 15-20' of steady state AE, 15' interval training); Intensity: light to moderate (steady state aerobic 60%-65% of HRmax) and moderate to vigorous (interval training 75%-80% HRmax); Mode: Warm-up included slow walks, easy movements of progressive intensity, steady state AE included continuous walking with arm movements and jogging, interval training included aerobic dance and jogging, cool-down included slow walks, easy movements, relaxation training Mixed exercise (aerobics, resistance, flexibility) (n = 21): combined supervised aerobic exercise and resistance exercise. Frequency: 2/wk; Duration: AE and RT same duration, which included 10' warm-up, 10-15' AE, 15-20' RT, 10' FX;Intensity: AE 65%-75% HRmax, RT weights 1-3 kg; Mode: RT 1 set of 8-10 reps for 8 different muscle groups with a load of 1-3 kg, FX 1 set of 3 reps of 8-9 different exercises, maintaining stretch position for 30 seconds, RT and FX exercises focused on main areas of pain in patients with fibromyalgia (deltoids, biceps, neck (trapezius), hops (gluteus, quadriceps), back/chest/torso (latissimus dorsi, pectoralis major, abdominals)) Control group (n = 21): received medical treatment for fibromyalgia and continued normal daily activities, which did not include structured exercise *For this review: only aerobic exercise and control group were considered
Outcomes	Health-related quality of life (FIQ Total), pain (SF-36), fatigue (SF-36), physical function (SF-36), CR submax (6-minute walk test) Other: muscle strength (grip strength), depression (Beck Depression Inventory)

Author and year	Sanudo 2010	
	Measurements taken at 0 and 24 weeks	
Adherence to exercise protocols	Monitoring methods: HR monitoring but unreported results and attendance; adherence criteria: not stated; adherence: attendance rate in 89% and in 86%	
Congruence with ACSM guidelines for aerobic training	No, based on frequency (only twice a week) for aerobics	
Notes	Country: Spain Language: English Study author contacted: yes, study author confirmed that data from 2 studies (J Rehabil Med 2011), although similar, were from 2 different groups of people Funding sources: none stated	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer random number generator was used
Allocation concealment (selection bias)	Low risk	Randomization by member not involved in recruitment or assessment of patients; randomization list kept at a separate location in a locked filing cabinet (page 1839)
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: health-related quality of life (FIQ Total), pain (SF-36), fatigue (SF-36), physical function (SF-36)
Blinding of objective outcome assessment (detection bias) All outcomes	Unclear risk	CR submax (6-minute walk test). No information provided on blinding
Blinding of participants and personnel (performance bias) All outcomes	High risk	Insufficient information on blinding of participants and personnel to permit judgment of risk
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data were analyzed by intention-to-treat
Selective reporting (reporting bias)	Low risk	Study protocol is available and all of the study's prespecified outcomes of interest in the review have been reported in the prespecified way

Author and year	Sanudo 2010	
Other bias	Low risk	Study appears to be free of other sources of bias

Author and year	Schachter 2003
Methods	3 groups: long bout (AE); short bout (AE); control Length: 16 weeks; follow-up: none Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 143:0 Age (years (SD)): 41.3 (8.7); 41.9 (8.6); 42.5 (6.7) Inclusion: diagnosis of fibromyalgia (ACR 1990), sedentary women, 20 to 55 years of age, willing to provide informed consent and be randomly assigned to treatment or control, permission from physician for participation Exclusion: more than 2 coronary artery disease risk factors outlined in 1995 ACSM, known cardiorespiratory or metabolic musculoskeletal or neurological conditions that could interfere with performance of moderate-intensity exercise Duration of illness (years (SD)): not specified for either group Baseline mean and SD (health-related quality of life 55 (1.3), pain 61 (1.97), stiffness 7 (1.9), and physical function 38 (1.86)
Interventions	Long bout aerobic exercise (n = 51): long bout of AE with rhythmical movements designed to use all major muscle groups of the lower extremities performed to music. Frequency: 3 up to 5/wk; Duration: 10' up to 30'; Intensity: moderate on week 1 (40%-50% HRR), vigorous by week 10 (65%-75% HRR) (modulated through changes in music tempo); Mode: home program of low-impact aerobics to videotaped instructor and music, rhythmical movements of lower body muscles Short bout aerobic exercise (n = 56): short bout of AE with rhythmical movements designed to use all major muscle groups of the lower extremities performed to music. Frequency: 3 up to 5/wk; Duration: 2/d 5' up to 15'; Intensity: moderate on week 1 (40%-50% HRR), vigorous by week 10 (65%-75% HRR) (modulated through changes in music tempo); Mode: home program of low-impact aerobics to videotaped instructor and music, rhythmical movements of lower body muscles Control (n = 36): Participants were asked to refrain from starting any new regular physical activity or exercise programs or other non-pharmacological interventions *For this review: All group interventions were considered
Outcomes	Health-related quality of life (FIQ Total), pain (VAS), fatigue (FIQ), stiffness (FIQ), physical function (FIQ impairment), CR max (peak VO2) Other: tenderness (mean myalgic score), clinician global rating (physician rating of global severity), depression (FIQ), anxiety (FIQ), self-efficacy (chronic pain self-efficacy scale), sleep (FIQ)

Author and year	Schachter 2003	
	Measurements taken at 0, 8, and 16 weeks	
Adherence to exercise protocols	Monitoring methods: HR monitoring but unreported results; adherence criteria: exercise adherence calculated in four 4-week phases by dividing the sum of the minutes of exercise performed within a phase (as recorded in the participant's exercise log) by the minimum number of minutes of exercise recommended for that period. Participants met the minimum recommended when they completed ≥ 11 of the 12 recommended sessions in ≥ 22 of the 24 recommended sessions for SBE in over 4 weeks; adherence in 46%, 40%, 42%, and 22% as compared with 68%, 74%, 54%, and 41% in those exercising at or above the minimum level across the 4 phases	
Congruence with ACSM guidelines for aerobic training	Yes	
Notes	Country: Canada Language: English Study author contacted: yes, study author provided additional information on outcome measures, risk of bias, and study procedures Funding source/declaration of interest: Work was supported by Saskatchewan Health Services Utilization and Research Commission, Canada	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random number sequence was prepared by a person not connected with the study
Allocation concealment (selection bias)	Low risk	Assignments were placed in opaque envelopes
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: health-related quality of life (FIQ Total), pain intensity (VAS), fatigue (FIQ), stiffness (FIQ), physical function (FIQ Impairment)
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	CR max (peak VO2). "One rheumatologist who was masked to group assignment conducted all tender point examinations and evaluated fibromyalgia severity of all participants before starting and after completing the study" (page 345)
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants were not blinded to the hypothesis and may have had contact with care providers who worked with other groups, although care providers for group meetings were trained and supervised regarding discussion of only specific topics with each group

Author and year	Schachter 2003	
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data were analyzed by intention-to-treat
Selective reporting (reporting bias)	Low risk	Study protocol is not available but published report includes all expected outcomes
Other bias	Low risk	Study appears to be free of other sources of bias

Author and year	Sencan 2004
Methods	3 groups: aerobic exercise; paroxetine; placebo transcutaneous electrical stimulation (TENS) Length: 6 weeks; follow-up at 26 weeks Study design: randomized clinical trial with parallel groups
Participants	Female:Male: 60:0 Age (years (SD)): 35.4 (9.6); 32.7 (9.4); 35.6 (7.9) Inclusion: diagnosis of fibromyalgia (ACR 1990), no other pharmacological treatment, other comorbid disease Exclusion: tumoral, infectious, metabolic, cardiovascular, or endocrine disease; drug dependency Duration of illness (years (SD)): 4.7; 6.5; 5.1
Interventions	Aerobic exercise (n = 20): aerobic exercise on stationary bicycle. Frequency: 3/wk; Duration: 40 minutes; not specified; Intensity: not specified; Mode: bicycle ergometer Paroxetine (n = 20): undertaken 20 mg/d paroxetine. Frequency: 1/d, home exercise for 6 months' follow-up (followed by telephone calls at 2 and 4 months); Duration: not specified; Intensity: not specified Placebo TENS (n = 20): given placebo TENS. Frequency: 3/wk; Duration: 20 minutes; Intensity: not specified; Mode: electrodes applied on the 2 most painful tender points (no current) *For this review: All interventions were considered
Outcomes	Pain intensity (VAS) Other outcomes not useable: tenderness (pressure algometry), depression (Beck Depression Inventory) Measurements taken at 0, 6, and 26 weeks
Adherence to exercise protocols	Monitoring methods: not specified; adherence criteria: not specified; adherence: unknown

Author and year	Sencan 2004	
Congruence with ACSM guidelines for aerobic training	Not enough information to judge	
Notes	Country: Turkey Language: English Study author contacted: no Funding source/declaration of interest: none stated	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Insufficient information on the method used to generate the allocation sequence to permit judgment of risk
Allocation concealment (selection bias)	Unclear risk	No description of the method used for allocation concealment to permit judgment of risk
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk (*Note: previous review rated as low risk of bias)*	Self-report instruments: pain intensity (VAS). Although this study includes a placebo control, it was not specified whether participants were aware of the assigned intervention, however this was deduced from interventions
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	Not applicable: Objective outcomes were not measured
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Insufficient information on blinding of participants and personnel to permit judgment of risk
Incomplete outcome data (attrition bias) All outcomes	Low risk	No missing outcome data at post-test
Selective reporting (reporting bias)	Low risk	Study protocol is not available but it is clear that the published report includes all expected outcomes
Other bias	Unclear risk	Insufficient information to assess whether an important risk of bias exists

Author and year	Wigers 1996
Methods	3 groups: aerobic exercises (AE); stress management; control

Author and year	Wigers 1996	
	Length: 14 weeks; follow-up: 4 years Study design: randomized clinical trial with parallel groups	
Participants	Female:Male: 55:5 Age (years (SD)): 43 (9); 44 (12); 46 (9) Inclusion: diagnosis of fibromyalgia (ACR 1990; Smythe 1979 + Yunus criteria 1981) Exclusion: none Duration of illness (years (SD)): 9 (5); 11 (10); 11 (9)	
Interventions	Aerobic exercise (n = 20): total duration (over 40 sessions) of aerobic exercise, focusing on the whole body and aimed at minimizing eccentric muscle strain, was 30 hours of active treatment. Frequency: 3/wk; Duration: 45' (23' music session comprising warming up and 2 peaks of high-intensity training, each 3-4', 15' aerobic games representing 2 high-intensity periods 5-6' with 4' calming down in between); Intensity: light to moderate (60%-70% HRmax); Mode: movement to music and games Stress management training (n = 20): 2 treatment groups of 10, with each totalling 20 sessions and 30 hours of active treatment; Frequency: 2/wk first 6 weeks, 1/wk remaining 8 weeks; Duration: 90' Control (n = 20): continued treatments being used at baseline For this review: All interventions were considered	
Outcomes	Pain (VAS), fatigue (VAS), CR max (ratio of max voluntary effort) Other: tenderness (tender point count), global rating (self-perceived change numerical rating scale), sleep (VAS), depression (VAS) Measurements taken at 0, 7 weeks (mid-test), 14 weeks (post-test), and 4 years	
Adherence to exercise protocols	Monitoring methods: self-monitored HR guidelines given to participants and attendance; adherence criteria: not stated; adherence: attendance rate 70%, 68%	
Congruence with ACSM guidelines for aerobic training	No, intensity too low, duration too short (only 18-20’ at HR 60%-70%)	
Notes	Country: Norway Language: English Study author contacted: no Funding source/declaration of interest: Work was supported by The Research Council of Norway and The Norwegian Fibromyalgia Association	
Risk of bias		
Bias	Authors' judgement	Support for judgement

Author and year	Wigers 1996	
Random sequence generation (selection bias)	Low risk	"After baseline registration the patients were randomized [by drawing lots] into an AE group, a SMT group or a TAU group" (page 78)
Allocation concealment (selection bias)	Unclear risk	No details on allocation concealment were provided
Blinding of self reported outcome assessment (detection bias) All outcomes	High risk	Self-report instruments: pain intensity (VAS), fatigue (VAS)
Blinding of objective outcome assessment (detection bias) All outcomes	Low risk	CR max (ratio of max voluntary effort). "Neither patients nor investigators had access to previous recordings on any test occasion" (page 78)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	"Patients were instructed not to reveal their group membership before treatment specific questions were asked at the very end of completion test. Neither patients nor investigators had access to previous recordings on any test occasion" (page 78)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data were analyzed by ITT
Selective reporting (reporting bias)	Low risk	Study protocol is not available but it is clear that the published reports include all expected outcomes
Other bias	Low risk	Study appears to be free of other sources of bias

ACR: American College of Rheumatology; AE: aerobic exercise; ANA: antinuclear antibody; CR submax: submaximal cardiorespiratory function; ESR: erythrocyte sedimentation rate; FIQ: Fibromyalgia Impact Questionnaire; FSS: Fatigue Severity Scale; FX: Flexibility; HR: heart rate; HRmax: maximum heart rate; HRR: heart rate reserve; ITT: intention to treat; LPA: lifestyle physical activity; RPE: rating of perceived exertion; RT: resistance exercise training; SBE: short bout exercise; SD: standard deviation; SF-36: Short Form 36; VAS: visual analogue scale; VO2: oxygen consumption

D.2.2 Busch 2013

Author and year	Bircan 2008
Methods	Randomized trial, 2 groups (aerobic exercise group, resistance exercise group), LENGTH: 8 wk.
Participants	FEMALE:MALE = 26:0, AGE (yrs (SD)): 46 (8.5) to 48.3 (5.3).

Author and year	Bircan 2008	
	DURATION OF ILLNESS (yrs (SD)): 3.85 (3.31) to 4.62 (5.22). INCLUSION: Women who met ACR 1990 diagnostic criteria for fibromyalgia (Wolfe 1990). EXCLUSION: Presence of serious cardiovascular, pulmonary, endocrine, neurologic or renal disease, inflammatory rheumatic disease, or participation in a physical therapy or exercise program in the last 6 months.	
Interventions	1) Resistance training group (randomized n = 15, completed and analyzed n = 13): frequency: 3/wk, duration: 40 min (30-min resistance exercise), intensity: unspecified 4-5 reps progressed to 12 reps, method: free weights or body weight resistance exercise in standing, sitting, and lying for upper and lower limb muscles and trunk muscles. 2) Aerobic training group (randomized n = 15, completed and analyzed n = 13): frequency: 3/wk; duration: 20 min progressing to 30 min; intensity: low to moderate; method: treadmill walking.	
Outcomes	Measurements: Pre- and post-intervention (8 wks): sleep disturbance (VAS), fatigue (VAS), tenderness (tender point count), cardio-respiratory function submaximal (6-min walk), anxiety (HAD Anxiety scale), depression (HAD Depression scale), self-reported physical function (SF-36 Physical functioning scale), mental health (SF-36 Mental Health Scale), pain (VAS)	
Congruence with ACSM Guidelines for Resistance Training (yes/no)	Guidelines for healthy adults: No (frequency - yes, type - yes, rep - no, starts too low, sets - unclear, intensity - unclear, progression - yes). Guidelines for older adults: Unclear (frequency - yes, type - yes, rep - yes, intensity - unclear, progression - yes)	
Notes	Adverse effects: page 529: "No patient experienced musculoskeletal injury or exacerbation of fibromyalgia related symptoms during the intervention". Attrition: Resistance training: n = 2 (13.33%), aerobic training: n = 2 (13.33%). Adherence: Not specified. Co-interventions: Both groups "were allowed to continue their medication at entry; however treatment had to remain stable for 1 month prior to entry to the study" (p. 528). Communication with author: Correction to data in table 2 confirming data for pain, sleep, fatigue are in centimeters (email 8 May 2013). Country: Turkey (paper published in English). Funding, conflict of interest: No information was available.	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Participants were randomly assigned to an AE group or a SE group" (AE: aerobic exercise; SE: strengthening exercise) Bircan 2008 (p. 528). In email communication with the author (29 June 2012), the authors clarified as follows, "The patients were assigned to groups by the random allocation rule. As the sample size was planned to

Author and year	Bircan 2008	
		be 30, special cards were prepared for each treatment (15 were labelled as A and 15 as B), the cards were inserted into opaque envelopes, and the envelopes were shuffled. Patients were assigned to groups during the study by drawing lots among these envelopes after the initial evaluations were done."
Allocation concealment (selection bias)	Low risk	Although no information was provided in the publication, in email communication with the author (29 June 2012), we learned that, "The patient's group was determined after all initial evaluations of the patient were done. The investigators did not know what the next treatment allocation would be."
Blinding (performance bias and detection bias) All outcomes	High risk	Although no information was provided in the publication, in email communication with the author (29 June 2012), we learned, "Participants, outcome assessors and people that delivered the intervention were not blind to study groups."
Blinding of outcome assessment (detection bias)	High risk	Only 1 variable was measured by an assessor (6-min walk) - in email communication (29 June 2012), we learned that this outcome was not blinded (see above).
Incomplete outcome data (attrition bias) All outcomes	Low risk	Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups. It is unclear why intention-to-treat analysis was not used.
Selective reporting (reporting bias)	Low risk	All outcomes specified on Bircan 2008, page 528 appear in data tables. According to email communication with the authors: "There were not any outcomes measured but not reported in the paper." (29 June 2012).
Other bias	Low risk	Based on the data provided, there is no indication that there are other important risks of bias.

Author and year	Hakkinen 2001
Methods	Randomized trial, 3 groups (fibromyalgia resistance exercise group, fibromyalgia control group, healthy resistance training group). LENGTH: 4-wk baseline control phase for all groups followed by a 21-wk intervention phase.
Participants	FEMALE:MALE = 33:0, AGE (yrs (SD)): 37 (6) to 39 (6). DURATION OF ILLNESS (yrs (SD)): 12 (4). INCLUSION: Diagnosis: fibromyalgia (ACR criteria; Wolfe 1990), pre-menopausal women. EXCLUSION: Unspecified.
Interventions	1) Fibromyalgia resistance training group (fibromyalgia: n = 11) frequency: 2/wk; duration: duration of each session not provided, intensity: moderate-to-heavy progressive resistance (15-20 reps at 40-60% of 1 RM progressing to 5-10 reps at 70-80% of 1 RM; from wk 7 on: 30% of leg exercise performed rapidly with 40-60% RM); method: 6-8 dynamic resisted exercises using David 200 dynamometer to upper extremity, lower extremity, and trunk muscle groups.

Author and year	Hakkinen 2001	
	2) Fibromyalgia control group (fibromyalgia: n = 10) Controls maintained their normal low-intensity recreational physical activities but did not participate in the strength training. 3) Healthy resistance training control group (healthy: n = 12) A training group made up of sedentary healthy women (without fibromyalgia) was also a part of this study. Data from this group were not analyzed in this review.	
Outcomes	Measurements: 4 wks pre-intervention, immediately pre-intervention, immediately post-intervention (21 wks). Patient-rated global well-being (VAS), pain (VAS), tenderness (tender point count), fatigue (VAS), muscle strength (maximum bilateral (1 RM) concentric leg extension), sleep (VAS), self-reported physical function (Health Assessment Questionnaire), muscle power (squat jump), muscle fiber activation (EMG), muscle size (cross-sectional area), depression (Beck Depression Index).	
Congruence with ACSM Guidelines for Resistance Training (yes/no)	Guidelines for healthy adults: Yes (frequency - yes, type - yes, reps - yes, sets - yes, intensity - yes, progression - yes). Guidelines for older adults: Yes (frequency - yes, type - yes, reps - yes, intensity - yes, progression - yes).	
Notes	Adverse effects: None reported. Attrition: n = 0 (0%), aerobic training: n = 0 (0%) Adherence to exercise protocol: Not specified Data for this study were extracted from 2 reports: Hakkinen 2001 (Primary); Hakkinen 2002 (Secondary). Additional data were obtained from the authors on the following outcome measures: maximum bilateral (1 RM) concentric leg extension, squat jump vertical, and tender points. The authors also clarified the timing of the assessments. The researcher reported that there were no dropouts. The author attributed this to intensive process for habituating participants to the study methods and cultural values unique to Finland where the study took place (personal communication). Also of note, prior to entry into the study, the "subjects in all groups were habitually active (such as walking, swimming, biking, skiing) but they had no background in strength training" (page 1288, Hakkinen 2002 (Secondary)). Co-interventions: No information was provided about co-interventions. Country: Finland. Funding, conflict of interest: As reported by the authors: "This study was supported in part by grants from Finnish Social Insurance Institution and the Yrjö Jahnsson Foundation". No information was available regarding conflict of interest.	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information regarding how participants were randomized.

Author and year	Hakkinen 2001	
Allocation concealment (selection bias)	Unclear risk	No procedure was described.
Blinding (performance bias and detection bias) All outcomes	High risk	Insufficient information, but it is unlikely that participants and care providers were blinded.
Blinding of outcome assessment (detection bias)	Unclear risk	No information on blinding of outcome assessors was provided.
Incomplete outcome data (attrition bias) All outcomes	Low risk	No dropouts reported. Table 1 in Hakkinen 2001 showed the sample size for both groups. We assume that these values are consistent for before and after treatment. Data on tenderness, which was not available in the research report, was provided by the study authors upon request.
Selective reporting (reporting bias)	Low risk	Although the study protocol was unavailable, between the primary, the companion paper and the response from the authors, all the variables measured have been accounted for.
Other bias	Low risk	Based on the data provided, there is no indication that there are other important risks of bias.

Author and year	Jones 2002
Methods	Randomized trial, 2 groups (resistance exercise group, flexibility exercise group). LENGTH: 12 wk.
Participants	FEMALE:MALE = 56:0, AGE (yrs (SD): 46.4 (8.6) to 49.2 (6.3). DURATION OF ILLNESS (yrs (SD)): 6.9 (6.6) to 7.7 (5.5). INCLUSION: Diagnosis: fibromyalgia (ACR criteria; Wolfe 1990), women only, ages 20-60 yrs. EXCLUSION: Current or past history of cardiovascular, pulmonary, neurologic, endocrine, or renal disease that would preclude exercise program; current use of medications that would affect normal physiologic response to exercise; current cigarette smoking, score = 29 on Beck Depression Scale modified for fibromyalgia, current participant in a regular exercise program.
Interventions	1) Resistance exercise group (n = 28): frequency: 2/wk; duration: 60 min; intensity: progressed from 4 to 12 reps; method: supervised dynamic resistance exercise for lower and upper limbs and trunk using hand weight (1-3 lb (0.45-1.36 kg)) and elastic tubing; minimization of eccentric work (a videotape to guide home practice of the strengthening exercise regimen was provided to participants). 2) Flexibility exercise group (n = 28): frequency: 2/wk; duration: 60 min; flexibility for lower limbs and trunk; intensity: n/a, method: supervised static stretches (a videotape to guide home practice of the flexibility exercise regimen was provided to participants).

Author and year	Jones 2002	
Outcomes	Measurement pre- and post-intervention (12 wks). Multidimensional function (FIQ total score), pain (FIQ VAS), tenderness (tender point count), fatigue (FIQ VAS), muscle strength (maximum isokinetic strength of nondominant knee extension), sleep (FIQ VAS), muscle/joint flexibility (hand-to-neck, hand-to-scapula movement), depression (Beck Depression Inventory), anxiety (Beck Anxiety Inventory), coping/self efficacy (Arthritis Self Efficacy Scale).	
Congruence with ACSM Guidelines for Resistance Training (yes/no)	Guidelines for healthy adults: No (F - yes, type - yes, reps - unclear, sets - unclear, I - no, progression - unclear). Guidelines for older adults: No (F- yes, type - yes, repetitions - unclear, I - unclear).	
Notes	Adverse effects: There were no occurrences of adverse events or injury during the intervention and incidence of worsening of pain or tenderness was the same in both groups (n = 3 in each group) (page 1045). Attrition: Authors stated that they had a low attrition rate (9%) (page 1045); however, following analysis of the data and communication with author (email 19 July 2010), the attrition from each group was not specified. The data were: 12/68 (17.64%) either dropped out or did not meet adherence criteria for inclusion. Resistance training n = 6 (17.64%), flexibility training n = 6 (17.64%). Adherence to exercise protocol: "Class attendance records by the exercise instructor indicated that 85% of the participants (n = 58) attended 13 or more classes" (page 1043); however, "the strengthening intervention was not monitored to assure that subjects progressively increased the load throughout the 12 weeks. Instead, participants were encouraged to listen to their bodies and increase the intensity as they thought they could tolerate it." (pages 1045, 1046). Co-interventions: No information was provided about co-interventions. Country: US. Communication with author: Additional data were obtained from the authors to clarify the content and delivery of the intervention (eg, videotapes, education, the exercise level at completion), the number randomized, and specifics related to dropouts. Funding, conflict of interest: As reported by the authors: "Supported by an Individual National Research Service Award (#1F31NR07337-01A1) from the National Institutes of Health, a doctoral dissertation grant (#2324938) from the Arthritis Foundation, and funds from the Oregon Fibromyalgia Foundation". No information was available regarding conflict of interest.	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization was accomplished with a coin flip" (page 1042).
Allocation concealment (selection bias)	Unclear risk	Insufficient information in the research report.

Author and year	Jones 2002	
Blinding (performance bias and detection bias) All outcomes	High risk	Insufficient information, but it is unlikely that participants and care providers were blinded.
Blinding of outcome assessment (detection bias)	Low risk	"Data were collected by an exercise science technician (strength and body fat) or the principal investigator (all other measures). Both were blinded to group assignment" (Jones 2002, page 1042).
Incomplete outcome data (attrition bias) All outcomes	Low risk	Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias). Authors stated that the participants who dropped out lived far from the fitness center (page 1045).
Selective reporting (reporting bias)	Low risk	The study protocol was not available but it was clear that the published reports included all expected outcomes, including those that were prespecified.
Other bias	Low risk	There may be a risk related to poor adherence to the exercise regimen. "85% of the participants attended only slightly more than 50% of the 24 supervised sessions" (Jones 2002, page 1043). The low attendance may have contributed to low power (ie, type 2 error).

Author and year	Kayo 2011
Methods	Randomized trial, 3 groups (walking group, strengthening exercise group, control group). LENGTH: 16 wks with follow-up for an additional 12 wks.
Participants	FEMALE:MALE = 90:0, AGE (yrs (SD)): 46.1 (6.4) to 47.7 (5.3). DURATION OF ILLNESS (yrs (SD)): 4 (3.1) to 5.4 (3.5). INCLUSION: women ages 30-55 yrs and agreed to participate in an exercise program 3 times/wk for 16 wks and to discontinue medications for fibromyalgia 4 wks before the start of the study and who had at least 4 yrs of schooling. EXCLUSION: women with any contraindications to exercise on the basis for clinical rheumatologic examination, and those involved in cases of medical litigation.
Interventions	1) Progressive aerobic exercise (n = 30): frequency: 3 times/wk x 16 wks; duration: ˜ 60 min (warm-up (5-10 min) conditioning stimulus, cool down (5 min); intensity: moderate to high intensity (40-50% to 60-70% heart rate reserve by wk 16); method: supervised indoor or outdoor walking monitored using heart rate monitor. 2) Resistance exercise training (n = 30): frequency: 3 times/wk x 16 wk; duration: ˜ 60 min; intensity: high intensity (4 on 10-point Borg scale)b, exercise load and intensity were increased every 2 wks (reps - wks 1 + 2: 3 sets of 10 reps with rest intervals of 1 min between sets, wks 3-16; load - wks 1-4, no load, wks 5-16 load was included), "The training load was individually and systematically adjusted every time the participant performed more than 15 repetitions with successfully"b; M: supervised exercise protocol

Author and year	Kayo 2011	
	consisting of 11 free active exercises for upper and lower limbs and trunk muscles, using free weights and body weight performed in the standing, sitting, and lying positions. 3) Control group (n = 30): control conditions not specified, except authors stated participants in all 3 groups were asked to discontinue tricyclic antidepressants but were allowed to use acetaminophen (paracetamol) for pain.	
Outcomes	Measurement pre-intervention, mid-intervention (8 wks), immediately post-intervention (16 wks), and follow-up (12 wks post-intervention). As reported in paper: multidimensional function (FIQ total), pain (VAS). As provided by author on request: fatigue (SF-36 - Vitality scale), tenderness (tender point pain), self-reported physical function (SF-36 Physical Function scale), mental health (SF36 Mental Health).	
Congruence with ACSM Guidelines for Resistance Training (yes/no)	Guidelines for healthy adults: No (frequency - yes, type - yes, reps - no, sets - yes, intensity - yes, according to description provided by authors regarding the scale, progression - yes). Guidelines for older adults: Yes (frequency - yes, type - yes, reps - yes, intensity - yes, progression - yes).	
Notes	Adverse effects: "No complications or adverse effects were observed during the study period among patients who completed the treatment protocols." Attrition: Aerobics training n = 1 (3.3%), resistance training n = 5 (16.6%), control n = 5 (16.6%). Adherence to exercise protocol: "We adopted Borg Scale (0-10) and the recommended intensity was 4 (somewhat severe) and all participants complied." From email communication (19 July 2012). 80% attendance rate - excluding those who dropped out for reasons of work or family illness, with only 1 participant assigned to the resistance training group that did not meet the attendance requirements of the study. Co-interventions: Exercise was administered in this study as a single modality; the timing of restarting medication was monitored. Country: Brazil Funding, conflict of interest: No information on funding of the study was found, but the authors stated there was no conflict of interest.	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The allocation sequence was based on a random number list (GraphPad Statmate version 1.0), which was organized by an investigator (MSP)" (online page 2).
Allocation concealment (selection bias)	Low risk	Opaque sealed envelopes were used.

Author and year	Kayo 2011	
Blinding (performance bias and detection bias) All outcomes	Low risk	No details provided in the report. "There was no contact among the groups"b.
Blinding of outcome assessment (detection bias)	High risk (*Note: previous review rated as low risk of bias)*	The study authors stated: "all patients were clinically examined by the same rheumatologist (CSM), who was blinded to group assignment throughout the study" (online page 2). However, participants not blinded (deduced from interventions)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Intention-to-treat analysis was used.
Selective reporting (reporting bias)	High risk	Outcome data for important variables (eg, tender points, SF-36 Physical Functioning, SF-36 Vitality, SF-36 Mental Health) were not provided in the published report, but the study authors provided these on requestb. An important shortcoming was that there were no performance tests for physical function applied in this study.
Other bias	Low risk	There did not appear to be any other serious sources of bias. Although the researchers found differences between groups in duration of disease at baseline (P value = 0.04, longer duration in control group than the intervention groups), no between-group differences were found in baseline levels of age, pain, tenderness, multidimensional function, SF-36 subscales, so we did not consider this a serious problem.

Author and year	Valkeinen 2004
Methods	Randomized trial, 3 groups (fibromyalgia resistance exercise group, fibromyalgia control group, healthy resistance exercise control group). LENGTH: 21 wk.
Participants	FEMALE:MALE = 36:0, AGE (yrs (SD)): 59.1 (3.5) to 60.2 (2.5). DURATION OF ILLNESS (yrs (SD)): 8.5 (4.3) to 6.6 (4.1). INCLUSION: Diagnosis: fibromyalgia (ACR criteria; Wolfe 1990), age = 55 yrs, women. EXCLUSION: No other diseases, no injuries, no experience of regular strength training exercises, willingness to participate in study protocol.
Interventions	1) Fibromyalgia resistance exercise group (fibromyalgia: n = 13): frequency: 2/wk; duration: 60-90 min, 80% strength 20% power, I: light- to high-intensity progressive resistance from 3 sets of 15-20 reps at 40-60% 1 RM to 3-5 sets of 5-10 reps at 70-80% 1 RM, for power (legs only) 2 sets of 8-12 reps at 40-50% 1 RM; method: resisted dynamic exercise to knee extensors x 2 plus 5-6 exercises for other main muscle groups of body (exercise equipment not specified).

Author and year	Valkeinen 2004	
	2) Fibromyalgia control group (fibromyalgia: n = 13): Control conditions were treatment as usual and physical activity as usual. 3) Healthy resistance exercise control group (healthy: n = 10): A group made up of sedentary women without fibromyalgia (n = 12) who carried out the exercise protocol was also a part of this study. Data from this group were not analyzed in this review.	
Outcomes	Measurements 4 wks pre-intervention, immediately pre-intervention, immediately post-intervention (21 wks). Tenderness (tender point count), muscle strength (Max concentric leg extension), self-reported function (Health Assessment Questionnaire), muscle fiber activation (EMG), muscle size (cross-sectional area). The study authors stated they measured 5 other variables (pain, fatigue, patient-rated global, depression, and sleep) but the data were not available in the report and they did not respond to our emails.	
Congruence with ACSM Guidelines for Resistance Training (yes/no)	Guidelines for healthy adults: Yes (frequency - yes, type - yes, reps - yes, sets - yes, intensity - yes). Guidelines for older adults: Yes (frequency - yes, type - yes, reps - yes, intensity - yes).	
Notes	Adverse effects: "After the initial phase of training, the patients did not complain of any unusual exercise-induced pain or muscle soreness" (Valkeinen 2004 (Primary) page 227). Attrition: Fibromyalgia resistance training n = 0 (0%), fibromyalgia control n = 0 (0%), healthy resistance training n = 0 (0%) Adherence to exercise protocol: The researchers did not specify if or how adherence to the exercise protocol was monitored; however, muscular function was measured at 7, 14, and 21 wks. They did state all fibromyalgia subjects "completed training". Co-interventions: "All subjects were allowed to continue their normal daily activities, to use their normal medication ... and to visit medical professionals if needed" (page 226). Country: Finland. Data for this study was extracted from 2 reports: Valkeinen 2004 (Primary), Valkeinen 2005 (Secondary). Funding, conflict of interest: As reported by the authors: "This study was supported in part by grants from the Central Hospital of Central Finland; Kuopio University Hospital, Peurunka-Medical Rehabilitation Foundation and The Ministry of Education, Finland". No information was available regarding conflict of interest.	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Described on page 225 Valkeinen 2004: "After inclusion, the fibromyalgia patients were randomly allocated by draw ..."
Allocation concealment (selection bias)	Unclear risk	No mention of allocation concealment.

Author and year	Valkeinen 2004	
Blinding (performance bias and detection bias) All outcomes	High risk	Insufficient information, but it is unlikely that participants and care providers were blinded.
Blinding of outcome assessment (detection bias)	High risk (*Note: previous review rated as low risk of bias)*	No information available but deduced from intervention
Incomplete outcome data (attrition bias) All outcomes	Low risk	Missing outcome data balanced in numbers across interventions groups, with similar reasons for missing data across groups.
Selective reporting (reporting bias)	High risk	Outcome of statistical analyses are reported for pain, fatigue, sleep, depression, perceived health (all non-significant) but point estimates for these outcome measures were not reported.
Other bias	Low risk	Based on the data provided, there is no indication that there are other important risks of bias.

a intention-to-treat analysis.

b based on email communication with the study author.

ACR: American College of Rheumatology, EMG: electromyography; FIQ: Fibromyalgia Impact Questionnaire; HAD: Hospital Anxiety and Depression; min: minute; rep: repetition; RM: repetition maximum; SD: standard deviation; SF: Short Form; VAS: visual analog scale; wk: week; yr: year.

D.2.3 Theodom 2015

Author and year	Bojner-Horwitz 2003
Methods	Randomised controlled trial
Participants	Female participants met the ACR criteria for fibromyalgia Total participants = 36 randomised (number withdrawn not stated) Mean age 57 years (SD 7.2 years)
Interventions	1) Dance and movement therapy consisted of four main themes including; awareness of the body; movement expressions; movement, feeling, image; and differentiation of feelings and integration 1 hour session, held weekly for 6 months 2) Control group participants received the intervention on completion of the study

Author and year	Bojner-Horwitz 2003	
Outcomes	Discontinuation Follow-up time points: baseline and month 14 (not able to be included in the review)	
Notes	The study was funded by the Order of Carpenters in Sweden	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Stated that patients were randomly allocated but details not provided
Allocation concealment (selection bias)	Unclear risk	Details of randomisation procedure not provided
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as unclear risk of bias)*	Details not provided but deduced from interventions
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Details not provided
Selective reporting (reporting bias)	High risk	Outcome data not reported for pain VAS and the Montgomery Asberg Depression Rating Scale

Author and year	Calandre 2009
Methods	Prospective randomised controlled trial
Participants	Patients who had a diagnosis of fibromyalgia according to the ACR criteria were recruited through a University Hospital Pain Unit Total participants = 81 randomised (57 completed) N = 73 female, N = 8 male Age range 32 to 69 years Exclusions: patients who had never attended a swimming pool as well as those suffering any co-concomitant disease susceptible to worsen with warm water exercise were excluded
Interventions	1) Tai chi was performed in a pool with water heated at 36 ° and was preceded by a shower with warm water to condition patients' bodies. A trained physiotherapist adjusted the movement intensity to meet individual needs and participants were taught the 16 movements which constitute tai chi therapy

Author and year	Calandre 2009	
	2) Stretching was facilitated using supportive aids such as long wooden sticks, flexible strings and tubes to stretch muscles in the cervical, upper and lower extremities and trunk Both groups received 18 sessions of 60 minutes, delivered 3 times per week for 6 weeks	
Outcomes	Measures relevant to this review: Fibromyalgia Impact Questionnaire, Pittsburghh Sleep Quality Index, Beck Depression Inventory, State and Trait Anxiety Inventory, SF12 Health Survey, tender point count Assessment time points: baseline, post-intervention, one and three month follow-up	
Notes	There was no reference to sources of funding or conflicts of interest declared in the article	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	Low risk	Computer generated table of random numbers
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk	Assessors were not blind to treatment allocation
Incomplete outcome data (attrition bias) All outcomes	High risk	A 29% total attrition rate; 3 adverse events were reported in the intervention group participants but not for controls, unclear if pain exacerbations directly related to intervention
Selective reporting (reporting bias)	Low risk	All of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way

Author and year	Carson 2010
Methods	Pilot randomised controlled trial
Participants	Women who had been diagnosed with fibromyalgia according to the ACR criteria for at least one year and were on a stable regimen of treatment Total participants = 53 randomised (48 completed) Mean age = 53.7 (SD 11.5) years

Author and year	Carson 2010	
	Exclusions: residing > 70 miles from the research site, unavailable to attend the intervention at one of the schedule times, currently engaged in yoga practice, actively contemplating suicide, currently undergoing disability application, or litigation, schedule for elective surgery during the study period, physically disabled in a manner that precluded meaningful participation in the intervention, unwilling to forgo changing any voluntary treatments for the length of this study and those unable to speak English	
Interventions	1) Yoga consisted of 2 hour sessions, held weekly for 8 weeks in a group based format led by a certified, experienced yoga teacher. The intervention included meditation, breathing exercises, study of the application of yoga principles to optimal coping and gentle stretching poses and group discussions 2) Usual care, wait list	
Outcomes	Measures relevant to this review: Fibromyalgia Impact Questionnaire, tender point score Assessment time points: baseline and post-intervention	
Notes	The study was supported by a grant from the Oregan Health and Science University Medical Research Foundation and resources supplied by the Fibromyalgia Information Foundation. The authors report no conflicts of interest	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	Low risk	Randomised assignments were generated by an individual not involved in the study using a random numbers table
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as low risk of bias)*	The outcome assessors were blinded to treatment allocation but participants aware of their interventions
Incomplete outcome data (attrition bias) All outcomes	Low risk	A 9% total attrition rate. There was no imbalance evident between groups
Selective reporting (reporting bias)	Low risk	All of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way

Author and year	Carson 2012
Methods	Randomised controlled trial

Author and year	Carson 2012	
Participants	Female participants who had been diagnosed according to the ACR criteria for fibromyalgia syndrome for at least one year. To be eligible participants needed to be on a stable regimen of pharmacological or non-pharmacological treatment for more than or equal to 3 months before study enrolment Total participants = 53 randomised (39 completed) Exclusions: residing > 70 miles from research site or unable to attend the intervention, engaged in intensive yoga practice, actively contemplating suicide, Undergoing disability assessment, or litigation, scheduled for elective surgery, physically disabled as to preclude meaningful participation in the intervention, unwilling to change treatment for duration of the study and non-English speaking	
Interventions	1) Yoga delivered within group sessions by a certified yoga instructor 120 minute sessions, delivered weekly over 8 weeks 2) Wait-list control group	
Outcomes	Measures relevant to this review: Fibromyalgia Impact Questionnaire Revised, tender point score Assessment time points: baseline and post-intervention	
Notes	The study was supported by a grant from the Oregan Health and Science University Medical Research Foundation and resources supplied by Fibromyalgia Information Foundation	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	Low risk	"Randomisation assignments were generated by an individual not involved in the study using a random number table. Assignments were concealed in envelopes until completion of the baseline assessment"
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as low risk of bias)*	"Research Assistants who collected assessment data were kept blind with regard to condition" but participants aware of their interventions
Incomplete outcome data (attrition bias) All outcomes	High risk (*Note: previous review rated as low risk of bias)*	A 24% total attrition rate, no imbalance evident between groups post-intervention
Selective reporting (reporting bias)	Low risk	All of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way

Author and year	Holmer 2004	
Methods	Randomised controlled trial	
Participants	Participants had been diagnosed with fibromyalgia based on the ACR criteria Total participants = 28 randomised (22 completed) Age range 18 to 65 years N = 26 female, N = 3 male Exclusions: none specified	
Interventions	1) Yoga delivered by a certified yoga instructor 2) Waiting list control	
Outcomes	Measures relevant to this review: Multidimensional Assessment of Fatigue Scale, Fibromyalgia Impact Assessment - pain scale, Arthritis Impact Measurement Scale - II, anxiety subscale, Center for Epidemiology Scale - Depression, Pittsburghh Sleep Quality Index, visual analog scale for pain Assessment time points: baseline and post-intervention	
Notes	There was no reference to sources of funding or conflicts of interest declared in the article	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	High risk	Alternate group assignment method was employed (informed by e-mail)
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk	Outcome assessors were not blind to treatment allocation (confirmed by e-mail)
Incomplete outcome data (attrition bias) All outcomes	High risk (*Note: previous review rated as low risk of bias)*	A 21% total attrition rate
Selective reporting (reporting bias)	Low risk	All of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way

Author and year	**Jones 2012**	
Methods	Randomised controlled trial	
Participants	Participants aged 40 years diagnosed with fibromyalgia syndrome or over were recruited with approval of a healthcare practitioner Total participants = 101 randomised (98 completed) Exclusions: practice of tai chi within past 6 months, exercised more than 30 minutes three times weekly for past 3 months, unable to ambulate without assistive devices, pain severity or interference scores less than 5, planned elective surgery in study period, actively involved in healthcare litigation, unwilling to keep all treatments stable throughout the study duration	
Interventions	1) Tai chi delivered in a group based format 90 minute sessions delivered twice weekly for 12 weeks 2) Education sessions delivered in a group based format on fibromyalgia , healthy eating, education based CBT strategies, sleep hygiene and lifestyle management 90 minute sessions delivered twice weekly for 12 weeks	
Outcomes	Measures relevant to this review: Fibromyalgia Impact Questionnaire, Brief Pain Inventory, Numerical Rating Scale for pain, Arthritis Self-Efficacy Scale, Pittsburghh Sleep Quality Index Assessment time points: baseline and post-intervention	
Notes	The study was funded by the National Institutes of Health/NIAMS grant number 5R21 AR053506, NIH/NCCAM1K23 AT006392-01. The authors report no conflicts of interest	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	Low risk	"computer generated table of random numbers with block stratification using age in 5-year intervals"
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as low risk of bias)*	No details provided but deduced from interventions
Incomplete outcome data (attrition bias) All outcomes	Low risk	A 3% attrition rate although all withdrawals occurred in the control group
Selective reporting (reporting bias)	High risk	Means and standard deviations not reported

Author and year	**Liu 2012**	
Methods	Randomised controlled trial	
Participants	Participants aged between 18 and 70 years with a diagnosis of FMS according to the ACR criteria were recruited from a neurology clinic and support group Total participants = 14 randomised (12 completed) Exclusions: severe psychiatric illness, significant suicide risk, alcohol abuse, use of benzodiazepines, history of behaviour that would prohibit compliance for the duration of the study, co-morbid medical conditions, severe sleep apnoea, pregnancy or breastfeeding	
Interventions	1) Qi-gong delivered in a group based format with home practice in between sessions 15 to 20 minute sessions, held weekly for 6 weeks 2) Sham qi-gong delivered in a group based format with no meditation or healing sounds 15 to 20 minute sessions, held weekly for 6 weeks	
Outcomes	Measures relevant to the review: Discontinuation Outcomes reported but not in useable format: Fibromyalgia Impact Questionnaire, McGill Pain Questionnaire, Multidimensional Fatigue Inventory, Pittsburghh Sleep Quality Index Assessment time points: baseline and post-intervention	
Notes	The authors report no conflicts of interest. No sources of funding were declared	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	Unclear risk	No details provided
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as low risk of bias)*	No details provided but deduced from interventions
Incomplete outcome data (attrition bias) All outcomes	High risk (*Note: previous review rated as low risk of bias)*	A 14% attrition, both withdrawals were in the treatment group
Selective reporting (reporting bias)	High risk	Means and standard deviations for outcome measures not reported

Author and year	**Lynch 2012**	
Methods	Randomised controlled trial	
Participants	Participants were recruited through advertisements in local newspapers. To be eligible participants were required to have a diagnosis of FMS according to the ACR criteria, have had a stable medication regime in the past 2 weeks, have an average weekly pain score more than 4 on an 11 point rating scale Total participants = 100 randomised (89 completed) Exclusions: significant medical disorder	
Interventions	1) Qi-gong delivered by a psychologist in a group based format in the community 3.5 day workshops held weekly with additional refresher sessions 2) Wait-list control	
Outcomes	Measures relevant to the review: Fibromyalgia Impact Questionnaire, 11 point numerical rating scale for pain, SF36 Health Survey, Pittsburghh Sleep Quality Index Assessment time points: baseline, post-intervention and 6 month follow-up	
Notes	The study was funded by a Pfizer Neuropathic Pain Research Award. Authors CH and DM provide qi-gong interventions in the community. The other co-authors report no conflicts of interest	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Study was described as a randomised controlled trial but no details of the sequence generation process provided
Allocation concealment (selection bias)	Low risk	"participants were assigned using computer generated numbers to an immediate Qigong training group or to a control group. Assignments were sealed in opaque white envelopes"
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as low risk of bias)*	No details specified but deduced from interventions
Incomplete outcome data (attrition bias) All outcomes	High risk (*Note: previous review rated as low risk of bias)*	An 11% attrition although more withdrawals occurred in the treatment group in comparison to control
Selective reporting (reporting bias)	High risk	Data were presented as change scores and were not able to be included in the analyses

Author and year	**Mannerkorpi 2004**	
Methods	A controlled randomised pilot study	
Participants	Women fulfilling the ACR criteria for fibromyalgia were recruited Total participants = 36 randomised (22 completed) Age range = 18 to 65 years Exclusions: unable to speak Swedish	
Interventions	1) Qi-gong + relaxation, 14 group sessions of 1.5 hours, were held weekly, delivered by a physiotherapist. The treatment included various breathing, relaxation and concentration techniques conducted in a supine or standing position including qi-gong movements. The movements were individually modified to match the functional limitations of the patients and there was an opportunity for discussion about the movements with the therapist. Participants were encouraged to practice the movements in between sessions 2) Usual care	
Outcomes	Measures relevant to this review: Fibromyalgia Impact Questionnaire Assessment time points: baseline and post-intervention	
Notes	The study was supported by grants from the Swedish Rheumatism Association and the Swedish Research Council	
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A random component is included in the sequence generation process used
Allocation concealment (selection bias)	Low risk	Independent person allocated patients to groups using sealed envelopes
Blinding of outcome assessment (detection bias) Questionnaire assessors blind?	High risk (*Note: previous review rated as low risk of bias)*	Outcome assessor was blinded to patients group membership but participants aware of their interventions
Incomplete outcome data (attrition bias) All outcomes	High risk (*Note: previous review rated as low risk of bias)*	A 39% total attrition rate

Author and year	Mannerkorpi 2004	
Selective reporting (reporting bias)	Low risk	All of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way

Appendix E: Forest plots

E.1 Aerobic exercise versus usual care

Figure 2: Pain at ≤3 months (VAS, final values, 0-100, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Sencan 2004	40.5	9.1	20	62	18.1	20	100.0%	-21.50 [-30.38, -12.62]
Total (95% CI)			**20**			**20**	**100.0%**	**-21.50 [-30.38, -12.62]**

Heterogeneity: Not applicable
Test for overall effect: Z = 4.75 (P < 0.00001)

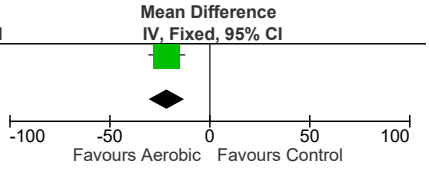


Figure 3: Pain at >3 months (VAS, FIQ pain subscale, final values, 0-100, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Andrade 2019	54	24	27	64	21	27	10.0%	-10.00 [-22.03, 2.03]
Kayo 2011	51	29.5	30	64.7	24.6	30	7.7%	-13.70 [-27.44, 0.04]
Mengshoel 1992	60	21.6	11	66	21.6	14	5.0%	-6.00 [-23.06, 11.06]
Sanudo 2010	67	15.6	22	80.5	18.1	21	14.1%	-13.50 [-23.62, -3.38]
Sanudo 2015	67	22	16	70	17	12	6.9%	-3.00 [-17.45, 11.45]
Schachter 2003	55.6	23.8	107	56	21.6	36	20.6%	-0.40 [-8.77, 7.97]
Sencan 2004	47.5	12.1	20	58.4	28.1	20	8.0%	-10.90 [-24.31, 2.51]
Van eijk-hustings 2013	53	21.25	47	57	20.79	48	20.2%	-4.00 [-12.46, 4.46]
Wigars 1996	62	21	20	72	24	20	7.4%	-10.00 [-23.98, 3.98]
Total (95% CI)			**300**			**228**	**100.0%**	**-6.97 [-10.77, -3.17]**

Heterogeneity: Chi² = 6.42, df = 8 (P = 0.60); I² = 0%
Test for overall effect: Z = 3.59 (P = 0.0003)

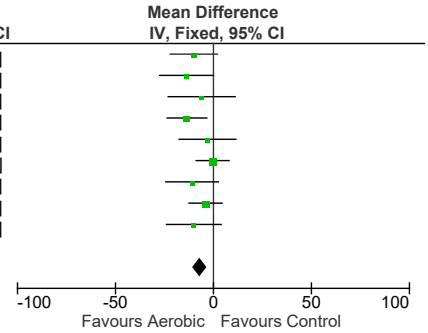
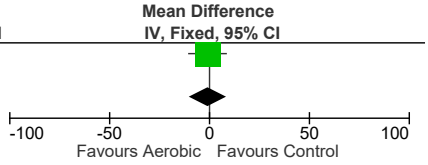


Figure 4: Pain at >3 months (FIQ pain subscale, final values, 0-100, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Van eijk-hustings 2013	52	25.37	47	53	20.79	48	100.0%	-1.00 [-10.34, 8.34]
Total (95% CI)			**47**			**48**	**100.0%**	**-1.00 [-10.34, 8.34]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.21 (P = 0.83)



Note: 18 month timepoint not meta-analysed with 12-24 week data.

Figure 5: Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
Gowans 2001	48.6	16.2	27	54.9	13	23	19.6%	-6.30 [-14.40, 1.80]
Kayo 2011	36.7	22.5	30	55.6	14.1	30	16.8%	-18.90 [-28.40, -9.40]
King 2002	49.6	14.7	42	54.3	12.6	34	24.3%	-4.70 [-10.84, 1.44]
Sanudo 2010	52.1	18.1	22	63.7	17.1	21	15.0%	-11.60 [-22.12, -1.08]
Schachter 2003	51.5	17.7	107	54	15.5	36	24.4%	-2.50 [-8.57, 3.57]
Total (95% CI)			**228**			**144**	**100.0%**	**-7.89 [-13.23, -2.55]**

Heterogeneity: Tau² = 20.77; Chi² = 9.42, df = 4 (P = 0.05); I² = 58%
Test for overall effect: Z = 2.90 (P = 0.004)

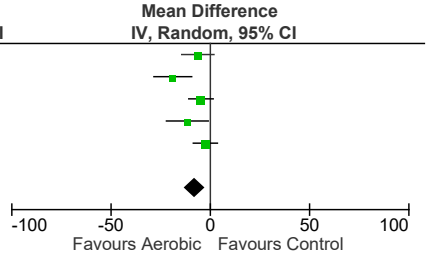


Figure 6: Quality of life at >3 months (SF-36 functional capacity subscale, 0-100, final values, high is good outcome)

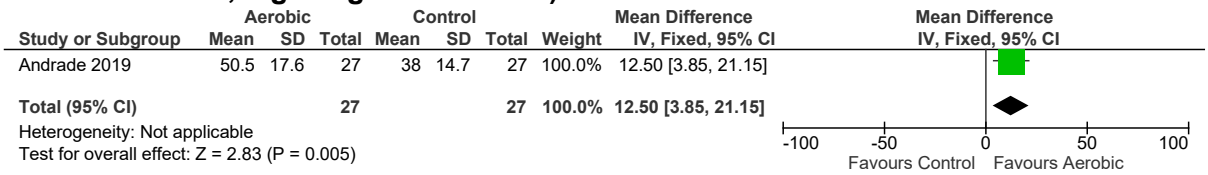


Figure 7: Quality of life at >3 months (SF-36 physical appearance subscale, 0-100, final values, high is good outcome)

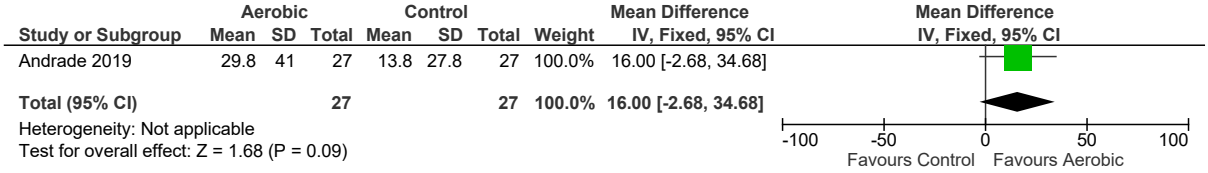


Figure 8: Quality of life at >3 months (SF-36 pain subscale, 0-100, final values, high is good outcome)

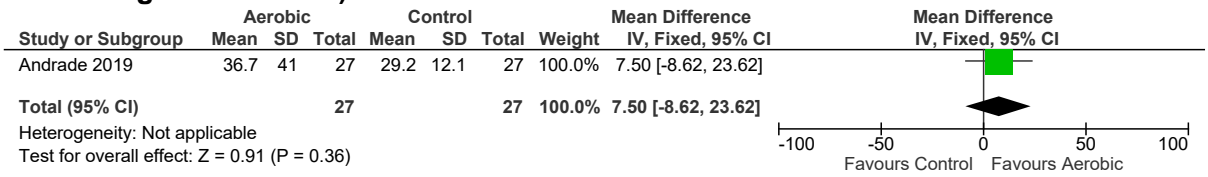


Figure 9: Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)

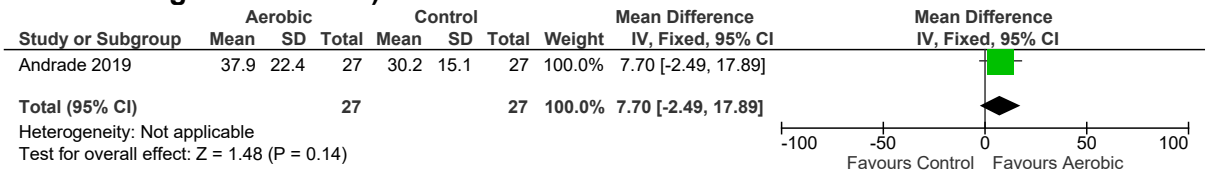


Figure 10: Quality of life at >3 months (SF-36 social aspects subscale, 0-100, final values, high is good outcome)

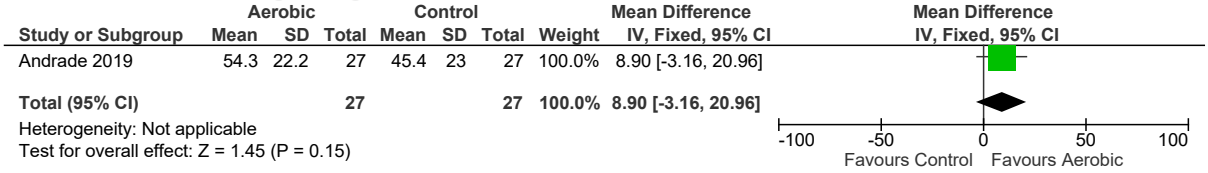


Figure 11: Quality of life at >3 months (SF-36 emotional aspects subscale, 0-100, final values, high is good outcome)

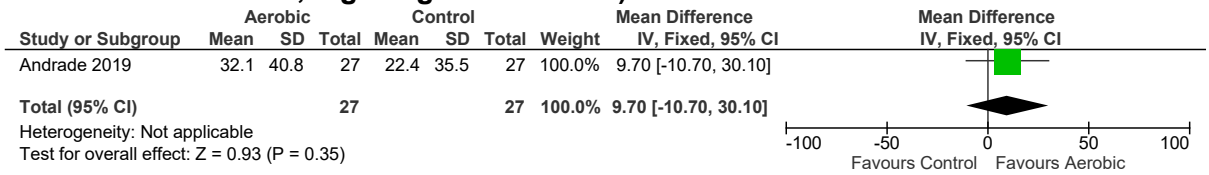


Figure 12: Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)

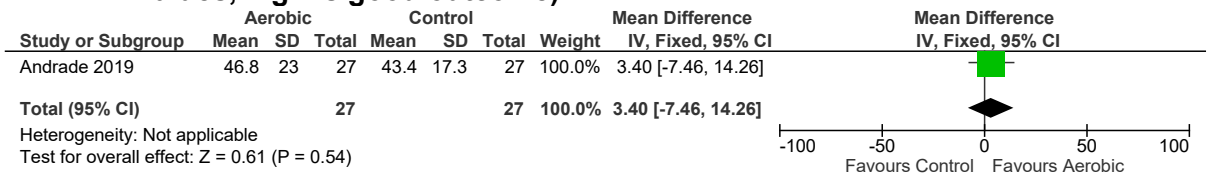


Figure 13: Quality of life at ≤3 months (EQ-5D, -0.594-1, final values, high is good outcome)

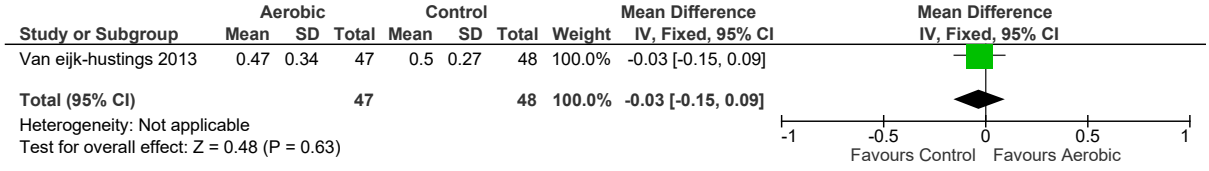


Figure 14: Quality of life at >3 months (EQ-5D, -0.594-1, final values, high is good outcome)

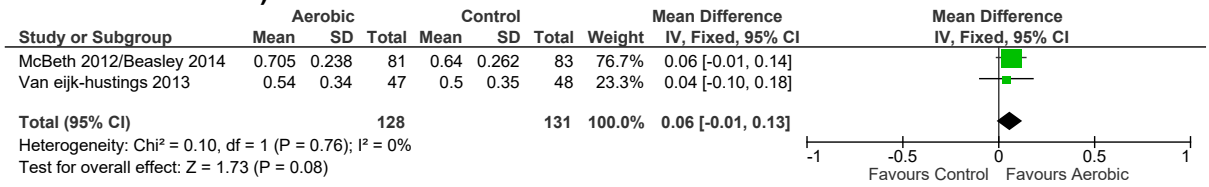


Figure 15: Quality of life at ≤3 months (EQ-5D-VAS, 0-100, final values, high is good outcome)

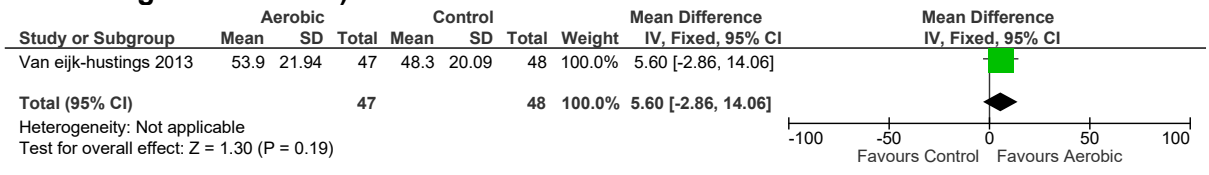


Figure 16: Quality of life at >3 months (EQ-5D-VAS, 0-100, final values, high is good outcome)

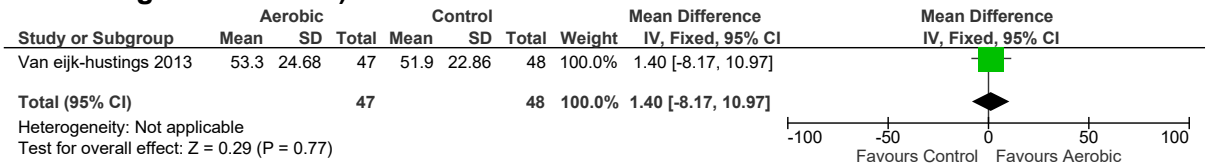


Figure 17: Physical function at ≤3 months (Timed up and go, seconds, high is good outcome)

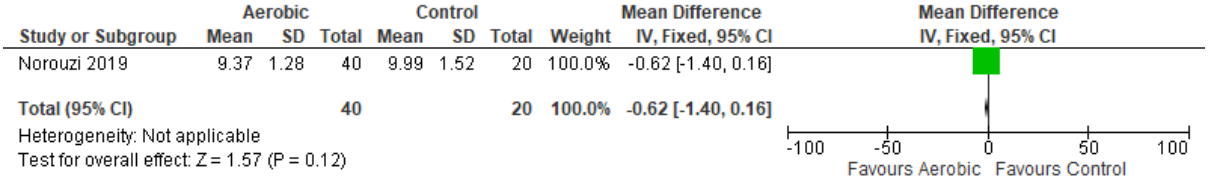


Figure 18: Physical function at ≤3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)

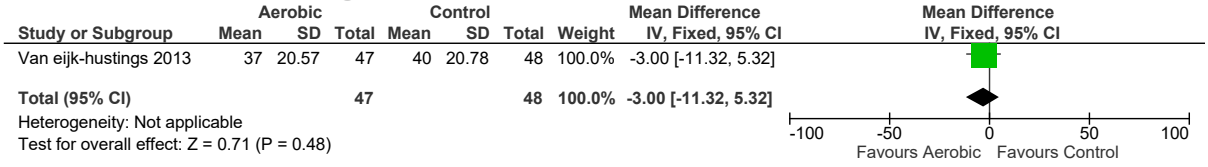


Figure 19: Physical function at >3 months (6 minute walking test, final values, metres)

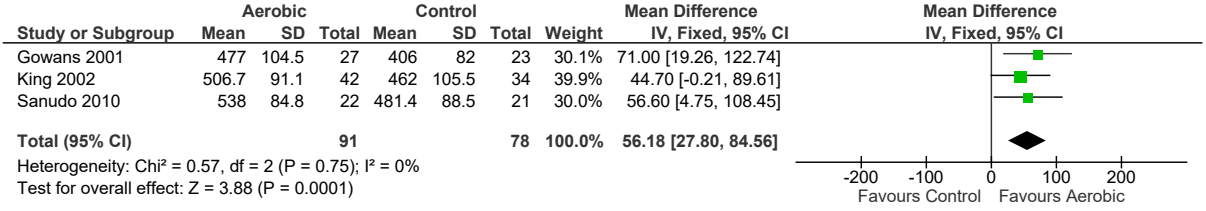


Figure 20: Physical function at >3 months (FIQ and SF-36 physical function subscales, 0-100, final values, high is poor outcome)

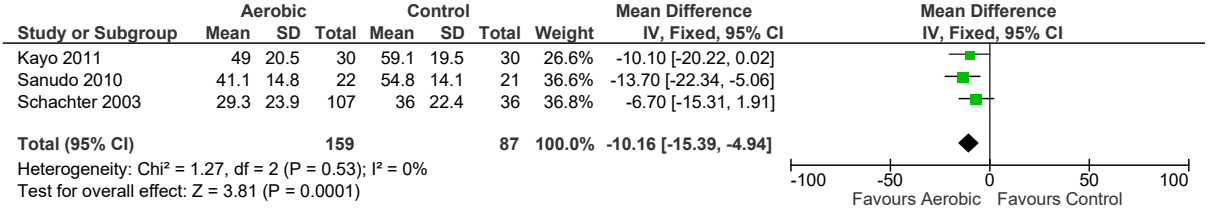
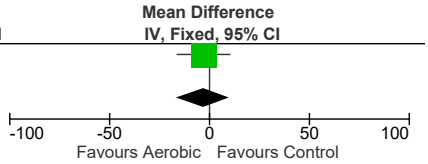


Figure 21: Physical function at >3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Van eijk-hustings 2013	36	41.1	47	39	20.78	48	100.0%	-3.00 [-16.14, 10.14]
Total (95% CI)			**47**			**48**	**100.0%**	**-3.00 [-16.14, 10.14]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.45 (P = 0.65)



Note: 18 month timepoint not meta-analysed with 16-24 week data.

Figure 22: Psychological distress at >3 months (Final values and change scores, BDI, 0-61, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Andrade 2019	15.8	9	27	19.6	8.6	27	35.5%	-3.80 [-8.50, 0.90]
Gowans 2001	13.6	7.9	15	19.4	10.8	16	17.8%	-5.80 [-12.43, 0.83]
Sanudo 2010	-8.5	8	18	-6.4	4	20	46.7%	-2.10 [-6.19, 1.99]
Total (95% CI)			**60**			**63**	**100.0%**	**-3.36 [-6.16, -0.56]**

Heterogeneity: Chi² = 0.92, df = 2 (P = 0.63); I² = 0%
Test for overall effect: Z = 2.36 (P = 0.02)

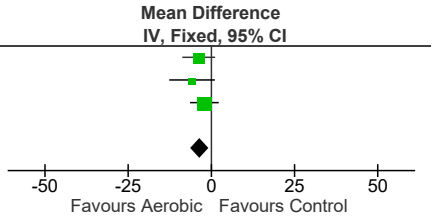


Figure 23: Psychological distress at >3 months (Final values, VAS and FIQ depression scale, 0-10, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Sanudo 2015	5.6	3.4	16	6.7	2.2	12	10.3%	-1.10 [-3.18, 0.98]
Schachter 2003	4.3	2.87	107	4.9	2.62	36	43.2%	-0.60 [-1.61, 0.41]
Van eijk-hustings 2013	4.6	2.74	47	4.5	2.77	48	36.2%	0.10 [-1.01, 1.21]
Wigars 1996	3.1	3.2	20	3.6	3.5	20	10.3%	-0.50 [-2.58, 1.58]
Total (95% CI)			**190**			**116**	**100.0%**	**-0.39 [-1.05, 0.28]**

Heterogeneity: Chi² = 1.37, df = 3 (P = 0.71); I² = 0%
Test for overall effect: Z = 1.14 (P = 0.25)

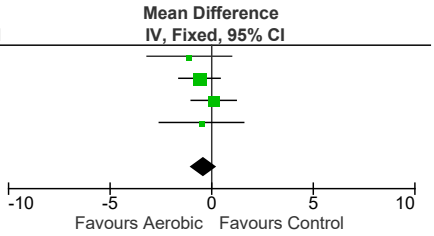


Figure 24: Psychological distress at >3 months (Final values, VAS and FIQ anxiety scales, BAI, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Andrade 2019	15.3	9.1	27	19.5	9	27	18.7%	-0.46 [-1.00, 0.08]
Sanudo 2015	5.7	3.3	16	7.5	2.5	12	9.3%	-0.58 [-1.35, 0.18]
Schachter 2003	4.76	2.62	107	5.2	2.6	36	38.3%	-0.17 [-0.55, 0.21]
Van eijk-hustings 2013	4.6	2.74	47	5.2	2.77	48	33.7%	-0.22 [-0.62, 0.19]
Total (95% CI)			**197**			**123**	**100.0%**	**-0.28 [-0.51, -0.04]**

Heterogeneity: Chi² = 1.46, df = 3 (P = 0.69); I² = 0%
Test for overall effect: Z = 2.32 (P = 0.02)

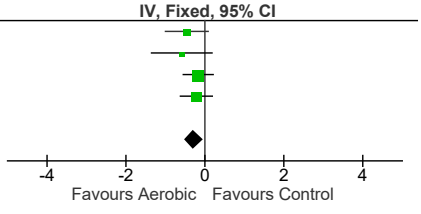


Figure 25: Psychological distress at >3 months (Change scores, STAI anxiety total scores, 0-21, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Fixed, 95% CI
Gowans 2002	-4.9	25	27	4.8	25	23	100.0%	-9.70 [-23.60, 4.20]
Total (95% CI)			**27**			**23**	**100.0%**	**-9.70 [-23.60, 4.20]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.37 (P = 0.17)

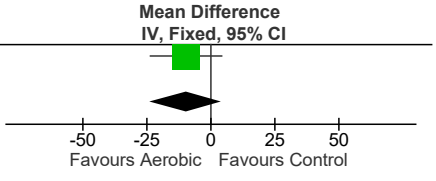
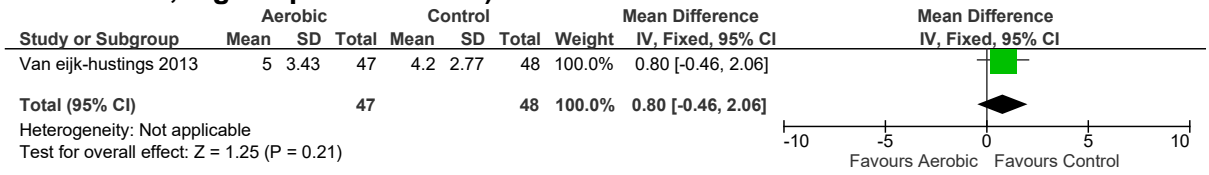
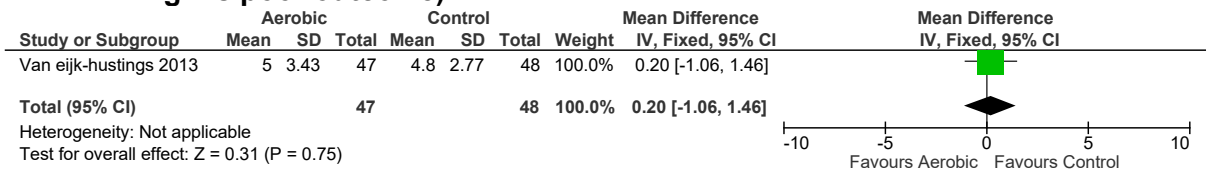


Figure 26: Psychological distress at >3 months (Final values, FIQ depression scale, 0-10, high is poor outcome)



Note: 18 month timepoint not meta-analysed with 12-24 week data.

Figure 27: Psychological distress at >3 months (Final values, FIQ anxiety scale, 0-10, high is poor outcome)



Note: 18 month timepoint not meta-analysed with 12-24 week data.

Figure 28: Psychological distress at ≤3 months (Final values, BDI dpression scale, high is poor outcome)

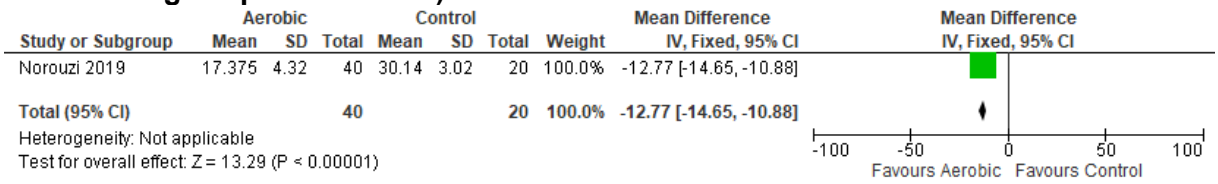


Figure 29: Use of healthcare services at 12 weeks (Number of GP contacts)

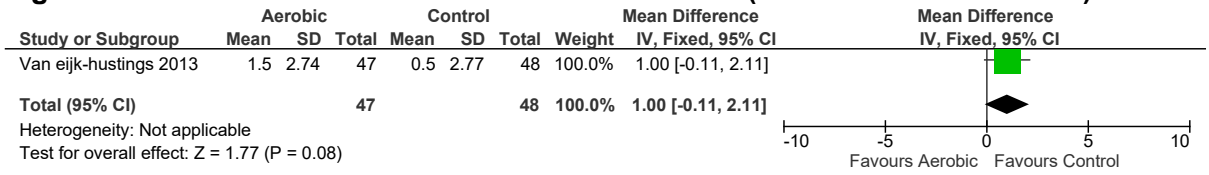


Figure 30: Use of healthcare services at 18 months (Number of GP contacts)

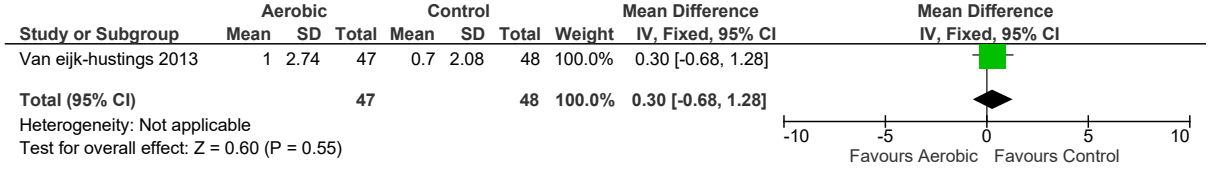


Figure 31: Use of healthcare services at 12 weeks (Number of medical specialist contacts)

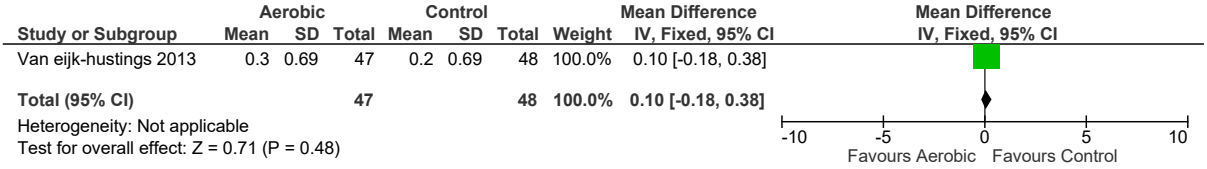


Figure 32: Use of healthcare services at 18 months (Number of medical specialist contacts)

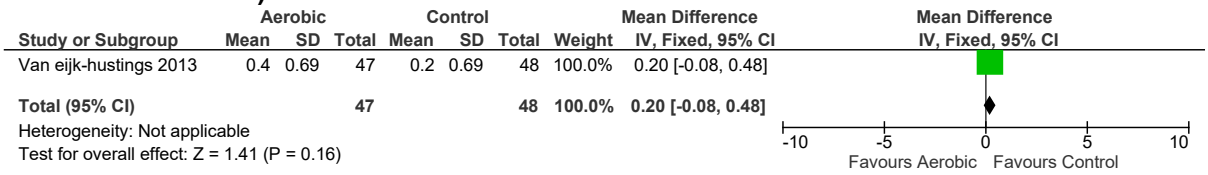


Figure 33: Use of healthcare services at 12 weeks (Number of physiotherapist contacts)

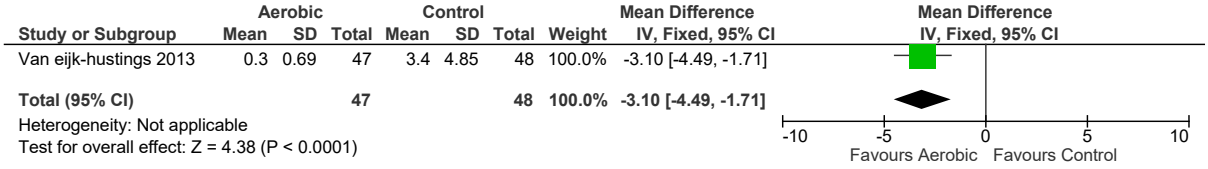


Figure 34: Use of healthcare services at 18 months (Number of physiotherapist contacts)

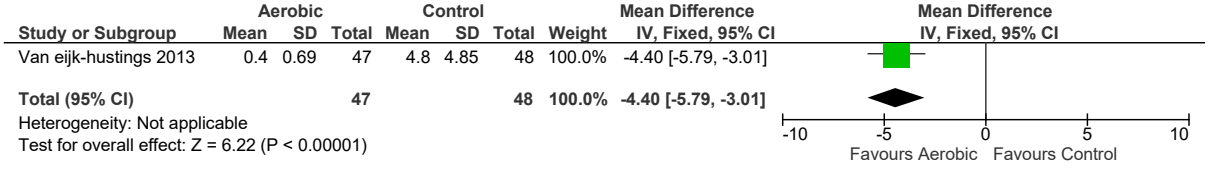


Figure 35:Sleep at >3 months (VAS sleep scale, PSQI, FIQ sleep subscale, final values, high is poor outcome)

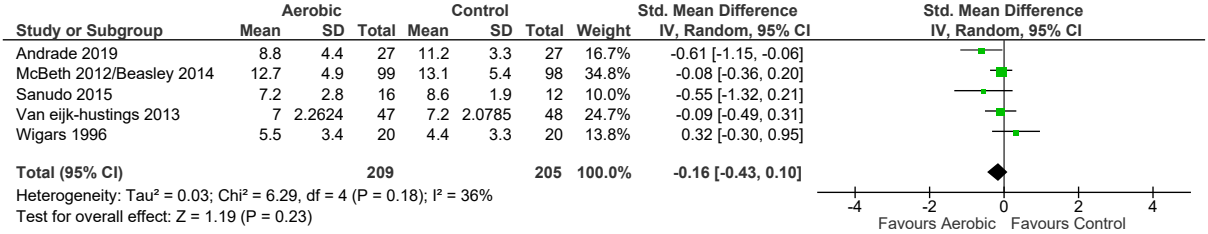
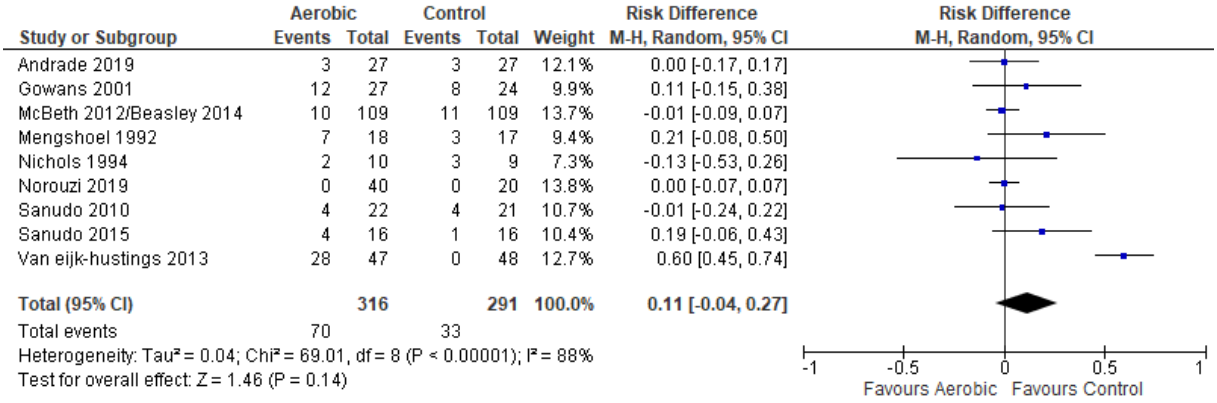


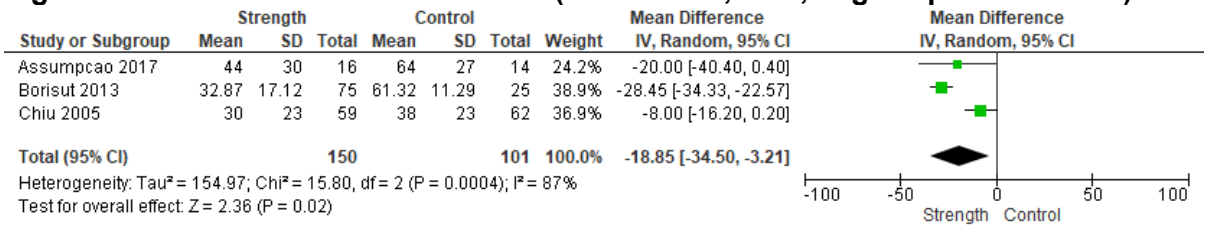
Figure 36: Discontinuation at >3 months



Heterogeneity not explained by subgroup analysis.

E.2 Strength training versus usual care

Figure 37: Pain reduction at ≤3 months (final values, VAS, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 38: Pain reduction at ≤3 months (change scores and final values, VAS, NRS, 0-100, high is poor outcome)

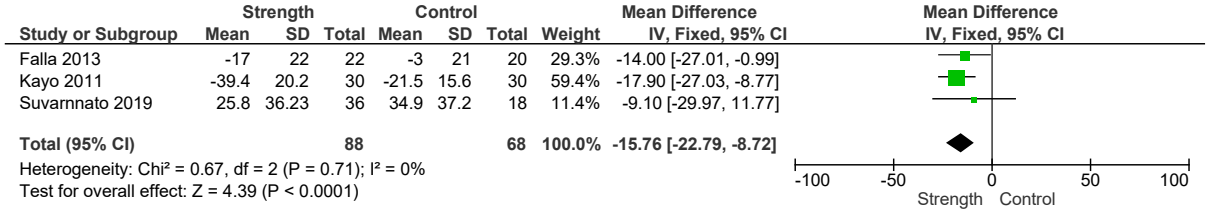
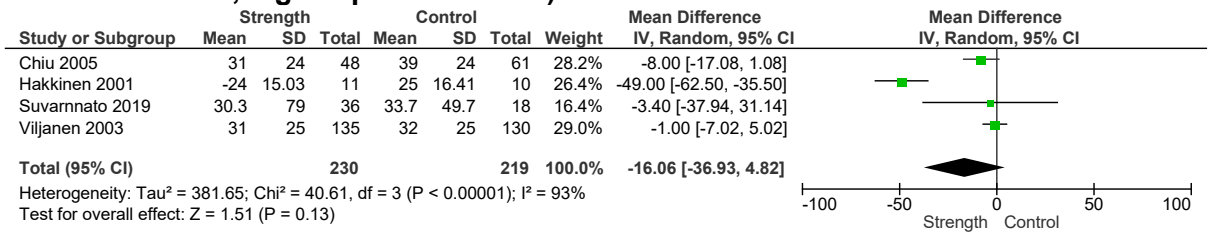


Figure 39: Pain reduction at >3 months (VAS, NRS, 0-100, final values and change scores, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 40: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)

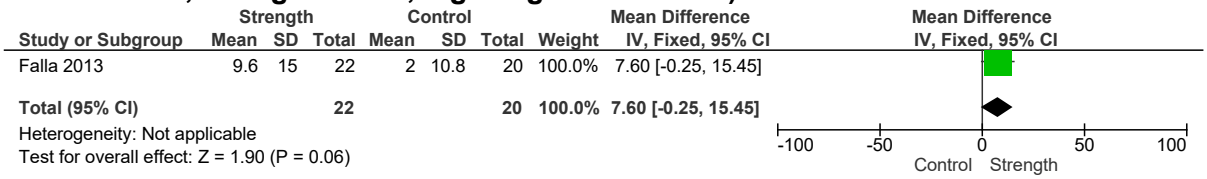


Figure 41: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)

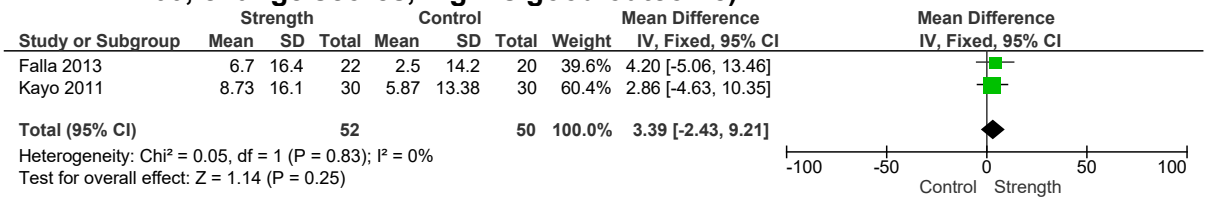
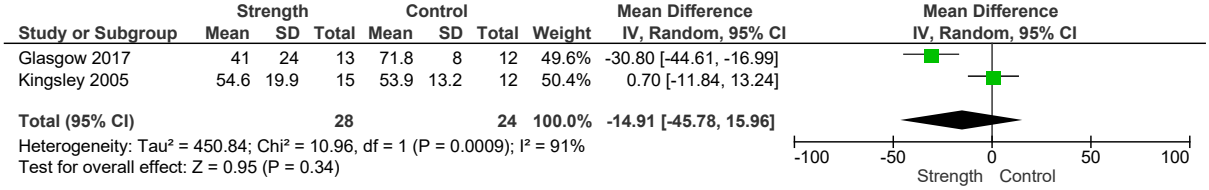
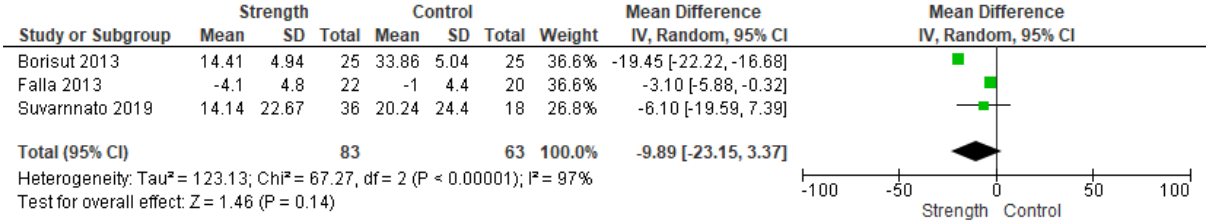


Figure 42: Quality of life at ≤3 months (FIQ, 0-100,final values, high is poor outcome)



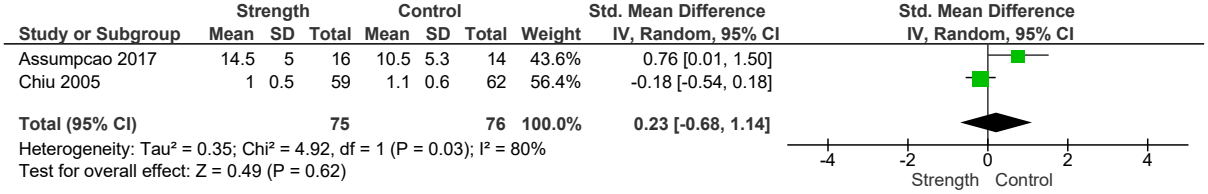
Heterogeneity not explained by subgroup analysis.

Figure 43: Physical function at ≤3 months (Neck disability index, change scores and final values, 0-100, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 44: Physical function at ≤3 months (final values, FIQ physical function subscale, Northwick pain questionnaire, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 45: Physical function at ≤3 months (6 minute walking test, final values, metres)

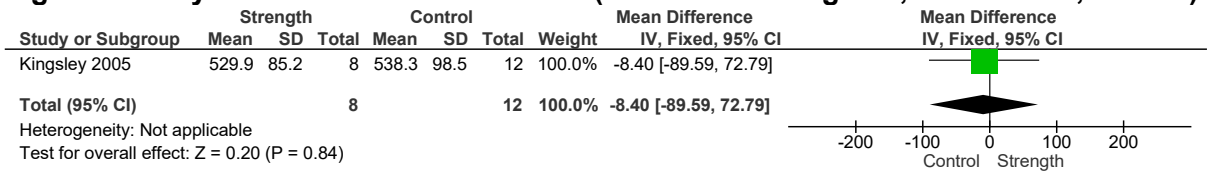


Figure 46:Physical function at >3 months (final values, Northwick Park questionnaire, Neck disability index, high is poor outcome)

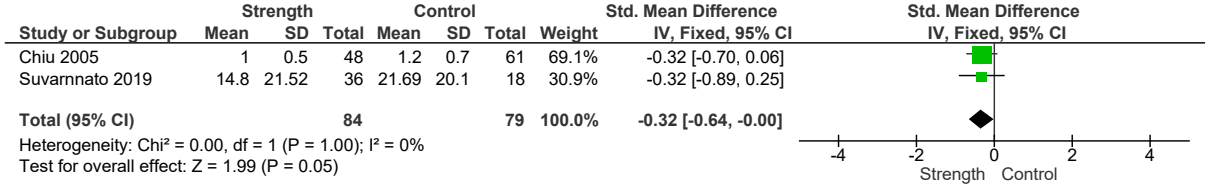


Figure 47: Physical function at >3 months (change scores, SF-36 physical function subscale, HAQ, 0-100, high is poor outcome)

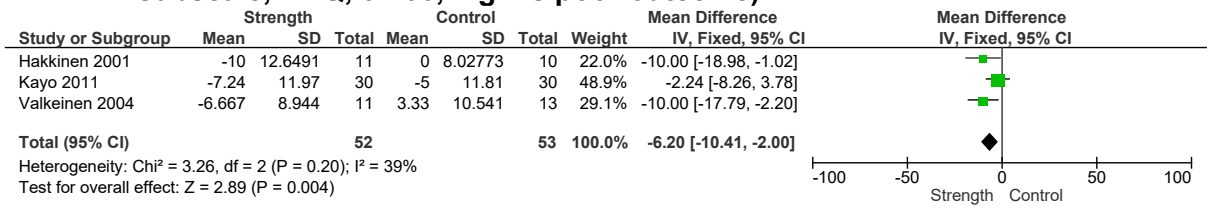


Figure 48: Psychological distress at ≤3 months (final scores, pain catastrophising scale, 0-100, high is poor outcome)

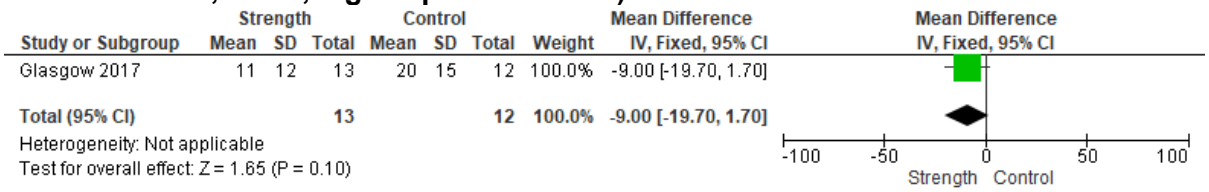


Figure 49: Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)

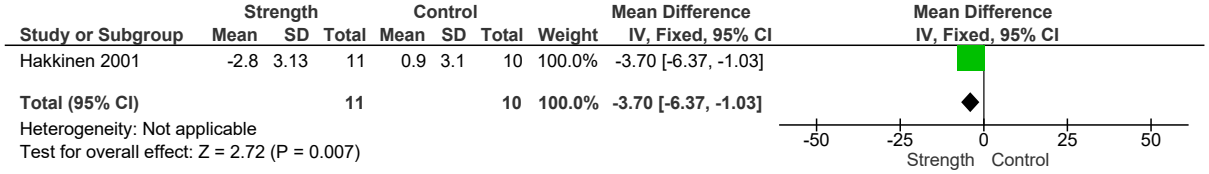


Figure 50: Use of healthcare services at >3 months

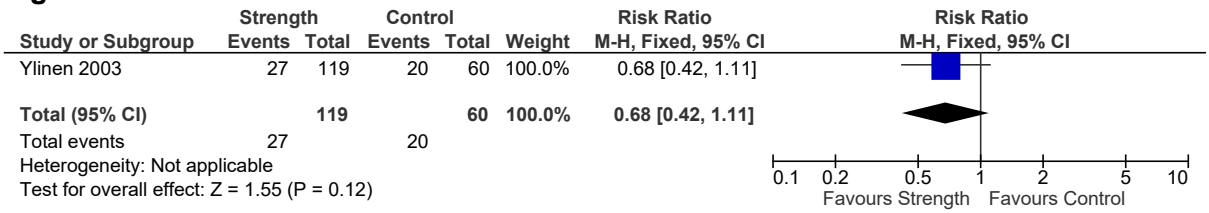


Figure 51: Sleep at >3 months (VAS sleep scale, 0-100, change scores, high is poor outcome)

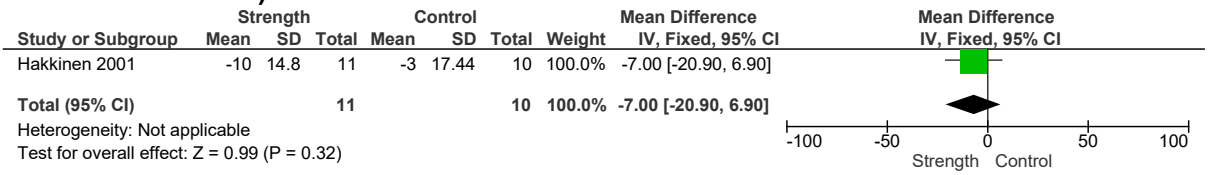


Figure 52: Discontinuation at ≤3 months

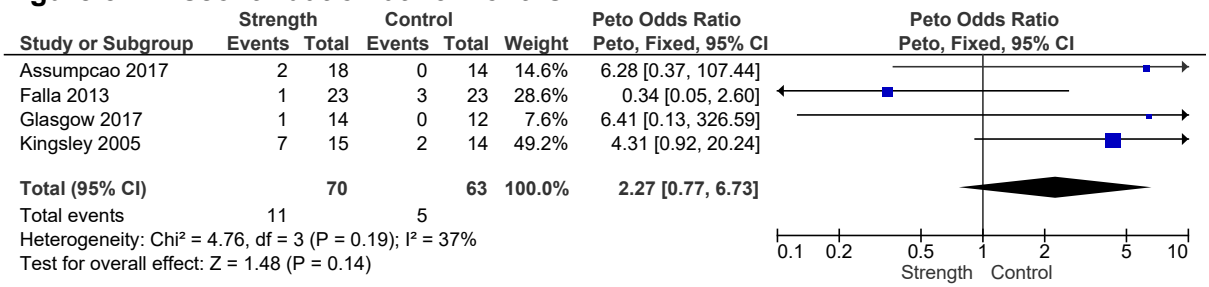
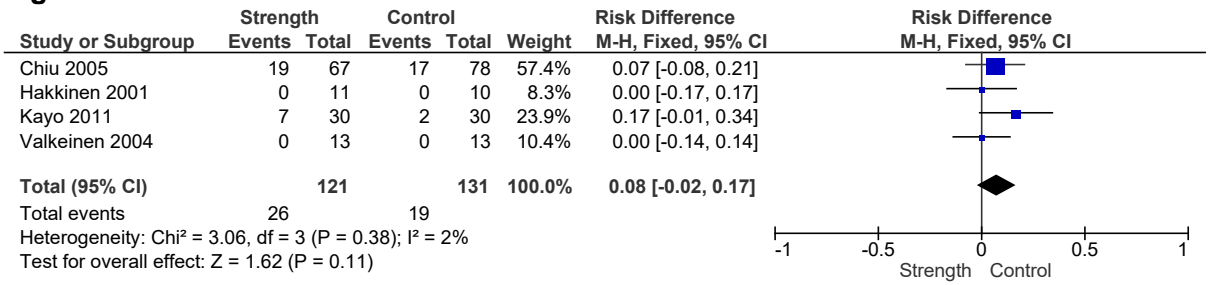
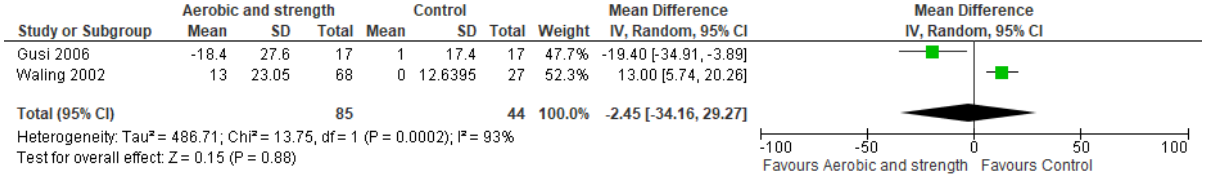


Figure 53: Discontinuation at >3 months



E.3 Aerobic and strength versus usual care

Figure 54: Pain at ≤3 months (VAS, 0-100, change scores, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 55: Pain at >3 months (VAS, FIQ pain subscale 0-100, final values, high is poor outcome)

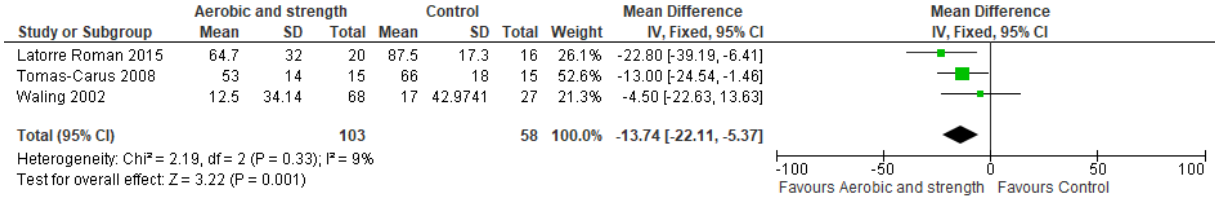


Figure 56: Quality of life at ≤3 months (EQ-5D, -0.594 to 1, final values, high is good outcome)

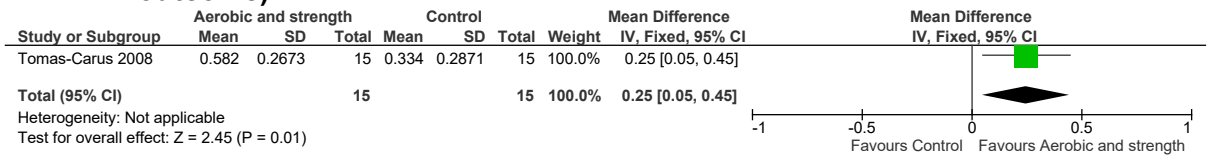


Figure 57: Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)

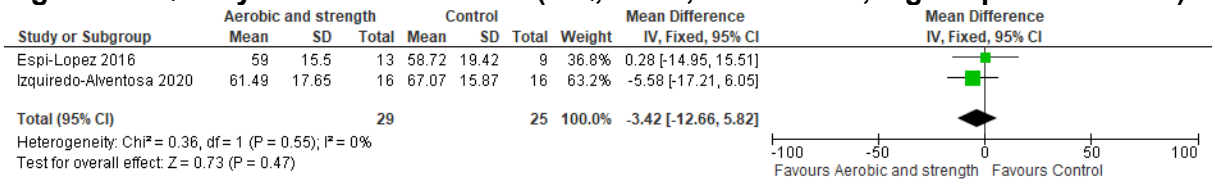
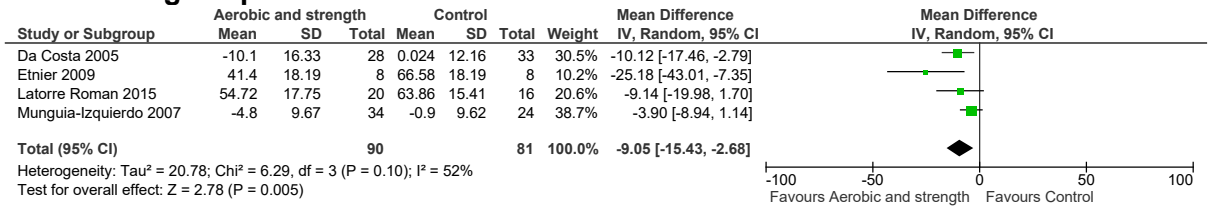


Figure 58: Quality of life at >3 months (FIQ, 0-100, final values and change scores, high is poor outcome



Heterogeneity not explained by subgroup analysis.

Figure 59: Quality of life at ≤3 months (EQ-5D, -0.594 to 1, final values, high is good outcome)

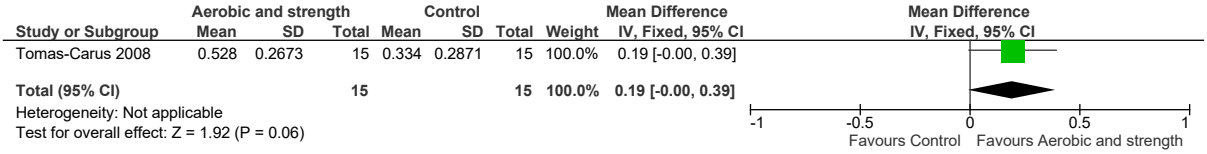


Figure 60: Quality of life at >3 months (SF-36 physical functioning subscale, 0-100, final values, high is good outcome)

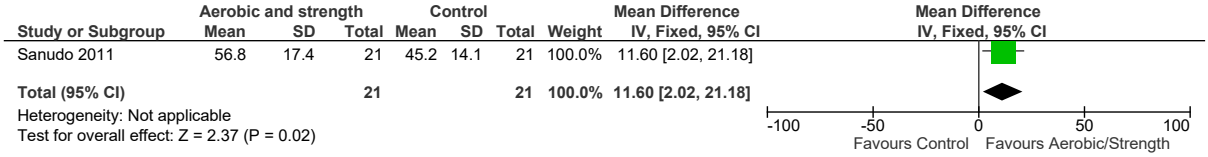


Figure 61: Quality of life at >3 months (SF-36 physical role subscale, 0-100, final values, high is good outcome)

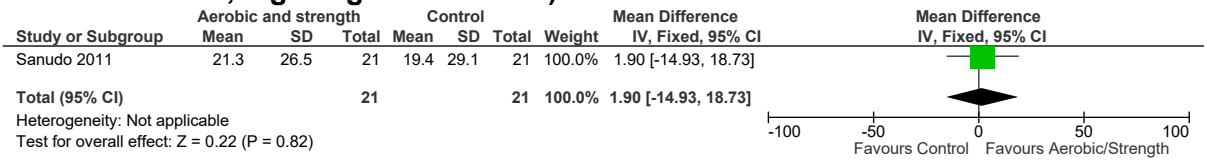


Figure 62: Quality of life at >3 months (SF-36 emotional role subscale, 0-100, final values, high is good outcome)

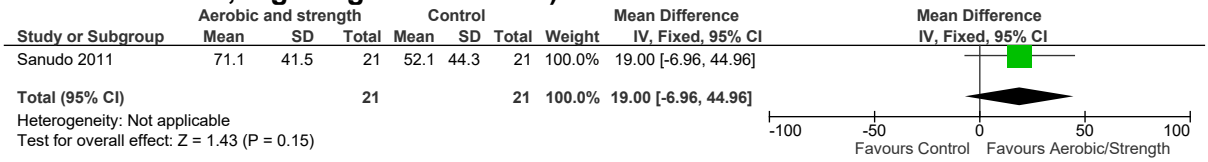


Figure 63: Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)

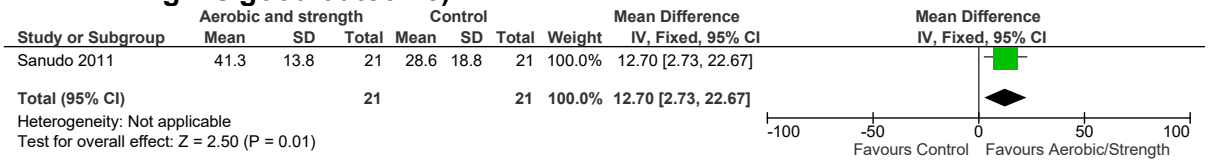


Figure 64: Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)

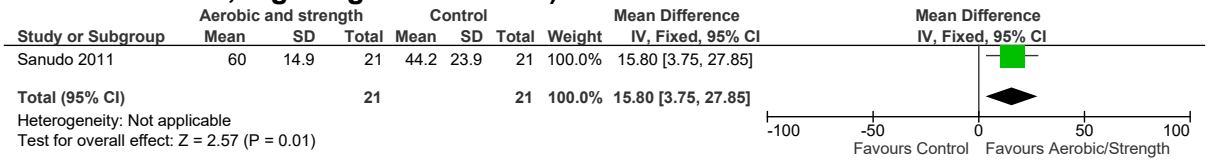


Figure 65: Quality of life at >3 months (SF-36 social role subscale, 0-100, final values, high is good outcome)

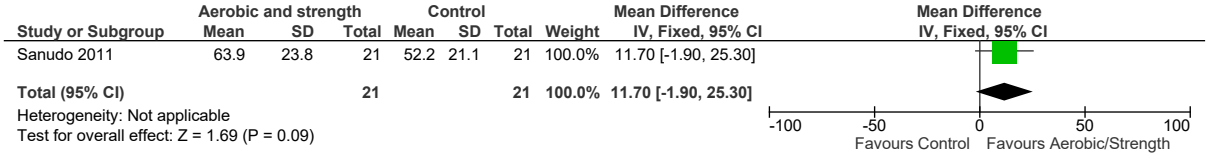


Figure 66: Quality of life at >3 months (SF-36 bodily pain subscale, 0-100, final values, high is good outcome)

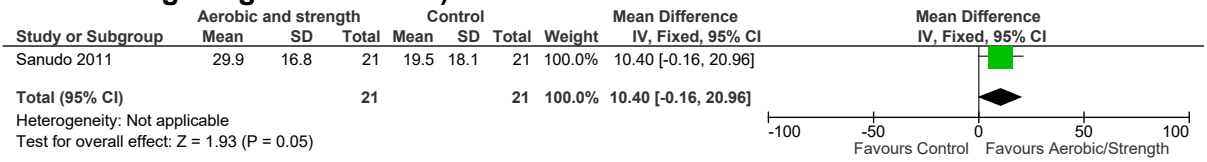


Figure 67: Quality of life at >3 months (SF-36 general health subscale, 0-100, final values, high is good outcome)

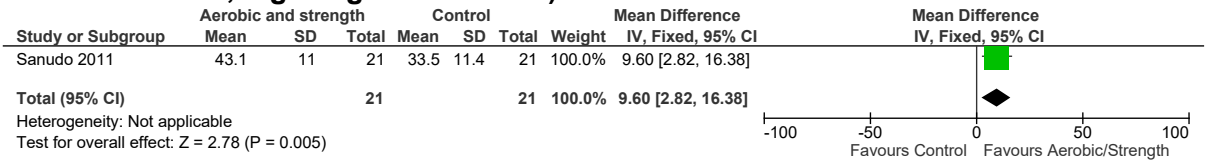


Figure 68: Physical function at >3 months (quarter mile walk test, seconds, final values, high is poor outcome`)

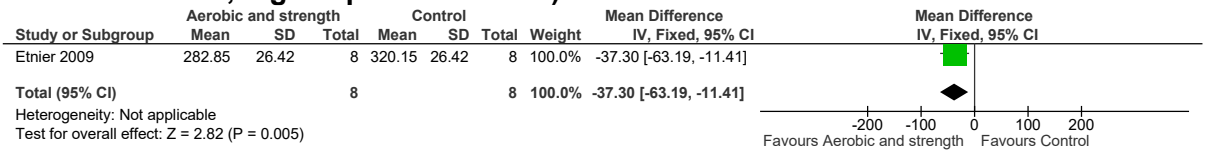


Figure 69: Physical function at >3 months (6 minute walk test, final values, metres)

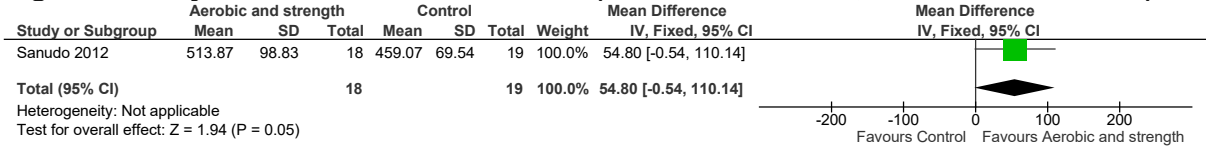


Figure 70: Physical function at ≤3 months (6 minute walk test, final values, metres)

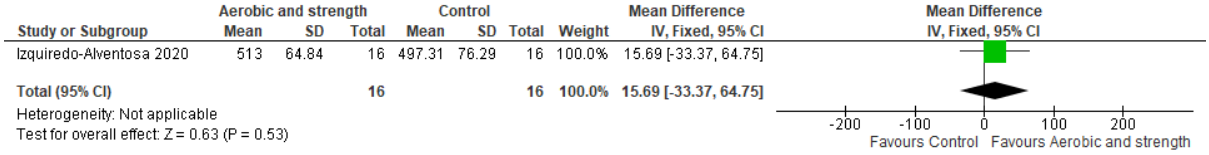


Figure 71: Physical function at >3 months (FIQ physical function subscale, 0-10, final values, high is poor outcome)

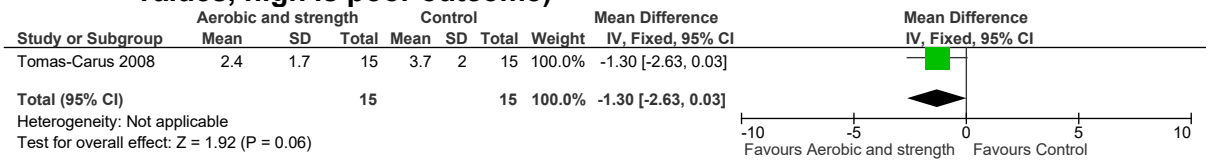


Figure 72: Psychological distress at ≤3 months (BDI, 0-30, final values, high is poor outcome)

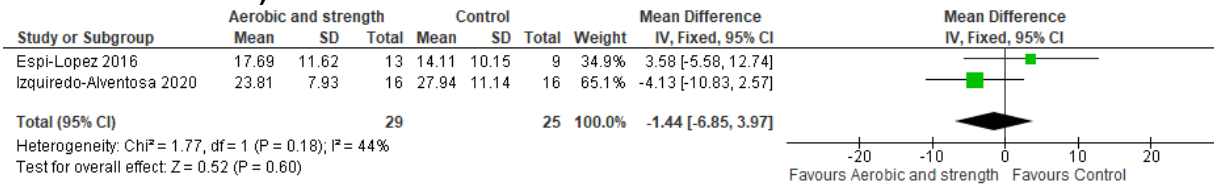


Figure 73: Psychological distress at ≤3 months (State anxiety inventory, 0-100, change scores, high is poor outcome)

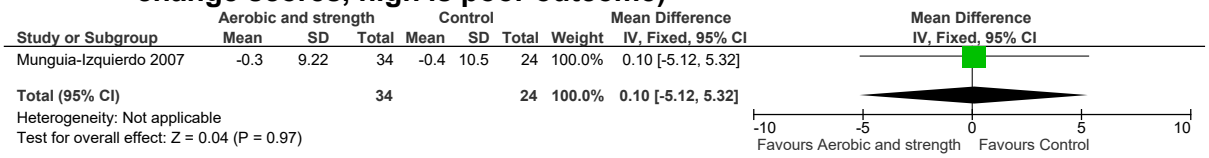


Figure 74: Psychological distress at ≤3 months (HADS anxiety, 0-21, high is poor outcome)

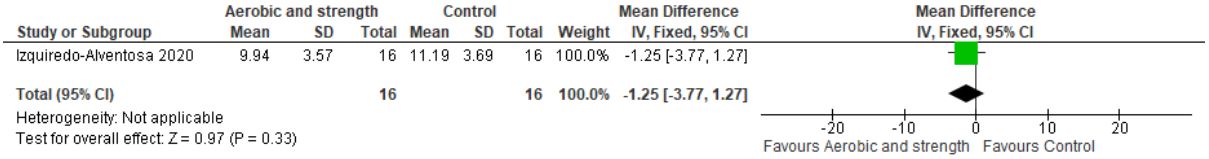


Figure 75: Psychological distress at >3 months (CES-D, BDI, FIQ depression subscale, final values, high is poor outcome)

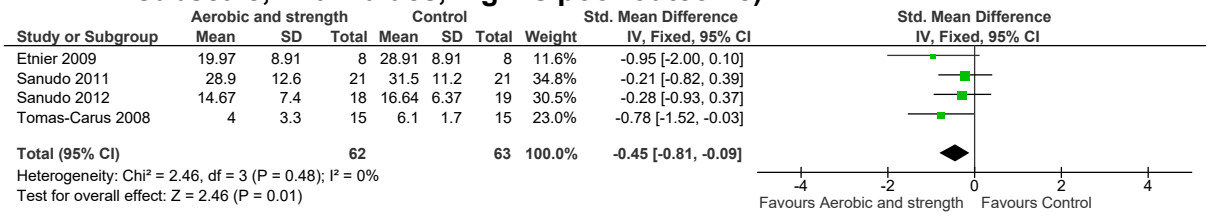


Figure 76: Psychological distress at >3 months (State anxiety inventory, 20-80, final values and change scores, high is poor outcome)

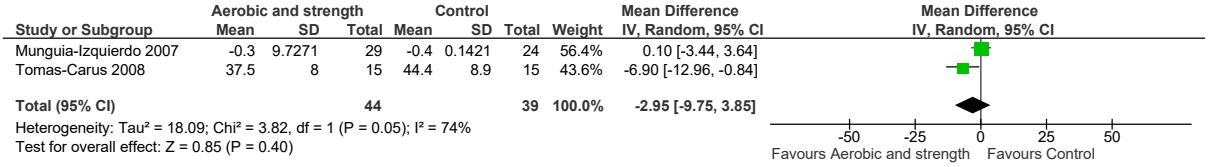


Figure 77: Sleep at >3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)

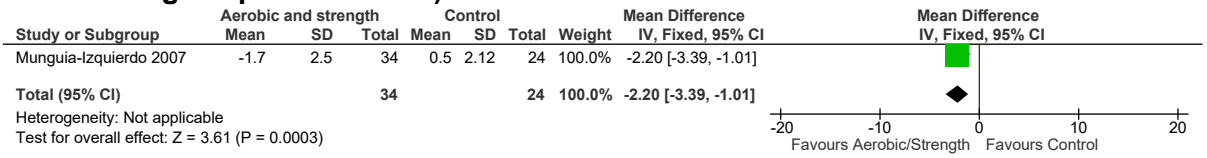


Figure 78: Health care utilisation at >3 months

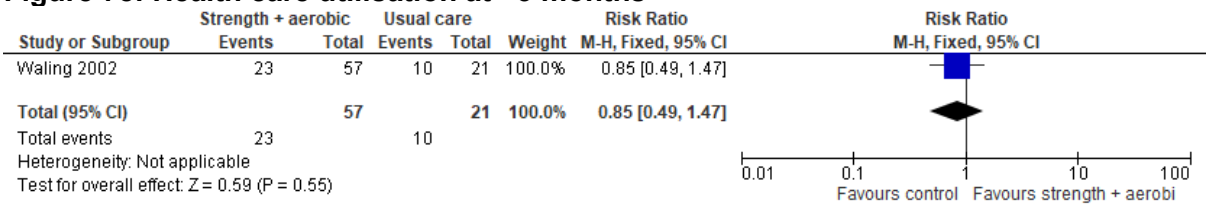


Figure 79: Discontinuation at ≤3 months

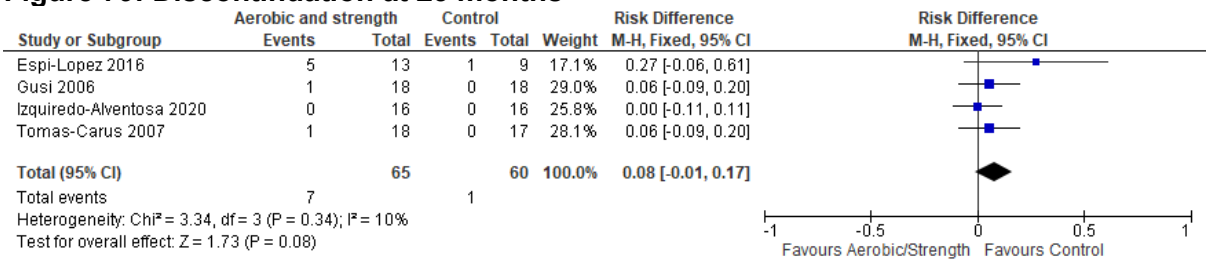
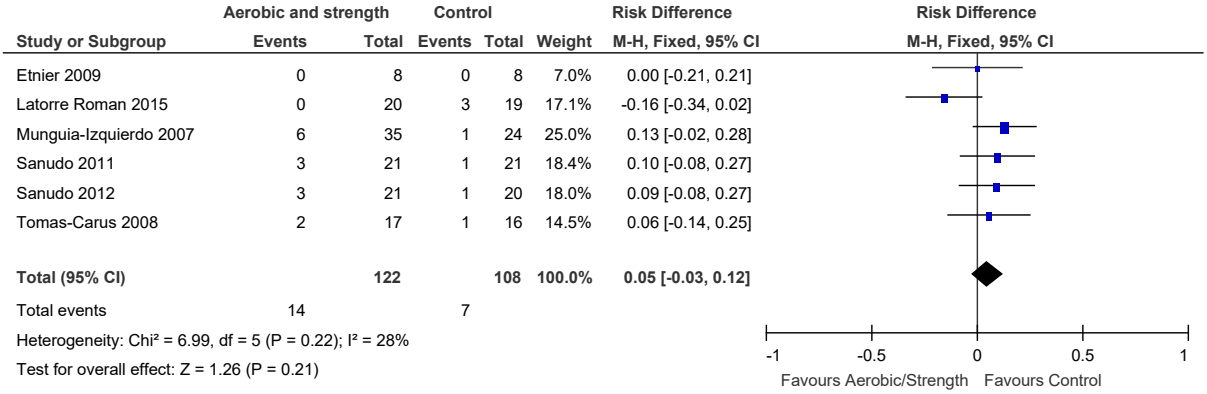


Figure 80: Discontinuation at >3 months



E.4 Aerobic, strength and flexibility versus usual care

Figure 81: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

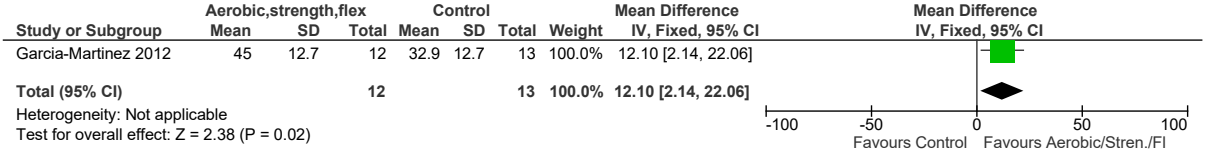
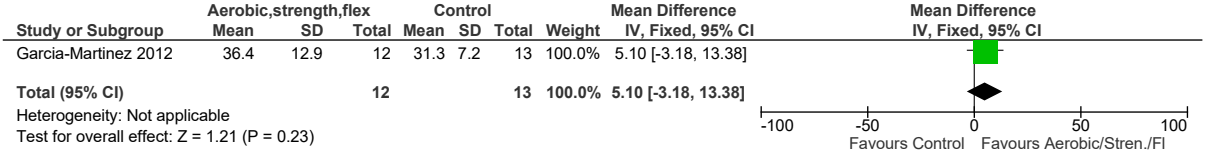


Figure 82: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)



E.5 Strength and flexibility versus usual care

Figure 83: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

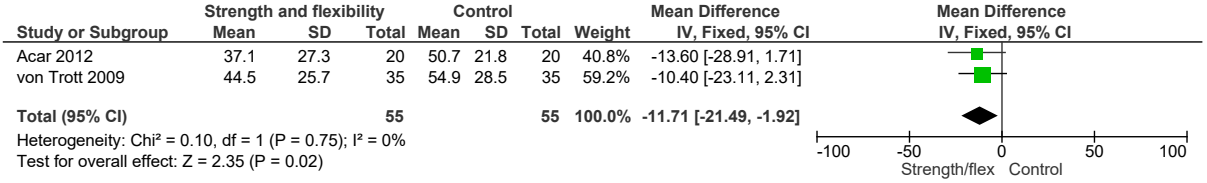


Figure 84: Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Rendant 2011	27.4	17.05	35	41	20.23	39	70.6%	-13.60 [-22.10, -5.10]
von Trott 2009	47.7	30.5	35	59.9	25.5	35	29.4%	-12.20 [-25.37, 0.97]
Total (95% CI)			**70**			**74**	**100.0%**	**-13.19 [-20.33, -6.05]**

Heterogeneity: Chi² = 0.03, df = 1 (P = 0.86); I² = 0%
Test for overall effect: Z = 3.62 (P = 0.0003)

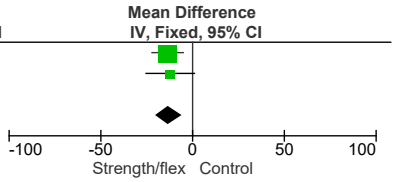


Figure 85: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
von Trott 2009	49.2	10.9	35	49.8	12.6	35	100.0%	-0.60 [-6.12, 4.92]
Total (95% CI)			**35**			**35**	**100.0%**	**-0.60 [-6.12, 4.92]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.21 (P = 0.83)

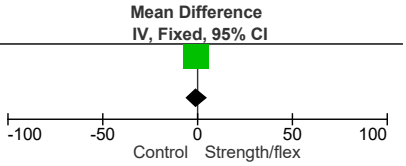


Figure 86: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Rendant 2011	47.8	8.75	35	45.4	8.76	39	61.4%	2.40 [-1.60, 6.40]
von Trott 2009	45.5	10.8	35	44.7	10.7	35	38.6%	0.80 [-4.24, 5.84]
Total (95% CI)			**70**			**74**	**100.0%**	**1.78 [-1.35, 4.91]**

Heterogeneity: Chi² = 0.24, df = 1 (P = 0.63); I² = 0%
Test for overall effect: Z = 1.12 (P = 0.26)

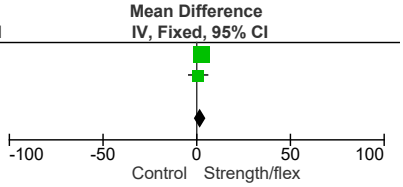


Figure 87: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
von Trott 2009	30.3	7.8	35	28.6	9.7	35	100.0%	1.70 [-2.42, 5.82]
Total (95% CI)			**35**			**35**	**100.0%**	**1.70 [-2.42, 5.82]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.81 (P = 0.42)

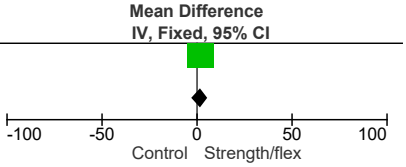


Figure 88: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Rendant 2011	44.7	7.55	35	43.1	7.17	39	53.8%	1.60 [-1.76, 4.96]
von Trott 2009	29.3	8.5	35	31.5	8.3	35	46.2%	-2.20 [-6.14, 1.74]
Total (95% CI)			**70**			**74**	**100.0%**	**-0.16 [-3.87, 3.56]**

Heterogeneity: Tau² = 3.73; Chi² = 2.07, df = 1 (P = 0.15); I² = 52%
Test for overall effect: Z = 0.08 (P = 0.93)

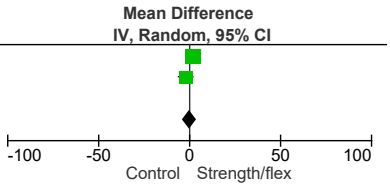


Figure 89: Physical function at ≤3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)

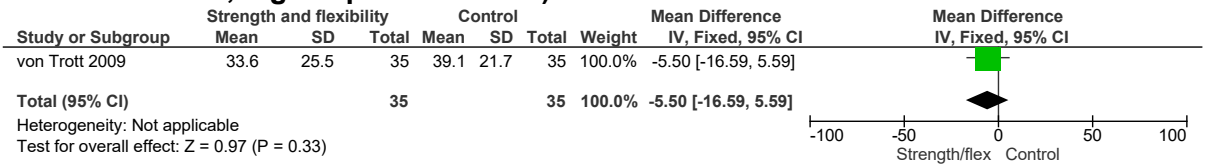


Figure 90: Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)

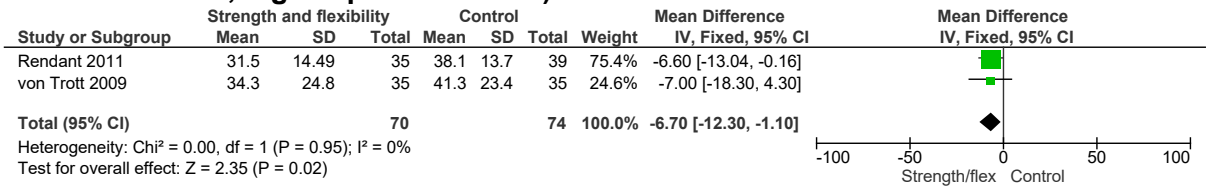


Figure 91: Psychological distress at ≤3 months (ADS depression scale, 0-60, final values, high is poor outcome)

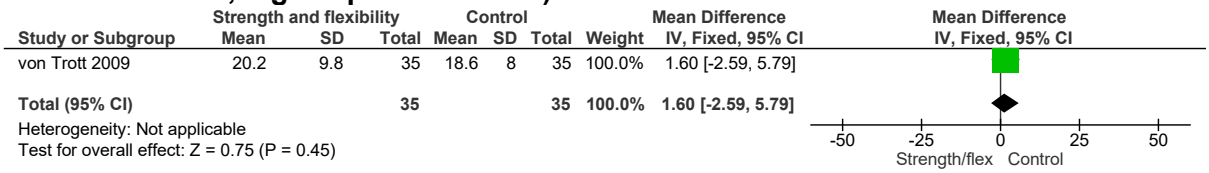


Figure 92: Psychological distress at >3 months (ADS depression scale, 0-60, final values, high is poor outcome)

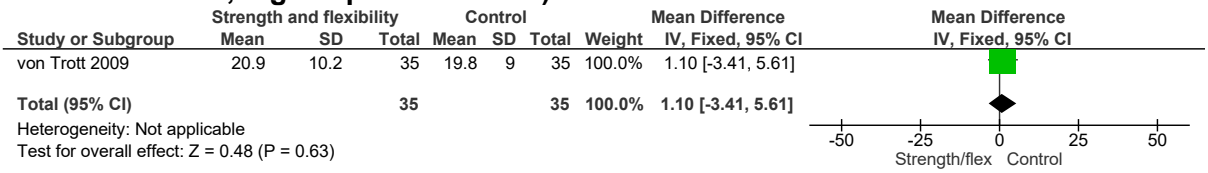
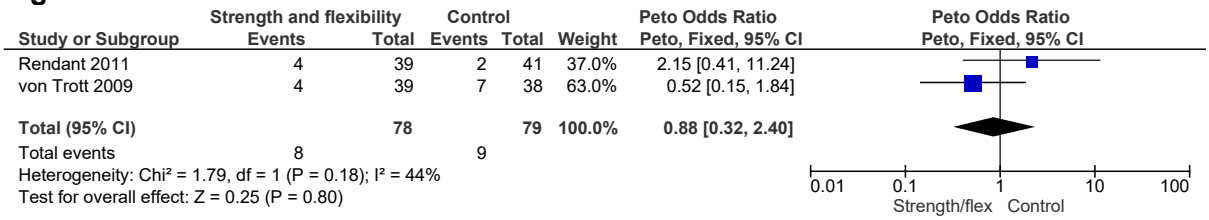


Figure 93: Discontinuation at >3 months



E.6 Strength, proprioception and flexibility versus usual care

Figure 94: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

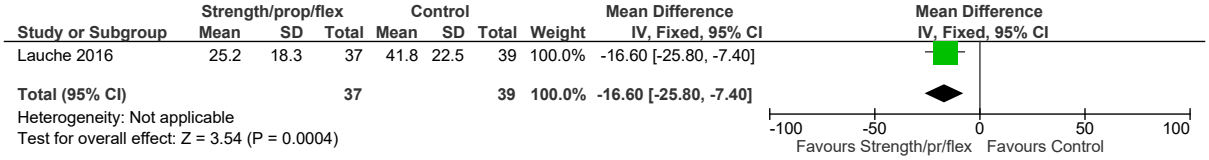


Figure 95: Pain at >3 months (VAS, 0-100, final values, high is poor outcome)

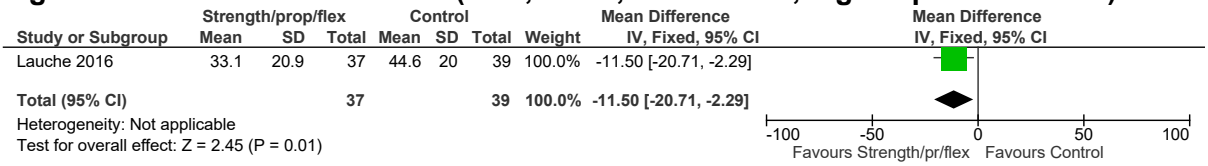


Figure 96: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

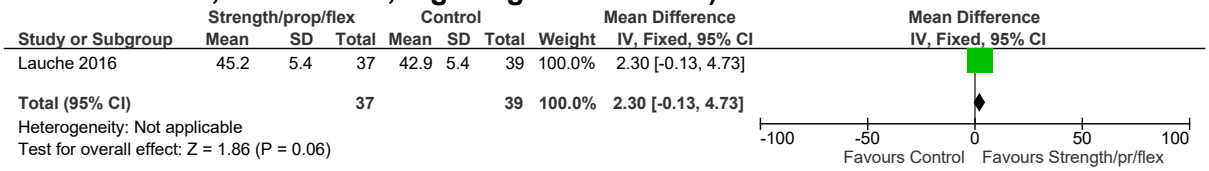


Figure 97: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

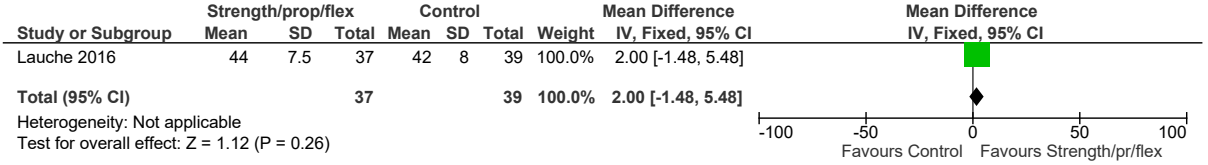


Figure 98: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

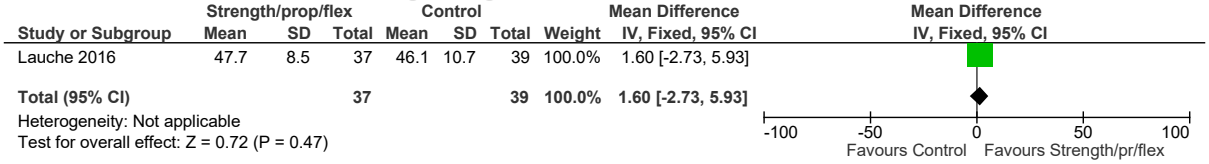


Figure 99: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

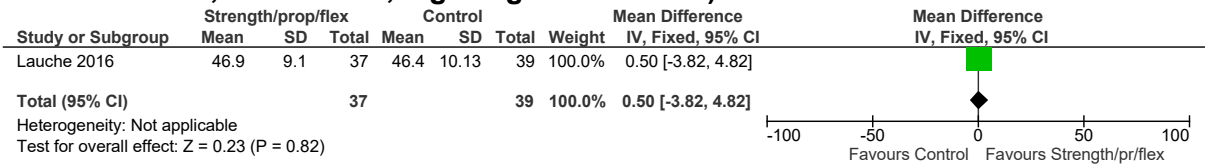


Figure 100: Psychological distress at ≤3 months (HADS anxiety, 0-21, final values, high is poor outcome)

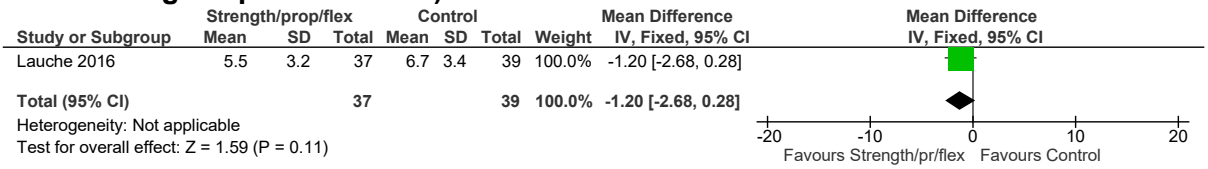


Figure 101: Psychological distress at >3 months (HADS anxiety, 0-21, final values, high is poor outcome)

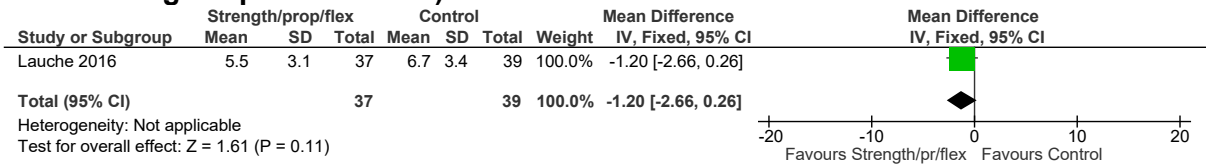


Figure 102: Psychological distress at ≤3 months (HADS depression, 0-21, final values, high is poor outcome)

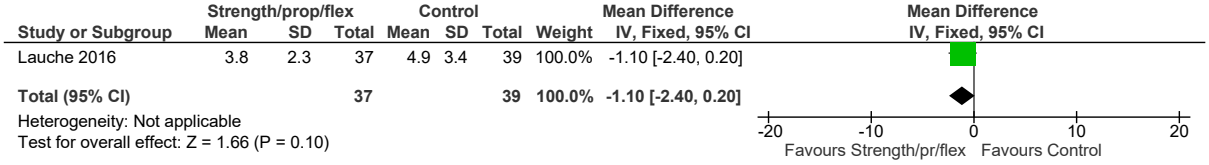


Figure 103: Psychological distress at >3 months (HADS depression, 0-21, final values, high is poor outcome)

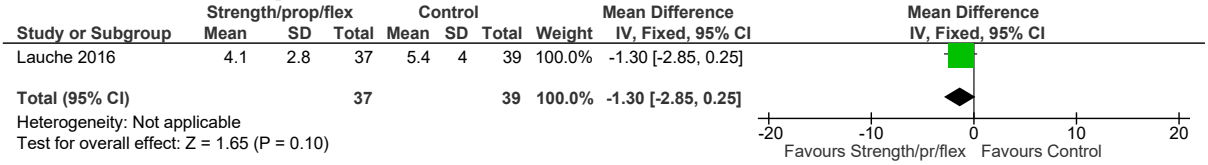


Figure 104: Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	22.7	9.3	37	27.5	11.4	39	100.0%	-4.80 [-9.47, -0.13]
Total (95% CI)			**37**			**39**	**100.0%**	**-4.80 [-9.47, -0.13]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.02 (P = 0.04)

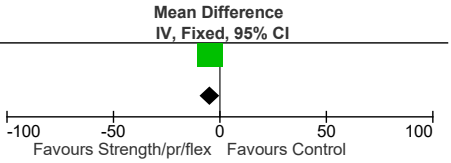


Figure 105: Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	25.1	12.9	37	29.4	12.7	39	100.0%	-4.30 [-10.06, 1.46]
Total (95% CI)			**37**			**39**	**100.0%**	**-4.30 [-10.06, 1.46]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.46 (P = 0.14)

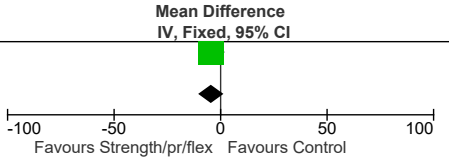
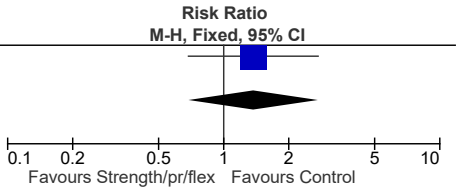


Figure 106: Discontinuation at ≤3 months

Study or Subgroup	Strength/prop/flex Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Lauche 2016	13	37	10	39	100.0%	1.37 [0.69, 2.73]
Total (95% CI)		**37**		**39**	**100.0%**	**1.37 [0.69, 2.73]**
Total events	13		10			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.89 (P = 0.37)



E.7 Proprioception versus usual care

Figure 107: Pain at ≤3 months (VAS, 0-10, final values, high is poor outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	5.81	2.7	24	5.63	1.62	22	100.0%	0.18 [-1.09, 1.45]
Total (95% CI)			**24**			**22**	**100.0%**	**0.18 [-1.09, 1.45]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.28 (P = 0.78)

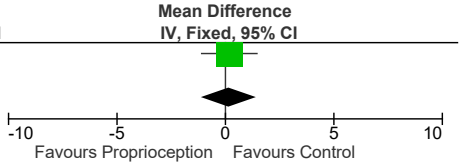


Figure 108: Pain at >3 months (VAS, 0-10, final values, high is poor outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	5.39	2.84	24	6.36	2.33	22	100.0%	-0.97 [-2.47, 0.53]
Total (95% CI)			**24**			**22**	**100.0%**	**-0.97 [-2.47, 0.53]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.27 (P = 0.20)

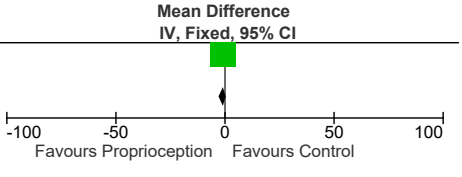


Figure 109: Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	48.29	19.4	24	50.17	11.95	22	100.0%	-1.88 [-11.11, 7.35]
Total (95% CI)			**24**			**22**	**100.0%**	**-1.88 [-11.11, 7.35]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.40 (P = 0.69)

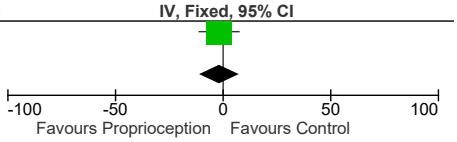


Figure 110: Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	49.37	20.35	24	52.96	16.92	22	100.0%	-3.59 [-14.37, 7.19]
Total (95% CI)			**24**			**22**	**100.0%**	**-3.59 [-14.37, 7.19]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.65 (P = 0.51)

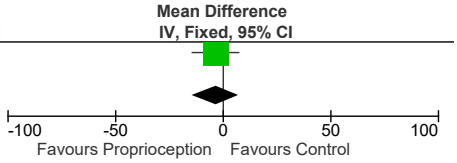


Figure 111: Physical function at ≤3 months (Sit to stand test, final values, high is good outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	24.21	3.82	24	28.59	4.56	22	100.0%	-4.38 [-6.82, -1.94]
Total (95% CI)			**24**			**22**	**100.0%**	**-4.38 [-6.82, -1.94]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.51 (P = 0.0004)

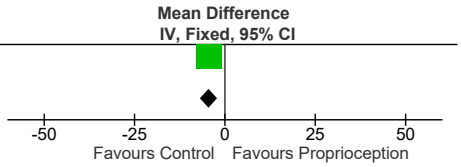


Figure 112: Physical function at >3 months (Sit to stand test, final values, high is good outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	24.91	2.87	24	25.77	4.82	22	100.0%	-0.86 [-3.18, 1.46]
Total (95% CI)			**24**			**22**	**100.0%**	**-0.86 [-3.18, 1.46]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.73 (P = 0.47)

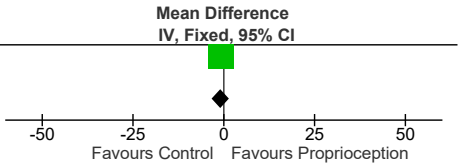


Figure 113: Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)

Study or Subgroup	Proprioception Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Altan 2004	9.21	6.97	24	13.95	5.79	22	100.0%	-4.74 [-8.43, -1.05]
Total (95% CI)			**24**			**22**	**100.0%**	**-4.74 [-8.43, -1.05]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.52 (P = 0.01)

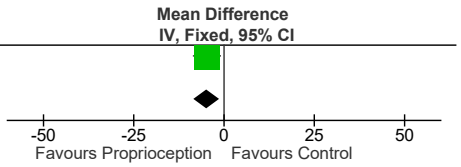


Figure 114: Psychological distress at >3 months (BDI, 0-61, final values, high is poor outcome)

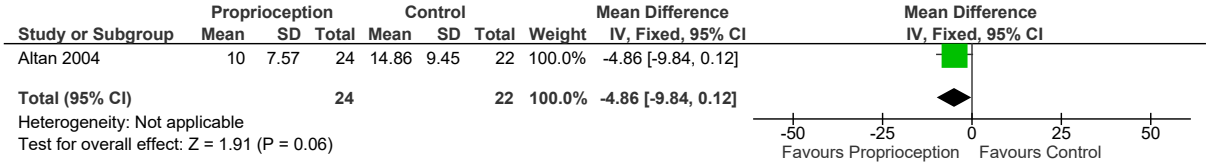
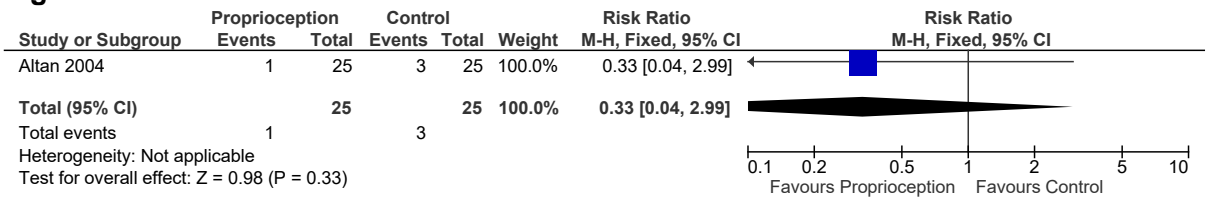


Figure 115: Discontinuation at >3 months



E.8 Mind-body versus usual care

Figure 116: Pain at ≤3 months (VAS, FIQ pain subscale, 0-100, final values and change scores, high is poor outcome)

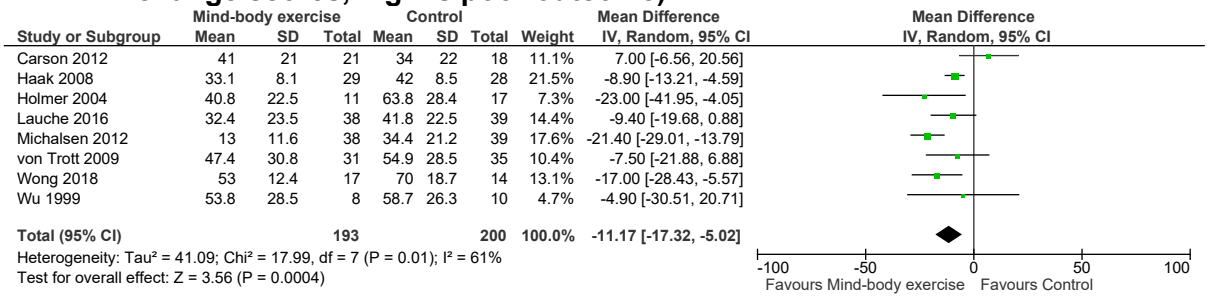


Figure 117: Pain improvement at <3 months (30% improvement on NRS)

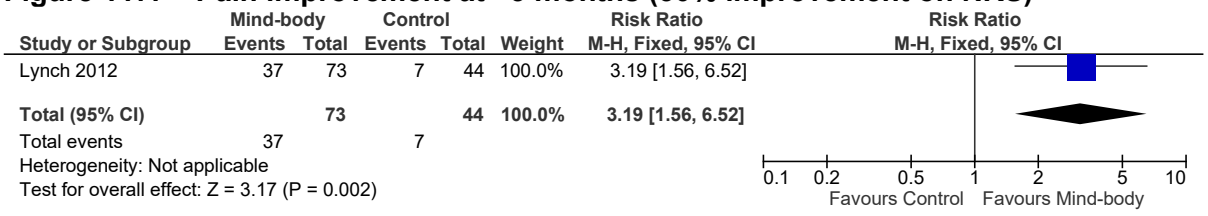


Figure 118: Pain improvement at >3 months (30% improvement on NRS)

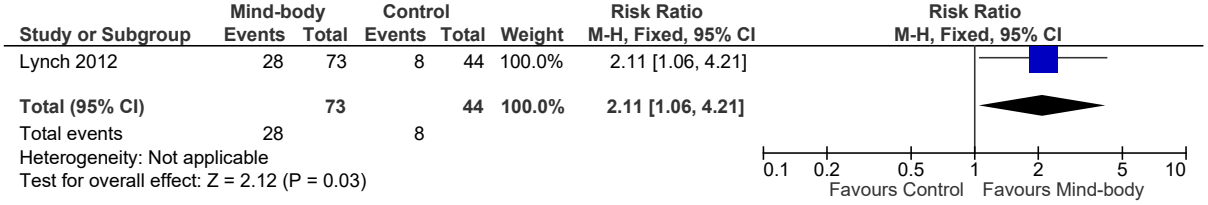
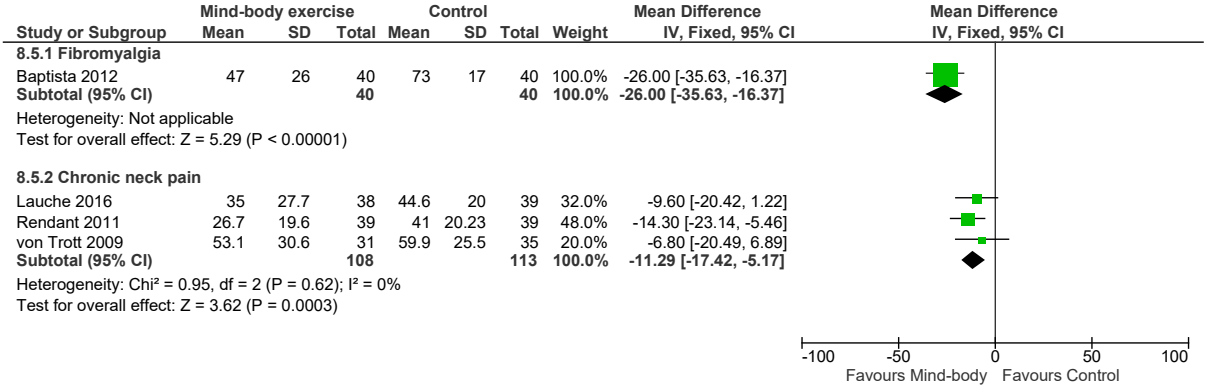


Figure 119: Pain at >3 months (VAS, SF-36 pain score, final values, 0-100, high is poor outcome)



NB: Heterogeneity explained by subgroup analysis

Figure 120: Quality of life at ≤3 months (WHOQOL-BREF, 0-5, final values, high is good outcome)

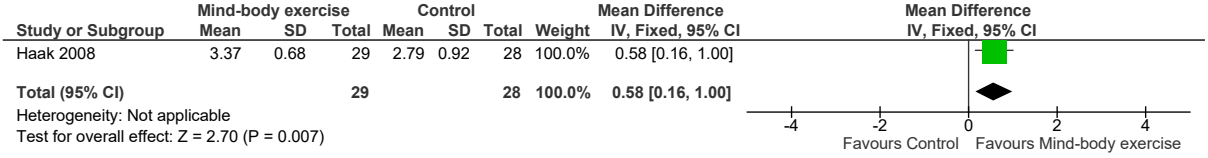


Figure 121: Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)

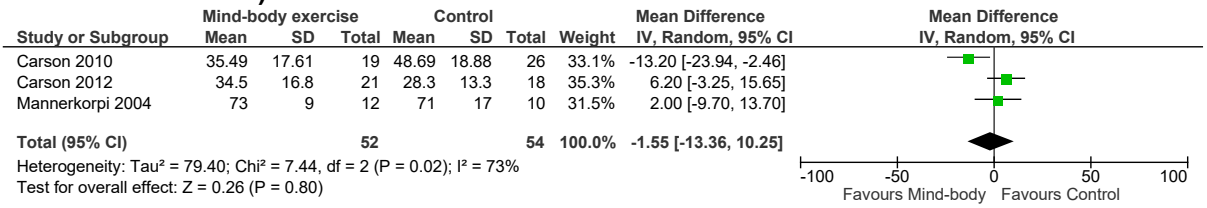


Figure 122: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

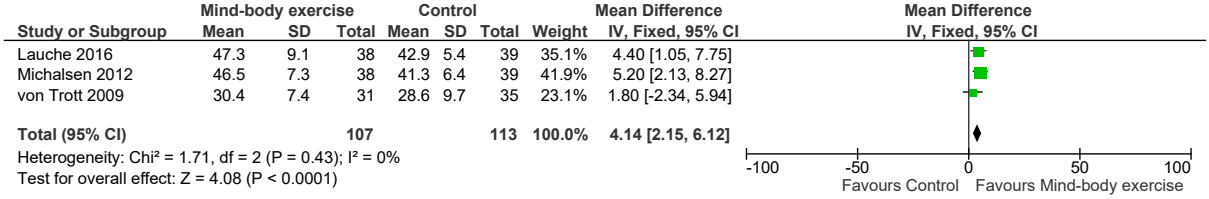
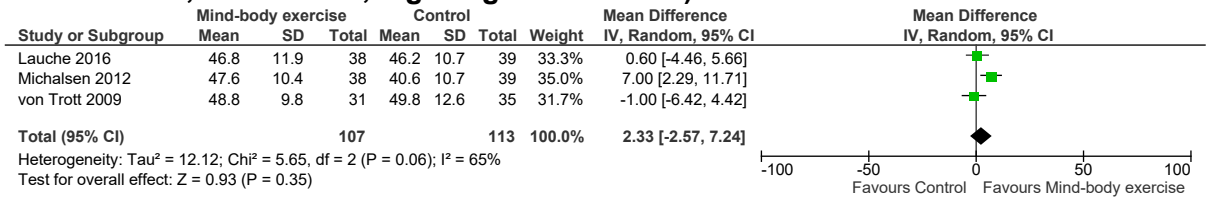
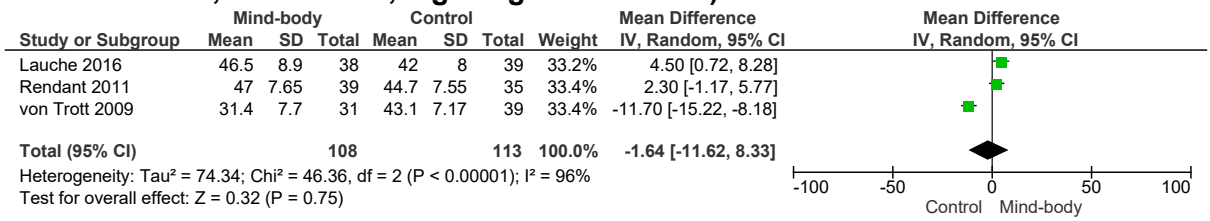


Figure 123: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)



Heterogeneity not explained by subgroup analysis

Figure 124: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)



Heterogeneity not explained by subgroup analysis

Figure 125: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

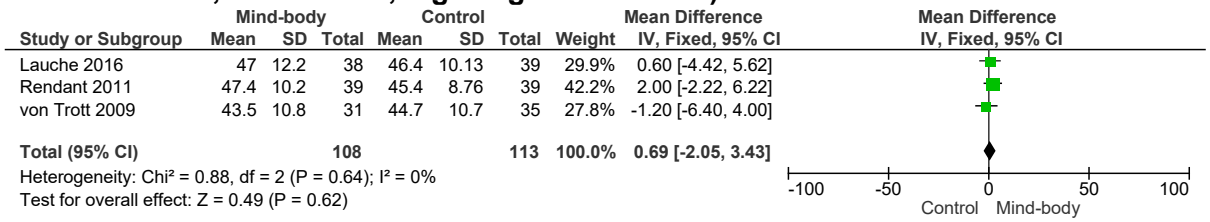


Figure 126: Quality of life at >3 months (SF-36 functional capacity subscale, 0-100, final values, high is good outcome)

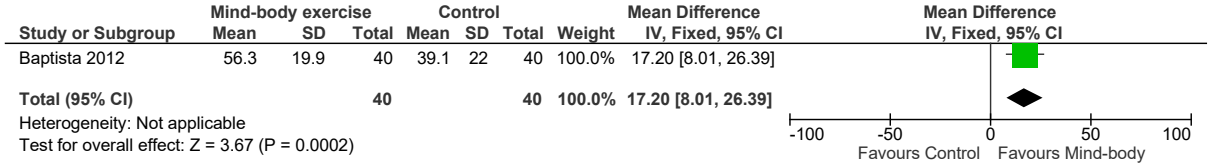


Figure 127: Quality of life at >3 months (SF-36 physical subscale, 0-100, final values, high is good outcome)

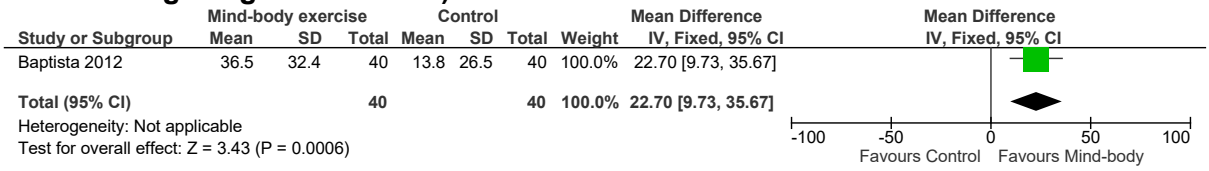


Figure 128: Quality of life at >3 months (SF-36 pain subscale, 0-100, final values, high is good outcome)

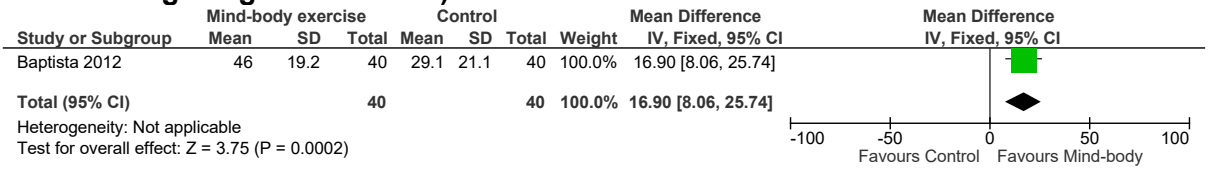


Figure 129: Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)

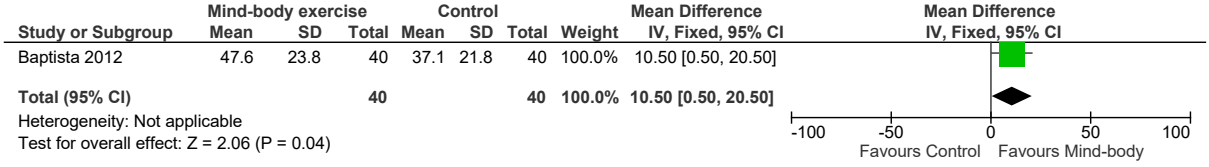


Figure 130: Quality of life at >3 months (SF-36 general health subscale, 0-100, final values, high is good outcome)

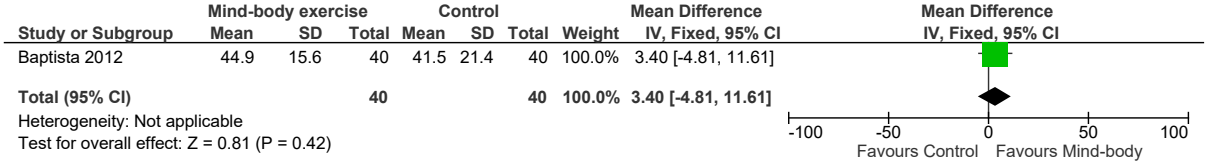


Figure 131: Quality of life at >3 months (SF-36 social subscale, 0-100, final values, high is good outcome)

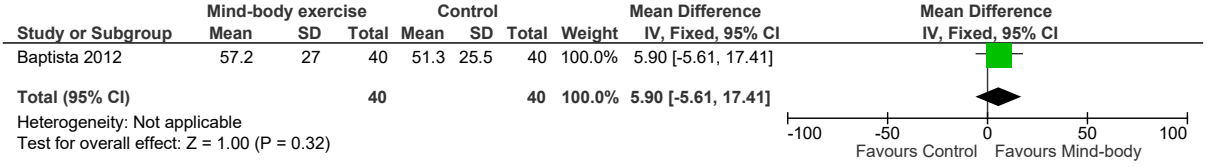


Figure 132: Quality of life at >3 months (SF-36 emotional subscale, 0-100, final values, high is good outcome)

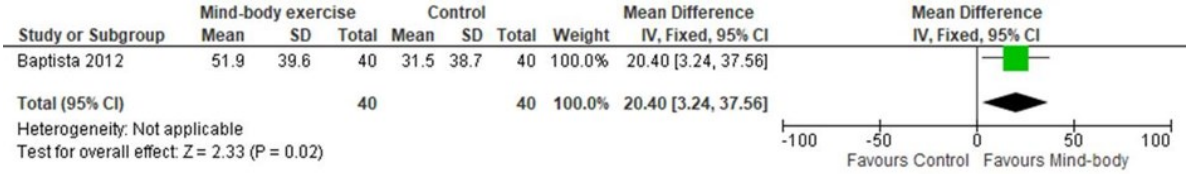


Figure 133: Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)

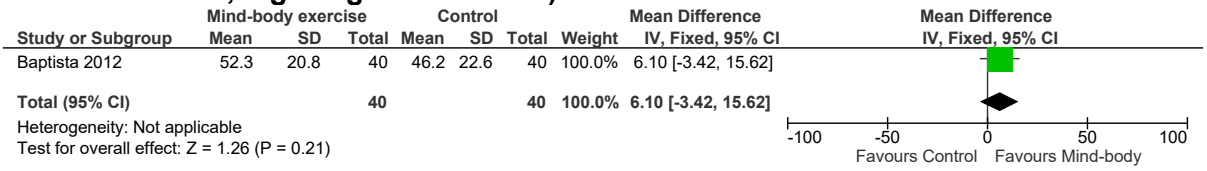
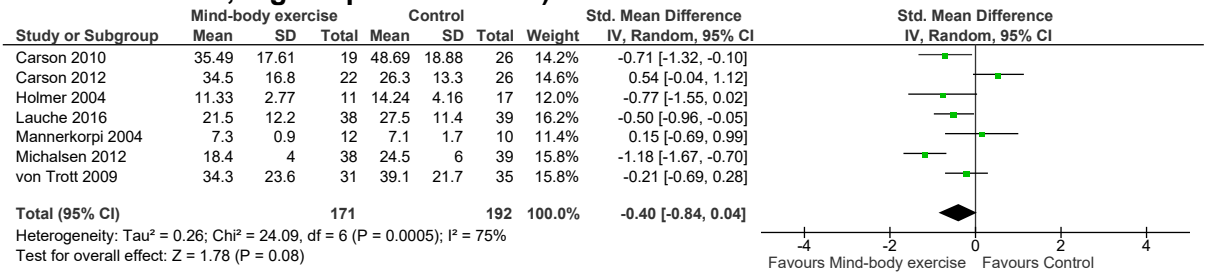


Figure 134: Physical function at >3 months (Neck pain disability scale, NDI, final values, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 135: Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)

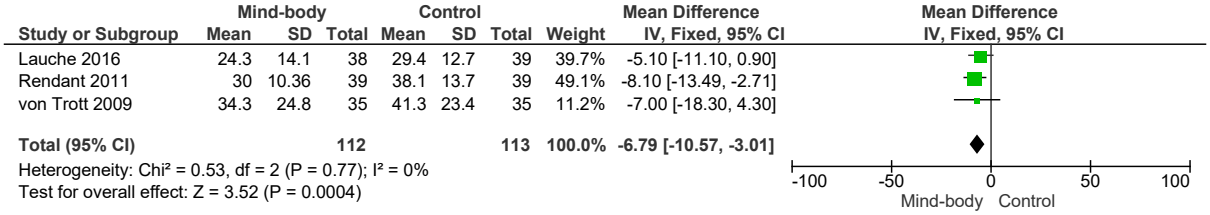


Figure 136: Physical function at >3 months (6 minute walk test, metres, final values, high is good outcome)

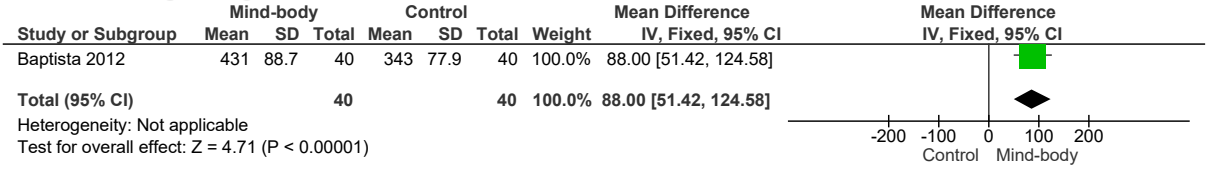
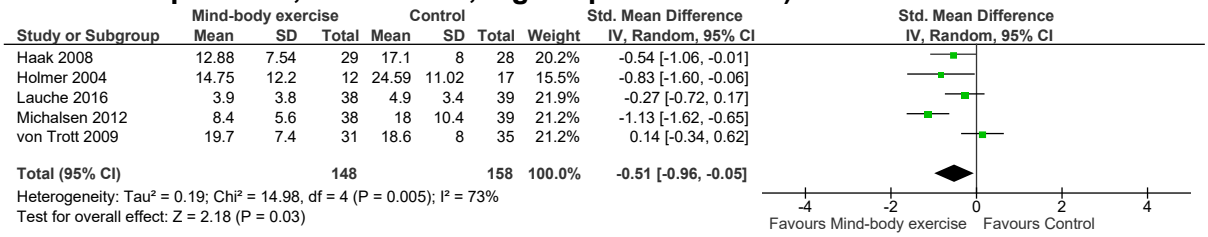


Figure 137: Psychological distress at ≤3 months (HADS:D, BDI, CES-D, ADS depression, final values, high is poor outcome)



Heterogeneity not explained by subgroup analysis.

Figure 138: Psychological distress at ≤3 months (HADS:A 0-61, STAI 0-21, final values, high is poor outcome)

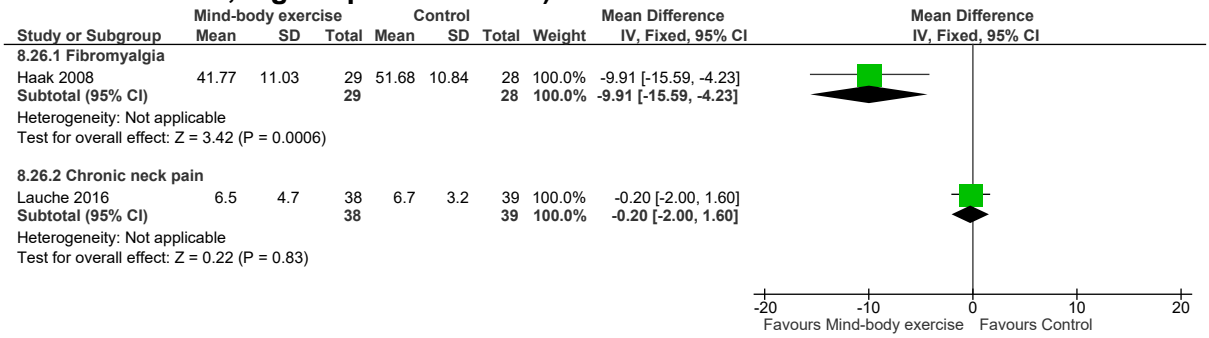


Figure 139: Psychological distress at >3 months (BDI, HADS:D, final values, high is poor outcome)

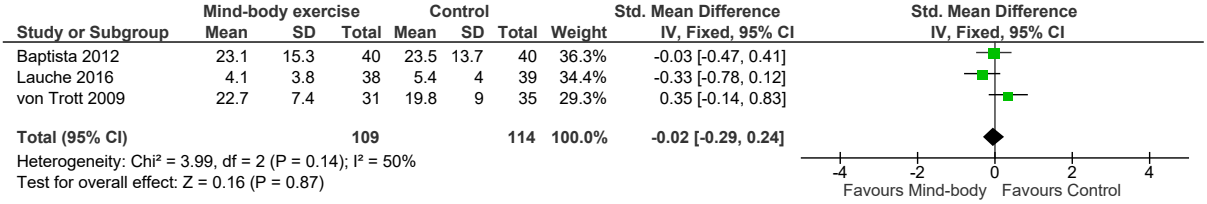


Figure 140: Psychological distress at >3 months (HADS:A, 0-21, final values, high is poor outcome)

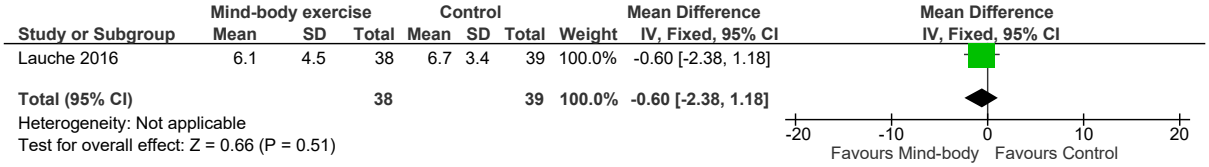


Figure 141: Sleep at ≤3 months (VAS sleep outcome, Pittsburgh sleep quality index, final values, high is poor outcome)

Study or Subgroup	Mind-body exercise Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Holmer 2004	9.06	5.23	12	13.76	3.78	17	48.8%	-1.03 [-1.82, -0.24]
Wong 2018	7.8	1.24	17	7.6	1.5	14	51.2%	0.14 [-0.57, 0.85]
Total (95% CI)			**29**			**31**	**100.0%**	**-0.43 [-1.58, 0.72]**

Heterogeneity: Tau² = 0.54; Chi² = 4.69, df = 1 (P = 0.03); I² = 79%
Test for overall effect: Z = 0.73 (P = 0.46)

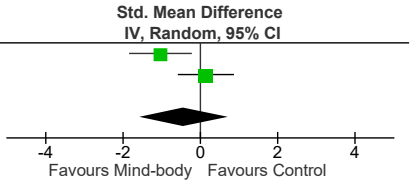
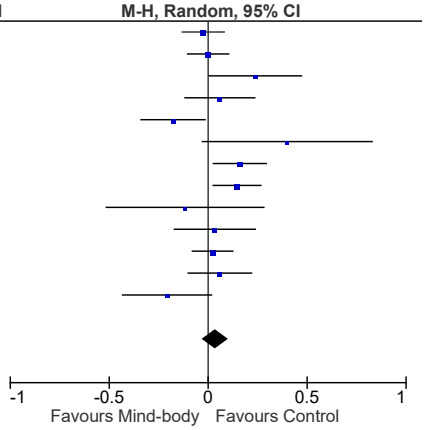


Figure 142: Discontinuation at >3 months

Study or Subgroup	Mind-body exercise Events	Total	Control Events	Total	Weight	Risk Difference M-H, Random, 95% CI
Baptista 2012	2	40	3	40	11.3%	-0.02 [-0.13, 0.08]
Bojner-Horwitz 2003	0	20	0	16	11.5%	0.00 [-0.10, 0.10]
Carson 2010	6	19	2	26	5.3%	0.24 [0.01, 0.47]
Carson 2012	3	22	2	26	7.5%	0.06 [-0.12, 0.24]
Lauche 2016	3	38	10	39	8.2%	-0.18 [-0.34, -0.02]
Liu 2012	2	5	0	7	2.1%	0.40 [-0.03, 0.83]
Lynch 2012	9	44	2	45	9.7%	0.16 [0.03, 0.29]
Lynch 2012	10	53	2	47	10.5%	0.15 [0.03, 0.27]
Mannerkorpi 2004	7	12	7	10	2.3%	-0.12 [-0.51, 0.28]
Michalsen 2012	12	38	11	39	6.3%	0.03 [-0.17, 0.24]
Rendant 2011	3	42	2	41	11.6%	0.02 [-0.08, 0.12]
von Trott 2009	7	38	5	40	8.2%	0.06 [-0.10, 0.22]
Wong 2018	1	18	5	19	5.6%	-0.21 [-0.43, 0.02]
Total (95% CI)		**389**		**395**	**100.0%**	**0.03 [-0.03, 0.10]**
Total events	65		51			

Heterogeneity: Tau² = 0.01; Chi² = 26.03, df = 12 (P = 0.01); I² = 54%
Test for overall effect: Z = 1.02 (P = 0.31)



Heterogeneity not explained by subgroup analysis

E.9 Flexibility versus usual care

Figure 143: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

Study or Subgroup	Flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Assumpcao 2017	46	26	16	64	27	12	100.0%	-18.00 [-37.89, 1.89]
Total (95% CI)			**16**			**12**	**100.0%**	**-18.00 [-37.89, 1.89]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.77 (P = 0.08)

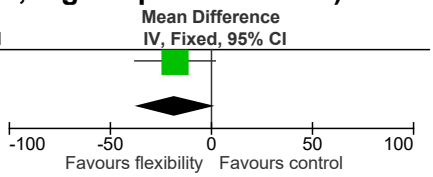


Figure 144: Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)

Study or Subgroup	Flexibility Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Assumpcao 2017	9	5.2	14	10.5	5.3	14	100.0%	-1.50 [-5.39, 2.39]
Total (95% CI)			**14**			**14**	**100.0%**	**-1.50 [-5.39, 2.39]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.76 (P = 0.45)

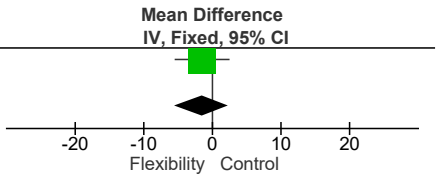
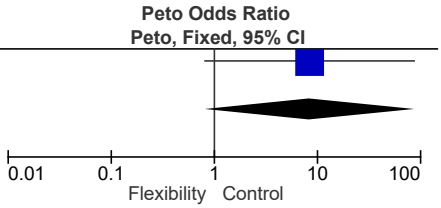


Figure 145: Discontinuation at ≤3 months

Study or Subgroup	Flexibility Events	Flexibility Total	Control Events	Control Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Assumpcao 2017	3	17	0	17	100.0%	8.41 [0.81, 86.84]
Total (95% CI)		**17**		**17**	**100.0%**	**8.41 [0.81, 86.84]**
Total events	3		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.79 (P = 0.07)

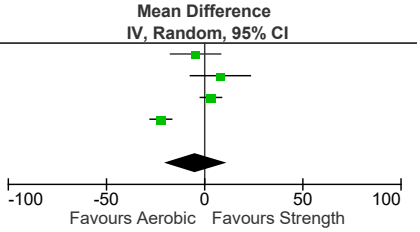


E.10 Aerobic exercise versus strength training

Figure 146: Pain at ≤3 months (VAS, FIQ pain subscale, MDPI, 0-100, final values and change scores, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Strength Mean	Strength SD	Strength Total	Weight	Mean Difference IV, Random, 95% CI
Bircan 2008	21.9	18.8	13	26.5	14.1	13	23.7%	-4.60 [-17.37, 8.17]
Ericsson 2016	-25	25.3	14	-33	13.4	12	22.3%	8.00 [-7.27, 23.27]
Hooten 2012	37.6	11.9	36	34.4	11.5	36	27.0%	3.20 [-2.21, 8.61]
Sevimli 2015	48.1	9.054	50	70.4	12.5	25	27.0%	-22.30 [-27.81, -16.79]
Total (95% CI)			**113**			**86**	**100.0%**	**-4.47 [-20.48, 11.54]**

Heterogeneity: Tau² = 239.05; Chi² = 46.90, df = 3 (P < 0.00001); I² = 94%
Test for overall effect: Z = 0.55 (P = 0.58)



Heterogeneity not explained by subgroup analysis

Figure 147: Pain at >3 months (VAS, 0-100, change scores, high is poor outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Strength Mean	Strength SD	Strength Total	Weight	Mean Difference IV, Fixed, 95% CI
Kayo 2011	-34.4	18.1	30	-27.7	19.5	30	100.0%	-6.70 [-16.22, 2.82]
Total (95% CI)			**30**			**30**	**100.0%**	**-6.70 [-16.22, 2.82]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.38 (P = 0.17)

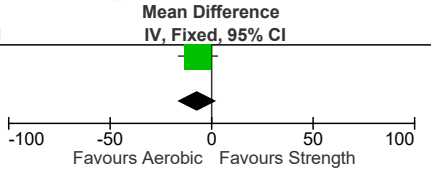
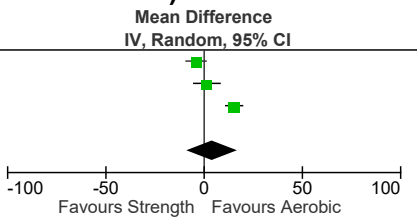


Figure 148: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values and change scores, high is good outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Strength Mean	Strength SD	Strength Total	Weight	Mean Difference IV, Random, 95% CI
Bircan 2008	38.92	6.11	13	43.01	7.02	13	33.6%	-4.09 [-9.15, 0.97]
Ericsson 2016	1.9	8.1	14	0.5	9.1	12	32.3%	1.40 [-5.27, 8.07]
Sevimli 2015	47.3	7.96	50	32.02	9.4	25	34.1%	15.28 [10.99, 19.57]
Total (95% CI)			**77**			**50**	**100.0%**	**4.29 [-8.40, 16.98]**

Heterogeneity: Tau² = 118.13; Chi² = 35.04, df = 2 (P < 0.00001); I² = 94%
Test for overall effect: Z = 0.66 (P = 0.51)

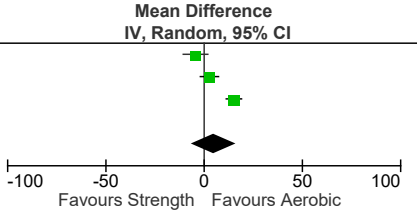


Heterogeneity not explained by subgroup analysis.

Figure 149: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values and change scores, high is good outcome)

Study or Subgroup	Aerobic Mean	Aerobic SD	Aerobic Total	Strength Mean	Strength SD	Strength Total	Weight	Mean Difference IV, Random, 95% CI
Bircan 2008	41.07	8.53	13	45.44	7.71	13	32.1%	-4.37 [-10.62, 1.88]
Ericsson 2016	4.9	6.2	14	2.2	5.8	12	33.6%	2.70 [-1.92, 7.32]
Sevimli 2015	51.95	7.4	50	36.8	8.4	25	34.2%	15.15 [11.27, 19.03]
Total (95% CI)			**77**			**50**	**100.0%**	**4.69 [-6.60, 15.97]**

Heterogeneity: Tau² = 92.95; Chi² = 32.93, df = 2 (P < 0.00001); I² = 94%
Test for overall effect: Z = 0.81 (P = 0.42)



Heterogeneity not explained by subgroup analysis.

Figure 150: Physical function at ≤3 months (multidimensional fatigue inventory reduced activity subscale, change scores, 0-20, high is poor outcome)

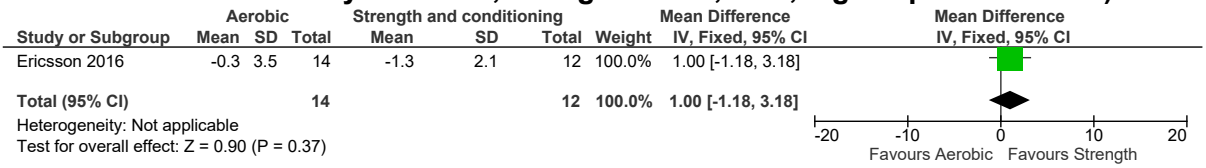


Figure 151: Physical function at ≤3 months (6 minute walking test, final values, metres)

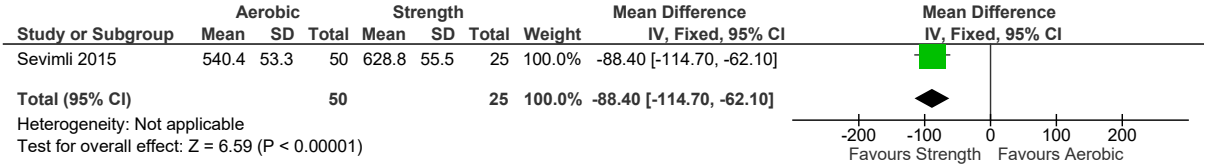


Figure 152: Physical function at >3 months (final values and change scores, SF-36 physical functioning subscale, 0-100, high is poor outcome)

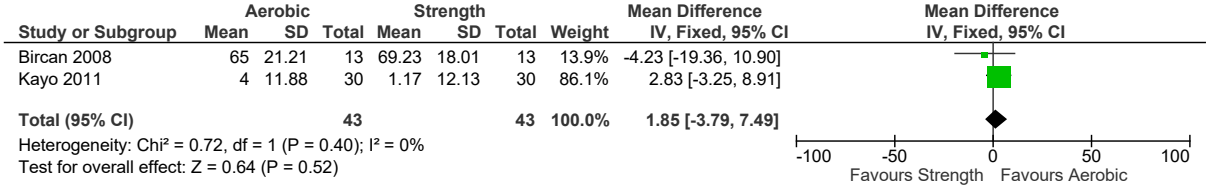


Figure 153: Psychological distress at ≤3 months (HADS: anxiety, 0-21, final values and change scores, high is poor outcome)

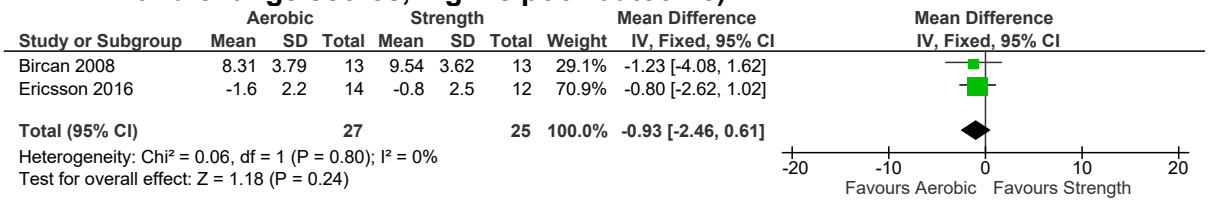


Figure 154: Psychological distress at ≤3 months (HADS: depression, 0-21, final values and change scores, high is poor outcome)

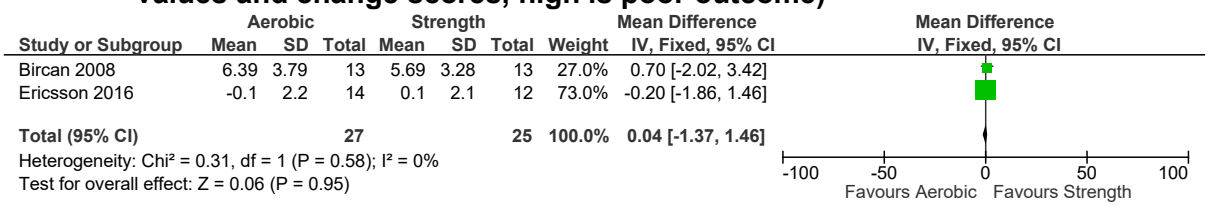


Figure 155: Psychological distress at ≤3 months (BDI, 0-60, final values, high is poor outcome)

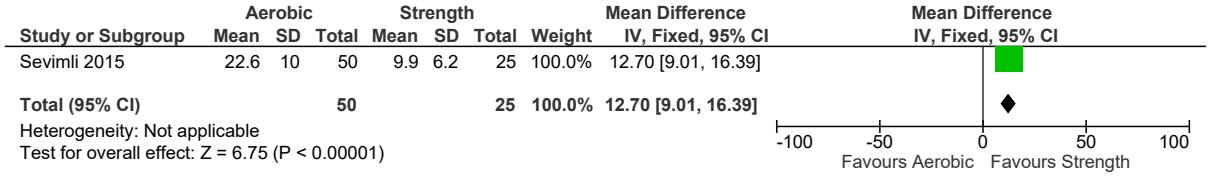


Figure 156: Sleep at ≤3 months (VAS sleep scale, 0-100, final values, high is poor outcome)

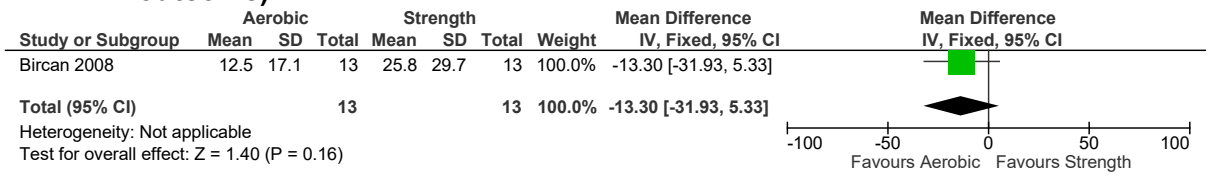
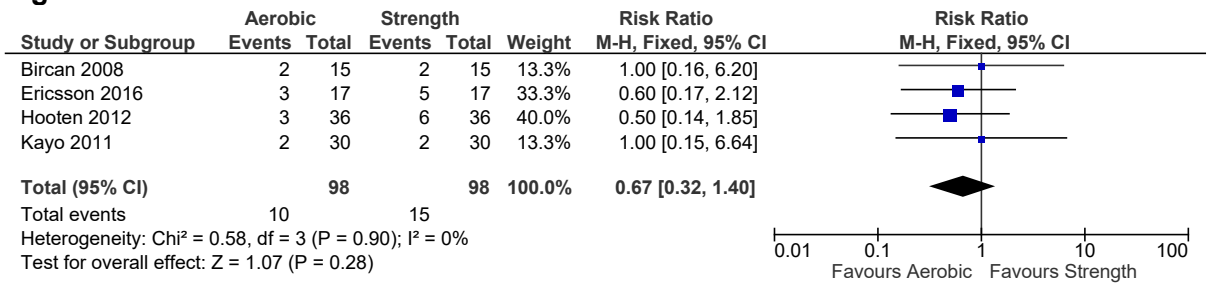


Figure 157: Discontinuation at >3 months



E.11 Aerobic versus flexibility

Figure 158: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

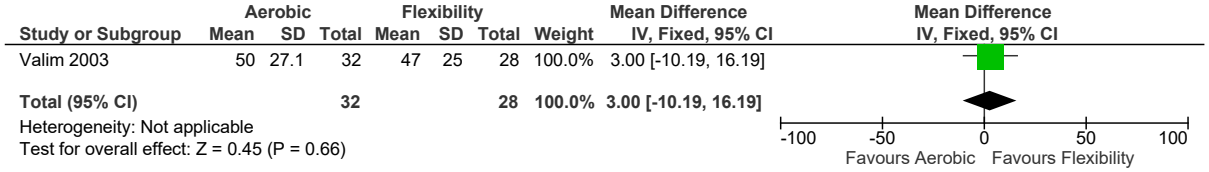


Figure 159: Pain at >3 months (VAS, 0-100, final values and change scores, high is poor outcome)

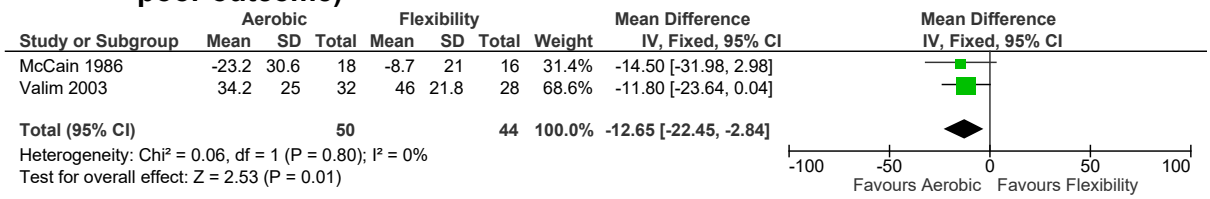


Figure 160: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

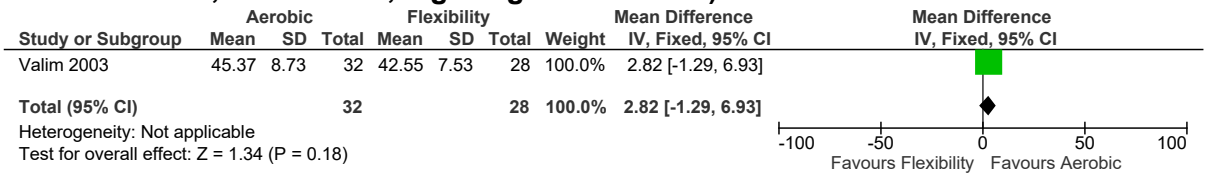


Figure 161: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

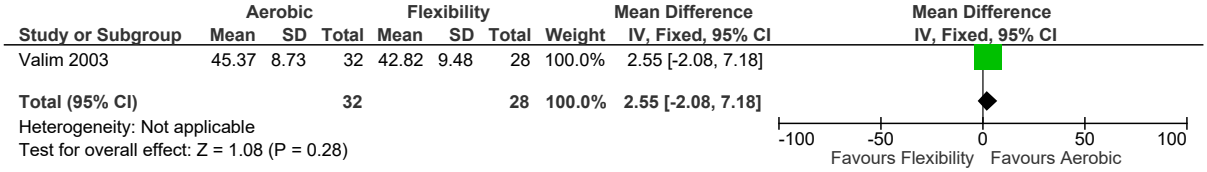


Figure 162: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

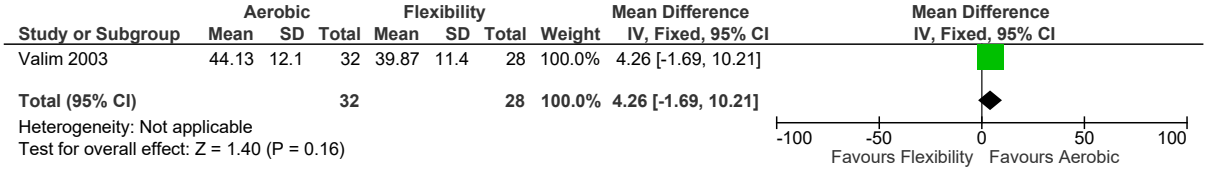


Figure 163: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

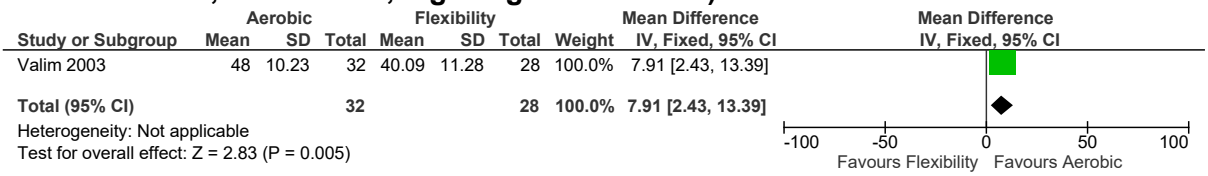


Figure 164: Psychological distress at ≤3 months (BDI, 0-21, final values, high is poor outcome)

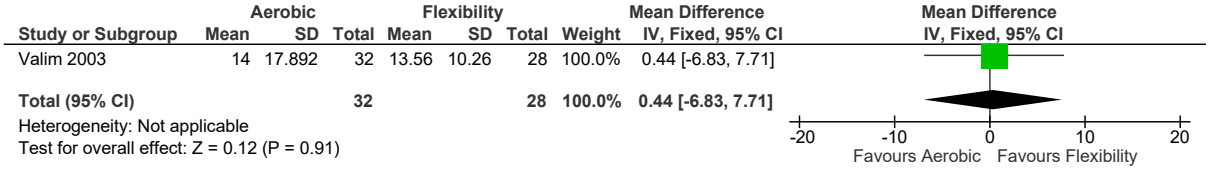


Figure 165: Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)

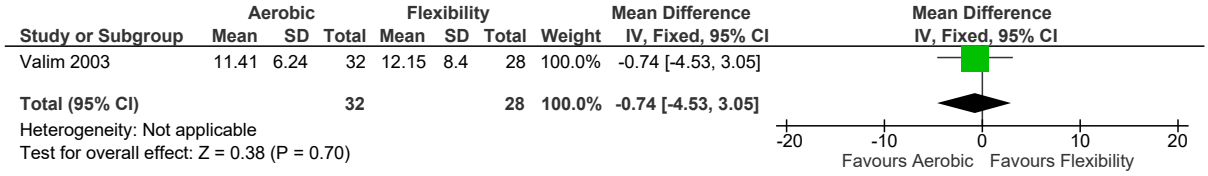


Figure 166: Psychological distress at ≤3 months (STAI anxiety, 0-100, final values, high is poor outcome)

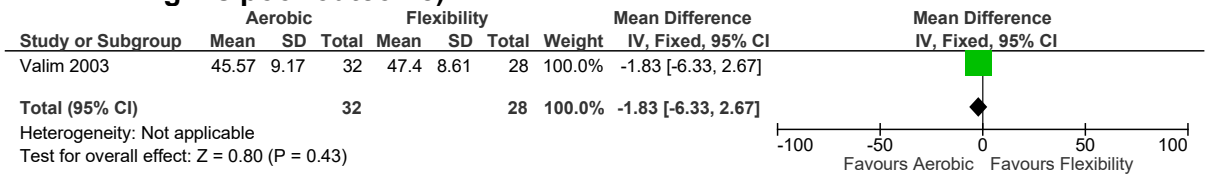


Figure 167: Psychological distress at >3 months (STAI anxiety, 0-100, final values, high is poor outcome)

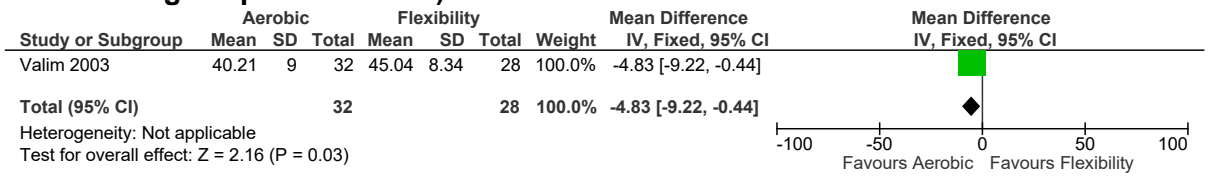
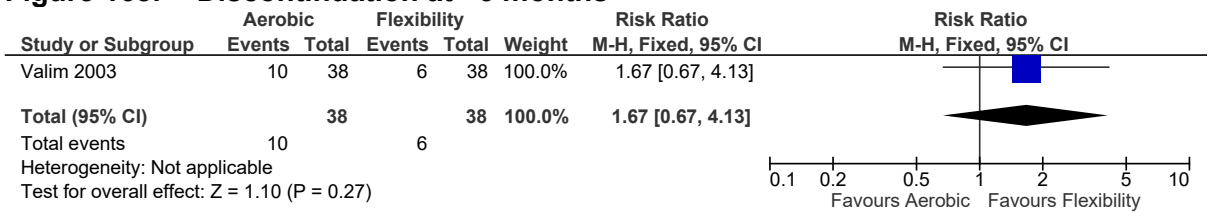


Figure 168: Discontinuation at >3 months



E.12 Aerobic exercise versus biomechanical exercise

Figure 169: Pain at ≤3 months (VAS, 0-10, high score is poor outcome)

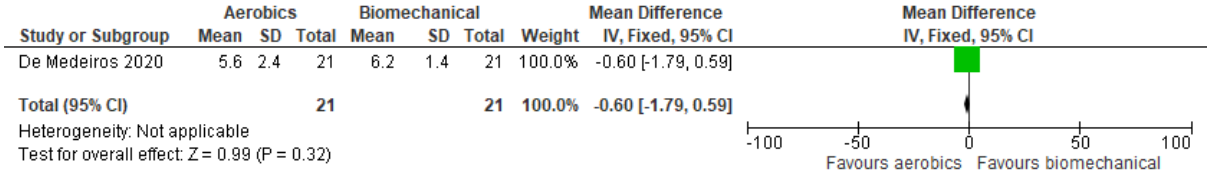


Figure 170: Quality of life at ≤3 months (SF36, 0-100, high score is good outcome)

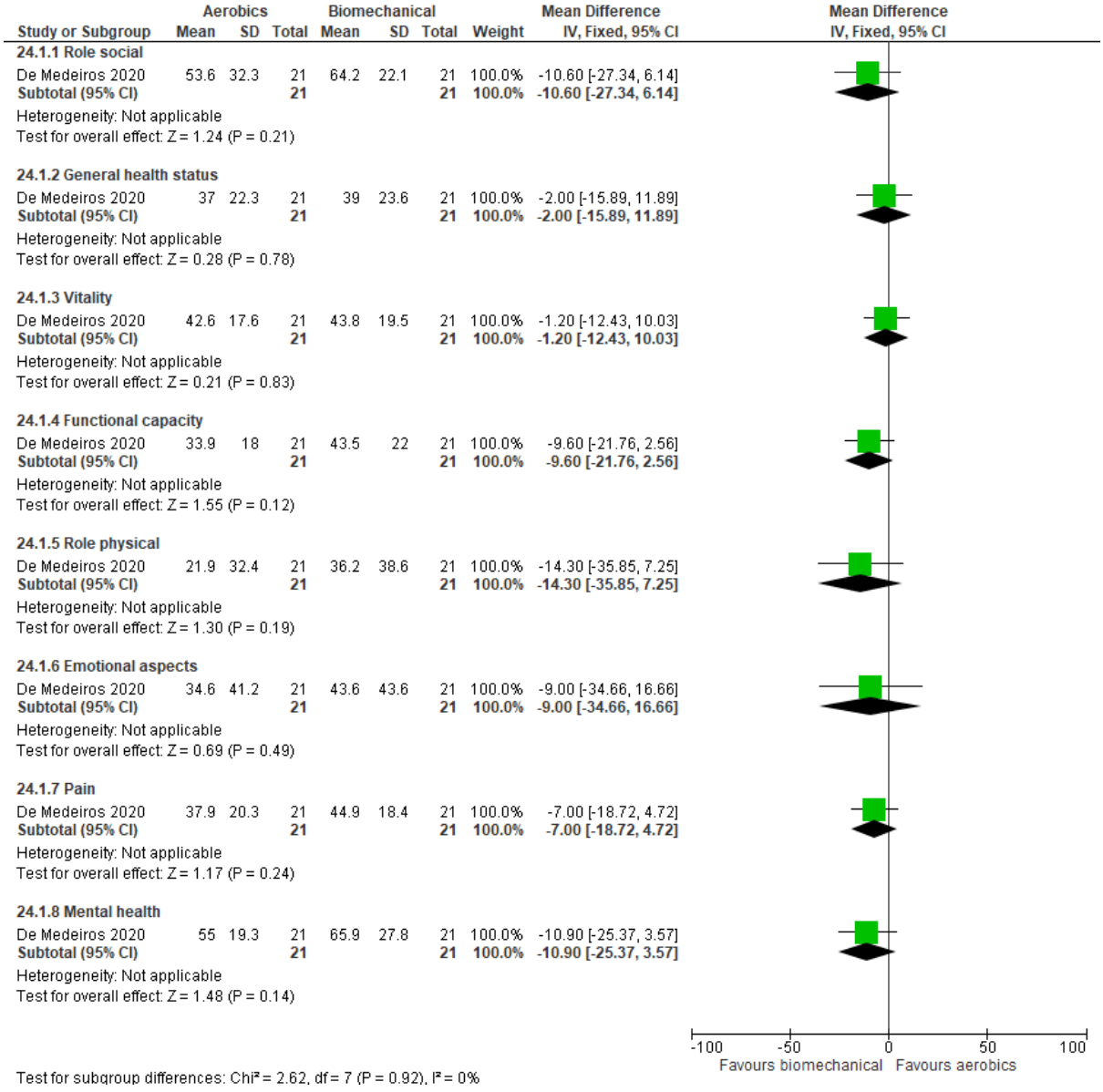


Figure 171: Psychological distress at ≤3 months (Scale of Catastropic Thoughts on Pain, 0-5, high score is poor outcome)

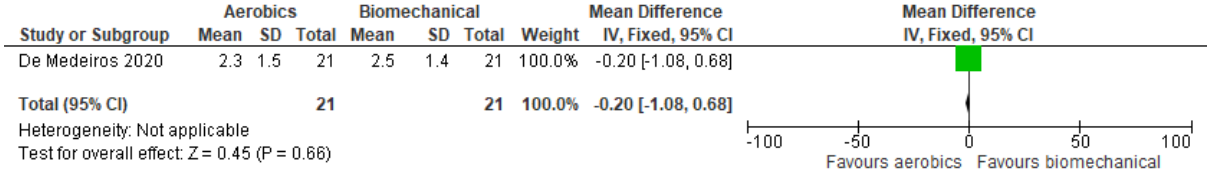


Figure 172: Sleep at ≤3 months (Pittsburgh Sleep Quality Index, 0-21, high score is poor outcome)

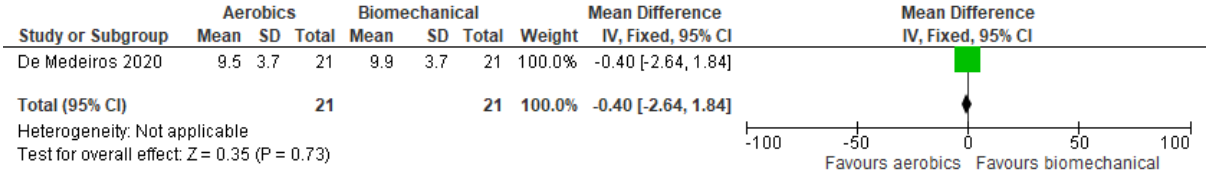
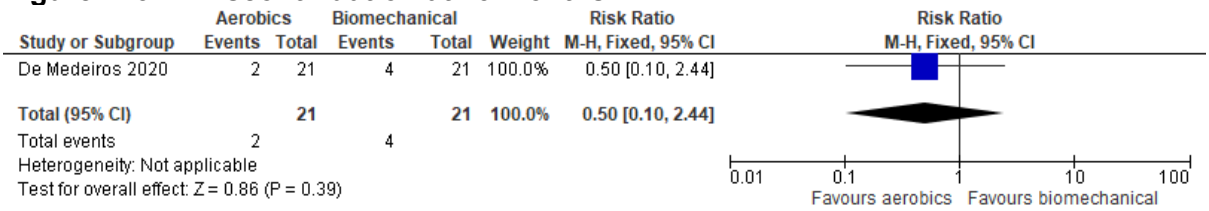


Figure 173: Discontinuation at ≤3 months



E.13 Aerobic and strength versus aerobic

Figure 174: Quality of life at >3 months (FIQ, 0-100, change scores, high is poor outcome)

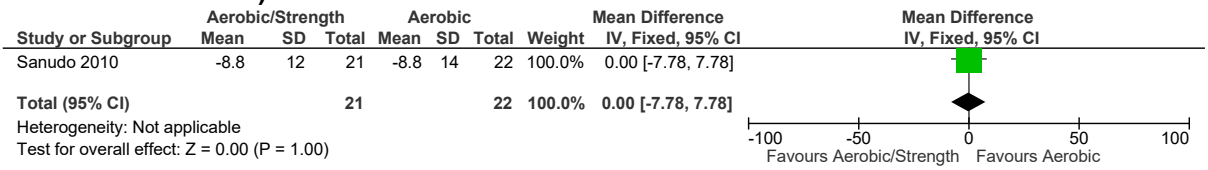


Figure 175: Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)

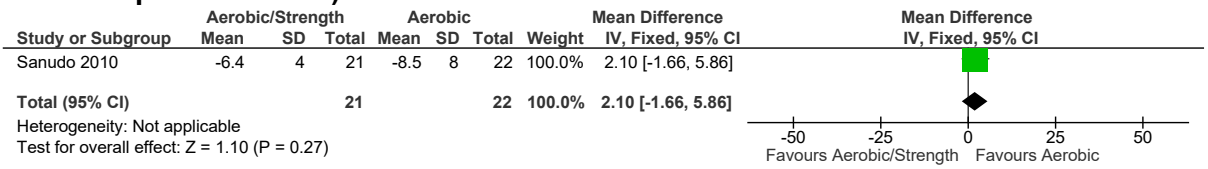
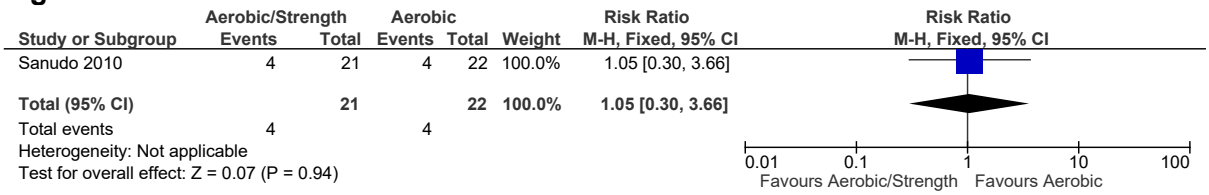


Figure 176: Discontinuation at >3 months



E.14 Aerobic and strength versus flexibility

Figure 177: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

Study or Subgroup	Aerobic and strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Giubilei 2007	43	14	41	47	14	44	100.0%	-4.00 [-9.96, 1.96]
Total (95% CI)			**41**			**44**	**100.0%**	**-4.00 [-9.96, 1.96]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.32 (P = 0.19)

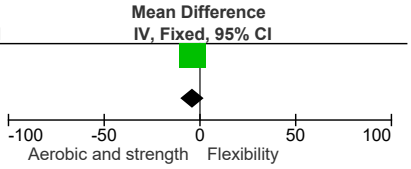


Figure 178: Pain at >3 months (VAS, 0-100, final values, high is poor outcome)

Study or Subgroup	Aerobic and strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Giubilei 2007	34	14	36	42	12	40	100.0%	-8.00 [-13.89, -2.11]
Total (95% CI)			**36**			**40**	**100.0%**	**-8.00 [-13.89, -2.11]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.66 (P = 0.008)

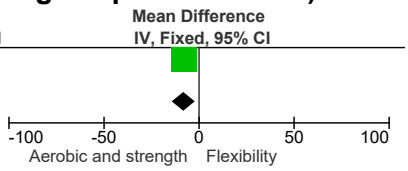


Figure 179: Quality of life at ≤3 months (NIS CPSI quality of life subscale 0-12, final values, high is poor outcome)

Study or Subgroup	Aerobic and strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Giubilei 2007	5.1	2.1	41	6.9	2.1	44	100.0%	-1.80 [-2.69, -0.91]
Total (95% CI)			**41**			**44**	**100.0%**	**-1.80 [-2.69, -0.91]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.95 (P < 0.0001)

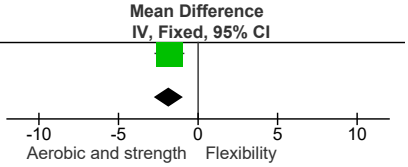


Figure 180: Quality of life at >3 months (NIS CPSI quality of life subscale 0-12, final values, high is poor outcome)

Study or Subgroup	Aerobic and strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Giubilei 2007	4.4	1.8	36	6.2	2.1	40	100.0%	-1.80 [-2.68, -0.92]
Total (95% CI)			**36**			**40**	**100.0%**	**-1.80 [-2.68, -0.92]**

Heterogeneity: Not applicable
Test for overall effect: Z = 4.02 (P < 0.0001)

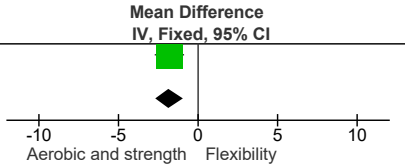


Figure 181: Psychological distress at ≤3 months (BDI, 0-21, final values, high is poor outcome)

Study or Subgroup	Aerobic and strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Giubilei 2007	9.8	4.3	41	9.3	4.3	44	100.0%	0.50 [-1.33, 2.33]
Total (95% CI)			**41**			**44**	**100.0%**	**0.50 [-1.33, 2.33]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.54 (P = 0.59)

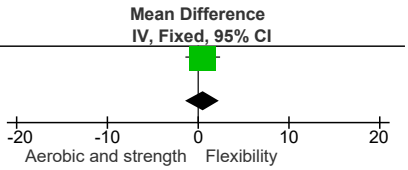


Figure 182: Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)

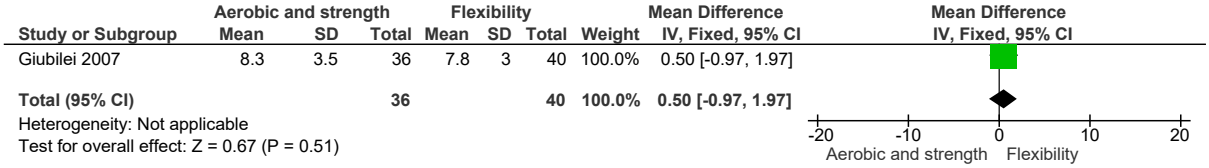
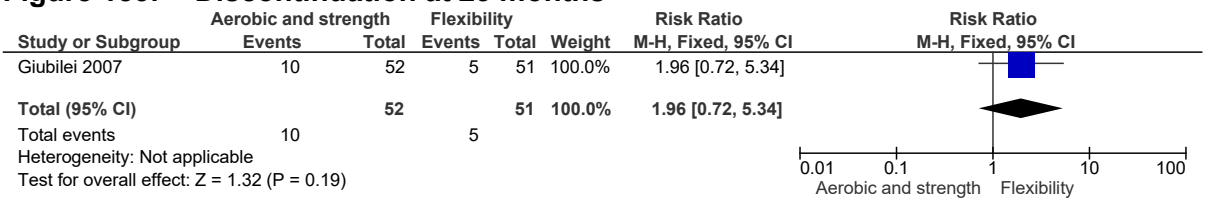


Figure 183: Discontinuation at ≤3 months



E.15 Aerobic and flexibility versus mind-body exercise

Figure 184: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)

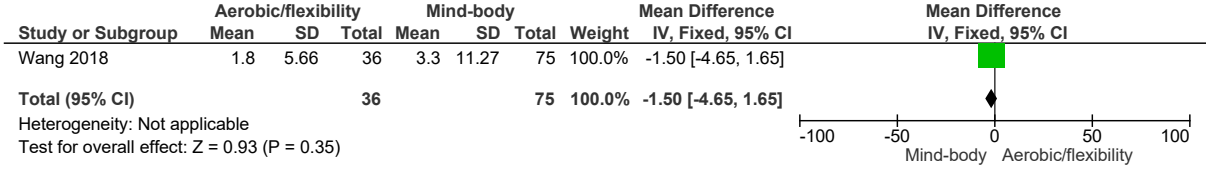


Figure 185: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)

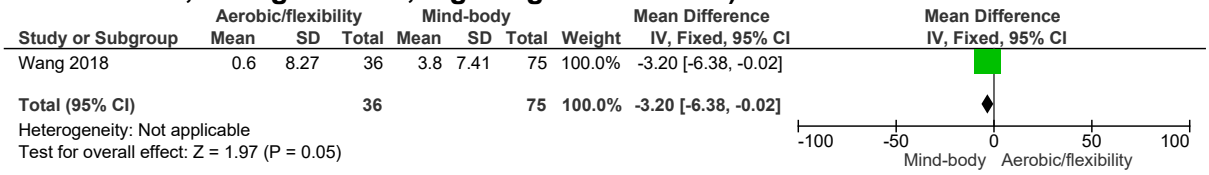


Figure 186: Quality of life at >3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)

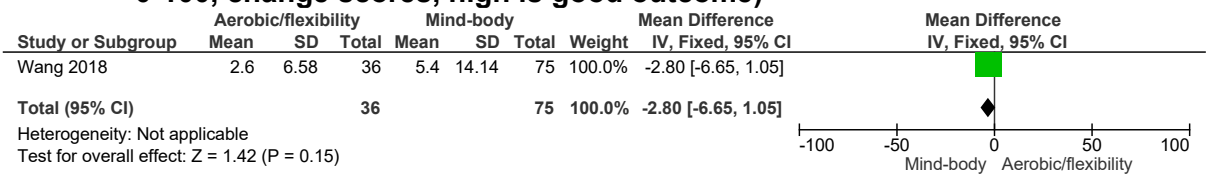


Figure 187: Quality of life at >3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)

Study or Subgroup	Aerobic/flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Wang 2018	3	9.34	36	5.4	20.1	75	100.0%	-2.40 [-7.88, 3.08]
Total (95% CI)			**36**			**75**	**100.0%**	**-2.40 [-7.88, 3.08]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.86 (P = 0.39)

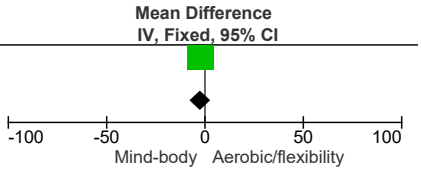


Figure 188: Physical function at ≤3 months (6 minute walking test, change scores, metres)

Study or Subgroup	Aerobic/flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Wang 2018	9.3	47.3	36	7.4	98.1	75	100.0%	1.90 [-25.15, 28.95]
Total (95% CI)			**36**			**75**	**100.0%**	**1.90 [-25.15, 28.95]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.14 (P = 0.89)

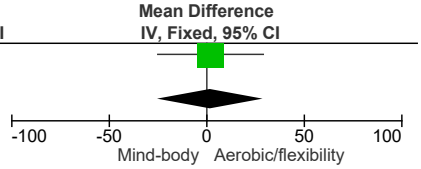


Figure 189: Physical function at >3 months (6 minute walking test, change scores, metres)

Study or Subgroup	Aerobic/flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Wang 2018	8	65.36	36	30.2	140.28	75	100.0%	-22.20 [-60.46, 16.06]
Total (95% CI)			**36**			**75**	**100.0%**	**-22.20 [-60.46, 16.06]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.14 (P = 0.26)

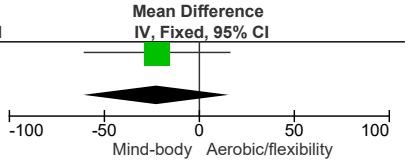


Figure 190: Psychological distress at ≤3 months (HADS depression, 0-21, change scores, high is poor outcome)

Study or Subgroup	Aerobic/flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Wang 2018	-0.5	2.45	36	-1.7	7.51	75	100.0%	1.20 [-0.68, 3.08]
Total (95% CI)			**36**			**75**	**100.0%**	**1.20 [-0.68, 3.08]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.25 (P = 0.21)

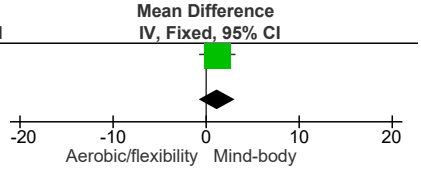


Figure 191: Psychological distress at ≤3 months (HADS anxiety, 0-21, change scores, high is poor outcome)

Study or Subgroup	Aerobic/flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Wang 2018	0.2	2.45	36	-1.6	5.08	75	100.0%	1.80 [0.40, 3.20]
Total (95% CI)			**36**			**75**	**100.0%**	**1.80 [0.40, 3.20]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.52 (P = 0.01)

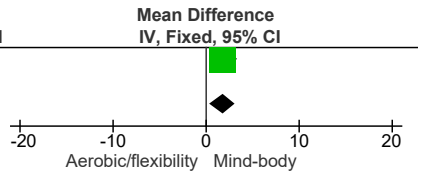


Figure 192: Psychological distress at >3 months (HADS anxiety, 0-21, change scores high is poor outcome)

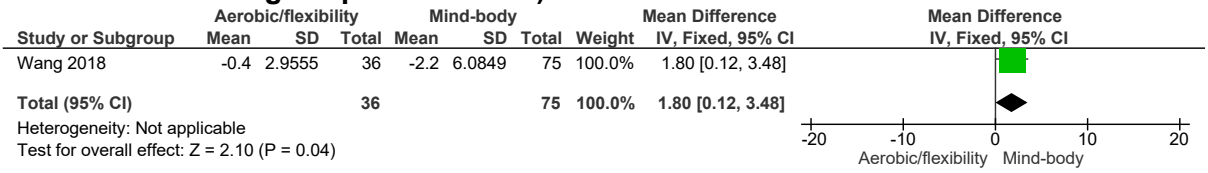


Figure 193: Psychological distress at >3 months (HADS depression, 0-21, change scores, high is poor outcome)

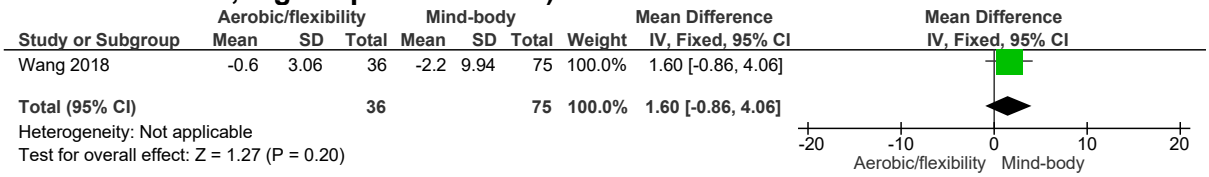


Figure 194: Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)

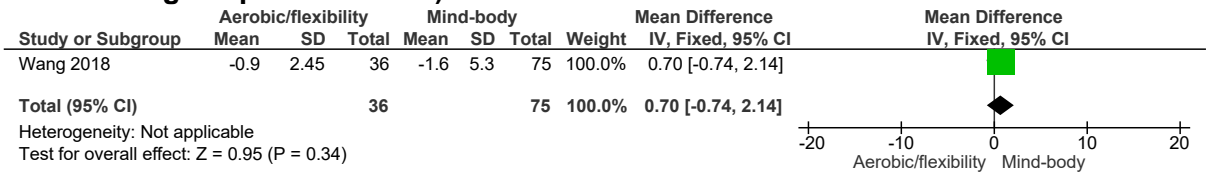


Figure 195: Sleep at >3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)

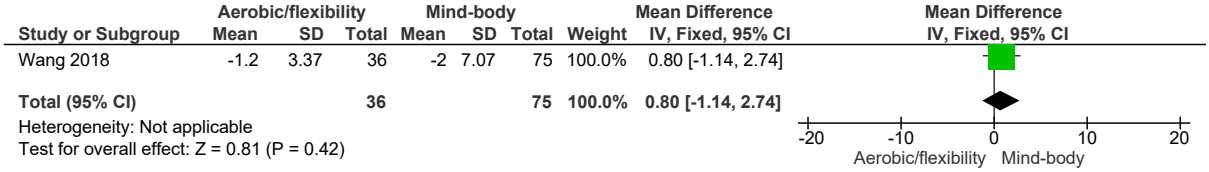
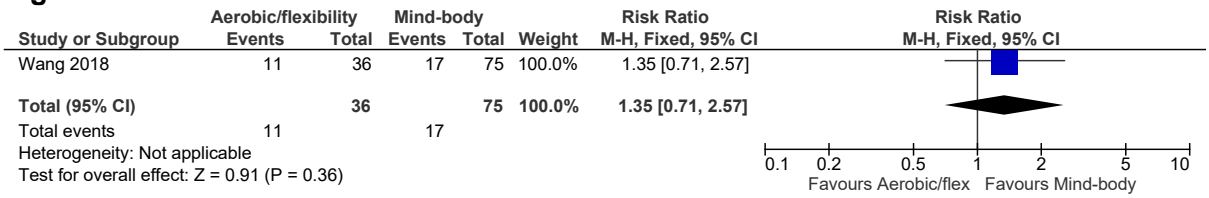


Figure 196: Discontinuation at ≤3 months



E.16 Aerobic exercise and flexibility versus aerobic exercise

Figure 197: Pain at 4 weeks (VAS, 0-100, high is poor outcome)

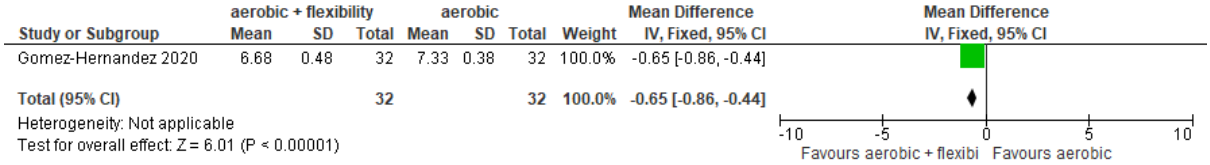


Figure 198: Pain at 12 weeks (VAS, 0-100, high is poor outcome)

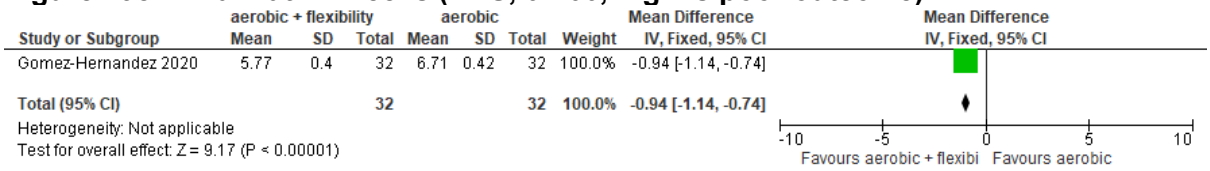


Figure 199: Quality of life at 4 weeks (FIQ, 0-100, high is poor outcome)

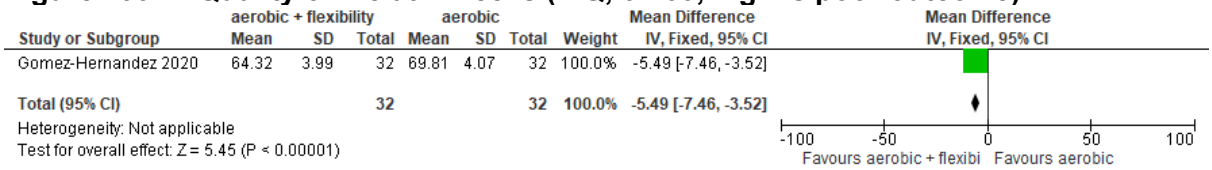


Figure 200: Quality of life at 12 weeks (FIQ, 0-100, high is poor outcome)

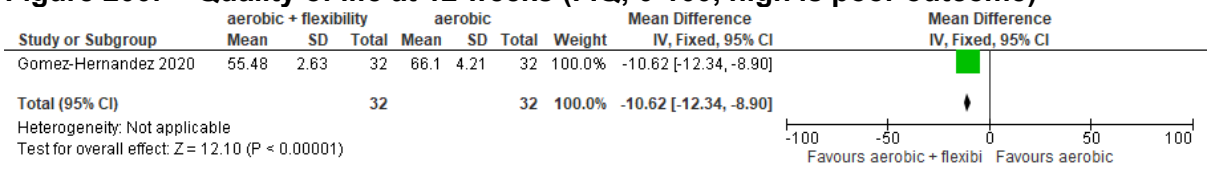


Figure 201: Sleep quality at 4 weeks (final score; Pittsburgh Sleep Quality Index)

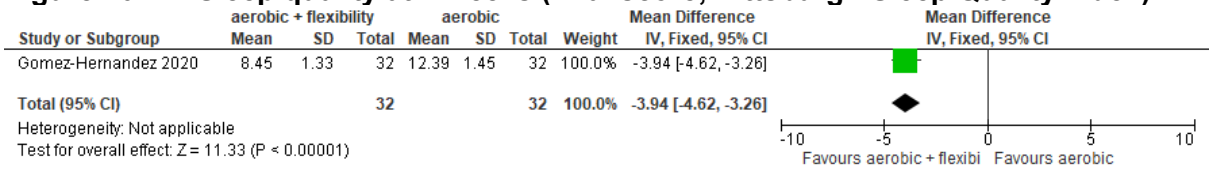


Figure 202: Sleep quality at 12 weeks (final score; Pittsburgh Sleep Quality Index)

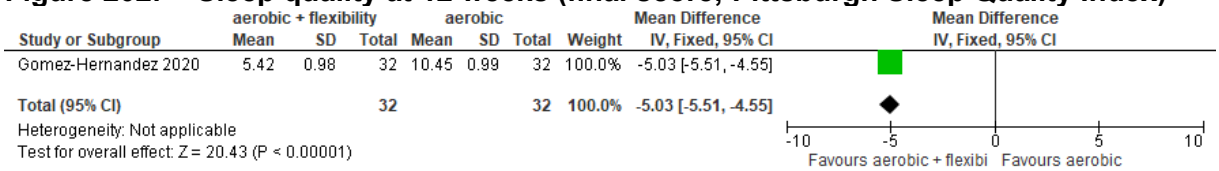
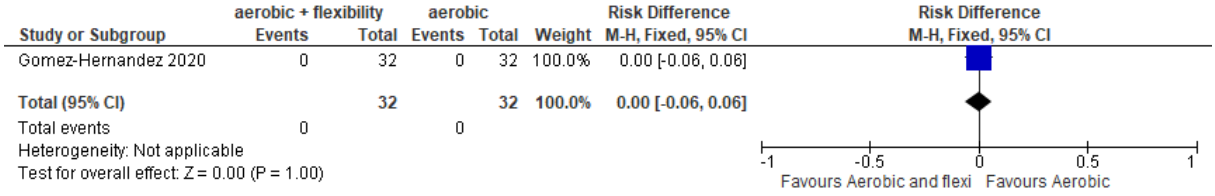


Figure 203: Discontinuation at 12 weeks



E.17 Aerobic, strength, mind-body and proprioception versus flexibility

Figure 204: Quality of life at ≤3 months (FIQ total score, high is poor outcome)

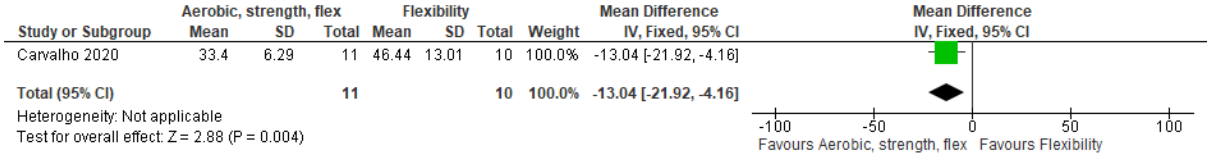


Figure 205: Physical function at ≤3 months (number of steps, high is good outcome)

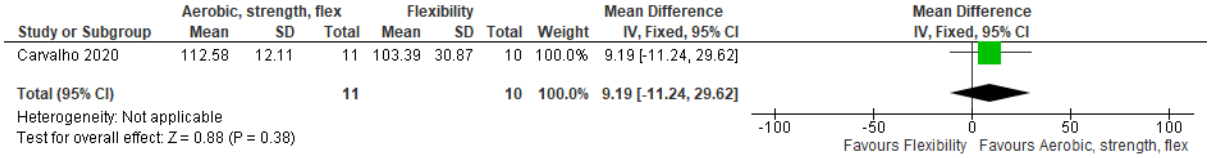
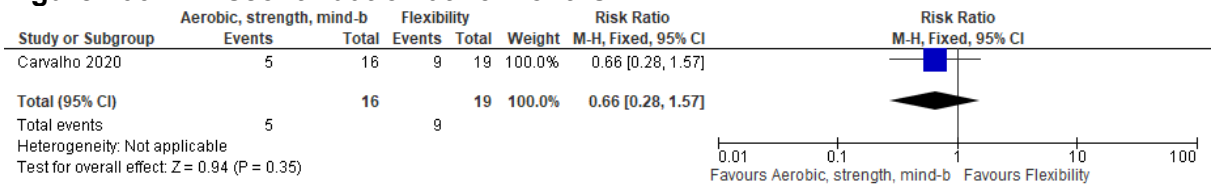


Figure 206: Discontinuation at ≤3 months



E.18 Strength training versus mind-body exercise

Figure 207: Pain at ≤3 months (VAS, 0-10, final values, high is poor outcome)

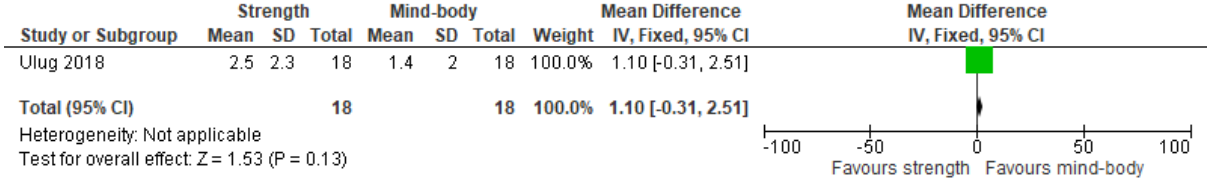


Figure 208: Quality of life at ≤3 months (Nottingham Health Profile, 0-600, final values, high is poor outcome)

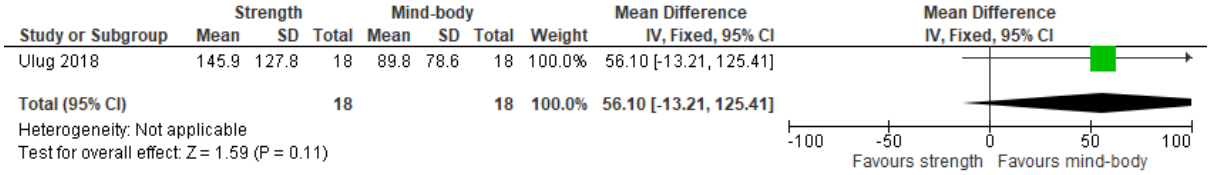


Figure 209: Physical function at ≤3 months (Neck Disability Index, 0-100, final values, high is poor outcome)

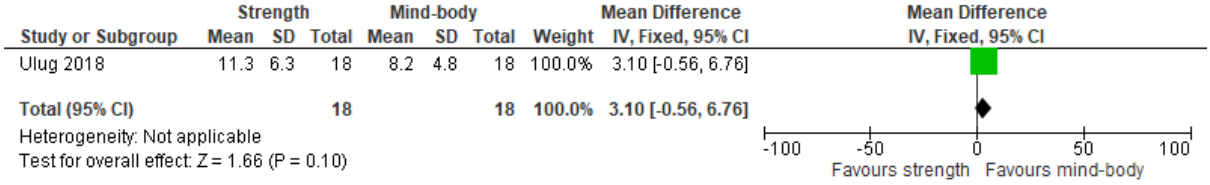


Figure 210: Psychological distress at ≤3 months (Beck Depression Inventory, 0-63, final values, high is poor outcome)

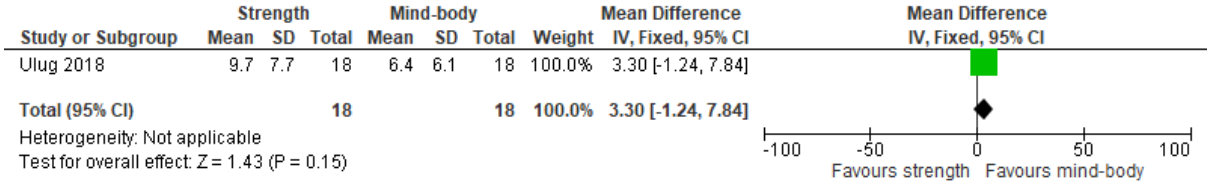
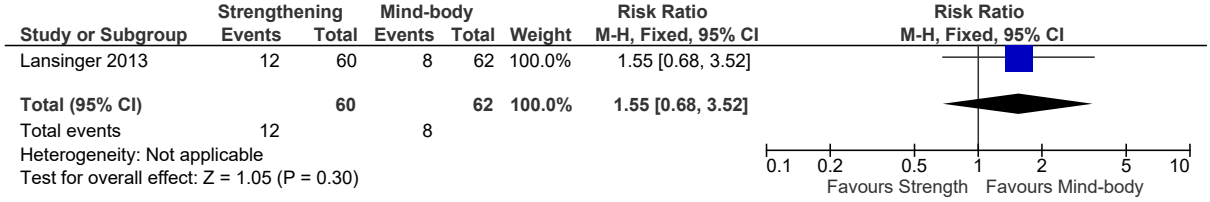


Figure 211: Discontinuation at <3 months



E.19 Strength training versus biomechanical exercise

Figure 212: Pain at ≤3 months (VAS, 0-10, final values, high is poor outcome)

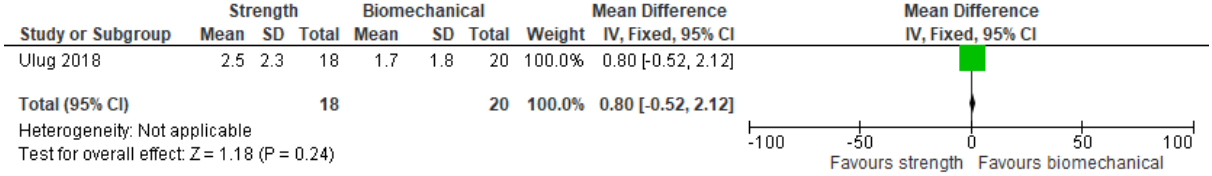


Figure 213: Quality of life at ≤3 months (Nottingham Health Profile, 0-600, final values, high is poor outcome)

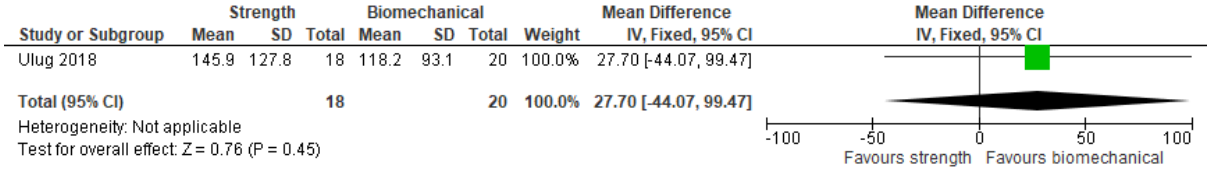


Figure 214: Physical function at ≤3 months (Neck Disability Index, 0-100, final values, high is poor outcome)

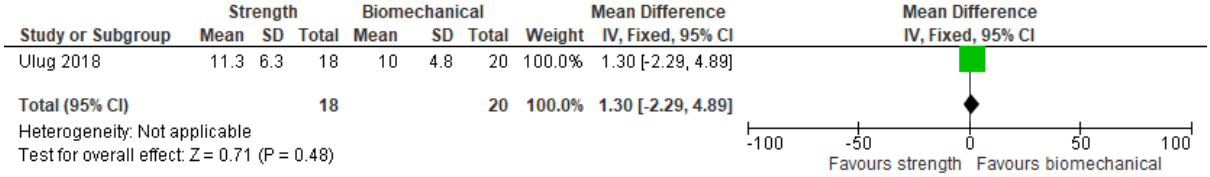
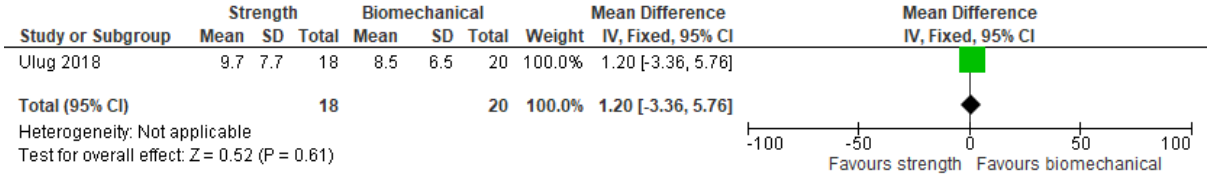


Figure 215: Psychological distress at ≤3 months (Beck Depression Inventory, 0-63, final values, high is poor outcome)



E.20 Strength training versus flexibility

Figure 216: Pain at ≤3 months (VAS, 0-100, change scores and final values, high is poor outcome)

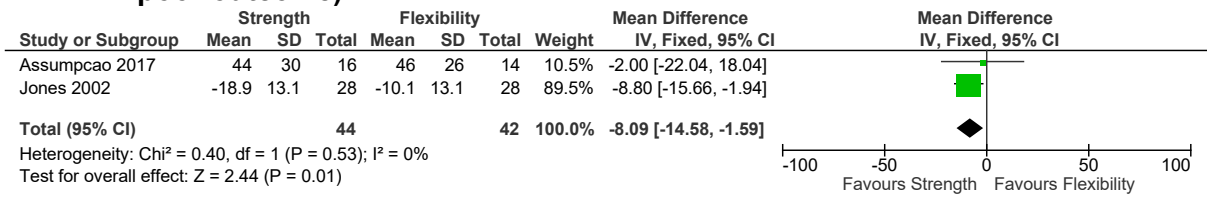


Figure 217: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

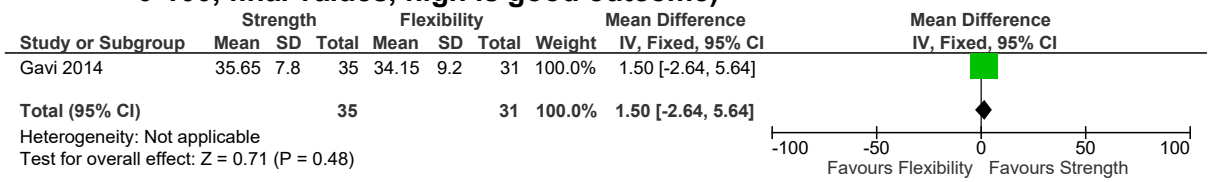


Figure 218: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Gavi 2014	39.16	12.64	35	44.55	13.6	31	100.0%	-5.39 [-11.75, 0.97]
Total (95% CI)			**35**			**31**	**100.0%**	**-5.39 [-11.75, 0.97]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.66 (P = 0.10)

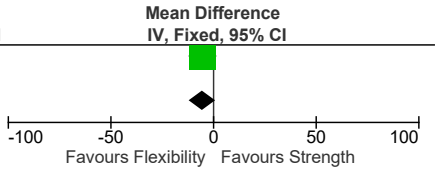


Figure 219: Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)

Study or Subgroup	Strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Assumpcao 2017	15.5	5	16	9.5	5.2	14	100.0%	6.00 [2.34, 9.66]
Total (95% CI)			**16**			**14**	**100.0%**	**6.00 [2.34, 9.66]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.21 (P = 0.001)

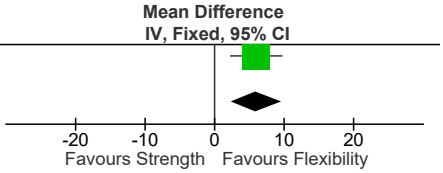


Figure 220: Psychological distress at ≤3 months (BDI, 0-61, change scores, high is poor outcome)

Study or Subgroup	Strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Jones 2002	-3.67	4.23	28	-1.84	4.03	28	100.0%	-1.83 [-3.99, 0.33]
Total (95% CI)			**28**			**28**	**100.0%**	**-1.83 [-3.99, 0.33]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.66 (P = 0.10)

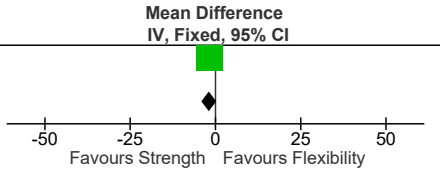


Figure 221: Psychological distress at ≤3 months (BAI, 0-61, change scores, high is poor outcome)

Study or Subgroup	Strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Jones 2002	-2.5	5.84	28	0.7	6.45	28	100.0%	-3.20 [-6.42, 0.02]
Total (95% CI)			**28**			**28**	**100.0%**	**-3.20 [-6.42, 0.02]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.95 (P = 0.05)

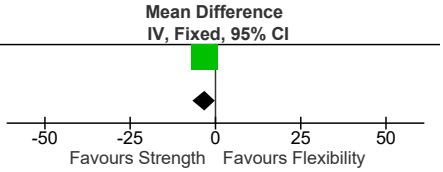


Figure 222: Sleep at ≤3 months (FIQ sleep subscale, 0-10, change scores, high is poor outcome)

Study or Subgroup	Strength Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Jones 2002	-2.3	1.65	28	-0.53	1.61	28	100.0%	-1.77 [-2.62, -0.92]
Total (95% CI)			**28**			**28**	**100.0%**	**-1.77 [-2.62, -0.92]**

Heterogeneity: Not applicable
Test for overall effect: Z = 4.06 (P < 0.0001)

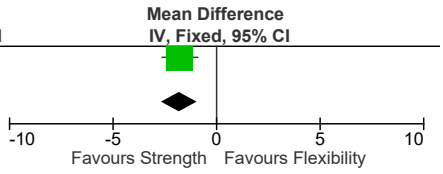
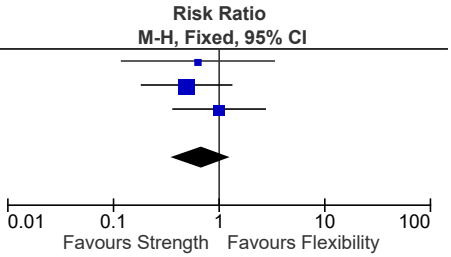


Figure 223: Discontinuation at >3 months

Study or Subgroup	Strength Events	Strength Total	Flexibility Events	Flexibility Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Assumpcao 2017	2	18	3	17	16.6%	0.63 [0.12, 3.32]
Gavi 2014	5	35	9	31	51.2%	0.49 [0.18, 1.31]
Jones 2002	6	28	6	28	32.2%	1.00 [0.37, 2.73]
Total (95% CI)		**81**		**76**	**100.0%**	**0.68 [0.36, 1.28]**
Total events	13		18			

Heterogeneity: Chi² = 0.99, df = 2 (P = 0.61); I² = 0%
Test for overall effect: Z = 1.19 (P = 0.23)



E.21 Strength and flexibility versus flexibility

Figure 224: Quality of life at >3 months (SF-36 physical functioning subscale, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Salo 2012	92	11.5	43	92.4	9.8	43	100.0%	-0.40 [-4.92, 4.12]
Total (95% CI)			**43**			**43**	**100.0%**	**-0.40 [-4.92, 4.12]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.17 (P = 0.86)

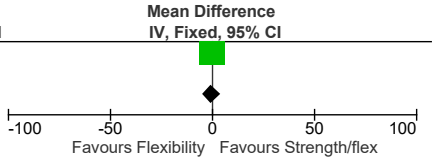


Figure 225: Quality of life at >3 months (SF-36 physical role subscale, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Salo 2012	78.3	36.1	43	79.4	33.9	43	100.0%	-1.10 [-15.90, 13.70]
Total (95% CI)			**43**			**43**	**100.0%**	**-1.10 [-15.90, 13.70]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.15 (P = 0.88)

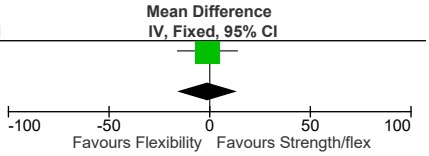


Figure 226: Quality of life at >3 months (SF-36 emotional subscale, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Salo 2012	89.1	23.8	43	87	31.5	43	100.0%	2.10 [-9.70, 13.90]
Total (95% CI)			**43**			**43**	**100.0%**	**2.10 [-9.70, 13.90]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.35 (P = 0.73)

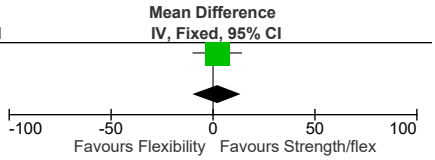


Figure 227: Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Flexibility Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Salo 2012	68.6	16.7	43	63.4	21.6	43	100.0%	5.20 [-2.96, 13.36]
Total (95% CI)			**43**			**43**	**100.0%**	**5.20 [-2.96, 13.36]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.25 (P = 0.21)

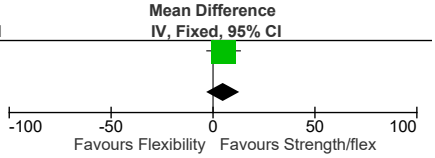


Figure 228: Quality of life at >3 months (SF-36 emotional wellbeing subscale, 0-100, final values, high is good outcome)

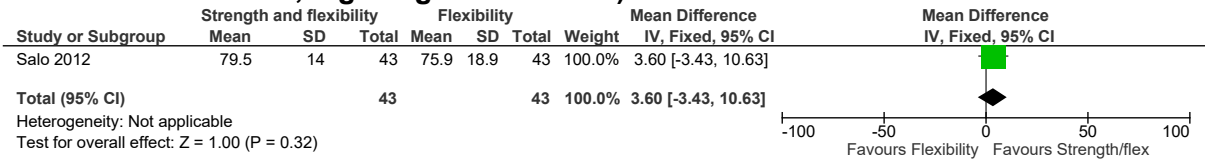


Figure 229: Quality of life at >3 months (SF-36 social functioning subscale, 0-100, final values, high is good outcome)

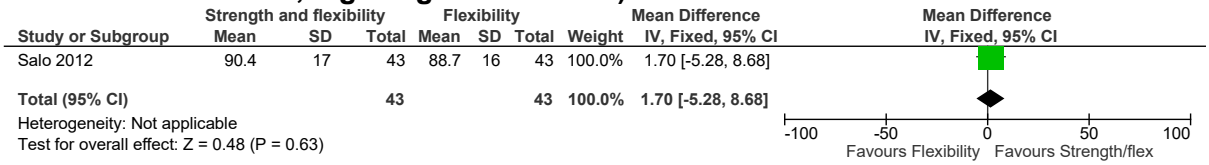


Figure 230: Quality of life at >3 months (SF-36 bodily pain subscale, 0-100, final values, high is good outcome)

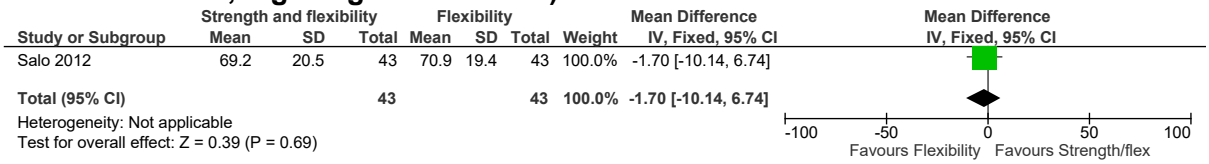


Figure 231: Quality of life at >3 months (SF-36 general health subscale, 0-100, final values, high is good outcome)

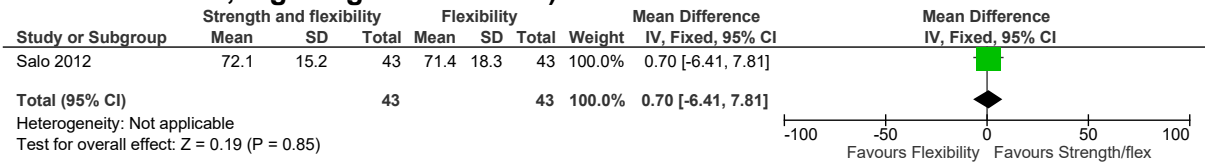
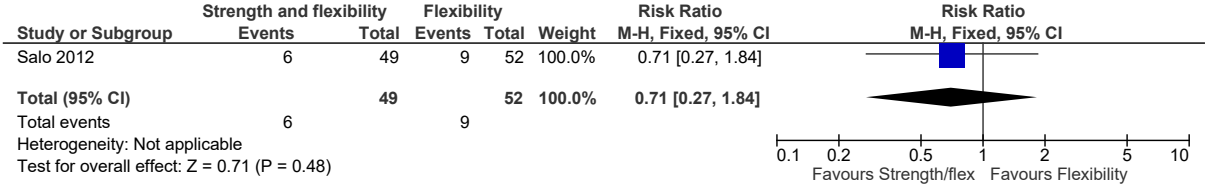


Figure 232: Discontinuation at >3 months



E.22 Strength and flexibility versus mind-body

Figure 233: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

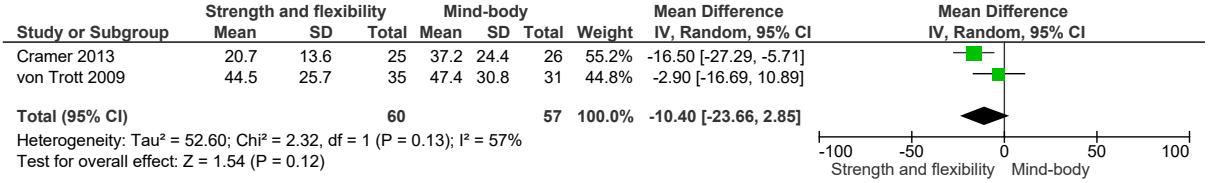


Figure 234: Pain at >3 months (VAS, 0-100, final values, high is poor outcome)

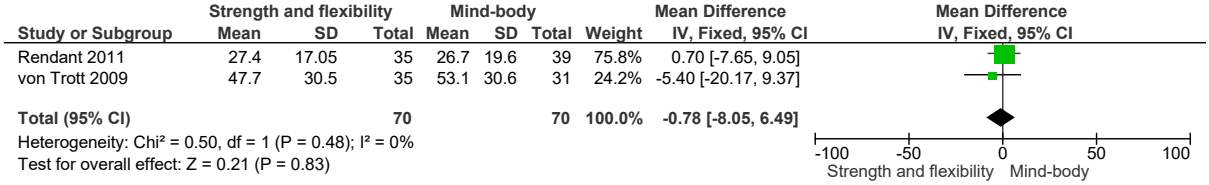


Figure 235: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

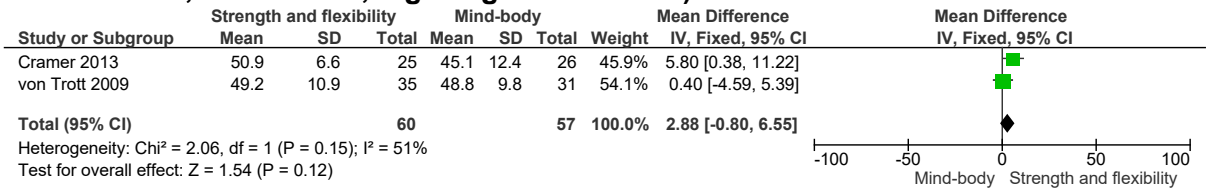


Figure 236: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

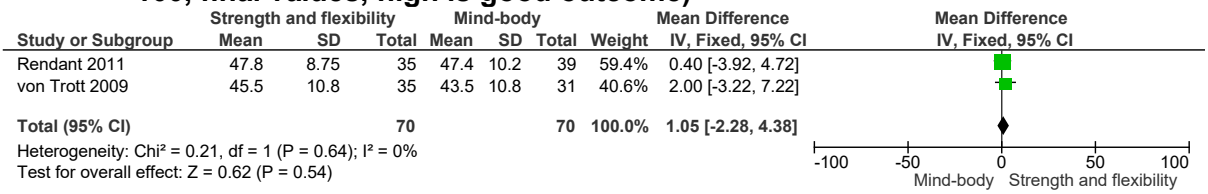


Figure 237: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

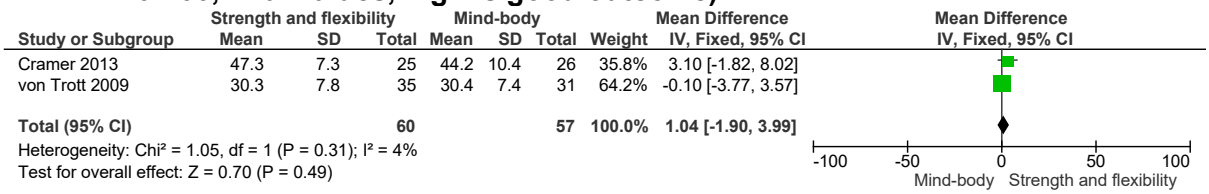


Figure 238: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Rendant 2011	44.7	7.55	35	47	7.65	39	56.0%	-2.30 [-5.77, 1.17]
von Trott 2009	29.3	8.5	35	31.4	7.7	31	44.0%	-2.10 [-6.01, 1.81]
Total (95% CI)			**70**			**70**	**100.0%**	**-2.21 [-4.81, 0.38]**

Heterogeneity: Chi² = 0.01, df = 1 (P = 0.94); I² = 0%
Test for overall effect: Z = 1.67 (P = 0.09)

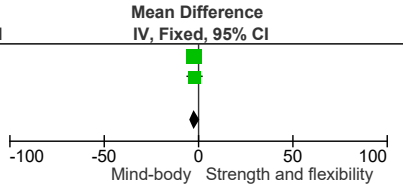


Figure 239: Physical function at ≤3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Cramer 2013	20	9.8	25	26.2	15	26	42.9%	-0.48 [-1.04, 0.08]
von Trott 2009	33.6	25.5	35	34.3	23.6	31	57.1%	-0.03 [-0.51, 0.46]
Total (95% CI)			**60**			**57**	**100.0%**	**-0.22 [-0.59, 0.14]**

Heterogeneity: Chi² = 1.44, df = 1 (P = 0.23); I² = 31%
Test for overall effect: Z = 1.19 (P = 0.23)

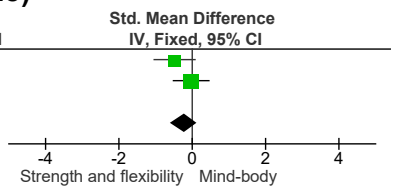


Figure 240: Physical function at >3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Rendant 2011	31.5	14.49	35	30	10.36	39	81.7%	1.50 [-4.30, 7.30]
von Trott 2009	34.3	24.8	35	39.8	25.8	31	18.3%	-5.50 [-17.75, 6.75]
Total (95% CI)			**70**			**70**	**100.0%**	**0.22 [-5.02, 5.46]**

Heterogeneity: Chi² = 1.03, df = 1 (P = 0.31); I² = 2%
Test for overall effect: Z = 0.08 (P = 0.93)

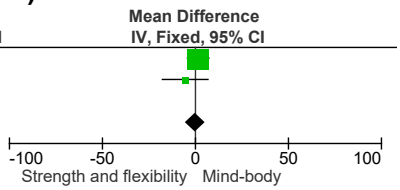


Figure 241: Psychological distress at ≤3 months (ADS depression scale, 0-60, final values, high is poor outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
von Trott 2009	20.2	9.8	35	19.7	7.4	31	100.0%	0.50 [-3.66, 4.66]
Total (95% CI)			**35**			**31**	**100.0%**	**0.50 [-3.66, 4.66]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.24 (P = 0.81)

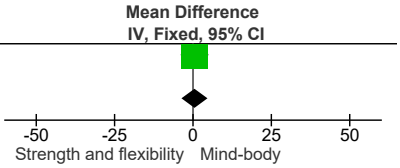


Figure 242: Psychological distress at >3 months (ADS depression scale, 0-60, final values, high is poor outcome)

Study or Subgroup	Strength and flexibility Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
von Trott 2009	20.9	10.2	35	22.7	7.4	31	100.0%	-1.80 [-6.07, 2.47]
Total (95% CI)			**35**			**31**	**100.0%**	**-1.80 [-6.07, 2.47]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.83 (P = 0.41)

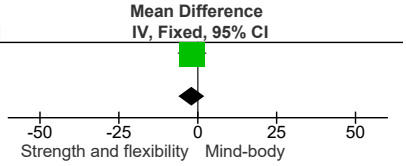
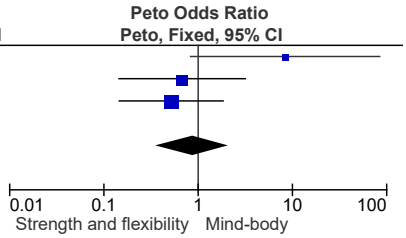


Figure 243: Discontinuation at >3 months

Study or Subgroup	Strength and flexibility Events	Total	Mind-body Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Cramer 2013	3	25	0	26	15.2%	8.37 [0.83, 84.38]
Rendant 2011	3	42	4	39	34.2%	0.68 [0.14, 3.16]
von Trott 2009	4	39	7	38	50.5%	0.52 [0.15, 1.84]
Total (95% CI)		**106**		**103**	**100.0%**	**0.87 [0.35, 2.14]**
Total events	10		11			

Heterogeneity: Chi² = 4.43, df = 2 (P = 0.11); I² = 55%
Test for overall effect: Z = 0.31 (P = 0.76)



E.23 Strength, flexibility and proprioception versus mind-body exercise

Figure 244: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	25.2	18.3	37	32.4	23.5	38	100.0%	-7.20 [-16.72, 2.32]
Total (95% CI)			**37**			**38**	**100.0%**	**-7.20 [-16.72, 2.32]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.48 (P = 0.14)

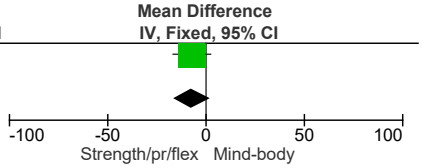


Figure 245: Pain at >3 months (VAS, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	33.1	20.9	37	35	27.7	38	100.0%	-1.90 [-12.99, 9.19]
Total (95% CI)			**37**			**38**	**100.0%**	**-1.90 [-12.99, 9.19]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.34 (P = 0.74)

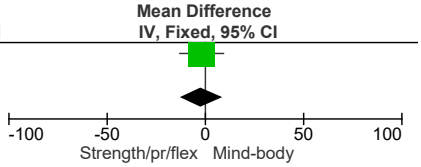


Figure 246: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	45.2	5.4	37	47.3	9.1	38	100.0%	-2.10 [-5.48, 1.28]
Total (95% CI)			**37**			**38**	**100.0%**	**-2.10 [-5.48, 1.28]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.22 (P = 0.22)

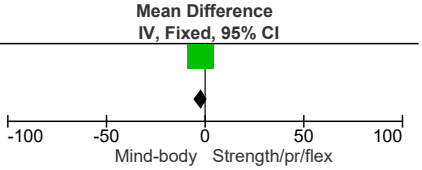


Figure 247: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	44	7.5	37	46.5	8.9	38	100.0%	-2.50 [-6.22, 1.22]
Total (95% CI)			**37**			**38**	**100.0%**	**-2.50 [-6.22, 1.22]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.32 (P = 0.19)

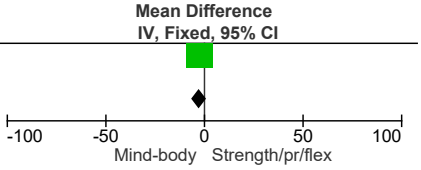


Figure 248: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	47.7	8.5	37	46.8	11.9	38	100.0%	0.90 [-3.77, 5.57]
Total (95% CI)			**37**			**38**	**100.0%**	**0.90 [-3.77, 5.57]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.38 (P = 0.71)

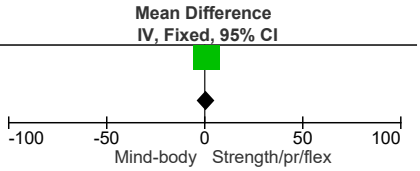


Figure 249: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	46.9	9.1	37	47	12.2	38	100.0%	-0.10 [-4.96, 4.76]
Total (95% CI)			**37**			**38**	**100.0%**	**-0.10 [-4.96, 4.76]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.04 (P = 0.97)

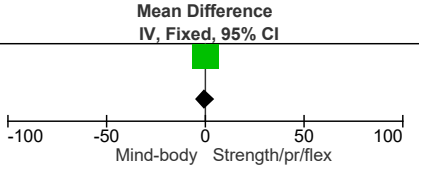


Figure 250: Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	22.7	9.3	37	21.5	12.2	38	100.0%	1.20 [-3.70, 6.10]
Total (95% CI)			**37**			**38**	**100.0%**	**1.20 [-3.70, 6.10]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.48 (P = 0.63)

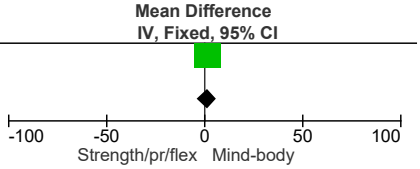


Figure 251: Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	25.1	12.9	37	24.3	14.1	38	100.0%	0.80 [-5.31, 6.91]
Total (95% CI)			**37**			**38**	**100.0%**	**0.80 [-5.31, 6.91]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.26 (P = 0.80)

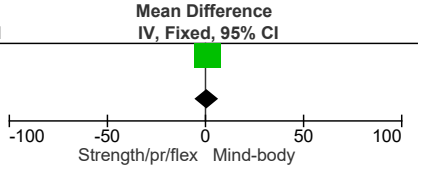


Figure 252: Psychological distress at ≤3 months (HADS anxiety, 0-21, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	5.5	3.1	37	6.5	4.7	38	100.0%	-1.00 [-2.80, 0.80]
Total (95% CI)			**37**			**38**	**100.0%**	**-1.00 [-2.80, 0.80]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.09 (P = 0.28)

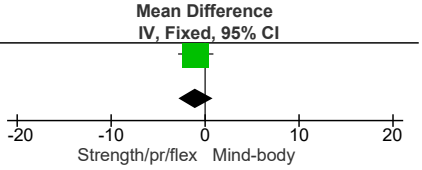


Figure 253: Psychological distress at >3 months (HADS anxiety, 0-21, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	5.5	3.1	37	6.1	4.5	38	100.0%	-0.60 [-2.34, 1.14]
Total (95% CI)			**37**			**38**	**100.0%**	**-0.60 [-2.34, 1.14]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.67 (P = 0.50)

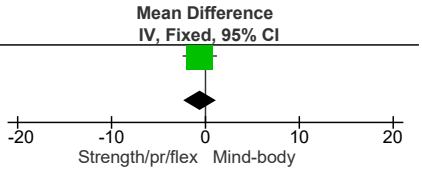


Figure 254: Psychological distress at ≤3 months (HADS depression, 0-21, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	3.8	2.3	37	3.9	3.8	38	100.0%	-0.10 [-1.52, 1.32]
Total (95% CI)			**37**			**38**	**100.0%**	**-0.10 [-1.52, 1.32]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.14 (P = 0.89)

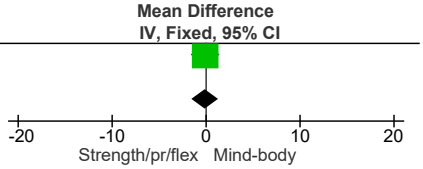


Figure 255: Psychological distress at >3 months (HADS depression, 0-21, final values, high is poor outcome)

Study or Subgroup	Strength/prop/flex Mean	SD	Total	Mind-body Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Lauche 2016	4.1	2.8	37	4.1	3.8	38	100.0%	0.00 [-1.51, 1.51]
Total (95% CI)			**37**			**38**	**100.0%**	**0.00 [-1.51, 1.51]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.00 (P = 1.00)

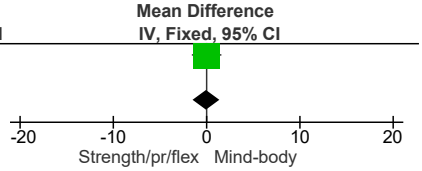
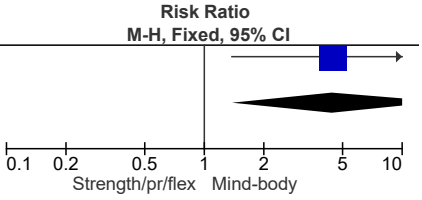


Figure 256: Discontinuation at ≤3 months

Study or Subgroup	Strength/prop/flex Events	Total	Mind-body Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Lauche 2016	13	37	3	38	100.0%	4.45 [1.38, 14.35]
Total (95% CI)		**37**		**38**	**100.0%**	**4.45 [1.38, 14.35]**
Total events	13		3			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.50 (P = 0.01)

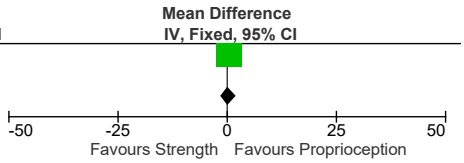


E.24 Strength training versus proprioception

Figure 257: Physical function at ≤3 months (Neck disability index, 0-50, final values, high is poor outcome)

Study or Subgroup	Strength Mean	SD	Total	Proprioception Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Gallego Izquierdo 2016	4.46	2.02	12	4.14	2.62	14	100.0%	0.32 [-1.47, 2.11]
Total (95% CI)			**12**			**14**	**100.0%**	**0.32 [-1.47, 2.11]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.35 (P = 0.73)



E.25 Mind-body exercise versus flexibility

Figure 258: Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)

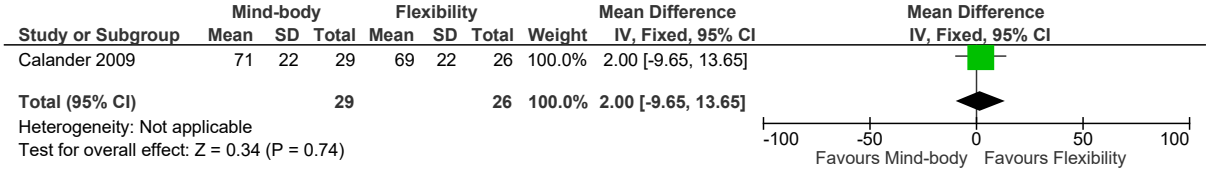


Figure 259: Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)

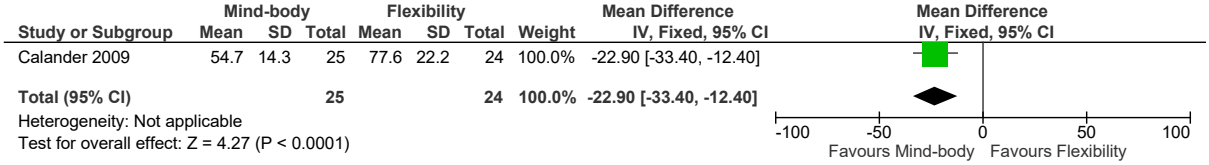


Figure 260: Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)

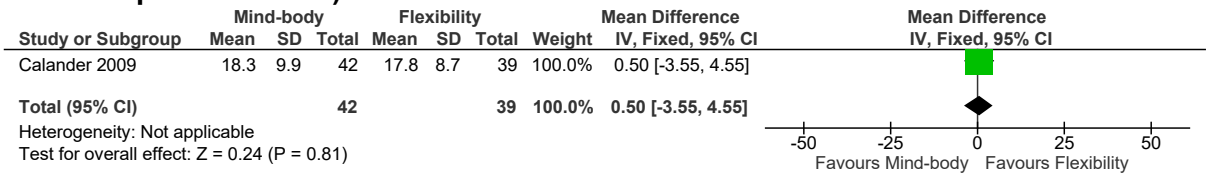


Figure 261: Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, final values, high is poor outcome)

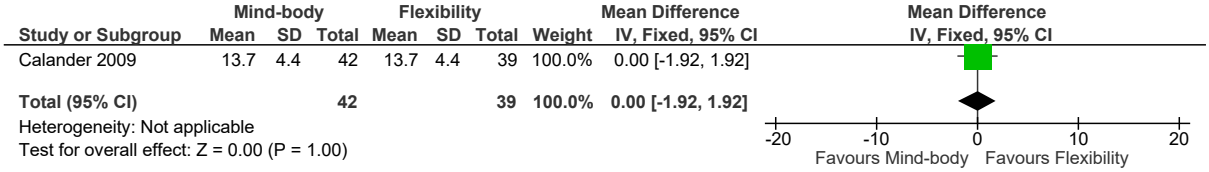
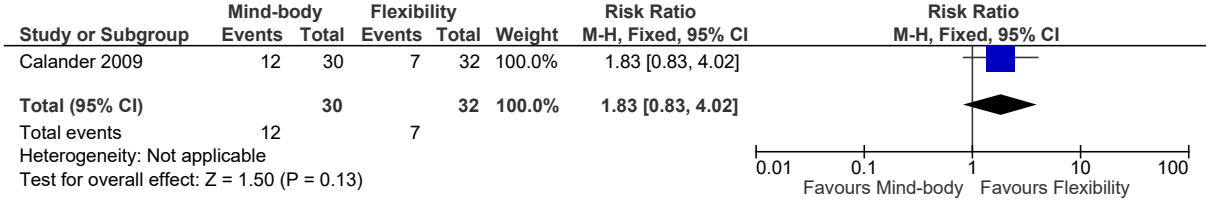


Figure 262: Discontinuation at ≤3 months



E.26 Mind-body exercise versus biomechanical

Figure 263: Pain at ≤3 months (VAS, 0-10, high is poor outcome)

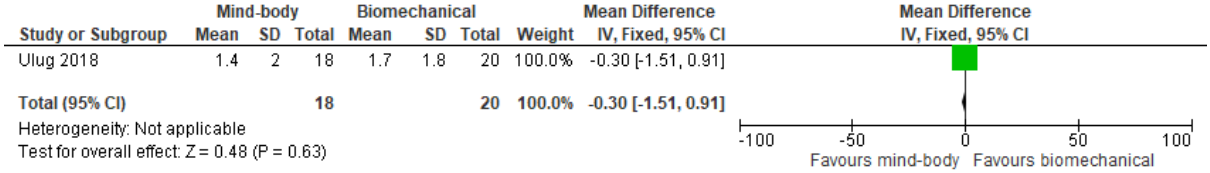


Figure 264: Quality of life ≤3 months (Nottingham Health Profile, 0-600, high is poor outcome)

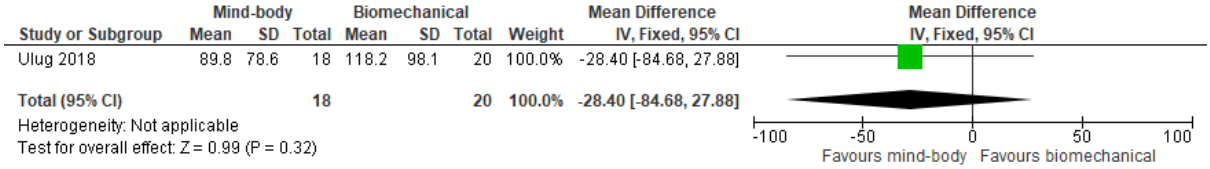


Figure 265: Physical function ≤3 months (Neck Disability Index, 0-100, high is poor outcome)

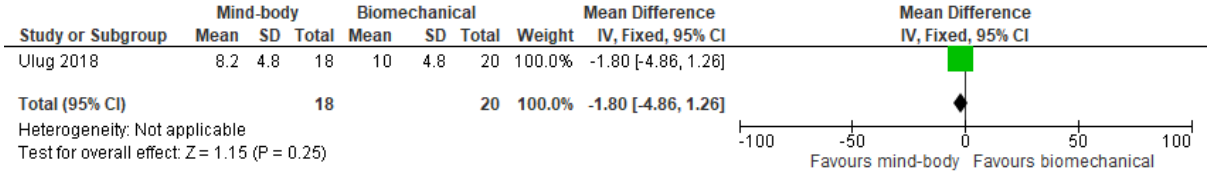
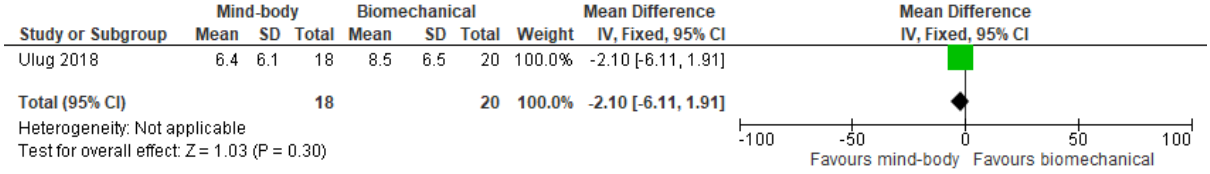


Figure 266: Psychological distress ≤3 months (Beck Depression Inventory, 0-63, high is poor outcome)



E.27 Flexibility and proprioception versus flexibility

Figure 267: Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome

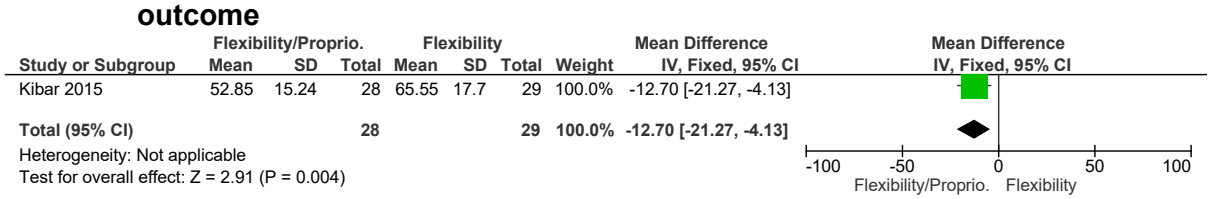


Figure 268: Psychological distress at ≤3 months (BDI, 0-63, final values, high is poor outcome)

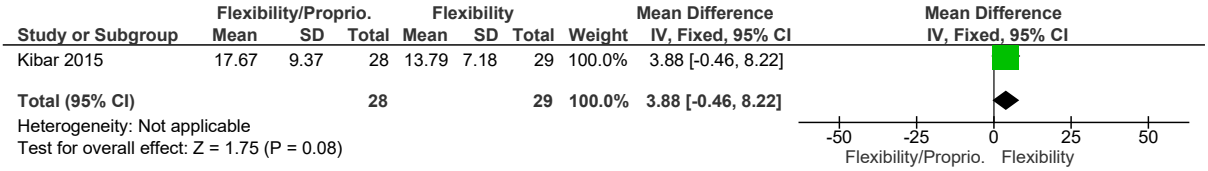
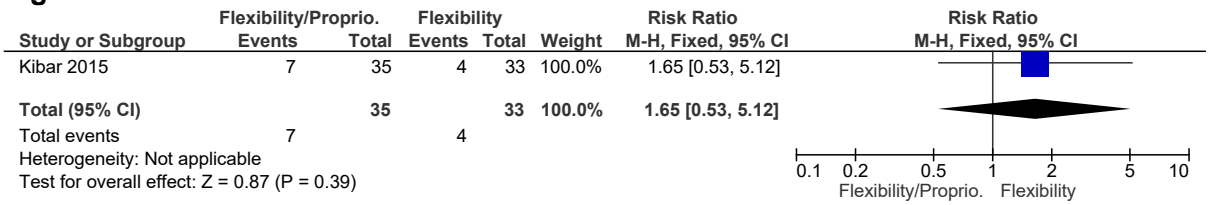


Figure 269: Discontinuation at ≤3 months



E.28 Flexibility and relaxation versus aerobic exercise

Figure 270: Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)

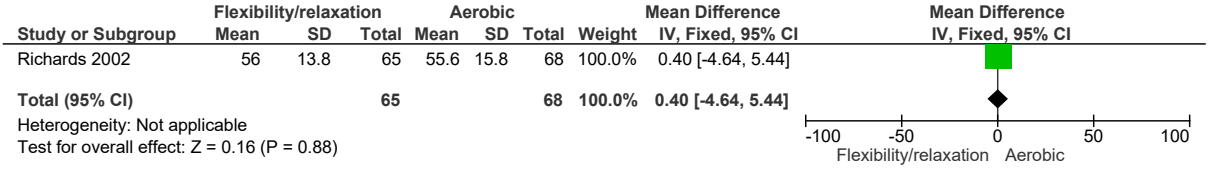
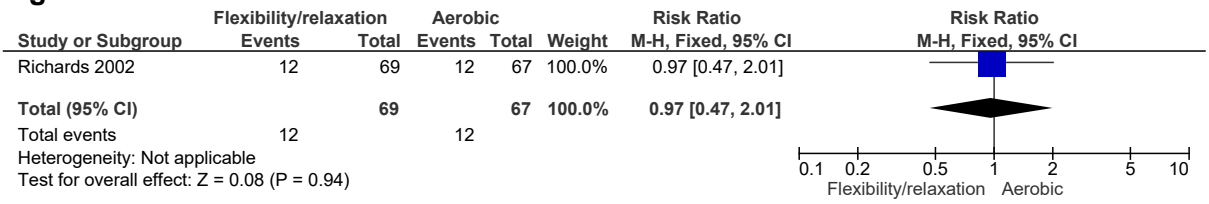
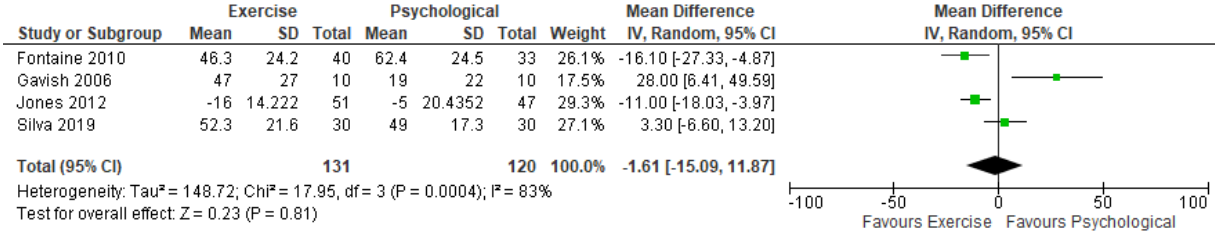


Figure 271: Discontinuation at ≤3 months



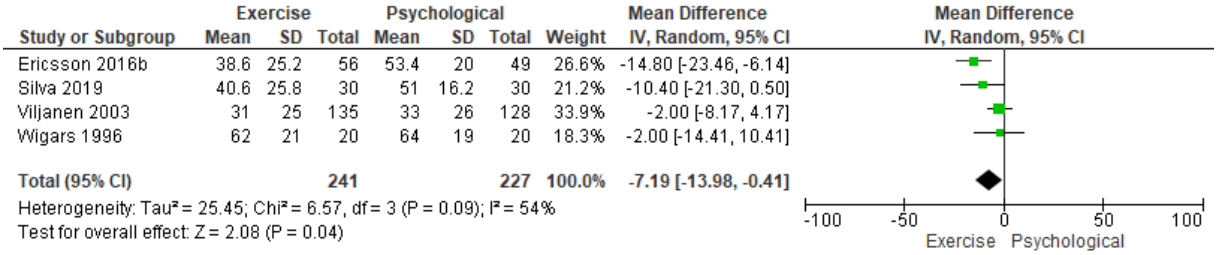
E.29 Exercise versus psychological therapies

Figure 272: Pain at ≤3 months (VAS, NRS, 0-100, final values, high is poor outcome)



NB: Heterogeneity not explained by subgroup analysis

Figure 273: Pain at >3 months (VAS, NRS, 0-100, final values, high is poor outcome)



NB: Heterogeneity not explained by subgroup analysis

Figure 274: Quality of life at ≤3 months (FIQ, 0-100, final values and change scores, high is poor outcome)

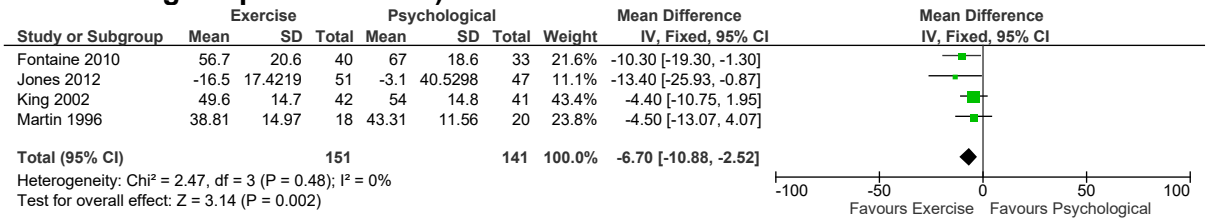


Figure 275: Quality of life at >3 months (EQ-5D, -0.594-1, high is good outcome, final values)

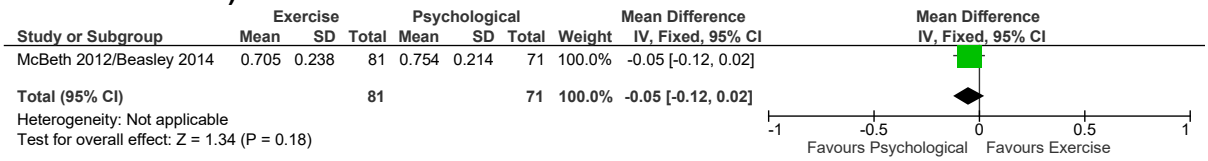


Figure 276: Quality of life at ≤3 months (SF36, 0-100, high score is good outcome

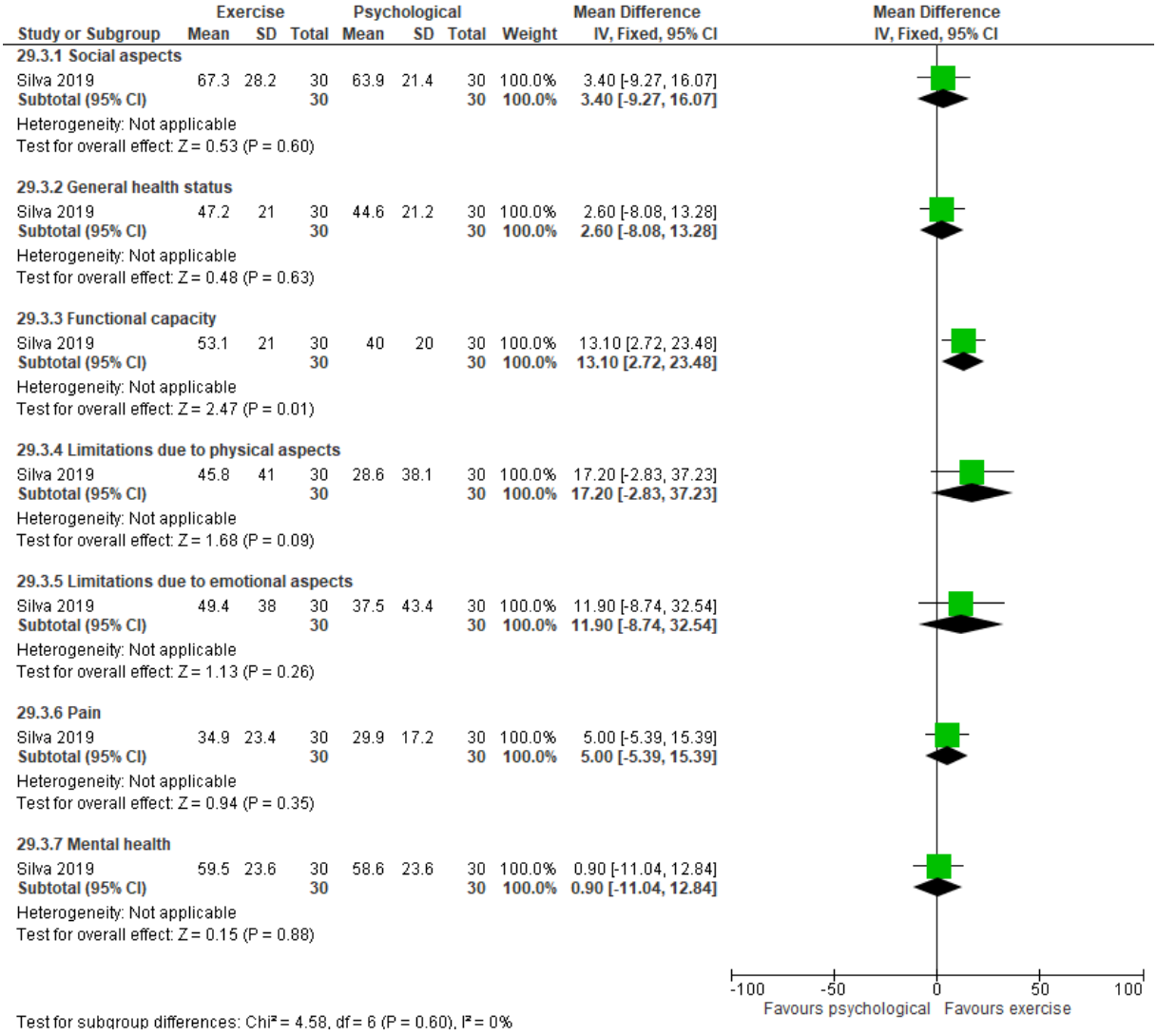


Figure 277: Physical function at ≤3 months (FIQ physical function subscale, 0-10, high is poor outcome, change scores)

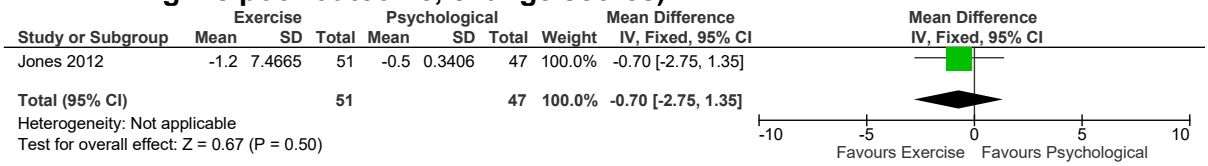


Figure 278: Physical function at ≤3 months (6 minute walking test, metres, high is good outcome, final values)

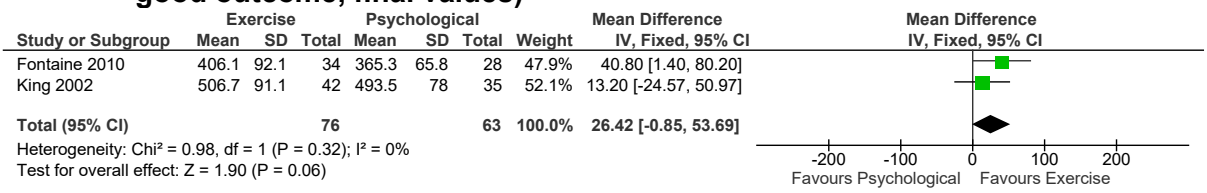


Figure 279: Physical function at >3 months (6 minute walking test, final values, metres)

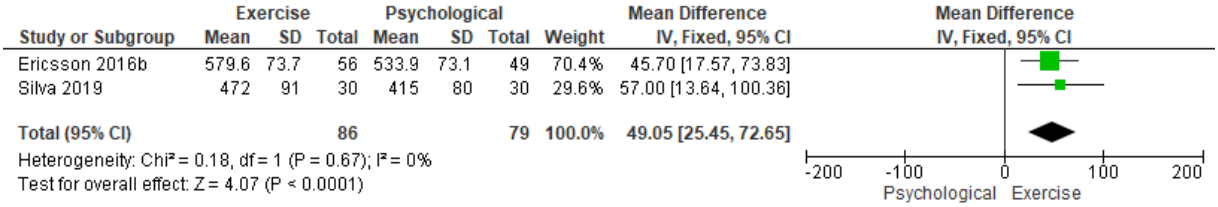


Figure 280: Psychological distress at ≤3 months (CES-D, 0-100, final values, high is poor outcome)

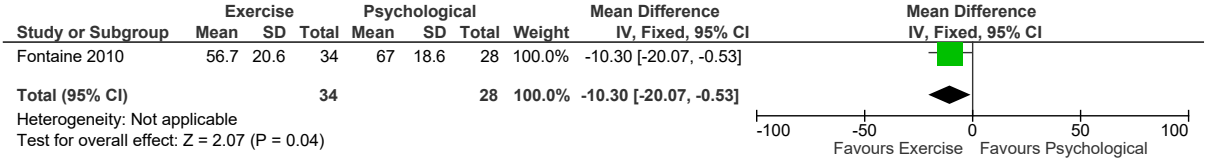


Figure 281: Psychological distress at >3 months (HADS depression, 0-21, change scores, high is poor outcome)

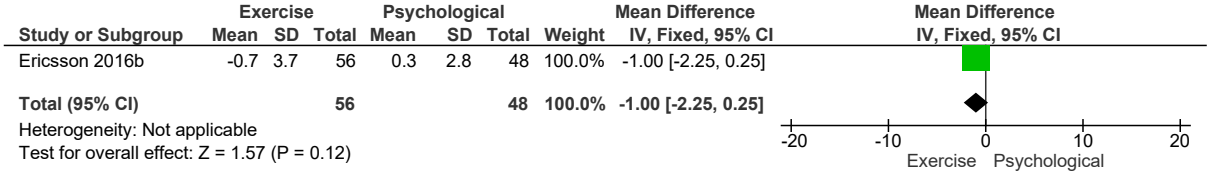


Figure 282: Psychological distress at >3 months (HADS anxiety, 0-21, change scores, high is poor outcome)

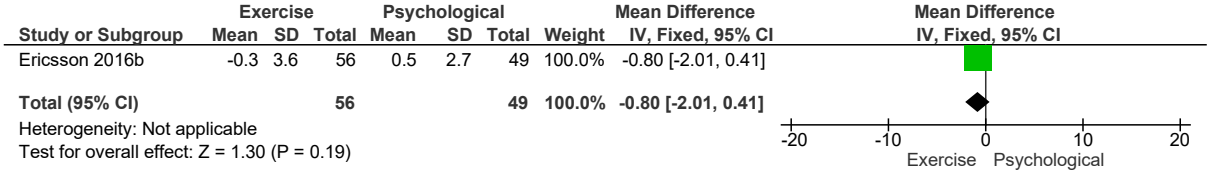


Figure 283: Sleep at >3 months (the sleep scale, 0-20, final values, high is poor outcome)

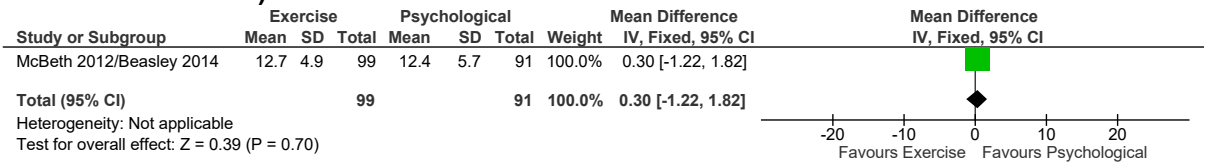


Figure 284: Sleep at >3 months (pittsburgh sleep quality index, 0-100, change scores, high is poor outcome)

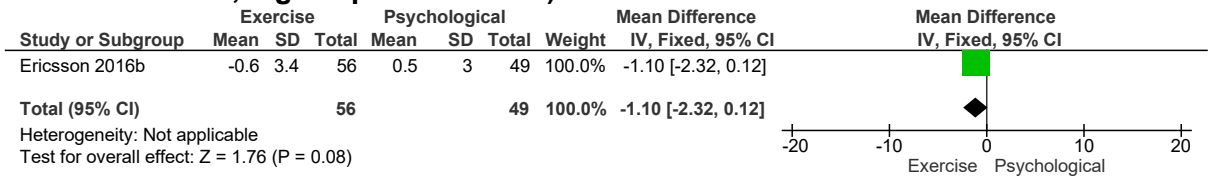
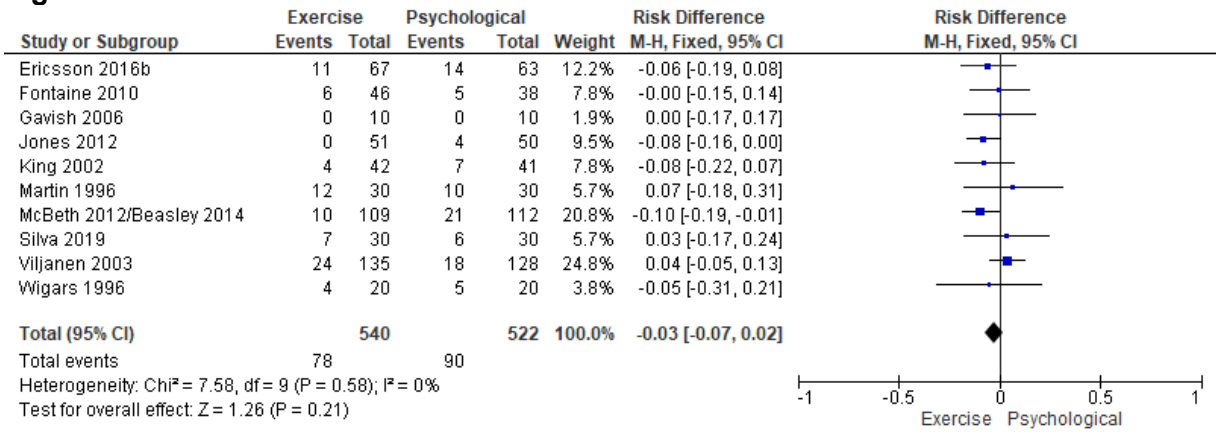


Figure 285: Discontinuation at >3 months



E.30 Manual therapy and exercise versus manual therapy

Figure 286: Pain at ≤3 months (NRS, high is poor outcome, 0-10, final values)

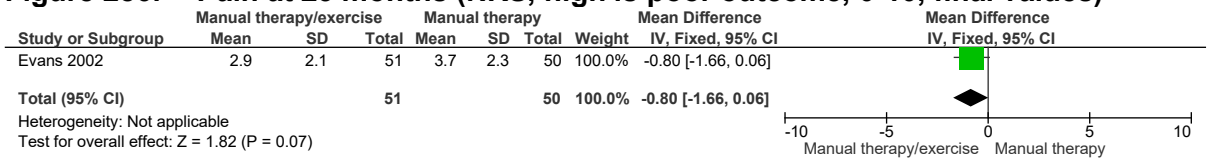


Figure 287: Pain at >3 months (NRS, high is poor outcome, 0-10, final values)

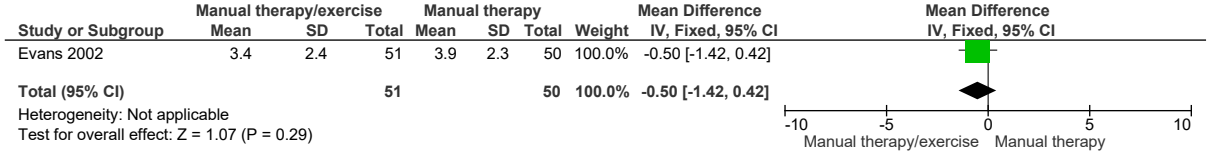


Figure 288: Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50)

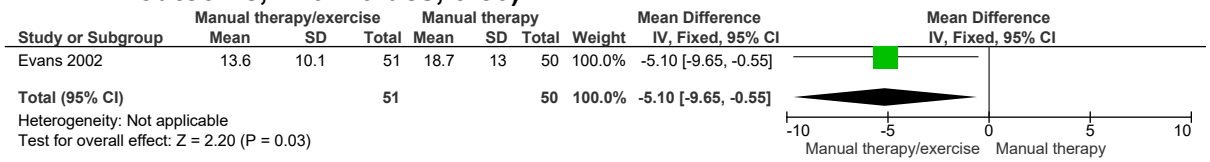


Figure 289: Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)

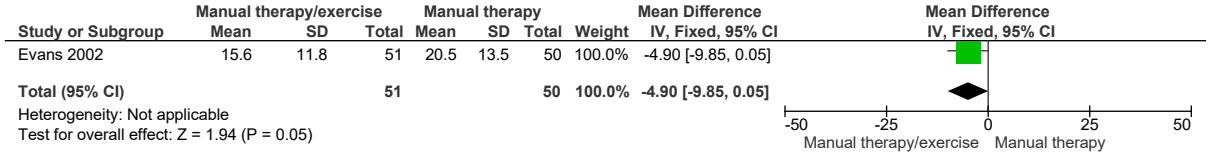
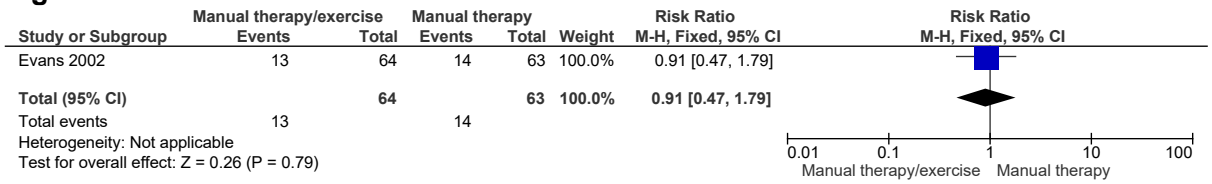
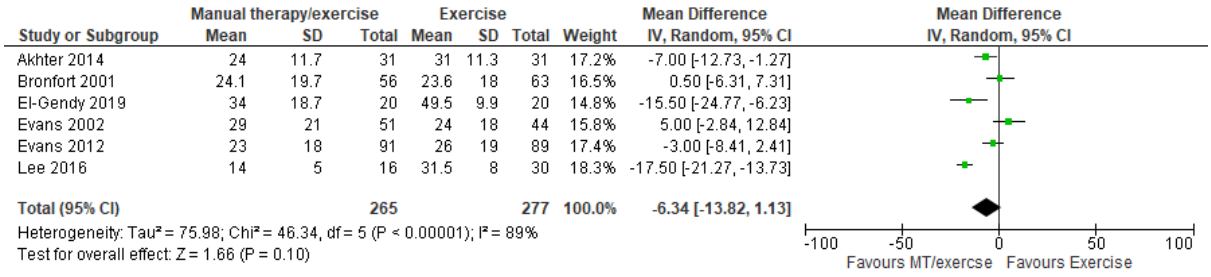


Figure 290: Discontinuation at ≤3 months



E.31 Manual therapy and exercise versus exercise

Figure 291: Pain at <3 months (VAS, NRS, high is poor outcome, final values, 0-100)



Heterogeneity not explained by subgroup analysis.

Figure 292: Pain at >3 months (NRS, VAS, 0-100, final values, high is poor outcome)

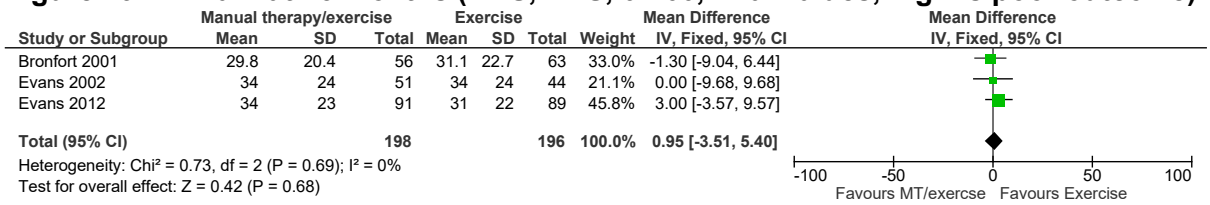


Figure 293: Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)

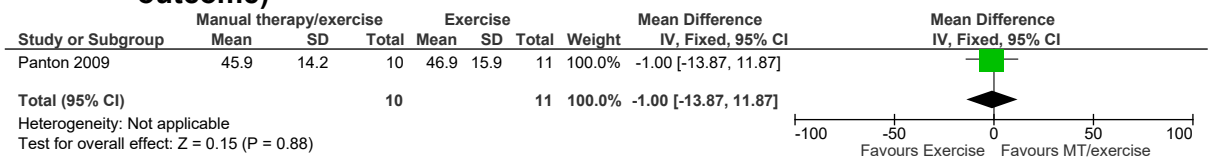


Figure 294: Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Manual therapy/exercise Mean	SD	Total	Exercise Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Evans 2012	50.7	6.7	91	50.1	6.6	89	100.0%	0.60 [-1.34, 2.54]
Total (95% CI)			**91**			**89**	**100.0%**	**0.60 [-1.34, 2.54]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.61 (P = 0.55)

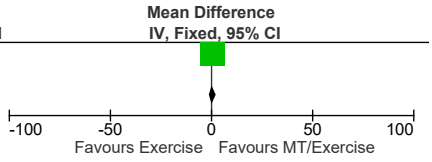


Figure 295: Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Manual therapy/exercise Mean	SD	Total	Exercise Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Evans 2012	50	6.4	91	49.8	7.2	89	100.0%	0.20 [-1.79, 2.19]
Total (95% CI)			**91**			**89**	**100.0%**	**0.20 [-1.79, 2.19]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.20 (P = 0.84)

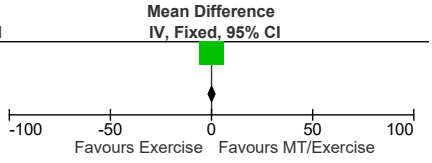


Figure 296: Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Manual therapy/exercise Mean	SD	Total	Exercise Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Evans 2012	53.9	9.8	91	54.6	9.7	89	100.0%	-0.70 [-3.55, 2.15]
Total (95% CI)			**91**			**89**	**100.0%**	**-0.70 [-3.55, 2.15]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.48 (P = 0.63)

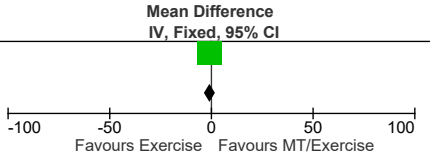


Figure 297: Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)

Study or Subgroup	Manual therapy/exercise Mean	SD	Total	Exercise Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Evans 2012	53	8.9	91	54.8	8.5	89	100.0%	-1.80 [-4.34, 0.74]
Total (95% CI)			**91**			**89**	**100.0%**	**-1.80 [-4.34, 0.74]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.39 (P = 0.17)

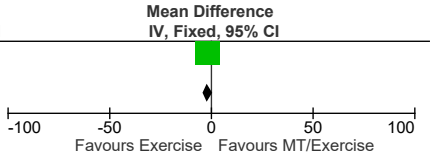
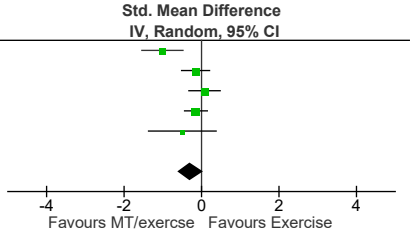


Figure 298: Physical function at >3 months (neck disability index, functional performance scale, final values, high is poor outcome)

Study or Subgroup	Manual therapy/exercise Mean	SD	Total	Exercise Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Akhter 2014	16.83	2.3	31	19.13	2.2	31	17.9%	-1.01 [-1.54, -0.48]
Bronfort 2001	17.1	10.3	56	18.6	9.2	63	23.7%	-0.15 [-0.51, 0.21]
Evans 2002	13.6	10.2	51	12.8	10.2	44	22.1%	0.08 [-0.33, 0.48]
Evans 2012	14.5	9.5	91	16	11.3	89	26.2%	-0.14 [-0.44, 0.15]
Panton 2009	61	14	10	67	9	11	10.0%	-0.49 [-1.37, 0.38]
Total (95% CI)			**239**			**238**	**100.0%**	**-0.29 [-0.62, 0.04]**

Heterogeneity: Tau² = 0.09; Chi² = 11.37, df = 4 (P = 0.02); I² = 65%
Test for overall effect: Z = 1.70 (P = 0.09)



Heterogeneity not explained by subgroup analysis.

Figure 299: Physical function at >3 months (neck disability index, high is poor outcome, final values, 0-100)

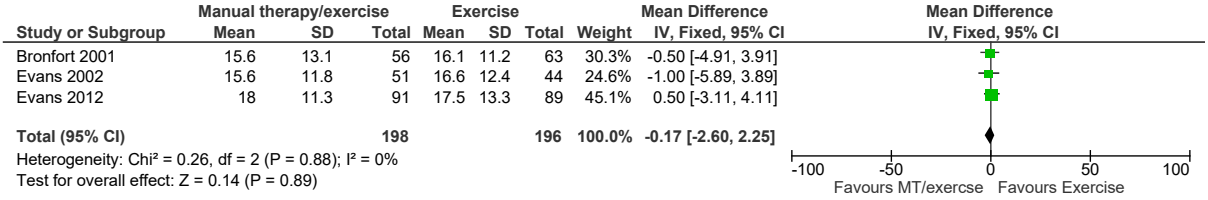


Figure 300: Physical function at ≤3 months (Neck disability index, high is poor outcome, 0-100)

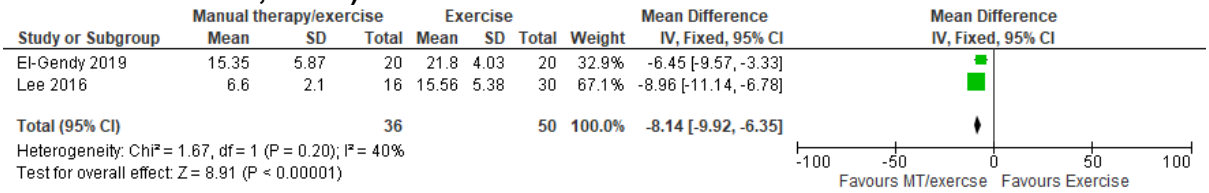
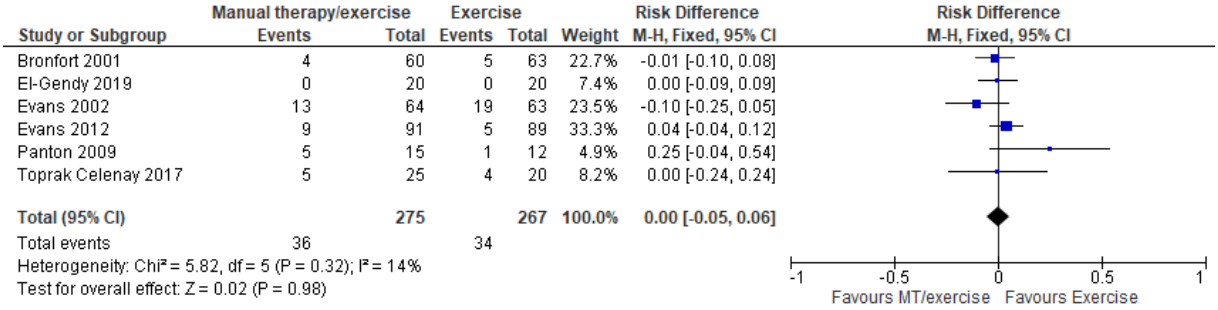


Figure 301: Discontinuation at >3 months



E.32 Exercise versus manual therapy

Figure 302: Pain at ≤3 months (NRS, 0-10, final values, high is poor outcome)

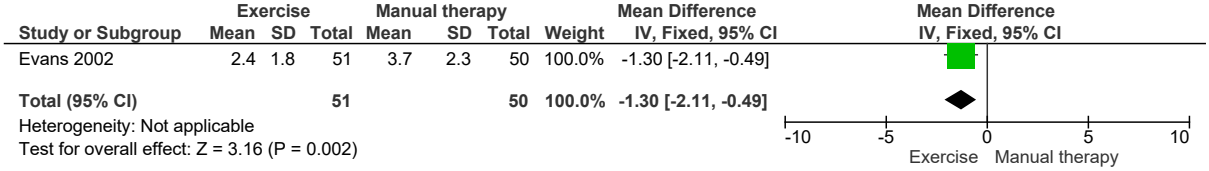


Figure 303: Pain at >3 months (NRS, 0-10, final values, high is poor outcome)

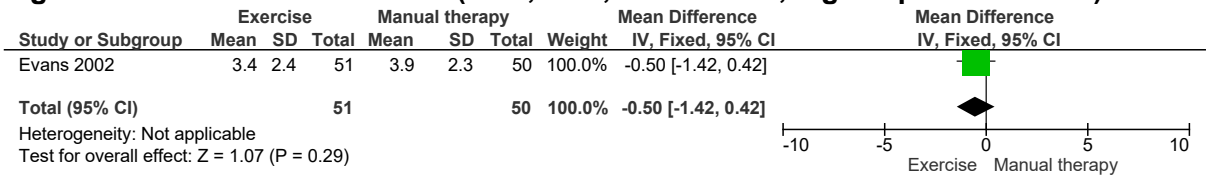


Figure 304: Physical function at ≤3 months (Neck disability index, 0-50, final values, high is poor outcome)

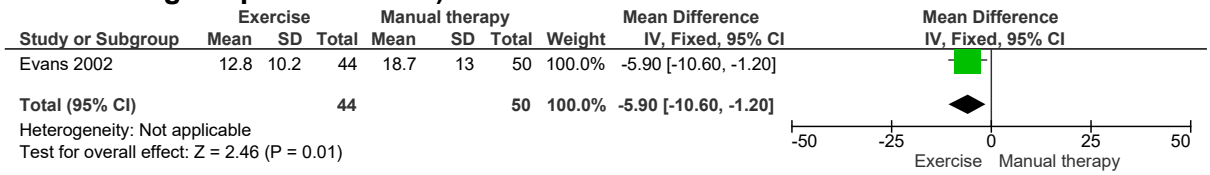


Figure 305: Physical function at >3 months (Neck disability index, 0-50, final values, high is poor outcome)

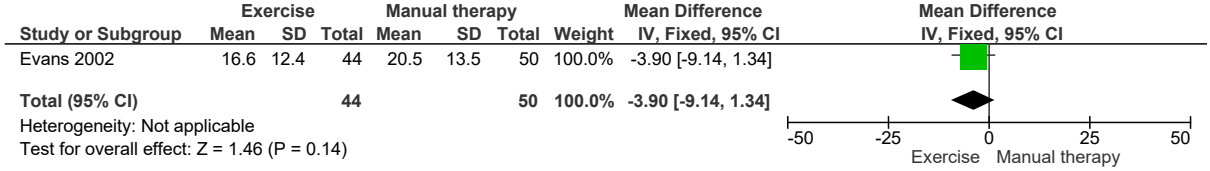
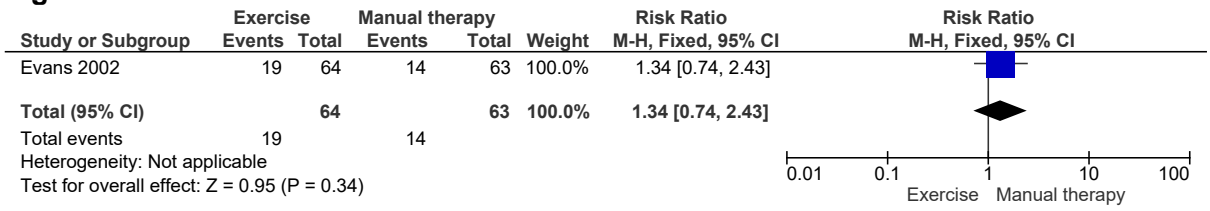


Figure 306: Discontinuation at ≤3 months



Appendix F: GRADE tables

Table 71: Clinical evidence profile: Aerobic versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic exercise	Control	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	20	20	-	MD 21.5 lower (30.38 to 12.62 lower)	⊕⊕⊕O MODERATE	CRITICAL
Pain at >3 months (VAS, FIQ pain subscale, 0-100, final values, high is poor outcome)												
9	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	300	228	-	MD 6.97 lower (10.77 to 3.17 lower)	⊕⊕OO LOW	CRITICAL
Pain at >3 months (FIQ pain subscale, 0-100, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	48	-	MD 1 lower (10.34 lower to 8.34 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)												
5	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[2]	none	228	144	-	MD 7.89 lower (13.23 to 2.55 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 functional capacity subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	27	27	-	MD 12.5 higher (3.85 to 21.15 higher)	⊕⊕OO LOW	CRITICAL

Quality of life at >3 months (SF-36 physical appearance subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	27	27	-	MD 16 higher (2.68 lower to 34.68 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 pain subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	27	27	-	MD 7.5 higher (8.62 lower to 23.62 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	27	27	-	MD 7.7 higher (2.49 lower to 17.89 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 social aspects subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	27	27	-	MD 8.9 higher (3.16 lower to 20.96 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 emotional aspects subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	27	27	-	MD 9.7 higher (10.7 lower to 30.1 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	27	27	-	MD 3.4 higher (7.46 lower to 14.26 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (EQ-5D, -0.594-1, high is good outcome, final values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 0.03 higher (0.15 lower to 0.09 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (EQ-5D, -0.594-1, high is good outcome, final values)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	128	131	-	MD 0.06 higher (0.01 lower to 0.13 higher)	⊕⊕OO LOW	CRITICAL

Quality of life at ≤3 months (EQ-5D VAS, 0-100. high is good outcome, final values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 5.6 higher (2.86 lower to 14.06 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (EQ-5D VAS, 0-100, high is good outcome, final values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	48	-	MD 1.4 higher (8.17 lower to 10.97 higher)	⊕⊕OO LOW	CRITICAL
Physical function at 12 weeks (Final values, timed up and go, seconds, high is good outcome) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	40	20	-	MD 0.62 lower (1.40 lower to 0.16 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 3 lower (11.32 lower to 5.32 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at >3 months (6 minute walking test, final values, metres, high is good outcome)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	91	78	-	MD 56.18 higher (27.8 to 84.56 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at >3 months (FIQ and SF-36 physical function subscales, 0-100, final values, high is poor outcome)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	159	87	-	MD 10.16 lower (15.39 to 4.94 lower)	⊕OOO VERY LOW	CRITICAL
Physical function at >3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 3 lower (16.14 lower to 10.14 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (Change scores and final values, beck depression inventory, 0-21, high is poor outcome)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	63	-	MD 3.36 lower (6.16 to 0.56 lower)	⊕⊕OO LOW	CRITICAL

Psychological distress at >3 months (Final values, VAS and FIQ depression scale, 0-10, high is poor outcome)												
4	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	190	116	-	MD 0.39 lower (1.05 lower to 0.28 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (Final values, VAS and FIQ anxiety scale, Beck anxiety inventory, final values, high is poor outcome)												
4	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	197	123	-	SMD 0.28 lower (0.51 lower to 0.04 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (Change scores, STAI anxiety total scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	27	23	-	MD 9.7 lower (23.6 lower to 4.2 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (final values, FIQ depression scale, 0-10, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 0.8 higher (0.46 lower to 2.06 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (final values, FIQ anxiety scale, 0-10, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	48	-	MD 0.2 higher (1.06 lower to 1.46 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at 12 weeks (Final values, BDI dpression scale, high is poor outcome) (Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	20	-	MD 12.77 lower (14.65 to 10.88 lower)	⊕⊕OO LOW	CRITICAL
Use of healthcare services ≤3 months (Number of GP contacts)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 1 higher (0.11 lower to 2.11 higher)	⊕OOO VERY LOW	CRITICAL
Use of healthcare services >3 months (Number of GP contacts)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 0.3 higher (0.68 lower to 1.28 higher)	⊕OOO VERY LOW	CRITICAL

Use of healthcare services ≤3 months (Number of medical specialist contacts)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 0.1 higher (0.18 lower to 0.38 higher)	⊕OOO VERY LOW	CRITICAL
Use of healthcare services >3 months (Number of medical specialist contacts)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 0.2 higher (0.08 lower to 0.48 higher)	⊕OOO VERY LOW	CRITICAL
Use of healthcare services at ≤3 months (Number of physiotherapist contacts)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	47	48	-	MD 3.1 lower (4.49 to 1.17 lower)	⊕OOO VERY LOW	CRITICAL
Use of healthcare services at >3 months (Number of physiotherapist contacts)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	47	48	-	MD 4.4 lower (5.79 to 3.01 lower)	⊕⊕OO LOW	CRITICAL
Sleep at >3 months (VAS sleep scale, PSQI, FIQ sleep subscale, final values, high is poor outcome)												
5	randomised trials	very serious[1]	serious[3]	no serious indirectness	no serious imprecision[2]	none	209	205	-	SMD 0.16 lower (0.43 lower to 0.1 higher)	⊕OOO VERY LOW	CRITICAL
Discontinuation at >3 months												
9	randomised trials	serious[1]	very serious[2]	no serious indirectness	serious[3]	none	70/316 (22.2%)	33/291 (11.3%)	RD 0.11 (-0.04 to 0.27)	110 more per 1000 (from 40 fewer to 270 more)	⊕OOO VERY LOW	CRITICAL

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias.
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis.

Table 72: Clinical evidence profile: Strength versus usual care

Quality assessment							No of patients		Effect		Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Strength	Control	Relative (95% CI)	Absolute		
Pain reduction at ≤3 months (final values, VAS, high is poor outcome)												
3	randomised trials	very serious[1]	very serious[2]	no serious indirectness	serious[3]	none	150	101	-	MD -18.85 (34.50 to 3.21 lower)	⊕OOO VERY LOW	CRITICAL
Pain reduction at ≤3 months (change scores and final values, VAS, NRS, 0-100, high is poor outcome)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	88	68	-	MD 15.76 lower (22.79 to 8.72 lower)	⊕OOO VERY LOW	CRITICAL
Pain reduction at >3 months (VAS, NRS, 0-100, final values and change scores, high is poor outcome)												
4	randomised trials	very serious[1]	serious[3]	no serious indirectness	serious[2]	none	230	219	-	MD 16.06 lower (36.93 lower to 4.82 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component summary, 0-100, change scores, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	22	20	-	MD 7.6 higher (0.25 lower to 15.45 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary, 0-100, change scores, high is good outcome)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	50	-	MD 3.39 higher (2.43 lower to 9.21 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (FIQ scale, 0-100, final values, high is poor outcome)												
2	randomised trials	very serious[1]	serious[3]	no serious indirectness	very serious[2]	none	28	24	-	MD 14.91 lower (45.78 lower to 15.96 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, change scores and final values, 0-100, high is poor outcome)												
3	randomised trials	serious[1]	very serious[2]	no serious indirectness	very serious[3]	none	83	63	-	MD 9.89 lower (23.15 lower to 3.37 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (final values, FIQ physical function subscale, Northwick Park Questionnaire, high is poor outcome)												
2	randomised trials	very serious[1]	Serious[3]	no serious indirectness	no serious imprecision	none	75	76	-	SMD 0.23 lower	⊕OOO VERY LOW	CRITICAL

										(0.68 lower to 1.14 higher)		
Physical function at ≤3 months (6 minute walking test, metres, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	8	12	-	MD 8.4 lower (89.59 lower to 72.79 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at >3 months (final values, Northwick Park Questionnaire, Neck Disability Index, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	84	79	-	SMD 0.32 lower (0.64 lower to 0.00 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (change scores, SF-36 physical function subscale, HAQ, 0-100, high is poor outcome)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	52	53	-	MD 6.2 lower (10.41 to 2 lower)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (pain catastrophising scale, 0-100, final scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	13	12		MD 9.00 lower (19.70 lower to 1.70 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	11	10	-	MD 3.7 lower (6.37 to 1.03 lower)	⊕⊕OO LOW	CRITICAL
Use of health care services at >3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	27/119 (22.7%)	20/60 (33.3%)	RR 0.68 (0.42 to 1.11)	107 fewer per 1000 (from 193 fewer to 37 more)	⊕⊕OO LOW	IMPORTANT
Sleep at >3 months (VAS sleep, 0-100, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	11	10	-	MD 7 lower (20.9 lower to 6.9 higher)	⊕⊕OO LOW	IMPORTANT
Discontinuation at ≤3 months												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	11/70 (15.7%)	6.5%	Peto OR 2.27 (0.77 to 6.73)	71 more per 1000 (from 14 fewer to 254 more)	⊕OOO VERY LOW	IMPORTANT

Discontinuation at >3 months												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision[2]	none	26/121 (21.5%)	3.3%	RD 0.08 (-0.02 to 0.17	33 fewer per 1000 (from 27 fewer to 34 fewer)	⊕⊕⊕O MODERATE	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 73: Clinical evidence profile: Aerobic and strength versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic and strength	Control	Relative (95% CI)	Absolute		
Pain reduction at ≤3 months (VAS, 0-100, change scores, high is poor outcome)												
2	randomised trials	serious[1]	very serious[3]	no serious indirectness	very serious[2]	none	85	44	-	MD 2.45 lower (34.16 lower to 29.27 higher)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, FIQ pain subscale, 0-100, final values, high is poor outcome)												
3	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	103	58	-	MD 13.74 lower (22.11 to 5.37 lower)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (EQ-5D, -0.594 to 1, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15	15	-	MD 0.25 higher (0.05 to 0.45 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (Fibromyalgia impact questionnaire, 0-100, final values, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	29	25	-	MD 3.42 lower (12.66 lower to 5.82 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (Fibromyalgia impact questionnaire, 0-100, final values and change scores, high is poor outcome)												
4	randomised trials	very serious[1]	Serious[3]	no serious indirectness	serious[2]	none	90	81	-	MD 9.05 lower (15.43 to 2.68 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (EQ-5D, -0.594 to 1, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15	15	-	MD 0.19 higher (0.00 to 0.39 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 physical functioning subscale, 0-100, final values, high is good outcome)												

1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 11.6 higher (2.02 to 21.18 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 physical role subscale, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	serious[2]	no serious imprecision	none	21	21	-	MD 1.9 higher (14.93 lower to 18.73 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 emotional role subscale, 0-100, final values, high is good outcome)												
1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 19 higher (6.96 lower to 44.96 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 vitality subscale, 0-100, final values, high is good outcome)												
1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 12.7 higher (2.73 to 22.67 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 mental health subscale, 0-100, final values, high is good outcome)												
1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 15.8 higher (3.75 to 27.85 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 24 weeks (SF-36 social role subscale, 0-100, final values, high is good outcome)												
1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 11.7 higher (1.9 lower to 25.3 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 bodily pain subscale, 0-100, final values, high is good outcome)												
1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 10.4 higher (0.16 lower to 20.96 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 general health subscale, 0-100,final values, high is good outcome)												

1	randomised trials	Very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	21	21	-	MD 9.6 higher (2.82 to 16.38 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (seconds, quarter mile walk test, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	8	8	-	MD 37.3 lower (63.19 to 11.41 lower)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (metres, 6-minute walk test, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	19	-	MD 54.8 higher (0.54 lower to 110.14 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (FIQ physical function subscale, final values, 0-10, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	15	15	-	MD 1.3 lower (2.63 lower to 0.03 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at 8 weeks (metres, 6-minute walk test, high is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16	16	-	MD 15.69 higher (33.37 lower to 64.75 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress ≤3 months (BDI, 0-30, final values, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	29	25	-	MD 1.44 lower (6.85 lower to 3.97 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress ≤3 months (State anxiety inventory, 0-10, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	34	24	-	MD 0.1 higher (5.12 lower to 5.32 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at 8 weeks (HADS, 0-21, high is poor outcome) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16	16	-	MD 1.25 lower (3.77 lower to 1.27 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (CES-D, BDI, FIQ depression subscale, final values, high is poor outcome)												

4	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	62	63	-	SMD 0.45 lower (0.81 to 0.09 lower)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (State anxiety inventory, 20-80, final values, high is poor outcome)												
2	randomised trials	very serious[1]	serious[3]	no serious indirectness	serious[2]	none	44	39	-	MD 2.95 lower (9.75 lower to 3.85 higher)	⊕OOO VERY LOW	CRITICAL
Sleep at >3 months (Pittsburg sleep quality index, high is poor outcome, final values, 0-21)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	34	24	-	MD 2.2 lower (3.39 to 1.01 lower)	⊕⊕OO LOW	CRITICAL
Healthcare utilisation at >3 months (follow-up 3 years)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	23/57 (40.4%)	10/21 (47.6%)	RR 0.85 (0.49 to 1.47)	71 fewer per 1000 (from 243 fewer to 224 more)	⊕OOO VERY LOW	CRITICAL
Discontinuation at ≤3 months												
4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	7/65 (10.8%)	1/60 (1.7%)	RD 0 (-0.01 to 0.17)	0 fewer per 1000 (from 10 fewer to 170 more)	⊕⊕OO LOW	IMPORTANT
Discontinuation at >3 months												
7	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	20/170 (11.8%)	4.9%	RD 0.05 (-0.03 to 0.12)	47 fewer per 1000 (from 43 fewer to 50 fewer)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded for heterogeneity, unexplained by subgroup analysis

Table 74: Clinical evidence profile: Aerobic, strength and flexibility versus usual care

Quality assessment	No of patients	Effect	Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic, strength and flexibility	Control	Relative (95% CI)	Absolute		
Quality of life at ≤3 months (SF-36 mental component, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	12	13	-	MD 12.1 higher (2.14 to 22.06 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	12	13	-	MD 5.1 higher (3.18 lower to 13.38 higher)	⊕⊕OO LOW	CRITICAL

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 75: Clinical evidence profile: Strength and flexibility versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Strength and flexibility	Control	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	55	55	-	MD 11.71 lower (21.49 to 1.92 lower)	⊕⊕OO LOW	CRITICAL
Pain at >3 months (VAS, SF-36 pain score, final values, 0-100, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	70	74	-	MD 13.19 lower (20.33 to 6.05 lower)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	35	-	MD 0.6 lower (6.12 lower to 4.92 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is poor outcome)												

2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	70	74	-	MD 1.78 higher (1.35 lower to 4.91 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	35	-	MD 1.7 higher (2.42 lower to 5.82 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is poor outcome)												
2	randomised trials	serious[1]	Serious[3]	no serious indirectness	no serious imprecision	none	70	74	-	MD 0.16 lower (3.87 lower to 3.56 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[1]	none	35	35	-	MD 5.5 lower (16.59 lower to 5.59 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	70	74	-	MD 6.7 lower (12.3 to 1.1 lower)	⊕⊕⊕O MODERATE	CRITICAL
Psychological distress at ≤3 months (ADS depression scale, 0-60, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	35	-	MD 1.6 higher (2.59 lower to 5.79 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (ADS depression scale, 0-60, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	35	-	MD 1.1 higher (3.41 lower to 5.61 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at >3 months												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	8/78 (10.3%)	11.7%	OR 0.88 (0.32 to 2.4)	13 fewer per 1000 (from 76 fewer to 124 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.
3 Downgraded due to heterogeneity, unexplained by subgroup analysis

Table 76: Clinical evidence profile: Strength, proprioception and flexibility versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Strength, proprioception and flexibility	Control	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 16.6 lower (25.8 to 7.4 lower)	⊕⊕OO LOW	CRITICAL
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 11.5 lower (20.71 to 2.29 lower)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 2.3 higher (0.13 lower to 4.73 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 2 higher (1.48 lower to 5.48 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 1.6 higher (2.73 lower to 5.93 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 0.5 higher (3.82 lower to 4.82 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS: anxiety, 0-21, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 1.2 lower (2.68 lower to 0.28 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (HADS: anxiety, 0-21, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 1.2 lower (2.66 lower to 0.26 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS: depression, 0-21, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 1.1 lower (2.4 lower to 0.2 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (HADS: depression, 0-21, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 1.3 lower (2.85 lower to 0.25 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 4.8 lower (9.47 to 0.13 lower)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	39	-	MD 4.3 lower (10.06 lower to 1.46 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[2]	none	13/37 (35.1%)	25.6%	RR 1.37 (0.69 to 2.73)	95 more per 1000 (from 79 fewer to 443 more)	⊕⊕OO LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias

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

Table 77: Clinical evidence profile: Proprioception versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Proprioception	Control	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	24	22	-	MD 0.18 higher (1.09 lower to 1.45 higher)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, 0-10, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24	22	-	MD 0.97 lower (2.47 lower to 0.53 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	24	22	-	MD 1.88 lower (11.11 lower to 7.35 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24	22	-	MD 3.59 lower (14.37 lower to 7.19 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (sit to stand test, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24	24	-	MD 4.38 lower (6.82 to 1.94 lower)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (sit to stand test, final values, high is good outcome)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24	22	-	MD 0.86 lower (3.18 lower to 1.46 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24	22	-	MD 4.74 lower (8.43 to 1.05 lower)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (BDI, 0-61, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	24	22	-	MD 4.86 lower (9.84 lower to 0.12 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at >3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	1/25 (4%)	3/25 (12%)	RR 0.33 (0.04 to 2.99)	80 fewer per 1000 (from 115 fewer to 239 more)	⊕⊕OO LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 78: Clinical evidence profile: Mind-body versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mind-body exercises**	**Control**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (VAS, Visual numeric scale, FIQ pain subscale, 0-100, final values and change scores, high is poor outcome)												
8	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	193	200	-	MD 11.17 lower (1717.3285 to 5.02 lower)	⊕OOO VERY LOW	CRITICAL
Pain improvement at ≤3 months (30% improvement on NRS)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	37/73 (50.7%)	15.9%	RR 3.19 (1.56 to 6.52)	348 more per 1000 (from 89 more to 878 more)	⊕⊕OO LOW	CRITICAL
Pain improvement at >3 months (30% improvement on NRS)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	28/73 (38.4%)	8/44 (18.2%)	RR 2.11 (1.06 to 4.21)	202 more per 1000 (from 11 more to 584 more)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome) - Fibromyalgia												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	40	-	MD 26 lower (35.63 to 16.37 lower)	⊕⊕OO LOW	CRITICAL
Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome) - Chronic neck pain												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	108	113	-	MD 11.29 lower (174219.52 to 5.17 lower)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (WHOQOL-BREF, 0-5, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	29	28	-	MD 0.58 higher (0.16 to 1 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)												
3	randomised trials	very serious[1]	serious[2]	no serious indirectness	very serious[3]	none	52	54	-	MD 1.55 lower (13.36 lower to 10.25 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision[3]	none	107	113	-	MD 4.14 higher (2.15 to 6.12 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												

3	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3]	none	107	113	-	MD 2.33 higher (2.57 lower to 7.24 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is poor outcome)												
3	randomised trials	serious[1]	serious[2]	no serious indirectness	very serious[3]	none	108	113	-	MD 1.64 lower (11.62 lower to 8.33 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is poor outcome)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	108	113	-	MD 0.69 higher (2.05 lower to 3.43 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, functional capacity scale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 17.2 higher (8.01 to 26.39 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, physical aspects subscale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 22.7 higher (9.73 to 35.67 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, pain subscale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 16.9 higher (9.19 to 24.61 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, vitality subscale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 10.5 higher (0.5 to 20.5 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, general health subscale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 3.4 higher (4.81 lower to 11.61 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, social subscale, final values, high is good outcome)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 5.9 higher (5.61 lower to 17.41 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, emotional subscale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	50	40	-	MD 20.4 higher (4.14 to 36.66 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36, 0-100, mental health subscale, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	40	40	-	MD 6.1 higher (3.42 lower to 15.62 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)												
7	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	171	192	-	SMD 0.40 lower (0.84 to 0.04 lower)	⊕OOO VERY LOW	CRITICAL
Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	112	113	-	MD 6.79 lower (10.57 to 3.01 lower)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (6 minute walk test, metes, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	40	40	-	MD 88 higher (51.42 to 124.58 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS:D, Beck depression inventory, CES-D, ADS depression scale, final values, high is poor outcome)												
5	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	148	158	-	SMD 0.51 lower (0.96 to 0.05 lower)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (State trace anxiety inventory, final values, high is poor outcome) - Fibromyalgia												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	29	28	-	MD 9.91 lower (15.59 to 4.23 lower)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS:A, final values, high is poor outcome) - Chronic neck pain												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	38	39	-	MD 0.2 lower (2 lower to 1.6 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (Beck depression inventory, HADS:D, final values, high is poor outcome)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	109	114	-	MD 0.02 lower (0.29 lower to 0.24 higher)	⊕⊕⊕O MODERATE	CRITICAL
Psychological distress at >3 months (HADS:A, 0-21, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	38	39	-	MD 0.6 lower (2.38 lower to 1.18 higher)	⊕⊕OO LOW	CRITICAL
Sleep at ≤3 months (VAS sleep outcome, pittsburgh sleep quality index, final values, high is poor outcome)												
2	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[2]	none	29	31	-	SMD 0.43 lower (1.58 lower to 0.72 higher)	⊕OOO VERY LOW	IMPORTANT
Discontinuation at >3 months												
12	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	65/389 (16.7%)	7.7%	RD 0.03 (-0.03 to 0.10)	40 more per 1000 (from 30 fewer to 100 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded for heterogeneity, unexplained by subgroup analysis
3 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 79: Clinical evidence profile: Flexibility versus usual care

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Flexibility**	**Control**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16	12	-	MD 18 lower (37.89 lower to 1.89 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	14	14	-	MD 1.5 lower (5.39 lower to 2.39 higher)	⊕OOO VERY LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	3/17 (17.6%)	0/17 (0%)	Peto OR 8.41 (0.81 to 86.84)	-	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 80: Clinical evidence profile: Aerobic versus strength

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Aerobic exercise**	**Strength**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (VAS, FIQ pain subscale, MDPI, 0-100, final values and change scores, high is poor outcome)												
4	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3]	none	113	86	-	MD 4.47 lower (20.48 lower to 11.54 higher)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, 0-100, change scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	30	30	-	MD 6.7 lower (16.22 lower to 2.82 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values and change scores, high is good outcome)												
3	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	77	50	-	MD 4.29 higher (8.4 lower to 16.98 higher)	⊕OOO VERY LOW	CRITICAL

Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values and change scores, high is good outcome)												
3	randomised trials	very serious[1]	serious[2]	no serious indirectness	serious[3]	none	77	50	-	MD 4.69 higher (6.6 lower to 15.97 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (Multidimensional fatigue inventory-20 reduced activity subscale, change scores, 0-20, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	14	12	-	MD 1 higher (1.18 lower to 3.18 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (6 minute walking test, metres, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	50	25	-	MD 88.4 lower (114.7 to 62.1 lower)	⊕⊕⊕O MODERATE	CRITICAL
Physical function at ≤3 months (Final values and change scores, SF-36 physical functioning subscale, 0-100, high is good outcome)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	43	-	MD 1.85 higher (3.79 lower to 7.49 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (Hospital anxiety and depression anxiety score, 0-21, final values and change scores, high is poor outcome)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	27	25	-	MD 0.93 lower (2.46 lower to 0.61 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (Final values and change scores, Hospital anxiety and depression scale, depression score, 0-21, high is poor outcome)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	27	25	-	MD 0.04 higher (1.37 lower to 1.46 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (Final values, BDI, 0-60, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	50	25	-	MD 12.7 higher (9.01 to 16.39 higher)	⊕OOO VERY LOW	CRITICAL
Sleep at ≤3 months (VAS Sleep scale, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	13	13	-	MD 13.3 lower (31.93 lower to 5.33 higher)	⊕OOO VERY LOW	IMPORTANT
Discontinuation at ≤3 months (due to other diagnoses, transportation problems)												

4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	10/98 (10.2%)	15%	RR 0.67 (0.32 to 1.4)	49 fewer per 1000 (from 102 fewer to 60 more)	⊕⊕OO LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded for heterogeneity, unexplained by subgroup analysis
3 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 81: Clinical evidence profile: Aerobic versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Aerobic exercise**	**Flexibility**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 3 higher (10.19 lower to 16.19 higher)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, 0-100, final values and change scores, high is poor outcome)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	50	44	-	MD 12.65 lower (22.45 to 2.84 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 2.82 higher (1.29 lower to 6.93 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 2.55 higher (2.08 lower to 7.18 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 4.26 higher (1.69 lower to 10.21 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 7.91 higher (2.43 to 13.39 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (BDI, 0-21, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	32	28	-	MD 0.44 higher (6.83 lower to 7.71 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 0.74 lower (4.53 lower to 3.05 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (State trace anxiety inventory, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 1.83 lower (6.33 lower to 2.67 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (State trace anxiety inventory, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	32	28	-	MD 4.83 lower (9.22 to 0.44 lower)	⊕OOO VERY LOW	CRITICAL
Discontinuation at >3 months												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/38 (26.3%)	15.8%	RR 1.67 (0.67 to 4.13)	106 more per 1000 (from 52 fewer to 495 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 82: Clinical evidence profile: Aerobic exercise versus biomechanical exercise

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic exercise versus biomechanical	Control	Relative (95% CI)	Absolute		
Quality of life at 12 weeks (SF36 role social subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 10.6 lower (27.34 lower to 6.14 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 general health status subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 2 lower (15.89 lower to 11.89 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 vitality subscale, 0-100, high score is good outcome) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 1.2 lower (12.43 lower to 10.03 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 functional capacity subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 9.6 lower (21.76 lower to 2.56 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 role physical subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 14.3 lower (35.85 lower to 7.25 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 emotional aspects subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 9 lower (34.66 lower to 16.66 higher)	⊕OOO VERY LOW	CRITICAL

Quality of life at 12 weeks (SF36 pain subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 7 lower (18.72 lower to 4.72 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 mental health subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	21	-	MD 10.9 lower (25.37 lower to 3.57 higher)	⊕OOO VERY LOW	CRITICAL
Sleep at 12 weeks (Pittsburgh Sleep Quality Index, 0-21, high score is poor outcome) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 0.4 lower (2.64 lower to 1.84 higher)	⊕⊕OO LOW	CRITICAL
Pain at 12 weeks (VAS, 0-10, high score is poor outcome) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	21	-	MD 0.6 lower (1.79 lower to 0.59 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at 12 weeks (Scale of Catastropic Thoughts on Pain, 0-5, high score is poor outcome) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[1]	none	21	21	-	MD 0.2 lower (1.08 lower to 0.68 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at 12 weeks												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	2/21 (9.5%)	4/21 (19%)	RR 0.50 (0.10 to 2.44)	95 fewer per 1000 (from 171 fewer to 274 more)	⊕OOO VERY LOW	IMPORTANT

1 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

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

Table 83: Clinical evidence profile: Aerobic and strength versus aerobic

Quality assessment							No of patients		Effect		
										Quality	Importance

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic and strength	Aerobic	Relative (95% CI)	Absolute		
Quality of life at >3 months (FIQ, 0-100, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	21	22	-	MD 0 higher (7.78 lower to 7.78 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	21	22	-	MD 2.1 higher (1.66 lower to 5.86 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at >3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	4/21 (19%)	18.2%	RR 1.05 (0.3 to 3.66)	9 more per 1000 (from 127 fewer to 484 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 84: Clinical evidence profile: Aerobic and strength versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic and strength	Flexibility	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	41	44	-	MD 4 lower (9.96 lower to 1.96 higher)	⊕OOO VERY LOW	CRITICAL

Pain at >3 months (VAS, 0-100, final values, high is poor outcome												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	40	-	MD 8 lower (13.89 to 2.11 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (NIH CPSI quality of life subscale, 0-12, final values, high is poor outcome												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	41	44	-	MD 1.8 lower (2.69 to 0.91 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (NIH CPSI quality of life subscale, 0-12, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	40	-	MD 1.8 lower (2.68 to 0.92 lower)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (BDI, 0-21, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	41	44	-	MD 0.5 higher (1.33 lower to 2.33 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	40	-	MD 0.5 higher (0.97 lower to 1.97 higher)	⊕OOO VERY LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10/52 (19.2%)	9.8%	RR 1.96 (0.72 to 5.34)	94 more per 1000 (from 27 fewer to 425 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 85: Clinical evidence profile: Aerobic and flexibility versus mind-body

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic and flexibility	Mind-body	Relative (95% CI)	Absolute		
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 1.5 lower (4.65 lower to 1.65 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	75	-	MD 3.2 lower (6.38 to 0.02 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 physical component summary score, 0-100, change scores, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 2.8 lower (6.65 lower to 1.05 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 mental component summary score, 0-100, change scores, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 2.4 lower (7.88 lower to 3.08 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (6 minute walking test change scores, metres, change scores, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 1.9 higher (25.15 lower to 28.95 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (6 minute walking test change scores, metres, change scores, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 22.2 lower (60.46 lower to 16.06 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS: depression, 0-21, change scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 1.2 higher (0.68 lower to 3.08 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS: anxiety, 0-21, change scores, high is poor outcome)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	75	-	MD 1.8 higher (0.4 to 3.2 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (HADS: anxiety, 0-21, change scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	36	75	-	MD 1.8 higher (0.12 to 3.48 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress at >3 months (HADS: depression, 0-21, change scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 1.6 higher (0.86 lower to 4.06 higher)	⊕⊕OO LOW	CRITICAL
Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 0.7 higher (0.74 lower to 2.14 higher)	⊕⊕OO LOW	IMPORTANT
Sleep at >3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	75	-	MD 0.8 higher (1.14 lower to 2.74 higher)	⊕⊕OO LOW	IMPORTANT
Discontinuation at ≤3 months												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	11/36 (30.6%)	22.7%	RR 1.35 (0.71 to 2.57)	79 more per 1000 (from 66 fewer to 356 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 86: Clinical evidence profile: Aerobic exercise and flexibility versus aerobic exercise

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Aerobic and flexibility versus aerobic**	**Control**	**Relative (95% CI)**	**Absolute**		

Pain perception at ≤3 months (Final score; VAS) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 0.65 lower (0.86 to 0.44 lower)	⊕⊕⊕O MODERATE	CRITICAL
Pain perception at >3 months (Final score; VAS) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 0.94 lower (1.14 to 0.74 lower)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at ≤3 months (final score; FIQ) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 5.49 lower (7.46 to 3.52 lower)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (final score; FIQ) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 10.62 lower (12.34 to 8.9 lower)	⊕⊕⊕O MODERATE	CRITICAL
Sleep quality at ≤3 months (final score; Pittsburgh Sleep Quality Index) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 3.94 lower (4.62 to 3.26 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Sleep quality at >3 months (final score; Pittsburgh Sleep Quality Index) (Copy) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	32	32	-	MD 5.03 lower (5.51 to 4.55 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Discontinuation at >3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/32 (0%)	0/32 (0%)	RD 0 (-0.06 to 0.06_	-	⊕⊕⊕O MODERATE	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias

Table 87: Clinical evidence profile: Aerobic, strength, mind-body and proprioception versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic, strength, mind-body and propioception versus flexibility	Control	Relative (95% CI)	Absolute		
Quality of life at 7 weeks (FIQ total score, high is poor outcome) (Better indicated by lower values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	11	10	-	MD 13.04 lower (21.92 to 4.16 lower)	⊕⊕OO LOW	CRITICAL
Physical function at 7 weeks (number of steps, high is good outcome) (Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	11	10	-	MD 9.19 higher (11.24 lower to 29.62 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at 7 weeks												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	5/16 (31.3%)	9/19 (47.4%)	RR 0.66 (0.28 to 1.57)	161 fewer per 1000 (from 341 fewer to 270 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 88: Clinical evidence profile: Strength versus mind-body

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Strength versus mind-body	Control	Relative (95% CI)	Absolute		
Pain (VAS, <3 months) (follow-up 6 weeks; range of scores: 0-10; Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	18	-	MD 1.1 higher (0.31 lower to 2.51 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life (Nottingham health profile, <3 months) (range of scores: 0-600; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	18	-	MD 56.1 higher (13.21 lower to 125.41 higher)	⊕OOO VERY LOW	
Physical function (NDI, <3 months) (follow-up 6 weeks; range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	18	-	MD 3.1 higher (0.56 lower to 6.76 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress (BDI, <3 months) (follow-up 6 weeks; range of scores: 0-63; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	18	-	MD 3.3 higher (1.24 lower to 7.84 higher)	⊕OOO VERY LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	12/60 (20%)	12.9%	RR 1.55 (0.68 to 3.52)	71 more per 1000 (from 41 fewer to 325 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 89: Clinical evidence profile: Strength versus biomechanical

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Strength versus biomechanical**	**Control**	**Relative (95% CI)**	**Absolute**		
Pain (VAS, <3 months) (follow-up 6 weeks; range of scores: 0-10; Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 0.8 higher (0.52 lower to 2.12 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life (Nottingham health profile, <3 months) (follow-up 6 weeks; range of scores: 0-600; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 27.7 higher (44.07 lower to 99.47 higher)	⊕OOO VERY LOW	CRITICAL
Physical function (NDI, <3 months) (follow-up 6 weeks; range of scores: 0-100; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 1.3 higher (2.29 lower to 4.89 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress (BDI, <3 months) (follow-up 6 weeks; range of scores: 0-63; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 1.2 higher (3.36 lower to 5.76 higher)	⊕OOO VERY LOW	CRITICAL

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 90: Clinical evidence profile: Strength versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Strength**	**Flexibility**	**Relative (95% CI)**	**Absolute**		
Pain reduction at ≤3 months (VAS, 0-100, change scores and final values, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	44	42	-	MD 8.09 lower (14.58 to 1.59 lower)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 physical component, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	31	-	MD 1.5 higher (2.64 lower to 5.64 higher)	⊕⊕OO LOW	CRITICAL

Quality of life at >3 months (SF-36 mental component, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	35	31	-	MD 5.39 lower (11.75 lower to 0.97 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	16	14	-	MD 6 higher (2.34 to 9.66 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (BDI, 0-61, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	28	28	-	MD 1.83 lower (3.99 lower to 0.33 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (BAI, 0-61, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[1]	none	28	28	-	MD 3.2 lower (6.42 lower to 0.02 higher)	⊕⊕OO LOW	CRITICAL
Sleep at ≤3 months (FIQ sleep subscale, 0-10, change scores, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	28	28	-	MD 1.77 lower (2.62 to 0.92 lower)	⊕⊕⊕O MODERATE	IMPORTANT
Discontinuation at >3 months												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	13/81 (16%)	18/76 (23.7%)	RR 0.68 (0.36 to 1.28)	76 fewer per 1000 (from 152 fewer to 66 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 91: Clinical evidence profile: Strength and flexibility versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Strength and flexibility**	**Flexibility**	**Relative (95% CI)**	**Absolute**		

Quality of life at >3 months (SF-36 physical functioning subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	43	43	-	MD 0.4 lower (4.92 lower to 4.12 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 role physical subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	43	-	MD 1.1 lower (15.9 lower to 13.7 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 role emotional subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	43	-	MD 2.1 higher (9.7 lower to 13.9 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 energy subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	43	43	-	MD 5.2 higher (2.96 lower to 13.36 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 emotional wellbeing subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	43	43	-	MD 3.6 higher (3.43 lower to 10.63 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 social functioning subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	43	-	MD 1.7 higher (5.28 lower to 8.68 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at 12 months (SF-36 bodily pain subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	43	43	-	MD 1.7 lower (10.14 lower to 6.74 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 general health subscale, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	43	43	-	MD 0.7 higher (6.41 lower to 7.81 higher)	⊕⊕⊕O MODERATE	CRITICAL
Discontinuation at >3 months												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	6/49 (12.2%)	17.3%	RR 0.71 (0.27 to 1.84)	50 fewer per 1000 (from 126 fewer to 145 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 92: Clinical evidence profile: Strength and flexibility versus mind-body

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Strength and flexibility**	**Mind-body**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (VAS, 0-100, high is poor outcome)												
2	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3]	none	60	57	-	MD 10.4 lower (23.66 lower to 2.85 higher)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, 0-100, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	70	70	-	MD 0.78 lower (8.05 lower to 6.49 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at ≤3 months (SF-36 mental component, 0-100, high is good outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	57	-	MD 2.88 higher (0.8 lower to 6.55 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 mental component, 0-100, high is good outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	70	70	-	MD 1.05 higher (2.28 lower to 4.38 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at 9-12 weeks (SF-36 physical component, 0-100, high is good outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	60	57	-	MD 1.04 higher (1.9 lower to 3.99 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 physical component, 0-100, high is good outcome)												

2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	70	70	-	MD 2.21 lower (4.81 lower to 0.38 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, neck pain disability scale, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	60	57	-	SMD 0.22 lower (0.59 lower to 0.14 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (Neck pain disability scale, high is poor outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	70	70	-	MD 0.22 higher (5.02 lower to 5.46 higher)	⊕⊕⊕O MODERATE	CRITICAL
Psychological distress at 12 weeks (Depression scale ADS, 0-60, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	35	31	-	MD 0.5 higher (3.66 lower to 4.66 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (Depression scale ADS, 0-60, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[3]	none	35	31	-	MD 1.8 lower (6.07 lower to 2.47 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at >3 months												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[3]	none	10/106 (9.4%)	10.3%	OR 0.87 (0.35 to 2.14)	12 fewer per 1000 (from 64 fewer to 94 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded for heterogeneity, unexplained by subgroup analysis
3 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 93: Clinical evidence profile: Strength, flexibility and proprioception versus mind-body

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Strength, flexibility and proprioception**	**mind-body**	**Relative (95% CI)**	**Absolute**		

Pain reduction at ≤3 months (VAS, 0-100, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 7.2 lower (16.72 lower to 2.32 higher)	⊕⊕OO LOW	CRITICAL
Pain reduction at >3 months (VAS, 0-100, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 1.9 lower (12.99 lower to 9.19 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 2.1 lower (5.48 lower to 1.28 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at >3 months (SF-36 physical component summary score, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 2.5 lower (6.22 lower to 1.22 higher)	⊕⊕OO LOW	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	37	38	-	MD 0.9 higher (3.77 lower to 5.57 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at >3 months (SF-36 mental component summary score, 0-100, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	37	38	-	MD 0.1 lower (4.96 lower to 4.76 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, 0-100, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 1.2 higher (3.7 lower to 6.1 higher)	⊕⊕OO LOW	CRITICAL

Physical function at >3 months (Neck disability index, 0-100, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	37	38	-	MD 0.8 higher (5.31 lower to 6.91 higher)	⊕⊕⊕O MODERATE	CRITICAL
Psychological distress at ≤3 months (HADS: anxiety, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 1 lower (2.8 lower to 0.8 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (HADS: anxiety, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	37	38	-	MD 0.6 lower (2.34 lower to 1.14 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (HADS: depression, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	37	38	-	MD 0.1 lower (1.52 lower to 1.32 higher)	⊕⊕⊕O MODERATE	CRITICAL
Psychological distress at >3 months (HADS: depression, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	37	38	-	MD 0 higher (1.51 lower to 1.51 higher)	⊕⊕⊕O MODERATE	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	none	13/37 (35.1%)	7.9%	RR 4.45 (1.38 to 14.35)	273 more per 1000 (from 30 more to 1000 more)	⊕⊕⊕⊕ HIGH	CRITICAL

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 94: Clinical evidence profile: Strength versus proprioception

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Strength	Proprioception	Relative (95% CI)	Absolute		
Physical function ≤3 months (Neck disability index, 0-50, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	12	14	-	MD 0.32 higher (1.47 lower to 2.11 higher)	⊕⊕⊕O MODERATE	CRITICAL

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias

Table 95: Clinical evidence profile: Mind-body versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mind-body	Flexibility	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, 0-100, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	29	26	-	MD 2 higher (9.65 lower to 13.65 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (FIQ, 0-100, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	25	24	-	MD 22.9 lower (33.4 to 12.4 lower)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (BDI, 0-61, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	42	39	-	MD 0.5 higher (3.55 lower to 4.55 higher)	⊕OOO VERY LOW	CRITICAL

Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	42	39	-	MD 0 higher (1.92 lower to 1.92 higher)	⊕⊕OO LOW	IMPORTANT
Discontinuation at ≤3 months												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	12/30 (40%)	21.9%	RR 1.83 (0.83 to 4.02)	182 more per 1000 (from 37 fewer to 661 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 96: Clinical evidence profile: Mind-body versus biomechanical

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Mindbody versus biomechanical**	**Control**	**Relative (95% CI)**	**Absolute**		
Pain (VAS, <3 months) (follow-up 6 weeks; range of scores: 0-10; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	18	20	-	MD 0.3 lower (1.51 lower to 0.91 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life (Nottingham health profile, <3 months) (follow-up 6 weeks; range of scores: 0-600; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 28.4 lower (84.68 lower to 27.88 higher)	⊕OOO VERY LOW	CRITICAL
Physical function (NDI, <3 months) (follow-up 6 weeks; Better indicated by lower values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 1.8 lower (4.86 lower to 1.26 higher)	⊕OOO VERY LOW	CRITICAL
Psychological distress (Depression, BDI, <3 months) (follow-up 6 weeks; range of scores: 0-63; Better indicated by lower values)												

1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	18	20	-	MD 2.1 lower (6.11 lower to 1.91 higher)	⊕OOO VERY LOW	CRITICAL

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 97: Clinical evidence profile: Flexibility and proprioception versus flexibility

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Flexibility and proprioception**	**Flexibility**	**Relative (95% CI)**	**Absolute**		
Quality of life at ≤3 months (FIQ, 0-100, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	28	29	-	MD 12.7 lower (21.27 to 4.13 lower)	⊕OOO VERY LOW	CRITICAL
Psychological distress at ≤3 months (BDI, 0-63, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	28	29	-	MD 3.88 higher (0.46 lower to 8.22 higher)	⊕OOO VERY LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	7/35 (20%)	12.1%	RR 1.65 (0.53 to 5.12)	79 more per 1000 (from 57 fewer to 499 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 98: Clinical evidence profile: Flexibility and relaxation versus aerobic

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Flexibility and relaxation	Aerobic	Relative (95% CI)	Absolute		
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	65	68	-	MD 0.4 higher (4.64 lower to 5.44 higher)	⊕⊕⊕O MODERATE	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	12/69	12/67	RR 0.97 (0.47 to 2.01)	10 fewer per 1000 (from 130 fewer to 120 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 99: Clinical evidence profile: Exercise versus psychological therapies

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Exercise	Psychological therapies	Relative (95% CI)	Absolute		
Pain at ≤3 months (VAS, FIQ pain scale, 0-100, high is poor outcome)												
4	randomised trials	serious[1]	very serious[2]	no serious indirectness	very serious[3]	none	131	120	-	MD 1.61 lower (15.09 lower to 11.87 higher)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (VAS, NRS, 0-100, high is poor outcome)												
4	randomised trials	serious[1]	serious[2]	no serious indirectness	serious[3]	none	121	110	-	MD 7.19 lower (13.98 to 0.41 lower)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (FIQ, 0-100, high is poor outcome)												

4	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	151	141	-	MD 6.7 lower (10.88 to 2.52 lower)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (EQ-5D, high is good outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	81	71	-	MD 0.05 lower (0.12 lower to 0.02 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 social aspects subscale, 0-100, high score is good outcome (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	30	30	-	MD 3.4 higher (9.27 lower to 16.07 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 general health status aspects subscale, 0-100, high score is good outcome (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	30	30	-	MD 2.6 higher (8.08 lower to 13.28 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 funcitonal capacity aspects subscale, 0-100, high score is good outcome (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	30	30	-	MD 13.1 higher (2.72 to 23.48 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 limitations due to physical aspects subscale, 0-100, high score is good outcome (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	30	30	-	MD 17.2 higher (2.83 lower to 37.23 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 limitations due to emotional aspects subscale, 0-100, high score is good outcome (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	30	30	-	MD 11.9 higher (8.74 lower to 32.54 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at 12 weeks (SF36 pain subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	30	30	-	MD 5 higher (5.39 lower to 15.39 higher)	⊕OOO VERY LOW	CRITICAL

Quality of life at 12 weeks (SF36 mental health subscale, 0-100, high score is good outcome) (Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	30	30	-	MD 0.9 higher (11.04 lower to 12.84 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (FIQ physical function subscale, 0-10, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	51	47	-	MD 0.7 lower (2.75 lower to 1.35 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at ≤3 months (6 minute walk test, metres, high is good outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	76	63	-	MD 26.42 higher (0.85 lower to 53.69 higher)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (6 minute walking test, metres, high is good outcome)												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	86	79	-	MD 49.05 higher (25.45 to 72.65 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at ≤3 months (CES-D, 0-100, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	34	28	-	MD 10.3 lower (20.07 to 0.53 lower)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (Hospital anxiety and depression scale, depression subscale, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	48	-	MD 1 lower (2.25 lower to 0.25 higher)	⊕⊕OO LOW	CRITICAL
Psychological distress at >3 months (Hospital anxiety and depression scale, anxiety subscale, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	49	-	MD 0.8 lower (2.01 lower to 0.41 higher)	⊕⊕OO LOW	CRITICAL
Sleep at >3 months (the sleep scale, 0-30, final values, high is poor outcome)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	99	91	-	MD 0.3 higher (1.22 lower to 1.82 higher)	⊕⊕⊕O MODERATE	IMPORTANT
Sleep at >3 months (Pittsburgh sleep quality index, 0-21, high is poor outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	56	49	-	MD 1.1 lower (2.32 lower to 0.12 higher)	⊕⊕OO LOW	IMPORTANT
Discontinuation at >3 months (due to increased pain, personal reasons, lost to follow up)												
10	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	78/540 (14.4%)	90/522 (17.2%)	RD -0.03 (-0.07 to 0.02)	30 fewer per 1000 (from 70 fewer to 20 more)	⊕⊕OO LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs.

Table 100: Clinical evidence profile: Manual therapy and exercise versus manual therapy

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Manual therapy and exercise**	**Manual therapy**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (NRS, high is poor outcome, final values, 0-10)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	51	50	-	MD 0.8 lower (1.66 lower to 0.06 higher)	⊕⊕OO LOW	CRITICAL
Pain at >3 months (NRS, high is poor outcome, final values, 0-10)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	51	50	-	MD 0.5 lower (1.42 lower to 0.42 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	51	50	-	MD 5.1 lower (9.65 to 0.55 lower)	⊕⊕OO LOW	CRITICAL

Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	51	50	-	MD 4.9 lower (9.85 lower to 0.05 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	13/64 (20.3%)	14/63 (22.2%)	RR 0.91 (0.47 to 1.79)	20 fewer per 1000 (from 118 fewer to 176 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Table 101: Clinical evidence profile: Manual therapy and exercise versus exercise

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Manual therapy and exercise**	**Exercise**	**Relative (95% CI)**	**Absolute**		
Pain at ≤3 months (VAS, NRS, high is poor outcome, final values, 0-100)												
6	randomised trials	serious[1]	very serious[2]	no serious indirectness	serious[3]	none	265	277	-	MD 6.34 lower (13.82 lower to 1.13)	⊕OOO VERY LOW	CRITICAL
Pain at >3 months (NRS, VAS, high is poor outcome, final values, 0-100)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	198	196	-	MD 0.95 higher (3.51 lower to 5.4 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (Fibromyalgia impact questionnaire, 0-100, final values, high is poor outcome)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	10	11	-	MD 1 lower (13.87 lower to 11.87 higher)	⊕OOO VERY LOW	CRITICAL
Quality of life at ≤3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												

1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	91	89	-	MD 0.6 higher (1.34 lower to 2.54 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 physical component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	91	89	-	MD 0.2 higher (1.79 lower to 2.19 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at ≤3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	91	89	-	MD 0.7 lower (3.55 lower to 2.15 higher)	⊕⊕⊕O MODERATE	CRITICAL
Quality of life at >3 months (SF-36 mental component summary score, 0-100, final values, high is good outcome)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	91	89	-	MD 1.8 lower (4.34 lower to 0.74 higher)	⊕⊕⊕O MODERATE	CRITICAL
Physical function at >3 months (Neck disability index, functional performance scale, final values, high is poor outcome, 0-100)												
5	randomised trials	serious[1]	serious[3]	no serious indirectness	serious[2]	none	239	238	-	SMD 0.29 lower (0.62 lower to 0.04 higher)	⊕OOO VERY LOW	CRITICAL
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-100)												
3	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	198	196	-	MD 0.17 lower (2.6 lower to 2.25 higher)	⊕⊕⊕O MODERATE	CRITICAL
Physical function at 4-10 weeks (Neck disability index, high is poor outcome, 0-100) (Better indicated by lower values)												
2	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	36	50	-	MD 8.14 lower (9.92 to 6.35 lower)	⊕⊕OO LOW	CRITICAL
Discontinuation at ≤3 months												
6	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[1]	none	36/275 (13.1%)	34/267 (12.7%)	RD 0 (-0.05 to 0.06)	0 fewer per 1000 (from 50 fewer to 60 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs
3 Downgraded for heterogeneity, unexplained by subgroup analysis

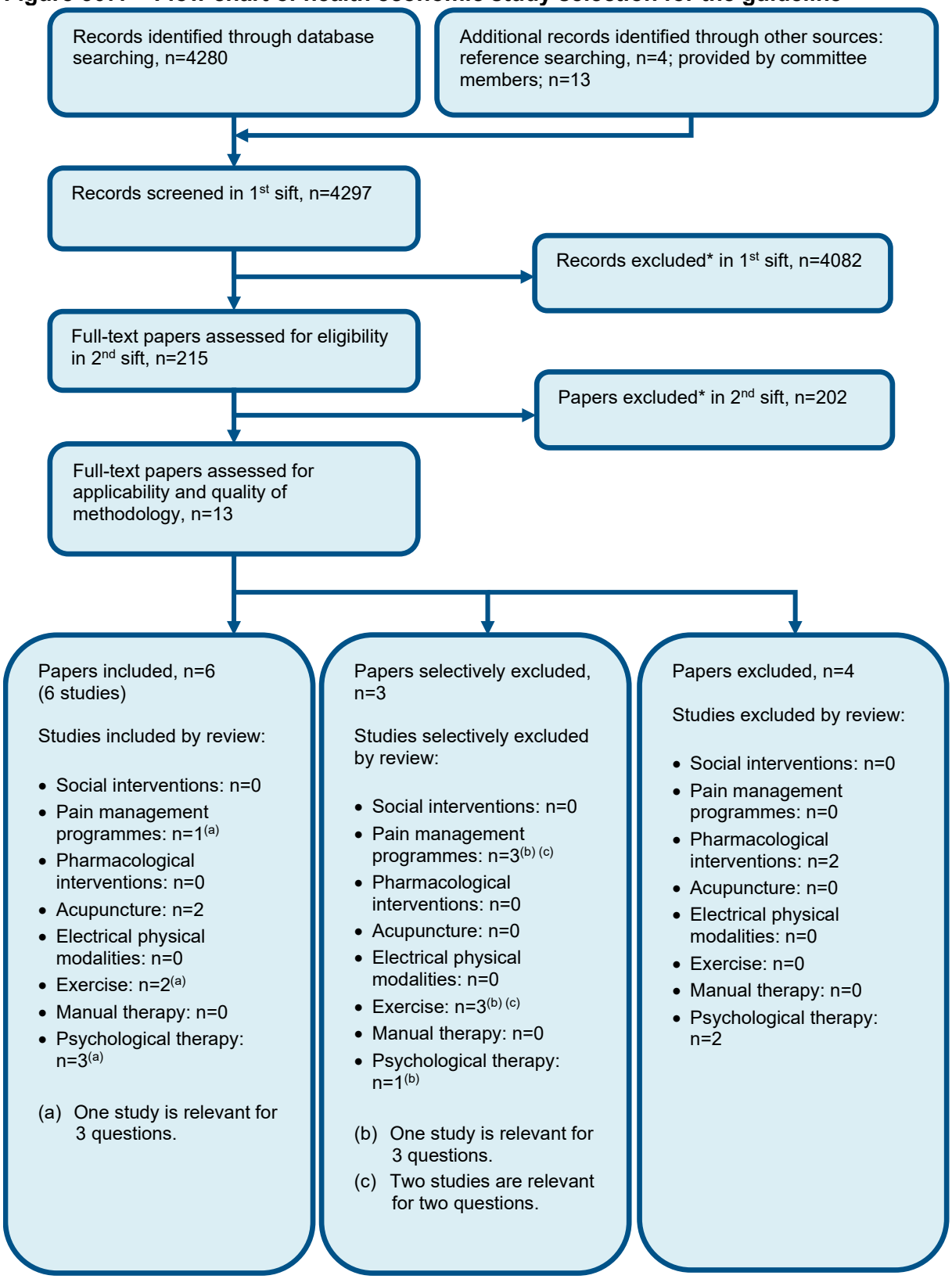
Table 102: Clinical evidence profile: Exercise versus manual therapy

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Exercise	Manual therapy	Relative (95% CI)	Absolute		
Pain at ≤3 months (NRS, high is poor outcome, final values, 0-10)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	51	50	-	MD 1.3 lower (2.11 to 0.49 lower)	⊕⊕OO LOW	CRITICAL
Pain at >3 months (NRS, high is poor outcome, final values, 0-10)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	51	50	-	MD 0.5 lower (1.42 lower to 0.42 higher)	⊕⊕OO LOW	CRITICAL
Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	44	50	-	MD 5.9 lower (10.6 to 1.2 lower)	⊕⊕OO LOW	CRITICAL
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	44	50	-	MD 3.9 lower (9.14 lower to 1.34 higher)	⊕⊕OO LOW	CRITICAL
Discontinuation at ≤3 months												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	19/64 (29.7%)	22.2%	RR 1.34 (0.74 to 2.43)	75 more per 1000 (from 58 fewer to 317 more)	⊕OOO VERY LOW	IMPORTANT

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, or by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed 1 MID or by 2 increments if the confidence interval crossed both MIDs

Appendix G: Health economic evidence selection

Figure 307: Flow chart of health economic study selection for the guideline



* Non-relevant population, intervention, comparison, design or setting; non-English language

Appendix H: Health economic evidence tables

<table>
<tr><th>Study</th><th colspan="4">Beasley (2015)[28]</th></tr>
<tr><th>Study details</th><th>Population & interventions</th><th>Costs</th><th>Health outcomes</th><th>Cost effectiveness</th></tr>
<tr>
<td>Economic analysis: CUA (health outcome: QALYs)

Study design: Within-trial analysis (RCT – clinical results in same paper)

Approach to analysis: Analysis of individual data for EQ-5D (adjusted for baseline differences in utility) and resource use. Unit costs applied.

Perspective: UK NHS

Follow-up: 30 months*

Discounting: Costs: 3.5%; Outcomes: 3.5%</td>
<td>Population:
People aged 25 years and over with chronic widespread pain according to the definition in the American College of Rheumatology (ACR) 1990 criteria for fibromyalgia, for which they have consulted their general practitioner in the previous year.

Patient characteristics:
N = 442 (in all four arms)
Age: 56.3
Male: 30.5%

Intervention 1:
Treatment as usual (from GP – precise care delivered not recorded)

Intervention 2:
Telephone-delivered cognitive behaviour therapy (TCBT): initial assessment (45-60mins) followed by 7 weekly sessions (30-45mins each), 1 session at three months, and 1 session at 6 months. Intervention delivered by 4 therapists</td>
<td>Incremental costs (mean per patient):

Intervention 1 is the reference.

<u>Complete cases</u>
Intervention 1: £0
Intervention 2: £574
Intervention 3: £1,924
Intervention 4: £1,778

<u>Multiple imputations</u>
Intervention 1: £0
Intervention 2: £554
Intervention 3: £1,256
Intervention 4: £1,453

Currency & cost year:
2010 UK pounds

Cost components incorporated:
• Intervention costs (for exercise this includes gym membership)</td>
<td>Incremental QALYs (mean per patient):

Intervention 1 is the reference.

<u>Complete cases</u>
Intervention 1: 0
Intervention 2: 0.097
Intervention 3: 0.025
Intervention 4: 0.047

<u>Multiple imputations</u>
Intervention 1: 0
Intervention 2: 0.140
Intervention 3: 0.071
Intervention 4: 0.096</td>
<td>ICER:
Full incremental analysis (complete cases, adjusted) (pa):
<table>
<tr><th>Int</th><th>Inc cost</th><th>Inc QALY</th><th>ICER</th><th>ICER (ruled out dominated options)</th></tr>
<tr><td>1</td><td>£0</td><td>£0</td><td>Reference</td><td>-</td></tr>
<tr><td>2</td><td>£574</td><td>0.097</td><td>£5,917</td><td>£5,917</td></tr>
<tr><td>3</td><td>£1,924</td><td>0.025</td><td>£76,960</td><td>Dominated</td></tr>
<tr><td>4</td><td>£1,778</td><td>0.047</td><td>£37,830</td><td>Dominated</td></tr>
</table>
Probability Intervention 2 cost effective (£20K threshold): approx. 75% (read off graph)

Full incremental analysis (multiple imputations, adjusted) (pa):
<table>
<tr><th>Int</th><th>Inc cost</th><th>Inc QALY</th><th>ICER</th><th>ICER (ruled out dominated options)</th></tr>
</table>
</td>
</tr>
</table>

1	£0	0	Reference	-
2	£554	0.140	£3,957	£3,957
3	£1,256	0.071	£17,690	Dominated
4	£1,453	0.096	£15,135	Dominated

Probability Intervention 2 cost effective (£20K/30K threshold): NR

Analysis of uncertainty: Used non-parametric bootstrapping. Multiple imputation was also used to assess the sensitivity of findings to missing data.

accredited by the British Association for Behaviour and Cognitive Psychotherapies. Therapists conducted a patient-centred assessment, developed shared understanding and formulation of the participants' problem(s) and identified two to three patient-defined goals. Patients also received a self-management CBT manual that included: behavioural activation, cognitive restructuring, unhelpful thinking and lifestyle changes.

Intervention 3:

Exercise therapy: leisure-facility-and-gym-based exercise program consistent with American College of Sport Medicine (ACSM) guidelines for improving cardiorespiratory fitness. Following an induction sessions, patients were offered 6 fitness instructor-led monthly appointments. Experienced fitness instructors delivered the intervention following a 1-day training session on exercise prescription for people with CWP. The specific exercises are negotiated between fitness instructor and patient, and can be changed while maintaining goal of improving cardio-respiratory fitness. Initial intensity was low to moderate,

- Routine health service (GP, nurse, physio, community visits, outpatient, inpatient, admission, primary care).

	patients were free to engage in additional exercises to those prescribed. Recommended session duration was 20-60 mins, patients were advised to attend at least twice a week and engage in 'everyday' activities on non-gym days. **Intervention 4:** Combination of Interventions 2 and 3.			

Data sources
*The follow up is 24 months post treatment, and given that the exercise and CBT interventions were about 6 months in length then that equates to a 30 month follow up. **Health outcomes:** Resource use was reported to 3 months post treatment, and at months 18-24 post treatment. Linear interpolation between reported health service costs at 3 and 24 months post treatment was used to impute an average cost per quarter for the 5 quarters not covered by data collection (i.e. months 3-6, 6-9, 9-12, 12-15 and 15-18 post treatment). **Quality-of-life weights:** EQ-5D UK tariff. QALYs calculated using patient response to EQ-5D at 24 months post-treatment. Additional QALYs accrued between 3 and 24 months post treatment were calculated for each person assuming a linear change in utility. **Cost sources:** Cost sources were the same as those used for the original McBeth 2012 economic evaluation that this paper is also based on, which are PSSU 2010, and NHS reference costs 2008/9
Comments
Source of funding: Arthritis Research UK. **Limitations:** Participation in study based on self-reported symptoms and recruited through primary care, may not necessarily be representative of general population with chronic widespread pain caused by fibromyalgia. Treatment as usual not defined, usual care provided by GP was not restricted and may not be the same across all participants in that group. Within-study analysis which may not reflect full body of evidence. The adjusted results are quite different to the unadjusted results for some of the interventions more than others (e.g. the QALYs for exercise are much lower in the adjusted analysis - lower than the combined intervention, whereas they are higher than the combined intervention in the unadjusted analysis. This can lead to a large change in the exercise ICER versus treatment as usual: making exercise cost effective in the unadjusted analysis). **Other:** Analyses were adjusted for: age, sex, baseline pain on CPG (chronic pain grade) scale, baseline GHQ (general health questionnaire) score and study centre.
Overall applicability:[a] Directly applicable **Overall quality:**[b] Potentially serious limitations

Abbreviations: 95% CI: 95% confidence interval; CUA: cost–utility analysis; da: deterministic analysis; 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; NR: not reported; pa: probabilistic analysis; 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	Gusi 2008[116]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: CUA (health outcome: QALYs) **Study design:** Within trial analysis **Approach to analysis:** Analysis of individual data for EQ-5D (adjusted for baseline differences in utility) and resource use. Unit costs applied. **Perspective:** Spanish healthcare perspective **Follow-up:** 8 months **Treatment effect duration:**[(a)] 8 months **Discounting:** Costs: NA; Outcomes: NA	**Population:** Women with fibromyalgia **Patient characteristics:** N: 33 Age: 50 **Intervention 1:** Usual care: included standard medical attention in the public system (hospital and outpatient clinic including primary care) and the social support of the local FM association. **Intervention 2:** Exercise + usual care: Exercise programme in a waist high pool of warm water (33°C). A qualified exercise leader instructed and trained the group three times a week for 1 h per session over a period of 8 months. Each session included 10 min of warm up with slow walking and easy movements of progressive	**Total costs (mean per patient):** Intervention 1: NR Intervention 2: NR Incremental (2−1): £475 (95% CI: NR; p=NR) **Currency & cost year:** 2005 Euros (presented here as 2005 UK pounds[(b)]) **Cost components incorporated:** - Programme cost (based on staff costs, renting the pool, management costs of the programme like insurance). - Health care costs (consultations, drug process).	**QALYs (mean per patient):** Intervention 1: 0.002 Intervention 2: 0.133 Incremental (2−1): 0.131 (95% CI: 0.011 to 0.290; p=NR)	**ICER (Intervention 2 versus Intervention 1):** £3,630 per QALY gained (bootstrapped estimate) 95% CI: £1,639 to £43,220 Probability Intervention 2 cost effective: Determined by reading off the graph based on the '2005 adjusted investment ceiling set at €34,729/QALY): approx. 97% **Analysis of uncertainty:** Calculated the 95% confidence interval using the non-parametric bootstrapping technique (1,000 iterations). Sensitivity analyses: From the health system perspective: - 30% less patients per group - 30% more patients per group - 30% lower salary (monitor and nurse) - 30% higher salary (monitor and nurse) - No additional salary of nurse - Best case scenario of salary, participation and effectiveness (rental + participation more persons per group + QALY differential at higher limit of 95% confidence interval).

	intensity, 10 min of aerobic exercises at 60–65% of maximal heart rate, 20 min of overall mobility and lower limb strength exercises using water resistance, another set of 10 min of aerobics at 60–65% of maximal heart rate, and 10 min cool down with low intensity exercises.			- Worst case scenario of salary, participation and effectiveness (opposite of above). All the above had ICERS below the threshold mentioned above (€34,729/QALY), except for the worst case scenario (€75,455/QALY). Similar analyses were also undertaken from the societal perspective.
Data sources				
Health outcomes: Based on the Tomas Carus 2008/2009 trials.[253],[252] **Quality-of-life weights:** EQ-5D Spanish tariff. Measured at baseline and 3 months and 8 months. To avoid bias, data were adjusted by regression analysis for differences in baseline EQ-5D scores. **Cost sources:** The unit costs are expressed in Euros (€) based on prices in 2005. The programme's cost based on: salaries at the level for a university graduate, cost of staff to run the programme, salaries at minimum wage for the patient's time (based on the 2005 official bulletin of the regional government), cost of renting a pool at a university at public prices without a grant, public bus prices, and private external management costs of the programme (insurance, monthly retrievals from patients and withdrawals to employees). Health care prices (consultations, etc.) were based on the 2005 official bulletin of the regional government. Drug prices were obtained from the Spanish version of Vademecum International. Costs were analysed from a healthcare and also from a social care perspective in a separate analysis (including patient costs like travel).				
Comments				
Source of funding: NR **Limitations:** Uses EQ-5D. Non-UK study. Only based on one study. Date and costs may not reflect current NHS context. Recruitment of participants was through local FM association, perhaps not representative of wider population with FM. **Other:**				
Overall applicability: Partially applicable(c) **Overall quality Potentially serious limitations**(d)				

Abbreviations: CUA= cost–utility analysis; da= deterministic analysis; 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; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years; FM = Fibromyalgia.

(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 2005 purchasing power parities[208]

(c) Directly applicable / Partially applicable / Not applicable

(d) Minor limitations / Potentially serious limitations / Very serious limitations

Appendix I: Excluded studies

I.1 Excluded clinical studies

Table 103: Studies excluded from the clinical review

Study	Exclusion reason
Acosta-Gallego, 2018 [2]	Incorrect comparison (land versus pool based exercises)
Actrn, 2018 [3]	Clinical trial registry
Adamse 2018 [4]	Systematic review with different PICO
Alentorn-Geli, 2008 [7]	Whole body vibration
Alentorn-Geli, 2009 [6]	No useable outcomes
Allende, 2018 [8]	No useable outcomes
Amanollahi 2013 [11]	Not in English
Amris 2014 [13]	Incorrect intervention: pain management programme
Andersen 2008 [14]	No useable outcomes
Andrade 2017 [16]	No useable outcomes
Andrade 2018 [15]	Systematic review, incorrect study design: non-randomised
Anonymous 2019 [76]	Incorrect comparison: both groups received TENS and hot packs in addition to interventions
Arami 2012 [18]	Not in English
Arcos-Carmona 2011 [19]	Not in English
Arimi 2017 [12]	Systematic review with different PICO
Asenlof 2005 [20]	Incorrect intervention: pain management programme
Asenlof 2009 [21]	Incorrect intervention: psychological
Assis 2006 [22]	Incorrect comparison: aerobic comparison
Assuncao Junior 2018 [24]	Incorrect study design: no comparator
Astin 2003 [25]	Incorrect comparison: exercise and meditation versus education
Bai 2015 [26]	Systematic review, incorrect population
Beltran-Alacreu 2015 [29]	Incorrect interventions: pain management programme
Bertozzi 2013 [31]	Systematic review with different PICO
Bidonde 2019 [32]	Cochrane review published after review finalised; references checked
Bjersing 2017 [35]	Subgroup analysis, not relevant
Bland 2010 [36]	Abstract
Bobos 2016 [37]	Incorrect comparison: different strength training protocols
Bowering 2013 [40]	Systematic review with different PICO
Brage 2015 [41]	Incorrect comparison: education
Bravo 2019 [42]	Incorrect intervention: body awareness therapy
Buckelew 1998 [44]	No useable outcomes
Burckhardt 1992 [45]	Abstract
Burckhardt 1994 [46]	No useable outcomes: no variation data
Busch 2007 [47]	Cochrane review, incorrect comparison
Busch 2008 [48]	Systematic review with different PICO
Cantarero-Villanueva 2012 [51]	Incorrect population
Carbonell-Baeza 2012 [52]	Protocol

Cerrillo-Urbina 2015 [56]	Systematic review with different PICO
Champagne 2018 [57]	Abstract
Chan 2012 [58]	Systematic review with different PICO
Cho 2016 [60]	Incorrect comparison: lymphatic drainage
Chung 2018 [61]	Incorrect comparison: different neck exercises
Collado-Mateo 2017 [62]	Incorrect intervention: virtual reality
Cramer 2013 [64]	Non-comparative follow up data
Cramer 2017 [63]	Systematic review with different PICO
Cramer 2017 [66]	Cochrane review, incorrect population: breast cancer pain
de Araujo Cazotti 2018 [68]	Incorrect comparison: pharmacological
Demir-Gocmen 2013 [70]	Incorrect comparison: supervised versus home exercises
Dobkin 2005 [71]	Incorrect study design: no comparator
Dunleavy 2016 [72]	Incorrect study design (not randomised)
Duray, 2018 [73]	Incorrect comparison (not relevant)
Duruturk 2015[74]	No useable outcomes
Dusunceli 2006 [75]	Incorrect comparison: TENS
Ekici 2008 [77]	Not in English
Emilson 2017 [79]	Incorrect comparison: pain management, follow-up study
Ernberg 2016 [82]	Incorrect comparison: healthy controls
Evcik 2008 [87]	Incorrect comparison: land versus water based, same exercises
Falla 2006 [88]	Incorrect comparison: different strength training protocols
Falla 2007 [89]	Incorrect comparison: different neck exercise protocols
Fernandes 2016 [91]	Incorrect comparison: swimming versus walking
Field 2003 [92]	Incorrect comparison: exercise and manual therapy versus relaxation
Fontaine 2007 [95]	Incorrect intervention (exercise and psychological therapy)
Fontaine 2011 [93]	No comparator
Galindez-ibarbengoetxea 2018 [96]	Unclear intervention time
Garcia-Hermoso 2015 [98]	Systematic review with different PICO
Geneen 2017 [102]	Cochrane review, incorrect population: chronic non-cancer pain
Ghaderi 2017 [103]	Incorrect comparison: neck stabilisation exercises versus neck strengthening, both interventions offer exercises to strengthen neck muscles
Ghodrati 2020 [104]	Incorrect comparison: manual therapy vs. manual therapy + exercise
Giannotti 2014 [105]	Incorrect interventions: physical and psychological elements, pain management programme
Gowans 1999 [111]	Not guideline condition. Not review population. No extractable data. Wrong study type: results are not extractable
Gowans 2004 [112]	No comparator
Gowans 2007 [109]	Systematic review with different PICO
Gross 2015 [114]	Cochrane review, incorrect population, different outcomes: with some overlap
GunendiZ 2008 [115]	Incorrect interventions: exercise combined with TENS and thermotherapy
Gutierrez-Espinoza 2019 [117]	Incorrect intervention: targeted at improving range of movement in the glenohumeral joint only and doesn't fall into any protocol categories of general exercise
Hakkinen 2002 [120]	No relevant outcomes

Hammond 2006 [121]	Incorrect comparison: relaxation versus exercise and education
Har 2000 [122]	Not available
Hoeger Bement 2011 [123]	Incorrect comparison (not relevant)
Humphreys 2002 [126]	Incorrect comparison: healthy controls
Iaroshevskyi, 2019 [127]	Incorrect study design
Ide 2008 [128]	Incorrect interventions: breathing exercises
Im 2013 [129]	Incorrect intervention, incorrect comparison: whirlpool therapy versus warm gel packs
Isomeri 1992 [130]	Abstract
Isomeri 1993 [131]	No useable outcomes
Jensen 2001 [133] (Bergstrom 2012 [30])	Incorrect population (low back pain)
Jentoft, 2001 [134]	Incorrect interventions: pool based versus land based, same exercise protocol
Jones 2011 [135]	Summary article
Jordan 1998 [137]	No useable outcomes
Jull 2009 [138]	Incorrect comparison: psychological therapies
Kalamir [139]	No relevant outcomes
Kaleth 2013 [140]	No useable outcomes
Kay 1992 [141]	No useable outcomes
Keel, 1998 [143]	Incorrect intervention: pain management programme
Kelley 2010 [144]	Systematic review with different PICO
Khan 2014 [146]	Incorrect comparison (both groups are different types of strength exercises)
Khan, 2018 [145]	Incorrect comparison (not relevant)
Kim 2019 [150]	Cochrane review published after review finalised; references checked
Kim 2016 [148]	Incorrect comparison: different neck exercise protocols
Kim 2016 [149]	Incorrect comparison: manual therapy versus ultrasound
Kim 2016 [151]	Systematic review with different PICO
Lagueux 2014 [154]	Conference abstract
Langhorst 2009 [156]	Systematic review, incorrect interventions: hydrotherapy, no exercise
Langhorst 2013 [155]	Systematic review with different PICO
Latorre 2013[158]	Incorrect study design (not randomised)
Lauche 2017 [161]	No useable outcomes
Law 2009 [162]	Incorrect study design: not randomised
Letafatkar 2020 [164]	Unclear population: inclusion citeria stated >3 months pain duration, but 50% had symptoms 6-12 weeks duration
Lima 2013 [165]	Systematic reviewwith different PICO
Lopez-de-Uralde-Villanueva 2020 [167]	Incorrect comparison: manual therapy vs. manual therapy + education vs. manual therapy + education + exercise
Lopez-Pousa 2015 [168]	Incorrect comparison: walking in a young vs. mature forest
Lopez-Rodriguez 2012 [169]	Not in English
López-Rodríguez 2013 [170]	Not in English
Lorena 2015 [171]	Not in English
Mannerkorpi 2000 [177]	Incorrect interventions: pain management programme
Mannerkorpi 2002 [173]	No comparator
Mannerkorpi 2009 [176]	Incorrect interventions: pain management programme
Mannerkorpi 2010 [175]	Incorrect comparison: different walking protocols

Martin-Martinez[178]	Incorrect interventions (virtual reality)
Matsutani 2007 [180]	Incorrect intervention: laser therapy
McDowell 2017 [184]	Systematic review with different PICO
McVeigh 2008 [185]	Systematic review, incorrect interventions: hydrotherapy, no exercise
Meiworm 1999 [187]	Not in English
Meiworm 2000 [186]	Incorrect study design: not randomised
Mendez-Rebolledo 2017 [188]	Systematic review with different PICO
Mesquita 2014 [190]	Abstract
Meyer, 2000 [191]	No useable outcomes
Miles 2014 [193]	Not available
Molinari 2018 [194]	Incorrect intervention: behavioural
Moseley 2004 [196]	No relevant outcomes
Moseley 2006 [195]	Incorrect population: phantom limb pain
Mosely 2005 [197]	Incorrect comparison
Moustafa 2015 [198]	Incorrect interventions: cervical manipulation, incorrect comparison
Nct, 2018 [202]	Clinical trial registry
Nct, 2018 [203]	Clinical trial registry
Nickel 2005 [205]	Incorrect comparison: pharmacological
Norregaard, 1997 [207]	No useable outcomes
Ote Karaca 2017 [209]	Incorrect population: low back pain
Perez-De la Cruz 2015 [211]	Not in English
Peters 2002 [212]	Incorrect population
Petersen 2015 [213]	Incorrect comparison, incorrect interventions: manual therapy with different neck exercises
Phattharasupharerk 2019 [214]	Incorrect population: low back pain
Pico-Espinosa 2020 [215]	Incorrect population: subacute and persistent pain included and results not reported separately
Pike 2015 [216]	Conference abstract
Plumbe 2016 [217]	Cochrane review: incorrect interventions, incorrect comparison: manipulation versus inactive control
Rajalaxmi, 2018 [218]	Unclear methods, no usable outcomes
Ramel 2009 [219]	Meta-analysis with different PICO
Ramsay 2000 [220]	Incorrect comparison: different types of aerobic exercise
Redondo 2004 [221]	Incorrect intervention: pain management programme
Reynolds 2020 [223]	Incorrect comparison: manual therapy + exercise vs. other manual therapy + exercise
Ris 2016 [225]	Incorrect comparison: pain management programme with and without training
Rivas Neira 2017 [226]	Protocol
Rolving 2014 [227]	No useable outcomes: unclear values
Ryan 2002 [228]	Not available
Saadat, 2019 [229]	Incorrect intervention (combination)
Salo 2010 [231]	No useable outcomes
Sarmento 2020 [236]	Incorrect comparator: sham Qigong
Sawynok 2013 [237]	No useable outcomes

Saxena 2017 [238]	Incorrect comparison: exercise versus medication
Segura-Jimenez 2013 [240]	No comparator, incorrect study design: not randomised
Skillgate 2015 [244]	Protocol
Skillgate 2020 [245]	Incorrect population: subacute and persistent pain included and results not reported separately
Song 2012 [246]	Conference abstract
Taggart 2003 [248]	No comparator
Taimela 2000 [249]	No useable outcomes
Thompson 2016 [251]	Incorrect comparison: exercises and psychological intervention versus exercises alone
Tomas-Carus 2007 [255]	Not in English
Valencia 2009 [258]	Incorrect comparison: different types of stretching
Valkeinen 2005 [261]	No relevant outcomes
van 2014 [262]	Cochrane review, incorrect population: medically unexplained symptoms
van Koulil 2011 [265]	Incorrect interventions: rehabilitation programme
Verstappen 1997 [266]	No relevant outcomes
Villafaina 2019 [269]	Incorrect interventions (virtual reality)
Villafaina 2019 [268]	Incorrect interventions (virtual reality)
Vitorino 2006 [270]	Incorrect comparison: same exercises on land versus water
Vonk 2009 [272]	Incorrect interventions: graded exercise therapy with psychological therapy
Wang 2010 [275]	Incorrect interventions (psychological combination)
Wiklund 2018 [277]	Incorrect population: chronic pain
Yang 2005 [280]	Incorrect population: general chronic pain, no useable outcomes
Ylinen 2004 [283]	Not in English
Ylinen 2005 [282]	No relevant outcomes
Ylinen 2006 [286]	No comparator
Zamuner 2015 [287]	Incorrect comparison: healthy controls
Zijlstra 2005 [288]	Incorrect study design. Intervention included flying to and staying in a luxurious hotel: with spa treatments, exercise therapy, relaxation
Zonneveld 2012 [289]	Incorrect population: multiple conditions causing unexplained physical symptoms

I.2 Excluded health economic studies

Table 104: Studies excluded from the health economic review

Reference	Reason for exclusion
McBeth 2012 [181]	This study was assessed as partially applicable with potentially serious limitations. However, other available evidence was of greater applicability and methodological quality and therefore this study was selectively excluded. This is the same study as the included economic evaluation but has shorter follow up period.
Van Eijk-Hustings 2016[263]	This study was assessed as partially applicable with potentially serious limitations. It has methodological limitations as it is a cost comparison study, based on an RCT included in the clinical review but also using additional data as it takes a period from diagnosis to

Reference	Reason for exclusion
	after the interventions (which includes before the interventions) and compares costs across the interventions. So slightly odd methodology.
Van Eijk-Hustings 2013[264]	This study was assessed as partially applicable with potentially serious limitations. However, other available evidence was of greater applicability as this was a cost consequences analysis.

Appendix J: MIDs for continuous outcomes

Table 105: MIDs for continuous outcomes: Aerobic exercise versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	9.05
Pain at >3 months (VAS, FIQ pain subscale, 0-100, final values, high is poor outcome)	10.8
Pain at >3 months (FIQ pain subscale, 0-100, high is poor outcome)	10.4
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)	7.05
Quality of life at ≤3 months (EQ-5D VAS, 0-100. high is good outcome, final values)	10.05
Quality of life at >3 months (EQ-5D VAS, 0-100, high is good outcome, final values)	11.43
Physical function at ≤3 months (timed up and go, seconds, final values, high is good outcome)	0.76
Physical function at ≤3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)	10.39
Physical function at >3 months (6 minute walking test, final values, metres, high is good outcome)	44.25
Physical function at >3 months (FIQ and SF-36 physical function subscales, 0-100, final values, high is poor outcome)	9.75
Physical function at >3 months (FIQ physical function subscale, 0-100, final values, high is poor outcome)	10.39
Psychological distress at >3 months (Change scores and final values, beck depression inventory, 0-21, high is poor outcome)	4.3
Psychological distress at >3 months (Final values, VAS and FIQ depression scale, 0-10, high is poor outcome)	1.35
Psychological distress at >3 months (Final values, VAS and FIQ anxiety scale, Beck anxiety inventory, final values, high is poor outcome)	0.5 (SMD)
Psychological distress at >3 months (Change scores, STAI anxiety total scores, high is poor outcome)	12.5
Psychological distress at >3 months (final values, FIQ depression scale, 0-10, high is poor outcome)	1.39
Psychological distress at >3 months (final values, FIQ anxiety scale, 0-10, high is poor outcome)	1.39
Psychological distress at ≤3 months (final values, BDI depression scale, high is poor outcome)	1.51
Use of healthcare services at ≤3 months (Number of GP contacts)	1.39
Use of healthcare services at >3 months (Number of GP contacts)	1.04
Use of healthcare services at ≤3 months (Number of medical specialist contacts)	0.35
Use of healthcare services at >3 months (Number of medical specialist contacts)	0.35

Outcomes	MID
Use of healthcare services at ≤3 months (Number of physiotherapist contacts)	2.43
Use of healthcare services at >3 months (Number of physiotherapist contacts)	2.43
Sleep at >3 months (VAS sleep scale, PSQI, FIQ sleep subscale, final values, high is poor outcome)	0.5 (SMD)

Table 106: MIDs for continuous outcomes: Strength training versus usual care

Outcomes	MID
Pain reduction at ≤3 months (final values, VAS, NRS, high is poor outcome)	10.75
Pain reduction at ≤3 months (change scores and final values, VAS, NRS, 0-100, high is poor outcome)	10.5
Pain reduction at >3 months (VAS, NRS, 0-100, final values and change scores, high is poor outcome)	12.25
Quality of life at ≤3 months (FIQ scale, 0-100, final values, high is poor outcome)	5.3
Physical function at ≤3 months (Neck disability index, change scores and final values, 0-100, high is poor outcome)	2.57
Physical function at ≤3 months (final values, FIQ physical function subscale, Northwick Park Questionnaire, high is poor outcome)	0.5 (SMD)
Physical function at ≤3 months (6 minute walking test, metres, final values, high is good outcome)	49.25
Physical function at >3 months months (final values, Northwick Park Questionnaire, Neck Disability Index, high is poor outcome)	0.5 (SMD)
Physical function at >3 months (change scores, SF-36 physical function subscale, HAQ, 0-100, high is poor outcome)	5.27
Psychological distress at ≤3 months (final values, pain catastrophising scale, high is poor outcome)	7
Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)	1.55
Sleep at >3 months (VAS sleep, 0-100, change scores, high is poor outcome)	8.72

Table 107: MIDs for continuous outcomes: Aerobic and strength versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, change scores, high is poor outcome)	7.5
Pain at >3 months (VAS, FIQ pain subscale, 0-100, final values, high is poor outcome)	4
Quality of life at ≤3 months (Fibromyalgia impact questionnaire, 0-100, final values, high is poor outcome)	8.82
Quality of life at >3 months (Fibromyalgia impact questionnaire, 0-100, final values and change scores, high is poor outcome)	6.89

Outcomes	MID
Physical function at >3 months (seconds, quarter mile walk test, final values, high is poor outcome)	13.21
Physical function at >3 months (metres, 6-minute walk test, final values, high is good outcome)	34.77
Physical function at >3 months (FIQ physical function subscale, 0-10, final values, high is poor outcome)	1
Physical function at ≤3 months (metres, 6-minute walk test, high is good outcome)	38.15
Psychological distress at ≤3 months (BDI, 0-30, final values, high is poor outcome)	5.32
Psychological distress at ≤3 months (State anxiety inventory, 0-10, change scores, high is poor outcome)	5.25
Psychological distress at ≤3 months (HADS anxiety, 0-21, high is poor outcome)	1.85
Psychological distress at >3 months (CES-D, BDI, FIQ depression subscale, final values, high is poor outcome)	0.5 (SMD)
Psychological distress at >3 months (State anxiety inventory, 20-80, final values and change scores, high is poor outcome)	2.26
Sleep at >3 months (Pittsburgh sleep quality index, high is poor outcome, change scores, 0-21)	1.06

Table 108: MIDs for continuous outcomes: Strength and flexibility versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	12.58
Pain at >3 months (VAS, SF-36 pain score, final values, 0-100, high is poor outcome)	11.43
Physical function at ≤3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)	10.85
Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)	9.28
Psychological distress at ≤3 months (ADS depression scale, 0-60, final values, high is poor outcome)	4
Psychological distress at >3 months (ADS depression scale, 0-60, final values, high is poor outcome)	4.5

Table 109: MIDs for continuous outcomes: Strength, proprioception and flexibility versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	11.25
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)	10
Psychological distress at ≤3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	1.7

Outcomes	MID
Psychological distress at >3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	1.7
Psychological distress at ≤3 months (HADS: depression, 0-21, final values, high is poor outcome)	1.7
Psychological distress at >3 months (HADS: depression, 0-21, final values, high is poor outcome)	2
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	5.7
Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)	6.35

Table 110: MIDs for continuous outcomes: Proprioception versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	0.81
Pain at >3 months (VAS, 0-10, final values, high is poor outcome)	1.17
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	5.98
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)	8.46
Physical function at ≤3 months (sit to stand test, final values, high is good outcome)	2.28
Physical function at >3 months (sit to stand test, final values, high is good outcome)	2.41
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	2.9
Psychological distress at >3 months (BDI, 0-61, final values, high is poor outcome)	4.73

Table 111: MIDs for continuous outcomes: Mind-body exercise versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, Visual numeric scale, FIQ pain subscale, 0-100, final values and change scores, high is poor outcome)	11.13
Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome) - Fibromyalgia	8.5
Pain at >3 months (VAS, SF-36 pain score, 0-100, final values, high is poor outcome) - Chronic neck pain	10.12
Quality of life at ≤3 months (WHOQOL-BREF, 0-5, final values, high is good outcome)	0.46
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	8.5
Physical function at ≤3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)	0.5 (SMD)
Physical function at >3 months (Neck pain disability scale, 0-100, final values, high is poor outcome)	6.85

Outcomes	MID
Physical function at >3 months (6 minute walk test, metes, final values, high is good outcome)	38.95
Psychological distress at ≤3 months (HADS:D, Beck depression inventory, CES-D, ADS depression scale, final values, high is poor outcome)	0.5 (SMD)
Psychological distress at ≤3 months (State trace anxiety inventory, final values, high is poor outcome) - Fibromyalgia	5.42
Psychological distress at ≤3 months (HADS:A, final values, high is poor outcome) - Chronic neck pain	1.6
Psychological distress at >3 months (Beck depression inventory, HADS:D, final values, high is poor outcome)	0.5 (SMD)
Psychological distress at >3 months (HADS:A, 0-21, final values, high is poor outcome)	1.7
Sleep at ≤3 months (VAS sleep outcome, pittsburgh sleep quality index, final values, high is poor outcome)	0.5 (SMD)

Table 112: MIDs for continuous outcomes: Flexibility versus usual care

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	13.5
Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)	2.65

Table 113: MIDs for continuous outcomes: Aerobic exercise versus strength

Outcomes	MID
Pain at ≤3 months (VAS, FIQ pain subscale, MDPI, 0-100, final values and change scores, high is poor outcome)	6.48
Pain at >3 months (VAS, 0-100, change scores, high is poor outcome)	9.75
Physical function at ≤3 months (Multidimensional fatigue inventory-20 reduced activity subscale, change scores, 0-20, high is poor outcome)	1.05
Physical function at ≤3 months (6 minute walking test, metres, final values, high is good outcome)	27.75
Psychological distress at ≤3 months (Hospital anxiety and depression anxiety score, 0-21, final values and change scores, high is poor outcome)	1.53
Psychological distress at ≤3 months (Final values and change scores, Hospital anxiety and depression scale, depression score, 0-21, high is poor outcome)	1.35
Psychological distress at ≤3 months (Final values, BDI, 0-60, high is poor outcome)	3.1
Sleep at ≤3 months (VAS Sleep scale, 0-100, final values, high is poor outcome)	14.85

Table 114: MIDs for continuous outcomes: Aerobic exercise versus flexibility

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	12.5
Pain at >3 months (VAS, 0-100, final values and change scores, high is poor outcome)	10.7
Psychological distress at ≤3 months (BDI, 0-21, high is poor outcome)	5.13
Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)	4.2
Psychological distress at ≤3 months (State trace anxiety inventory, 0-100, final values, high is poor outcome)	4.31
Psychological distress at >3 months (State trace anxiety inventory, 0-100, final values, high is poor outcome)	4.17

Table 115: MIDs for continuous outcomes: Aerobic exercise versus biomechanical exercise

Outcomes	MID
Pain at ≤3 months (VAS, 0-10, high score is poor outcome)	0.7
Psychological distress at ≤3 months (Scale of Catastrophic Thoughts on Pain, 0-5, high score is poor outcome)	0.7
Sleep at ≤3 months (Pittsburgh Sleep Quality Index, 0-21, high score is poor outcome)	1.85

Table 116: MIDs for continuous outcomes: Aerobic and strength versus aerobic exercise

Outcomes	MID
Quality of life at >3 months (FIQ, 0-100, change scores, high is poor outcome)	7
Psychological distress at >3 months (BDI, 0-61, change scores, high is poor outcome)	4

Table 117: MIDs for continuous outcomes: Aerobic and strength versus flexibility

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	7
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)	6
Quality of life at ≤3 months (NIH CPSI quality of life subscale, 0-12, final values, high is poor outcome)	1.05
Quality of life at >3 months (NIH CPSI quality of life subscale, 0-12, final values, high is poor outcome)	1.05
Psychological distress at ≤3 months (BDI, 0-21, final values, high is poor outcome)	2.15
Psychological distress at >3 months (BDI, 0-21, final values, high is poor outcome)	1.5

Table 118: MIDs for continuous outcomes: Aerobic and flexibility versus mind-body exercise

Outcomes	MID
Physical function at ≤3 months (6 minute walking test change scores, metres, change scores, high is good outcome)	49.05
Physical function at >3 months (6 minute walking test change scores, metres, change scores, high is good outcome)	70.14
Psychological distress at ≤3 months (HADS: depression, 0-21, change scores, high is poor outcome)	3.76
Psychological distress at ≤3 months (HADS: anxiety, 0-21, change scores, high is poor outcome)	2.54
Psychological distress at >3 months (HADS: anxiety, 0-21, change scores, high is poor outcome)	3.04
Psychological distress at >3 months (HADS: depression, 0-21, change scores, high is poor outcome)	4.97
Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)	2.65
Sleep at >3 months (Pittsburgh sleep quality index, 0-21, change scores, high is poor outcome)	3.54

Table 119: MIDS for continuous outcomes: Aerobic exercise and flexibility versus aerobic exercise

Outcomes	MID
Pain perception at <3 months (Final score; VAS 0-10; high is poor outcome)	0.19
Pain perception at >3 months (Final score; VAS, 0-10; high is poor outcome)	0.21
Quality of life at <3 months (final score; FIQ, 0-100, high is poor outcome)	2.04
Quality of life at >3 months (final score; FIQ, 0-100, high is poor outcome)	2.11
Sleep quality at <3 months (final score; Pittsburgh Sleep Quality Index, 0-21, high is poor outcome)	0.73
Sleep quality at >3 months (final score; Pittsburgh Sleep Quality Index, 0-21, high is poor outcome)	0.5

Table 120: MIDs for continuous outcomes: Aerobic, strength, mind-body and proprioception versus flexibility

Outcomes	MID
Quality of life at ≤3 months (FIQ total score, 0-100, high is poor outcome)	6.51
Physical function at ≤3 months (number of steps, high is good outcome)	15.44

Table 121: MIDs for continuous outcomes: Strength versus mind-body

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	0.45
Quality of life at ≤3 months (NHP, 0-600, final values, high is poor outcome)	59.05
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	2.65
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	3.1

Table 122: MIDs for continuous outcomes: Mind-body versus biomechanical

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	0.65
Quality of life at ≤3 months (NHP, 0-600, final values, high is poor outcome)	48.95
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	3.3
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	3.8

Table 123: MIDs for continuous outcomes: Strength versus flexibility

Outcomes	MID
Pain reduction at ≤3 months (VAS, 0-100, change scores and final values, high is poor outcome)	9.78
Physical function at ≤3 months (FIQ physical function subscale, 0-30, final values, high is poor outcome)	2.6
Psychological distress at ≤3 months (BDI, 0-61, change scores, high is poor outcome)	2.02
Psychological distress at ≤3 months (BAI, 0-61, change scores, high is poor outcome)	3.23
Sleep at ≤3 months (FIQ sleep subscale, 0-10, change scores, high is poor outcome)	0.81

Table 124: MIDs for continuous outcomes: Strength and flexibility versus mind-body

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	13.8
Pain at >3 months (VAS, 0-100, final values, high is poor outcome)	12.55
Physical function at ≤3 months (Neck disability index, neck pain disability scale, final values, high is poor outcome)	0.5 (SMD)
Physical function at >3 months (Neck pain disability scale, final values, high is poor outcome)	9.04
Psychological distress at ≤3 months (Depression scale ADS, 0-60, final values, high is poor outcome)	3.7

Outcomes	MID
Psychological distress at >3 months (Depression scale ADS, 0-60, final values, high is poor outcome)	3.7

Table 125: MIDs for continuous outcomes: Strength, flexibility and proprioception versus mind-body

Outcomes	MID
Pain reduction at ≤3 months (VAS, 0-100, final values, high is poor outcome)	11.75
Pain reduction at >3 months (VAS, 0-100, final values, high is poor outcome)	13.85
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	6.1
Physical function at >3 months (Neck disability index, 0-100, final values, high is poor outcome)	7.05
Psychological distress at ≤3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	2.35
Psychological distress at >3 months (HADS: anxiety, 0-21, final values, high is poor outcome)	2.25
Psychological distress at ≤3 months (HADS: depression, 0-21, final values, high is poor outcome)	1.9
Psychological distress at >3 months (HADS: depression, 0-21, final values, high is poor outcome)	1.9

Table 126: MIDs for continuous outcomes: Strength versus proprioception

Outcomes	MID
Physical function at ≤3 months (Neck disability index, 0-50, final values, high is poor outcome)	1.31

Table 127: MIDs for continuous outcomes: Mind-body versus flexibility

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	11
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	11.1
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	4.35
Sleep at ≤3 months (Pittsburgh sleep quality index, 0-21, final values, high is poor outcome)	2.2

Table 128: MIDs for continuous outcomes: Mind-body versus biomechanical

Outcomes	MID
Pain at ≤3 months (VAS, 0-100, final values, high is poor outcome)	0.65

Outcomes	MID
Quality of life at ≤3 months (NHP, 0-600, final values, high is poor outcome)	48.95
Physical function at ≤3 months (Neck disability index, 0-100, final values, high is poor outcome)	3.3
Psychological distress at ≤3 months (BDI, 0-61, final values, high is poor outcome)	3.8

Table 129: MIDs for continuous outcomes: Flexibility and proprioception versus flexibility

Outcomes	MID
Quality of life at ≤3 months (FIQ, 0-100, final values, high is poor outcome)	8.85
Psychological distress at ≤3 months (BDI, 0-63, final values, high is poor outcome)	3.59

Table 130: MIDs for continuous outcomes: Flexibility and relaxation versus aerobic exercise

Outcomes	MID
Quality of life at >3 months (FIQ, 0-100, final values, high is poor outcome)	7.9

Table 131: MIDs for continuous outcomes: Exercise versus psychological therapies

Outcomes	MID
Pain at ≤3 months (VAS, FIQ pain scale, 0-100, high is poor outcome, final values and change scores) - Fibromyalgia	10.61
Pain at >3 months (VAS, NRS, 0-100, high is poor outcome, final values)	9.75
Quality of life at ≤3 months (FIQ, 0-100, high is poor outcome, final values and change scores)	8.35
Physical function at ≤3 months (FIQ physical function subscale, 0-10, high is poor outcome, change scores)	0.17
Physical function at ≤3 months (6 minute walk test, metres, high is good outcome, final values)	35.95
Physical function at >3 months (6 minute walking test, metres, high is good outcome, final values)	39
Psychological distress at ≤3 months (CES-D, 0-100, high is poor outcome, final values)	9.3
Psychological distress at >3 months (Hospital anxiety and depression scale, depression subscale, 0-21, high is poor outcome, change scores)	1.4
Psychological distress at >3 months (Hospital anxiety and depression scale, anxiety subscale, 0-21, high is poor outcome, change scores)	1.35
Sleep at >3 months (the sleep scale, 0-30, final values, high is poor outcome)	2.85

Outcomes	MID
Sleep at >3 months (Pittsburgh sleep quality index, 0-21, high is poor outcome, change scores)	1.5

Table 132: MIDs for continuous outcomes: Manual therapy and exercise versus manual therapy

Outcomes	MID
Pain at ≤3 months (NRS, high is poor outcome, final values, 0-10, final values)	1.15
Pain at >3 months (NRS, high is poor outcome, final values, 0-10, final values)	1.15
Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50, final values)	6.5
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)	6.75

Table 133: MIDs for continuous outcomes: Manual therapy and exercise versus exercise

Outcomes	MID
Pain at ≤3 months (VAS, NRS, high is poor outcome, final values, 0-100, final values)	4.25
Pain at >3 months (NRS, VAS, high is poor outcome, final values, 0-100)	11.35
Quality of life at >3 months (Fibromyalgia impact questionnaire, 0-100, final values, high is poor outcome)	7.95
Physical function at >3 months (Neck disability index, functional performance scale, final values, high is poor outcome, 0-100)	0.5 (SMD)
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-100)	6.2
Physical function at ≤3 months (Neck disability index, high is poor outcome, 0-100)	8.14

Table 134: MIDs for continuous outcomes: Exercise versus manual therapy

Outcomes	MID
Pain at ≤3 months (NRS, high is poor outcome, final values, 0-10)	1.15
Pain at >3 months (NRS, high is poor outcome, final values, 0-10)	1.15
Physical function at ≤3 months (Neck disability index, high is poor outcome, final values, 0-50)	6.5
Physical function at >3 months (Neck disability index, high is poor outcome, final values, 0-50)	6.75