National Institute for Health and Care Excellence

Draft

Multiple sclerosis in adults: management

[C] Evidence review for non-pharmacological management of fatigue

NICE guideline <number>

Evidence reviews underpinning recommendations 1.5.2 to 1.5.10 and research recommendations in the NICE guideline

December 2021

Draft for Consultation

These evidence reviews were developed by the National Guideline Centre, hosted by 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/or their carer or guardian.

Local commissioners and/or 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN:

Contents

1 Non-pharmacological management of fatigue

1.1 Review question

For adults with MS, including people receiving palliative care, what is the clinical and cost effectiveness of non-pharmacological interventions for fatigue?

1.1.1 Introduction

MS related fatigue is not well understood and may be as a direct result of damage to myelin, necessitating alternative nerve pathways to be developed in the central nervous system. Perhaps it is secondary to extra exertion due to weakness, stiffness, spasticity, tremor and disturbed sleep or a combination of all of the above factors.

Many sufferers, manage fatigue with alternative methods such as planning and prioritising of tasks, incorporating time to rest, modifying diet and re-organising living or workspaces to conserve energy. Prior to the current update of this review, interventions covered by recommendations included mindfulness-based training, cognitive behavioural therapy (CBT) and fatigue management programmes. Previously, aerobic exercise and moderate progressive resistance activity combined with CBT in those with moderately impaired mobility (an EDSS score of greater than or equal to 4) has been suggested as helpful for fatigue. There is, however, uncertainty of the advised duration of this approach and the effectiveness of exercise programmes on those with lower EDSS scores. With the recent modifications to lifestyle as a result of the COVID pandemic, there may be more virtual opportunities that have been trialled and have a stronger evidence base. Finding an alternative and effective therapeutic strategy is essential for optimising care and quality of life in people with MS.

1.1.2 Summary of the protocol

For full details see the review protocol in appendices.

Table 1: PICO characteristics of review question

Population	Inclusion: Adults (≥18 years) with MS, including people receiving palliative care, who are experiencing fatigue. Exclusion: Children and young people (≤18 years).
Interventions	Any non-pharmacological intervention for fatigue, for example: • Multidisciplinary rehabilitation/programmes including progressive resistance training • Energy conservation programs • Mindfulness based training • Exercise including aerobic exercise training • Resistance training – (distinguish it from balance and vestibular rehab) • Vestibular rehab • Getting To Grips • Gym prescription

	• Self-management programmes • Fatigue management programmes • FACETS (Fatigue: Applying Cognitive behavioural and Energy effectiveness Techniques to lifeStyle) • FatiMa (Fatigue management in MS– patient education programme) • Diet (ketogenic, intermittent fasting and George Jelinek* which is plant based, wholefood diet, excluding dairy and minimising saturated fat intake) • Yoga, • Tai chi • Pilates • Relaxation • Cognitive behavioural therapy • Hyperbaric oxygen Combinations may be included if relevant to clinical practice (to be checked with guideline committee if unsure)
Comparisons	Interventions will be compared to each other placebo/sham, usual care or no treatment.

Outcomes	All outcomes are considered equally important for decision making and therefore have all been rated as critical.
	
	• Patient-reported outcome measures to assess MS fatigue, including MFIS Fatigue Severity Scale (FSS), National Fatigue Index (NFI), MS-specific FSS (MFSS), Modified Fatigue Impact Scale (MFIS), and Visual Analogue Scale (VAS)
	• Health-related Quality of Life, for example EQ-5D, SF-36, Leeds MS quality of life scale, MS Impact Scale.
	• Impact on carers.
	• Functional scales that quantify level of disability, such as the Expanded Disability Status Scale (EDSS), the Multiple Sclerosis Functional Composite (MSFC), the Cambridge Multiple Sclerosis Basic Score (CAMBS), or the Functional Assessment of Multiple Sclerosis (FAMS).
	• Cognitive functions, such as memory and concentration
	• Psychological symptoms assessed by validated and disease-specific scales, questionnaire or similar instruments.
	• Adverse effects of treatment for example:
	  ○ Incidence of adverse events
	  ○ Adverse events leading to withdrawal
	• Outcomes measuring how acceptable the intervention was. These may be measured in terms of how acceptable it was to patients, completion rates, response to follow up, adherence, engagement or disengagement.
	
	<u>Follow up:</u>
	• 3-6 months (minimum of 3 months but can include 1-3 months and downgrade)
	• >6 months – 1 year (can include > 2years for diet, include >12 months but downgrade)

Study design	Systematic reviews of RCTs and RCTs will be considered for inclusion.

	Cross-over trials will also be considered for inclusion if they have an appropriate washout period. Published NMAs and IPDs will be considered for inclusion.

1.1.3 Methods and process

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

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

1.1.4 Effectiveness evidence

1.1.4.1 Included studies

Eighty-nine randomised controlled trials (from ninety-four papers) were included in the review; these are summarised in Table 2 below. Note that the total number of studies presented in this table is higher than the number included as studies are separated by type of intervention covered, with some studies covering more than one of the interventions listed in the protocol.

Evidence from these studies is summarised in the clinical evidence summary below (Tables 3-5). Evidence that could not be analysed using GRADE can be found in Tables 55-60.

The majority of included studies were parallel randomised controlled trials, though there were also five crossover trials and one cluster-randomised trial included.

Population

As this review question is specific to the treatment of fatigue in MS, only studies that had fatigue as one of their aims of treatment were included in the review. In line with the previous version of this review, a study was determined to be relevant in terms of treating fatigue if any of the following applied:

- The study used a threshold for fatigue as an inclusion criterion in the study (e.g. only those with score ≥4.0 on the Fatigue Severity Scale)
- The study did not use a threshold for fatigue for inclusion, but it was clear from the paper that fatigue was the primary outcome or one of the primary outcomes
- The study did not use a threshold for fatigue for inclusion and it was listed as a secondary outcome, but it was clear from the paper that fatigue was one of the focuses of the paper
- The study did not focus on any particular MS symptom and fatigue was emphasised as an important outcome

Of studies that reported the proportion of people with different types of MS, most had relapsing-remitting MS as the most common type of MS among people included in the studies. The range of Expanded Disability Status Scale (EDSS) scores included in studies varied. Some were more focused on those with lower scores (less disability) and used a certain threshold for EDSS as an inclusion criterion, while others included a wider range of EDSS scores and did not focus on a particular range.

Interventions and comparisons covered by the evidence

Evidence was identified for the following interventions and comparisons:

- Exercise, where **aerobic exercise** was the main component (n=15 studies)
  - vs. control (waitlist control, control with no intervention, control with education/physician contact only) in n=14 studies
  - vs. neurological rehabilitation (focus on respiratory postural and motor synergies and stretching exercises) in n=1 study

- Exercise, where **resistance training** was the main component (n=5 studies)
  - vs. control (waitlist control, control with no intervention, control with education/social contact only) in n=4 studies
  - vs. aerobic (endurance) training in n=1 study

- Exercise, where **vestibular balance rehabilitation** was the main component (n=8 studies)
    - vs. control (waitlist control, routine usual care, control with social contact only) in n=6 studies
    - vs. progressive resistance training in n=1 study
    - vs. aerobic exercise in n=3 studies
    - vs. standard neurorehabilitation (stretching, postural alignment, mobilisations, balance training and neuromuscular facilitations) in n=1 study
- Exercise, **progressive resistance training + balance exercises** (n=2 studies)
    - vs. control (no intervention, waitlist control group) in n=2 studies
    - vs. primarily balance exercises in n=1 study

- Exercise, **progressive resistance training + aerobic exercise** (n=4 studies)
    - vs. control (no intervention, waitlist control group, control with social contact only) in n=4 studies
    - vs. yoga in n=1 study

- Exercise, **balance training + aerobic exercise** (n=1 study)
    - vs. control (control with education/social contact only) in n=1 study

- Exercise, **progressive resistance training + balance training + aerobic exercise** (n=6 studies)
    - vs. control (no intervention and continue usual care) in n=6 studies
    - vs. massage in n=1 study
    - vs. conventional rehabilitation (muscle strength, balance, gait or upper limb function depending on treatment plan; n=1 study)
    -

- Exercise (**resistance training + aerobic exercise) + cognitive behavioural therapy (CBT)** (n=1 study)
    - vs. control (waitlist control group) in n=1 study

- Standard exercises **(progressive resistance training + aerobic exercise + balance training) + high-intensity lower limb resistance training** (n=1 study)
    - vs. standard exercises alone (progressive resistance training + aerobic exercise + balance training) in n=1 study

- **Multidisciplinary rehabilitation (medical, exercise, counselling and fatigue self-management)** + methylprednisolone (n=1 study)
    - vs. standard procedure (not to offer rehabilitation following intravenous methylprednisolone treatment) in n=1 study

- **Massage + exercise (strength, stretching, aerobic/endurance and balance)** in n=1 study
    - vs. control (continue standard medical care) in n=1 study
    - vs. massage alone in n=1 study
    - vs. exercise (strength, stretching, aerobic/endurance and balance) alone in n=1 study

- **Balance training + Pilates** in n=1 study
  - vs. control (relaxation) in n=1 study

- **Multidisciplinary rehabilitation – physical activity with/without other rehabilitation + fatigue self-management programme (n=3 studies)**
  - vs. control (relaxation) in n=1 study
  - vs. control (information or nurse consultation only) in n=2 studies
  - vs. physical activity only in n=1 study

- **Fatigue management** and **energy conservation management programmes (n=9 studies)**
  - vs. control (waitlist control, no intervention, continue usual care, information/social contact only) in n=7 studies
  - vs. control (relaxation) in n=1 study
  - vs. general self-management in MS intervention (MS: Take Control programme) in n=1 study

- **FACETS** (**Fatigue: Applying Cognitive behavioural and Energy Effectiveness Techniques to lifestyle)** programme (n=1 study)
  - vs. control (current local practice, varying between centres) in n=1 study

- **Diets (n=6 studies)**
  - vs. control (usual care, no dietary intervention, educational only) in n=3 studies
  - vs. standard World Health Organisation healthy diet recommendations or standard healthy diet in accordance with US Department of Agriculture dietary guidelines for Americans 2010 (instead of personalised diet plan) in n=2 studies
  - Comparison of two diets (Wahls modified Palaeolithic elimination diet vs. Swanks low-saturated fat diet) in n=1 study

- **Mindfulness training (n=1 study)**
  - vs. control (usual care) in n=1 study

- **Self-management programmes** (n=3 studies)
  - vs. control (no intervention, waitlist control, education only) in n=3 studies

- **Self-management programme + exercise** (n=1 study)
  - vs. control (waitlist control) in n=1 study

- **Functional electrical stimulation + exercise (n=1 study)**
  - vs. control (waitlist control group) in n=1 study

- **Yoga** (n=7 studies)
  - vs. control (waitlist control, no intervention, education or social contact only) in n=6 studies
  - vs. aerobic exercise in n=3 studies

  - vs. aerobic exercise + resistance training in n=1 study
  - vs. sports climbing in n=1 study

- **Pilates** (n=5 studies)
  - vs. control (relaxation) in n=1 study
  - vs. control (no intervention, waitlist control) in n=3 studies
  - vs. traditional exercise (strength, balance and coordination training) in n=1 study

- **Relaxation – including relaxation, reflexology, massage and acupressure** (n=9 studies)
  - Acupressure vs. touch only/sham (n=2 studies)
  - Reflexology or relaxation vs. routine treatment (n=2 studies)
  - Reflexology vs. non-specialised foot massage (n=1 study)
  - Relaxation vs. waitlist control (n=1 study)
  - Massage vs. routine treatment/control (n=3 studies)

- **CBT or motivational interviewing** (n=7 studies)
  - Motivational interviewing vs. control (waitlist control, no intervention/usual care) in n=2 studies
  - CBT vs. control (no intervention/usual care, waitlist control, education/social contact only) in n=4 studies
  - CBT vs. control (relaxation) in n=1 study

- **Motivational interviewing + exercise** (n=1 study)
  - vs. control (usual care/no intervention) in n=1 study

No relevant randomised controlled trials including the following interventions were identified:

- 'Getting to Grips' programme
- Gym prescription
- 'FatiMa' patient education programme
- Tai Chi
- Hyperbaric oxygen

See also the study selection flow chart, study evidence tables, forest plots and GRADE tables in appendices.

1.1.4.2 Excluded studies

Eleven Cochrane reviews[6, 7, 15, 49, 60, 63, 67, 83, 84, 94, 95] were ordered and reviewed to assess relevance to this review.

These reviews were not included in the review for the following reasons:

- Review not specific to fatigue and many of the included studies do not reported fatigue as an outcome[6]
- Review focuses on those with MS and chronic pain rather than fatigue – those not reporting pain outcomes excluded meaning some studies focused on fatigue would not be included[7]

- Population of the review is not limited to those with MS and fatigue outcomes are not reported[15]
- Comparisons in this review of exercise interventions for fatigue in MS are broader whereas they are split into more specific interventions in our review protocol and evidence review. All studies included in this Cochrane review were already included in the previous version of the NICE evidence review[49]
- Not all included studies were specifically targeted at fatigue in MS[60, 94]
- Fatigue not an outcome analysed in the review[63]
- Interventions not relevant to the review protocol[67]
- Not limited to RCTs and not all studies were specific to fatigue as fatigue considered a secondary outcome[83]
- Study withdrawn from publication and not specific to the MS population[84]
- Main focus of the interventions was on cognitive outcomes not fatigue[95]

Despite not being included in the review, all of these reviews were checked to identify any references that were relevant for inclusion in the current evidence review.

See the excluded studies list in the appendices.

1.1.5 Summary of studies included in the effectiveness evidence

Table 2: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Comments
Exercise - aerobic			
Ahmadi 2013[4] Associated papers: Ahmadi 2010[3] and Ahmadi 2010[5] N=20 randomised across these two groups Conducted in Iran	**Treadmill training** Supervised treadmill training exercises three times a week for 8 consecutive weeks. Each training session consisted of 30 min of treadmill exercise. The exercise class began and ended with 10 min stretching of muscles and flexion and rotation movements of the trunk and the lower limb. Training intensity was 40-75% age predicted maximal heart rate. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean FSS score 3.5-4.2 across the two groups. Threshold for fatigue used for inclusion (yes/no): no Age: mean 37.0 years for both groups Type of MS: not reported EDSS: score between 1.0 and 4.0 was an inclusion criterion. Mean score 2.3-2.4 across the two groups	Included in previous guideline version. Additional comparisons of yoga vs. exercise and control groups included in a separate section.
Feys 2019[33] N=42 randomised Conducted in Belgium	**12-week start to run programme** Individualised training instructions based on their baseline aerobic capacity received 3 times weekly by email for 12 weeks to be performed in community and with the aim of participating in a running event. Gradual programme starting with walking and increasing the amount of running over the course of the programme. Two group sessions (weeks 4 and 8) at a running track arranged where they performed their individual programme and were observed to discuss progress and potential for injuries. Also allowed education sessions and learning from others.	Multiple sclerosis Fatigue at baseline: mean Fatigue Scale for Motor and Cognitive Function – Physical subscale 29.0-32.0 across the two groups. Threshold for fatigue used for inclusion (yes/no): no	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	vs. **Control** Waitlist control group.	Age: mean 36.0-44.0 years across the two groups Type of MS: not reported EDSS: not reported	
Geddes 2009[41] N=12 randomised Conducted in USA	**Home walking programme** Home walking programme that was individualised based on the results of the 6-Minute Walk Test at baseline. Instructed to walk 3 times weekly for 12 weeks. For the first 2 weeks, walking was 5 min below the lower limits of their training heart rate range, followed by 15 min of walking within their training heart rate range, and then a 5 min cool down below this range. For weeks 3-12, subjects increased their training time in the training heart rate range to 20-30 min. vs. **Control** Asked to refrain from any regular exercise during the 12-week period.	Multiple sclerosis Fatigue at baseline: not reported Threshold for fatigue used for inclusion (yes/no): no Age: mean 35.0-51.0 years across the two groups Type of MS: majority had not been classified EDSS: score ≤6.0 was an inclusion criterion	Included in previous guideline version.
Hasanpour Dehkordi 2016[45] N=60 randomised into these two groups Conducted in Iran	**Aerobic exercise** Three sessions weekly for 12 weeks. Each session lasted 40 min, with 5-10 min warm-up, 25-30 min exercise (walking) and 5 min cooling down. Exercise aimed to reach 60% of heart rate reserve. After 6 sessions duration of exercise increased to 30-35 min at a heart rate of 70% heart rate reserve. vs.	Multiple sclerosis Fatigue at baseline: mean 3.8-4.9 across the two groups on 'Rhoten Fatigue Test' Threshold for fatigue used for inclusion (yes/no): no	New study published since previous guideline version Additional comparisons of yoga vs. aerobic exercise and control groups included in a separate section.

Study	Intervention and comparison	Population	Comments
	Control Educational support with no exercise protocol. Asked to continue medications and usual lifestyle.	Age: mean 31.9 years for the whole population (including a third group not included here) Type of MS: not reported EDSS: score not reported	
Hebert 2011[47] N=26 randomised into these two groups Conducted in USA	**Exercise** Bicycle endurance training and stretching exercises. Twice weekly for 6 weeks. Endurance exercise consisted of stationary bicycling: 5-min warm-up, two 15 min sessions and 2 to 5 min cool down. Training intensity during the 15 min sessions was 65% to 75% of peak heart rate. For stretches, these were held for 30 seconds. Also received 5 min fatigue management session including discussions of daily rest intervals, self-monitoring of exertion levels, workstation ergonomics and heat tolerance education. vs. **Control** Waitlist control group. Received usual medical care.	Multiple sclerosis Fatigue at baseline: mean MFIS total score 51.0-56.0 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥84 on MFIS total Age: mean 43.0-50.0 years across the two groups Type of MS: majority relapsing-remitting (>80% for both groups) EDSS: score not reported	Included in previous guideline version. Additional comparisons from this study of vestibular rehabilitation vs. exercise and control groups are included in a separate section.
Heine 2017[50] N=90 randomised Conducted in The Netherlands	**Aerobic exercise** Aerobic interval training performed three times a week for 16 weeks (12 sessions supervised in outpatient clinic and 36 sessions home-based using identical equipment). Each session included 30 min aerobic interval training on cycle ergometer. Involved 6 interval cycles of 3 min at 40%, 1 min at 60% and 1 min at 80% of peak power.	Multiple sclerosis Fatigue at baseline: mean CIS20r fatigue subscale score of 43.0 for the whole population Threshold for fatigue used for inclusion (yes/no): yes – severe	New study published since previous guideline version. TREFAMS-AT trial.

Study	Intervention and comparison	Population	Comments
	vs. **Control** Three 4 min consultations with an MS nurse over the 16-week period. Consisted of reliable information on MS-related fatigue and guidance from the nurse to reassure the patient that their concerns or questions were being taken seriously. Referral of the patient to any other outpatient or inpatient facilities for the treatment of fatigue was not permitted.	fatigue (score ≥35 on CIS20r fatigue subscale) Age: mean 46.0 years for the whole population Type of MS: majority relapsing-remitting (>70% for the whole population) EDSS: score ≤6.0 was an inclusion criterion. Median score 3.0 for the whole population.	
McCullagh 2008[72] N=30 randomised Conducted in Ireland	**Exercise** Twice weekly sessions for 12 weeks: 5 min warm-up and cool down, and 40 min exercise. 4-6 participants per class. Each session involved four stations lasting 10 min each. Varied between treadmill walking/running, cycling, stair-master training, arm strengthening exercises, volleyball and outdoor walking including steps and slopes. Encouraged to maintain exertion levels between fairly light and somewhat hard. Also required to perform one home-based exercise programme for 40-60 min and the type of exercise could be of their choice. vs. **Control** Asked to maintain normal activity levels. Visited physiotherapist once monthly to discuss any issues.	Multiple sclerosis Fatigue at baseline: median MFIS total 26.0-27.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 34.0-41.0 years across the two groups Type of MS: all had relapsing-remitting or secondary progressive MS (majority relapsing-remitting) EDSS: not reported	Included in previous guideline version.
Mostert 2002[74] N=37 randomised	**Aerobic exercise** Attended 5 supervised training sessions per week over 3-4 weeks. Each	Multiple sclerosis	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
Conducted in Switzerland	training session consisted of 30 min bicycle exercise training. vs. **Control** No intervention. Avoid regular physical exercise, which could improve aerobic fitness.	Fatigue at baseline: mean FSS 5.1-5.2 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 44.0-45.0 years across the two groups Type of MS: relapsing-remitting (31-39% across the two groups); chronic-progressive (23-31% across the two groups); and relapsing-progressive (23-46% across the two groups) EDSS: mean score 4.5-4.6 across the two groups.	
Oken 2004[80] N=43 randomised into these two groups Conducted in USA	**Aerobic exercise** One session per week along with home exercise. Cycling on recumbent or dual-action stationary bicycles. The weekly exercise class began and ended with 5 min stretching of cycling muscles. Intensity was very light to moderate. Sometimes option of using Swiss ball and arm, trunk and balance work, though cycling main form of exercise. Encouraged to exercise regularly at home in addition to in-person sessions. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean score on Multidimensional Fatigue Inventory – Physical subscale was 13.0-14.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 48.0-49.0 years across the two groups Type of MS: not reported	Excluded previously but on review deemed relevant. Additional comparisons of yoga vs. exercise and control groups are included under a separate section.

Study	Intervention and comparison	Population	Comments
		EDSS: score ≤6.0 was an inclusion criterion. Mean score 2.9-3.1 across the two groups.	
Pazokian 2013[85] N=100 randomised Conducted in Iran	**Aerobic exercise, with or without stretching** 12-week exercise intervention. Aerobic exercises performed 3 times weekly with each session lasting 30 min: walking, cycling and treadmill exercise. For group that also performed stretching exercises, stretching of upper and lower limbs and trunk muscles was performed for 15 min prior to aerobic exercises. vs. **Control** No intervention performed.	Multiple sclerosis Fatigue at baseline: mean score on FSS 43-51 across the three groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 35 years for whole population Type of MS: relapsing-remitting MS was an inclusion criterion. EDSS: 1.0-5.5 was inclusion criterion.	Two similar groups of those performing aerobic exercise, some with and some without stretching exercises, included together as a single intervention compared to control Not included previously but identified as part of the new search
Plow 2019[86] N=138 randomised to these two groups Conducted in USA	**Physical activity only** Delivered via phone for 12 weeks. 3 group teleconference sessions and 4 individually tailored phone calls. Taught how to engage in pedometer-based walking programme, set goals, overcome barriers and self-monitor progress. vs. **Control** Delivered via phone for 12 weeks. Information on health topics relevant to MS. Designed as a contact control group.	Multiple sclerosis Fatigue at baseline: mean score on FIS 68-71.0 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – had to have moderate-severe fatigue (defined as score ≥4.0 on FSS) Age: mean age 51-52 years across the two groups	Additional comparisons from this study involving combined physical activity + fatigue self-management are included under a separate section. New study published since last version of guideline

Study	Intervention and comparison	Population	Comments
		Type of MS: majority had relapsing-remitting MS (>80%). EDSS: not reported – on PDDS, required score of 1.0-5.0 for inclusion.	
Rampello 2007[90] N=19 randomised Crossover study Conducted in Italy	**Aerobic training** 3 sessions per week on leg cycle ergometer for 8 weeks, with 30 min at 60% max work rate. Stretching of lower limbs and trunk muscles then performed for 15 min. vs. **Neurological rehabilitation** 3 sessions per week, with each session lasting 60 min. Aimed at improving respiratory postural and respiratory-motor synergies and of stretching exercises.	Multiple sclerosis Fatigue at baseline: median score on MFIS 30-36 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 41 years for whole population Type of MS: not reported. EDSS: ≤6.0 was inclusion criterion. Median score 3.5.	8-week washout period Included in previous guideline version
Sadeghi Bahmani 2019[97] N=62 randomised into these two groups Conducted in Iran	**Endurance exercise** Three supervised group sessions (30-45 min per session) weekly for 8 weeks. 5 min warm-up and stretching, 25-35 min exercise on treadmill, exercise bicycles or walking/jogging with 1-2 min rest periods, followed by 5 min of cooling down. Aim was for participants to feel slightly exhausted but not severely exhausted. vs. **Control**	Multiple sclerosis Fatigue at baseline: mean score on FSS 39.0-43.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 38.0 years in both groups	Additional comparisons of coordinative training (balance) vs. endurance exercise and control groups are included under a separate section. New study published since previous guideline version

Study	Intervention and comparison	Population	Comments
	Met three times weekly for 30-45 min sessions at the hospital centre to control for social contact elements of the other interventions.	Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Mean score 2.0-2.5 across the two groups.	
Schulz 2004[100] N=46 randomised Conducted in Germany	**Aerobic exercise** 8-week bicycle ergometer training programme tailored to individual capacities. Sessions were twice weekly, including 30 min at a maximal intensity of 75% of the maximal watts taken from the ergometry results at baseline. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 23.0-37.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 41.0-42.0 years across the two groups Type of MS: not reported – likely that majority was relapsing-remitting but not mentioned specifically for this analysis group EDSS: score <5.0 was an inclusion criterion. Mean score 2.5-2.7 across the two groups.	Excluded previously but on review deemed relevant.
van den Berg 2006[109] N=19 randomised Conducted in UK	**Exercise** Supervised treadmill training three times weekly for 4 weeks. Walking duration was increased during training period as tolerated, up to a maximum of 30 min with a maximum of three rest periods. Once maximum walking duration was attained, intensity was increased by increasing walking speed. Encouraged to train at an intensity of 55-85% of age-predicted maximum heart rate.	Multiple sclerosis Fatigue at baseline: mean score on FSS 31.0-32.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
	vs. **Control** Waitlist control group. No intervention for 4 weeks.	Age: range 30-65 years for the whole population Type of MS: not reported EDSS: not reported	
Exercise – resistance training			
Callesen 2020[21] N=43 randomised into these two groups Conducted in Denmark	**Progressive resistance training** A total of 12 1-h training sessions over 10 weeks (2 sessions per week). 10 min warm-up on a stationary bicycle or treadmill. Focused on knee and hip flexion, with exercises progressing from 3 sets of 10 repetitions at 15 RM to 4 sets of 8 repetitions at 8 RM. Conducted using machines targeting specific muscle groups. vs. **Control** Waitlist control group. Encouraged to maintain usual care and level of physical activity.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 42.0-44.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: median 52-56 years across the two groups Type of MS: majority relapsing-remitting (>80% in both groups) EDSS: score 2.0-6.5 was an inclusion criterion. Median 3.5-4.0 across the two groups	New study published since previous guideline version. Additional comparisons of balance and motor control vs. progressive resistance training and control groups are included under a separate section.
Dalgas 2010[27] N=38 randomised Conducted in Denmark	**Progressive resistance training** 12 weeks (2 sessions per week) of resistance training of the lower extremities performed. 5 min warm-up on stationary bicycle followed by 5 different exercises (leg press, knee extension, hip flexion, hamstring curl and hip extension). The programme progressed from three sets of 10 repetitions at 15 RM (week 1-2) in	Multiple sclerosis Fatigue at baseline: mean score on FSS 5.5-5.8 across the two groups Threshold for fatigue used for inclusion (yes/no): no	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
	increments every two weeks up until three sets of 8 repetitions at a load of 8 RM in weeks 11-12. Rests periods of 2-3 min between sets and exercises were allowed. Most sessions in groups of 2-4 participants. vs. **Control** Continue previous daily activity level.	Age: mean 48.0-49.0 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score 3.0-5.5 was an inclusion criterion. Mean 3.7-3.9 years across the two groups	
Dodd 2011[30] N=76 randomised Conducted in Australia	**Progressive resistance training** 10-week progressive resistance programme (45 min per session, two times weekly), with exercises targeting key lower limb muscles for support body weight and walking. Core exercises but also some could be individualised in terms of starting position or being replaced. All completed on weight machines. Two sets of 10-12 repetitions for each exercise. Weight lifted was increased when two sets of 12 repetitions could be completed. Rest periods were 2 min between each exercise set. Up to 12 participants attending each session. vs. **Control** Usual care in addition to a social programme. Could include normal exercise they participated in or therapy as long as it did not include progressive resistance training. Attention and social programme conducted for 1 h each week for 10 weeks – leisure and social activities not expected to have an effect on fitness or training, such as massage, lunches and educational sessions.	Multiple sclerosis Fatigue at baseline: >50% in both groups were considered to be fatigued at baseline (MFIS total score >38) Threshold for fatigue used for inclusion (yes/no): no Age: mean 48.0-50.0 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: not reported	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
Grubic Kezele 2019[44] N=19 randomised Conducted in Croatia	**Upper limb and breathing exercises** Two sessions per week (60 min per session) under physiotherapist supervision in addition to independent home exercise three days a week (at least 20 min per session) for 4 weeks. Exercises performed sitting in chair. Range of motions, resistance level and exercise speed was individualised to each person. 30-60 second pause after each exercise. Began with 15 min warm-up of breathing and active mobility of upper limbs. Breathing aimed to strengthen abdominal muscles, diaphragm and intercostal muscles. Exercises included range movement, coordination and strengthening with minimal resistance. vs. **Control** No exercise. Required to visit centre two times weekly (up to 60 min) where they could socialise and have contact with the investigators. Any existing exercise unchanged.	Multiple sclerosis Fatigue at baseline: mean score on MFIS Physical subscale 12.0-19.0 Threshold for fatigue used for inclusion (yes/no): no Age: not reported – 18.0-70.0 years was an inclusion criterion Type of MS: not reported EDSS: score between 0.0 and 8.0 was an inclusion criterion. Mean/median score not reported.	New study published since previous guideline version.
Sabapathy 2011[96] N=21 randomised Crossover study Conducted in Australia	**Resistance exercise** Two weekly sessions for 8 weeks. Intensity was intended to be moderate-hard. Three upper body and three lower body exercises as well as one core-strength and one stability exercise. Performed 2-3 sets, comprised of 6-10 repetitions of each exercise per set. Minimum 30-60 seconds rest between each exercise set. Resistance increased throughout using Therabands and/or weights used on applicable exercises. vs.	Multiple sclerosis Fatigue at baseline: mean score on MFIS Physical subscale 18.0-20.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 55.0 years for the whole population	Not previously included but identified from the new search 8-week washout period

Study	Intervention and comparison	Population	Comments
	Endurance exercise Two weekly sessions for 8 weeks. Intensity was intended to be moderate-hard. Circuit of eight exercise stations involving six different activities. 5 min at each station and rested for 2 min every 10 min. Exercise stations were: step ups, arm cranking, upright cycling, arm cranking, recumbent cycling, cross-trainer, treadmill walking and arm cranking. Exercise intensity increased throughout the program by adjusting resistance and/or cadence.	Type of MS: majority (63% of whole population) relapsing remitting EDSS: not reported	
Exercise – vestibular balance rehabilitation			
Callesen 2020[21] N=71 randomised across the three groups Conducted in Denmark	**Balance and motor control training** A total of 20 1 h training sessions over 10 weeks (2 sessions per week). All started with 10 min warm-up on stationary bicycle or treadmill. Followed task-oriented approach and covered tasks such as sitting, standing, stepping, walking and eye-movement training. To ensure they remained challenging, complexity level was maintained by variation and progression, for example by altering base of support or changing movement speed. Cognitive multitask challenges were added to some exercises. vs **Progressive resistance training** A total of 21 1 h training sessions over 10 weeks (2 sessions per week). 10 min warm-up on a stationary bicycle or treadmill. Focused on knee and hip flexion, with exercises progressing from 3 sets of 10 repetitions at 15 RM to 4 sets of 8 repetitions at 8 RM. Conducted using machines targeting specific muscle groups.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 41.0-44.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: median 52 years for the whole population Type of MS: majority relapsing-remitting (>80% in the whole population) EDSS: score 2.0-6.5 was an inclusion criterion. Median 3.5 for the whole population.	New study published since previous guideline version. Additional comparison of progressive resistance training vs. the control group is included under a separate section.

Study	Intervention and comparison	Population	Comments
	Balance and motor control training As described above. vs. **Control** Waitlist control group. Encouraged to maintain usual care and level of physical activity.		
Dettmers 2009[28] N=30 randomised Conducted in Germany	**Non-aerobic training (including balance)** Warming up, sensory training, stretching, balance, coordination training and periods of relaxation. Any training involving the heart and circulation was avoided. Training sessions lasted 3 weeks. vs. **Exercise intervention – primarily aerobic** Three weekly 45-min sessions. Warming up, mild strength training, repetitive endurance exercise, followed by relaxation and feedback. Attempts to camouflage training difficulties by including games and other playful elements. There were 3-5 participants per session, enabling an individual training plan. Generally, patients were trained to keep their own comfortable speed and not to compete too hard with other individuals.	Multiple sclerosis Fatigue at baseline: mean MFIS total score 37.0-42.0 across the two groups. Threshold for fatigue used for inclusion (yes/no): yes – complained of fatigue at baseline Age: mean 40.0-46.0 years across the two groups Type of MS: majority relapsing-remitting (77% of the whole population) EDSS: score <4.5 was an inclusion criterion. Mean score 2.6-2.8 across the two groups.	Included in previous guideline version.
Hebert 2011[47]	**Vestibular rehabilitation** Twice weekly for 6 weeks. Consisted of upright postural control and eye movement exercises. Each item was performed for 1-2 min for a total of 55 min.	Multiple sclerosis	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
N=38 randomised across the three groups Conducted in USA	Specific items were selected for a daily independent home exercise programme, which was to be performed throughout the intervention and follow-up phase. Also received 5 min fatigue management session including discussions of daily rest intervals, self-monitoring of exertion levels, workstation ergonomics and heat tolerance education. vs. **Exercise** Bicycle endurance training and stretching exercises. Twice weekly for 6 weeks. Endurance exercise consisted of stationary bicycling: 5-min warm-up, two 15 min sessions and 2 to 5 min cool down. Training intensity during the 15 min sessions was 65% to 75% of peak heart rate. For stretches, these were held for 30 seconds. Also received 5 min fatigue management session including discussions of daily rest intervals, self-monitoring of exertion levels, workstation ergonomics and heat tolerance education.	Fatigue at baseline: mean MFIS total score 51.0-56.0 across the three groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥45 on MFIS total Age: mean 43.0-50.0 years across the three groups Type of MS: majority relapsing-remitting (>80% for all groups) EDSS: score not reported	Additional comparison from this study of exercise vs. control is included under a separate section.
	Vestibular rehabilitation As described above. vs. **Control** Waitlist control group. Received usual medical care.		
Hebert 2018[48] N=88 randomised Conducted in USA	**Balance and eye movement exercises** Two times weekly with supervision and daily home exercise for 6 weeks (phase 1) followed by once weekly supervised session with daily home exercise for 8 weeks (phase 2). Three main elements are standing balance on different surfaces, mobility-based	Multiple sclerosis Fatigue at baseline: mean MFIS total score 49.0-50.0 across the two groups	New study published since previous guideline version

Study	Intervention and comparison	Population	Comments
	balance in walking with and without head movements and visual stability. vs. **Control** Waitlist control group.	Threshold for fatigue used for inclusion (yes/no): yes – score ≥22 on MFIS total Age: mean 43.0-47.0 years across the two groups Type of MS: not reported EDSS: mean score 3.3-3.5 across the two groups	
Karami 2018[56] N=75 randomised across the three original groups Conducted in Iran	**Vestibular rehabilitation or Frenkel exercises (coordination/balance)** Three exercise sessions performed on alternate days over 12 weeks. Sessions lasted ~60 min, including two 30 min sessions and two 15 min rest periods. Vestibular rehabilitation exercise was performed based Cawthorne and Cooksey methods. Performed in both the sitting and the upright position. Performed once with open eyes subsequently with eyes closed. Frenkel exercise group performed Frenkel exercises in sitting up, lying down and standing positions. vs. **Control** Routine usual care only.	Multiple sclerosis Fatigue at baseline: mean Fatigue Impact Scale total score of 89.0-93.0 across the three original groups Threshold for fatigue used for inclusion (yes/no): yes – score of 54-107 on the Fatigue Impact Scale Age: mean 33.0 years for the whole population Type of MS: majority relapsing-remitting (>90% for the whole population) EDSS: not reported	New study published since previous guideline version. Vestibular rehabilitation and Frenkel exercise group combined as both are types of balance intervention.
Sadeghi Bahmani 2019[97]	**Coordinative (balance) training** Three supervised group sessions weekly for 8 weeks (30-45 min per session). 5 min warm up, followed by exercises focused on coordination such as balancing	Multiple sclerosis	Additional comparison endurance exercise vs. control group is included under a separate section.

Study	Intervention and comparison	Population	Comments
N=92 randomised across the three groups Conducted in Iran	on a small bar, mirroring and imitating instructors' movements (such as dancing steps), balancing balls, mirroring participants' bouncing balls of different size, surface and weight, 'football-tennis', balancing with closed eyes on a rope on the floor and similar exercises. Aim was for participants to feel slightly exhausted but not severely exhausted. vs. **Endurance exercise** Three supervised group sessions (30-45 min per session) weekly for 8 weeks. 5 min warm-up and stretching, 25-35 min exercise on treadmill, exercise bicycles or walking/jogging with 1-2 min rest periods, followed by 5 min of cooling down. Aim was for participants to feel slightly exhausted but not severely exhausted.	Fatigue at baseline: mean score on FSS 39.0-43.0 across the three groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 38.0-39.0 years across the three groups Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Mean score 2.0-3.4 across the three groups.	New study published since previous guideline version
	Coordinative (balance) training As described above. vs. **Control** Met three times weekly for 30-45 min sessions at the hospital centre to control for social contact elements of the other interventions.		
Tramontano 2018[107] N=30 randomised	**Vestibular rehabilitation** Five sessions (20 min each) per week for 4 weeks to improve gaze stability (exercises performed while holding gaze on a firm target) and postural control (hold positions while standing on a foam cushion) by vestibular rehabilitation. This was in addition to two	Multiple sclerosis Fatigue at baseline: mean score on FSS 48.0-55.0 across the two groups	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
Conducted in Italy	daily 40 min sessions (5 times weekly) of conventional neurorehabilitation therapy for MS (muscle stretching, postural alignment, active-assisted mobilisations, neuromuscular facilitations and balance training)., which was also performed in the control group. vs. **Control – standard neurorehabilitation therapy** Two daily 40 min sessions (5 times weekly) of conventional neurorehabilitation therapy for MS (muscle stretching, postural alignment, active-assisted mobilisations, neuromuscular facilitations and balance training).	Threshold for fatigue used for inclusion (yes/no): no Age: mean age 46.0-51.0 years across the two groups Type of MS: not reported EDSS: score between 5.0 and 7.0 was an inclusion criterion. Mean score 6.3-6.7 across the two groups.	
Yazgan 2019[113] N=47 randomised Conducted in Turkey	**Video game-based balance exercises** Assigned to either Nintendo Wii Fit balance games or the Balance Trainer device. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean score on FSS 41.0-48.0 across the three original groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 41.0-48 years across the three original groups Type of MS: majority relapsing-remitting (>60% in the three original groups) EDSS: score between 2.5 and 6.0 was an inclusion criterion. Mean score 3.8-4.2 across the three original groups.	New study published since previous guideline version. Two balance training groups reported in this study combined and compared with the control group.

Study	Intervention and comparison	Population	Comments
Exercise - progressive resistance training + balance exercises			
Cakit 2010[20] N=45 randomised Conducted in Turkey	**Cycling progressive resistance training + balance exercises** Twice weekly progressive resistance training on static bicycle ergometer for 2 months, with each session followed by 5 min warm-up activities (walking) and stretching, and 20-25 mins of balance exercises, followed by 5 mins of whole body stretching vs. **Home-based exercise programme** Focus on lower limb muscle strengthening and balance. Twice weekly sessions for 2 months. Identical to group described above but without progressive resistance cycling: 5 min warm-up activities (walking) and stretching, and 20-25 mins of balance exercises, followed by 5 mins of whole body stretching	Multiple sclerosis Fatigue at baseline: mean score on FSS 40-50 across the three groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 36-43 years across the three groups Type of MS: relapsing-remitting or secondary progressive – proportion of each not reported EDSS: ≤6.0 for inclusion	Supervised and home-based interventions analysed separately as appear to be other differences in addition to whether home-based or supervised. Included in previous guideline version.
	Cycling progressive resistance training + balance exercises As described above. vs. **Control** No participation in any exercise programme and asked to continue their normal living		
	Home-based exercise programme As described above. vs. **Control**		

Study	Intervention and comparison	Population	Comments
	As described above.		
Tarakci 2013[103] N=110 randomised Conducted in Turkey	**Exercise** Three sessions (60 min each) per week for 12 weeks. Groups of up to 6/7 participants with similar age and EDSS score. Included flexibility, range of motion, strengthening with/without Theraband for lower extremity, core stabilisation, balance and coordination exercises and functional activities. Performed on alternate days. vs. **Control** Waitlist control group. Advised to continue usual routine but avoid beginning any new exercise during the study.	Multiple sclerosis Fatigue at baseline: mean score on FSS 39.0-40.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 40.0-42.0 years across the two groups Type of MS: majority relapsing-remitting (>60% in both groups) EDSS: score between 2.0 and 6.5 was an inclusion criterion. Mean score 4.2-4.4 across the two groups.	Included in previous guideline version.
Exercise – progressive resistance training + aerobic exercise			
Correale 2021[23] N= 27 randomised Conducted in Italy	**Endurance + resistance training** Endurance and resistance training at training facility twice weekly for 12 weeks. Sessions between 45- and 60-min. Sessions involved 5 min warm-up on motorised treadmill or cycle ergometer followed by 25 min aerobic training at moderate-vigorous intensity. Exercise progressively increased or decreased every 2 weeks based on heart rate responses. Endurance training followed by resistance training (calisthenics, dumbbells and elastic band exercises at 8-12 repetitions for each exercise). Load increased when sets of repetitions could be easily completed. vs.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total score was 39.9 and 44.8 in the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 45.4 and 48.3 years in the two groups	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	Control No further details provided, assume no intervention.	Type of MS: had to have relapsing-remitting MS to be included EDSS: score <4.0 was an inclusion criterion, mean scores at baseline not reported	
Garrett 2013[39] Associated studies: Garrett 2013[40] N=228 randomised across the three interventions Conducted in Ireland	**Resistance + aerobics class – led by physiotherapist** 1 h sessions for 10 weeks of circuit-style exercises that were resisted by body weight or free weights. In addition, they were advised to perform ~30 mins of aerobic exercise (of their own choice) twice weekly at an intensity of 65% max heart rate. Additional self-directed progressive resistance training and aerobic session was added from week 6. vs. **Control** Asked not to change their exercise habits during the 10-week treatment period	Multiple sclerosis Fatigue at baseline: mean score on MFIS 36-40 across the three groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 49-52 years across the three groups Type of MS: majority relapsing remitting (>50%) EDSS: not reported	Another comparison of yoga vs. control included under a separate section. Included in previous guideline version.
	Resistance + aerobics class – led by physiotherapist As described above. vs. **Yoga** Not pre-defined and differed depending on which yoga instructor gave the class.		
Razazian 2016[91]	**Exercise** Three weekly sessions of aquatic exercise for 8 weeks, including a series of water activities (1 h per	Multiple sclerosis	Additional comparisons of yoga vs. exercise and control groups are included under a separate section.

Study	Intervention and comparison	Population	Comments
N=36 randomised into these two groups Conducted in Iran	session). Included warming up, 10 min walking, stretching and gymnastics, 40 min power endurance activities, strength training and 10 min cooling down, relaxing, stretching and breathing exercises. vs. **Control** Non-exercise control group. Met 2-3 times weekly for 60-90 min. Able to talk to physicians and hospital staff, to complete everyday duties, to participate in occupational therapy and to meet and to talk to other patients.	Fatigue at baseline: mean score on FSS 39.6-48.7 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 33.0-35.0 years across the two groups Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Mean score 3.3-3.4 across the two groups.	New study published since previous guideline version.
Maurer 2018[71] N=178 randomised Conducted in Germany	**Internet-based exercise programme** Web-based app used to deliver adaptive and individualised exercise protocol. Home-based programme. Target intensity was moderate. Involved strengthening exercises twice weekly and endurance training once a week. Balance and core stability exercises could be added. vs. **Control** Waitlist control group – no exercise intervention.	Multiple sclerosis Fatigue at baseline: mean MFIS total 32.4 for the whole population Threshold for fatigue used for inclusion (yes/no): yes – MFIS score ≥14.0 Age: mean 40.2 years for the whole population Type of MS: majority relapsing-remitting MS was an inclusion criterion	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
		EDSS: score ≤3.5 was an inclusion criterion. Mean score 2.2 for the whole population.	
Exercise - aerobic + balance exercises			
Kargarfard 2018[58] N=40 randomised Conducted in Iran	**Exercise** 8-week aquatic exercise training. Three sessions per week (60 min per session), Including 10 min warm-up, 40 min exercise and a 10 min cool down. Intensity was 50-75% of maximum heart rate. Aquatic exercises included activities focused on joint mobility, functional exercises, balance and walking at different intensities. Met 2-3 times weekly with physical therapists for 30-40 min and received weekly educational sessions explaining topics such as the nature of MS, diagnosis and treatment, stress reduction techniques etc. vs. **Control** Maintain current treatment and behaviour throughout the 8-week treatment period. Met 2-3 times weekly with physical therapists for 30-40 min and received weekly educational sessions explaining topics such as the nature of MS, diagnosis and treatment, stress reduction techniques etc.	Multiple sclerosis Fatigue at baseline: mean MFIS total score of 43.8 for the whole population Threshold for fatigue used for inclusion (yes/no): no Age: mean 36.4 years for the whole population Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score ≤3.5 was an inclusion criterion. Mean score 3.6 for the whole population.	New study published since previous guideline version.
Exercise - progressive resistance training + aerobic + balance			
Gervasoni 2014[42] N=30 randomised Conducted in Italy	**Treadmill training** 15 min sessions of treadmill training in addition to 30 min conventional therapy (aimed to increase joint range of motion, muscle strength, balance, gait or upper limb function according to treatment plan). 12 treatment sessions over 2 weeks.	Multiple sclerosis Fatigue at baseline: median FSS score 4.5 for both groups Threshold for fatigue used for inclusion (yes/no): no	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
	vs. **Control** Conventional therapy. 45 min sessions of conventional therapy (aimed to increase joint range of motion, muscle strength, balance, gait or upper limb function according to treatment plan). 12 treatment sessions over 2 weeks.	Age: mean 46.0-50.0 years across the two groups Type of MS: majority relapsing-remitting (38-55% across the two groups) or secondary progressive (27-38% across the two groups) EDSS: median 5.0-5.5 across the two groups	
Kargarfard 2012[57] N=32 randomised Conducted in Iran	**Exercise** 8-week aquatic exercise training. Three sessions per week (60 min per session), Including 10 min warm-up, 40 min exercise and a 10 min cool down. Intensity was 50-75% of maximum heart rate. Aquatic exercises included activities focused on joint mobility, flexor and extensor muscle strength, balance movements, posture, functional activities and intermittent walking vs. **Control** Maintain current treatment and behaviour throughout the 8-week treatment period.	Multiple sclerosis Fatigue at baseline: mean MFIS total score of 42.0-46.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 32.0-34.0 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score ≤3.5 was an inclusion criterion. Mean score 2.9-3.0 across the two groups.	Included in previous guideline version
Kooshiar 2015[62] N=40 randomised	**Exercise** Aquatic exercise three times weekly for 8 weeks (45 min sessions). Shallow section of indoor pool. Included 36 movements including warm-up, stretching,	Multiple sclerosis Fatigue at baseline: mean MFIS total and FSS scores of 41.0-44.0 and	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
Conducted in Iran	endurance, balance/coordination, strengthening and cool down exercises. vs. **Control** No intervention and asked to continue usual routine and treatments.	38.0-42.0, respectively, across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 29.0 years for the whole population Type of MS: majority relapsing-remitting MS (>75% of whole population) EDSS: score 1.0-5.5 was an inclusion criterion. Mean score 2.5 for the whole population.	
Learmonth 2012[68] N=32 randomised Conducted in UK	**Aerobic, resistance and balance training exercises** Leisure-centred based group exercises. Twice weekly for 12 weeks led by physiotherapist and fitness instructor. 10 min warm up followed by 30-40 mins of circuit exercises designed to train aerobic endurance, resistance and balance, and cooling down exercises. vs. **Control** Usual care. Advised to continue their usual routine, seeking healthcare as required and to avoid beginning any new exercise regimes for the 12 weeks.	Multiple sclerosis Fatigue at baseline: mean score on FSS 5.3-5.7 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 51-52 years across the two groups Type of MS: not reported EDSS: mean 5.8-6.1 across the two groups	Included in previous guideline version.
Sangelaji 2014[99]	**Combination exercises – aerobic, strengthening and balance exercises**	Multiple sclerosis	New study published since last version of the guideline

Study	Intervention and comparison	Population	Comments
N=84 randomised Conducted in Iran	10 weeks of exercises 3 times weekly including stretching and 20-40 min aerobic (bicycle and treadmill) exercises, 10-15 min strengthening exercises and 10-20 min balancing exercises. vs. **Control** Not well defined – assume continued usual routine with no additional exercise.	Fatigue at baseline: mean score on FSS 34-38 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 32-33 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: 0-4 to be included. Mean 1.7-2.0 across the two groups.	
Negahban 2013[79] N=36 randomised across the three groups Conducted in Iran	**Exercise alone** Combined set of strength, stretch, endurance and balance training exercises vs. **Control** Continue standard medical care and asked to avoid participation in any new exercise programme or change usual activities for 5 weeks	Multiple sclerosis Fatigue at baseline: mean score on FSS 41.3-47.6 across the three groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 36.3-36.8 years across the three groups Type of MS: relapsing-remitting MS or secondary progressive MS was an inclusion criterion. Proportion with each not reported. EDSS: mean 3.5-3.8 across the three groups.	Additional comparisons from this study are included under separate sections. Included in previous guideline version.
	Exercise alone As described above. vs. **Massage alone** Three 30 min supervised intervention sessions per week for 5 weeks Swedish massage by trained massage therapist.		

Study	Intervention and comparison	Population	Comments
Straudi 2014[102] N=24 randomised Conducted in Italy	**Task-oriented circuit training** Sessions performed 5 times weekly for 2 weeks. Used six different stations where exercises performed for 3 min each. Two laps per session lasting 60 min. Walking endurance also trained at each session for 30 min on treadmill. Subsequently, brochure given so they could continue independently for 3 months – gait training, stretching and strengthening exercises. Some exercises appear to have balance components as well. vs. **Control** Usual care – no specific rehabilitation for gait performance and mobility improvement.	Multiple sclerosis Fatigue at baseline: median score on FFS 5.4-5.8 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 52.6 years for the whole population Type of MS: relapsing-remitting, primary progressive or secondary progressive MS was an inclusion criterion. Most had relapsing-remitting (42%), with 33% secondary progressive and 25% primary progressive. EDSS: score between 4.0 and 5.5 was an inclusion criterion. Mean 4.89 for whole population.	New study published since previous guideline version.
Exercise (aerobic + resistance training) + cognitive behavioural therapy			
Carter 2014[22] N=120 randomised Conducted in UK	**Exercise intervention** EXIMS. 3-month exercise intervention in addition to usual care. Two supervised sessions per week during weeks 1-6 at a university exercise research facility and one additional self-directed exercise session in home environment. Supervised exercise sessions involved up to three participants and lasted ~1 h. Aerobic exercise was core exercise modality (short bouts of low-moderate intensity aerobic exercise e.g., cycling, rowing at 50-69% of predicted maximum heart rate). Where appropriate, participants	Multiple sclerosis Fatigue at baseline: mean MFIS total score 43.0-45.0 across the two groups. Threshold for fatigue used for inclusion (yes/no): no	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
	also performed exercises for strength and control (majority did take part in this and typically involved 2-6 types of resistance exercises). Balance exercises included where control and coordination were a problem and static stretching exercises for large skeletal muscle groups were also included if appropriate. Cognitive behavioural techniques also incorporated in terms of goal setting, finding social support and the costs and benefits of exercising. vs. **Control** Usual care. Waitlist control group.	Age: mean 46.0 years for both groups Type of MS: majority relapsing-remitting (>75% in both groups) EDSS: score between 1.0 and 6.5 was an inclusion criterion. Mean score 3.8 for both groups.	
Standard exercises (progressive resistance training + aerobic + balance) + high-intensity lower limb resistance training vs. standard exercises alone			
Hayes 2011[46] N=22 randomised Conducted in USA	**Standard exercises + high-intensity lower limb resistance training** Standard exercises included aerobic training on a recumbent stepper, lower extremity stretching, upper extremity strength training and balance exercises using a wobble board. These were performed 3 times weekly for 45-60 min each time for 12 weeks. In this group, additional high-intensity resistance training using an ergometer was performed 3 times a week for 12 weeks. **Standard exercises only** Standard exercises included aerobic training on a recumbent stepper, lower extremity stretching, upper extremity strength training and balance exercises using a wobble board. These were performed 3 times weekly for 45-60 min each time for 12 weeks.	Multiple sclerosis Fatigue at baseline: mean score on FSS 5.8-6.1 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 49 years for whole population Type of MS: not reported EDSS: mean 5.24 for whole population	Comparison differs to other similar studies in terms of the components of the two interventions so not meta-analysed Included in previous guideline version.
Multidisciplinary rehabilitation (medical, exercise, counselling and fatigue self-management) + methylprednisolone			

Study	Intervention and comparison	Population	Comments
Nedeljkovic 2016[78] N=39 randomised Conducted in Serbia	**Multidisciplinary rehabilitation programme + methylprednisolone** Intravenous methylprednisolone 1 g/day for 5 days. Provision of mobility aids, bladder management and instruction on basic exercises performed in the clinic and for 5 days at home. Followed by outpatient rehabilitation programme 1-3 days after intravenous treatment, which was multidisciplinary and contained the following components: medical treatment for symptoms, exercise therapy (individually tailored – 5 times weekly for 1 h, including aerobic activity and possibly other types), fatigue self-management and counselling, and occupational therapy visits. vs. **Control** Treated in accordance with standard procedure, which does not recommend inclusion in rehabilitation following intravenous methylprednisolone treatment	Multiple sclerosis Fatigue at baseline: mean score on FSS 41-43 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 40-42 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: mean 4.2-4.4 across the two groups.	Very different population to others included – specifically in a relapse population requiring methylprednisolone treatment New study published since last version of the guideline
Massage + exercise (strength, stretching, endurance and balance)			
Negahban 2013[79] N=36 randomised across the three groups Conducted in Iran	**Massage + exercise** Combined exercise and massage (15 minutes of each per 30-minute session). Massage consisted of three supervised intervention sessions per week for 5 weeks Swedish massage by trained massage therapist. Exercise involved combined set of strength, stretch, endurance and balance training exercises. vs. **Exercise alone** Combined set of strength, stretch, endurance and balance training exercises	Multiple sclerosis Fatigue at baseline: mean score on FSS 41.3-47.6 across the four groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 36.3-36.8 years across the four groups	Additional comparisons from this study are included under separate sections. Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
	Massage + exercise As described above. vs. **Control** Continue standard medical care and asked to avoid participation in any new exercise programme or change usual activities for 5 weeks	Type of MS: relapsing-remitting MS or secondary progressive MS was an inclusion criterion. Proportion with each not reported. EDSS: mean 3.5-3.8 across the four groups	
	Massage + exercise As described above. vs. **Massage alone** Three 30 min supervised intervention sessions per week for 5 weeks Swedish massage by trained massage therapist.		
Balance training + Pilates			
Ozkul 2020[82] N=54 randomised Conducted in Turkey	**Balance training + Pilates** Pilates-based core stability training for 30 min. Followed by 10 min rest and 20 min of either immersive virtual reality balance or non-virtual reality balance training. Twice weekly for 8 weeks. **Control - relaxation** Jacobson's progressive relaxation exercise taught by physiotherapist once and asked to practice at home for 15-20 min twice weekly for 8 weeks.	Multiple sclerosis Fatigue at baseline: median score on FSS 46.0-49.0 across the three groups originally randomised Threshold for fatigue used for inclusion (yes/no): no Age: median 29-34 across the three groups originally included	Study randomised to three groups but for the purpose of this review the two balance training groups were combined and compared as a single group against the control. New study published since last version of the guideline

Study	Intervention and comparison	Population	Comments
		Type of MS: relapsing-remitting MS was an inclusion criterion. EDSS: score <6.0 an inclusion criterion. median 1.0-2.0 across the groups	
Multidisciplinary rehabilitation – physical activity with/without other rehabilitation + fatigue self-management programme			
Hersche 2019[51] N=47 randomised Conducted in Switzerland	**Inpatient energy management education** Fatigue management group-based education – 6.5 h over 7 face-to-face sessions in 3 weeks, delivered by an occupational therapist. In addition to usual 3-week rehabilitation (physio – endurance and reinforcement training, occupational therapy, speech therapy, neuropsychological training and counselling, if relevant). No fatigue management was discussed as part of the occupational therapy sessions in the usual rehabilitation component. vs. **Control – relaxation** Progressive muscle relaxation – Jacobson's. Delivered by trained physiotherapist. 6 1 h face-to-face group sessions over 3 weeks. In addition to usual 3-week rehabilitation (physio – endurance and reinforcement training, occupational therapy, speech therapy, neuropsychological training and counselling, if relevant). No fatigue management was discussed as part of the occupational therapy sessions in the usual rehabilitation component.	Multiple sclerosis Fatigue at baseline: mean score on FFS 9.8-10.1 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – score on FSS >4.0 Age: mean age 51.0-52.0 across the two groups Type of MS: majority were relapsing-remitting (32%), primary progressive (23%) or secondary progressive (32%) EDSS: score ≤6.5 was an inclusion criterion. Mean score 4.8-5.3 across the two groups.	New study published since previous guideline version.
Plow 2019[86] N=208 randomised	**Physical activity + fatigue self-management** Delivered via phone for 12 weeks. 3 group teleconference sessions and 4 individually tailored phone calls. Taught how to engage in pedometer-	Multiple sclerosis	An additional comparison from this study of physical activity vs. control is included under a separate section.

Study	Intervention and comparison	Population	Comments
Conducted in USA	based walking programme, set goals, overcome barriers and self-monitor progress. Additionally received components of the 'Managing Fatigue' intervention, a 6-week energy conservation course. vs. **Physical activity only** Delivered via phone for 12 weeks. 3 group teleconference sessions and 4 individually tailored phone calls. Taught how to engage in pedometer-based walking programme, set goals, overcome barriers and self-monitor progress.	Fatigue at baseline: mean score on FIS 68.0-71.0 across the three groups Threshold for fatigue used for inclusion (yes/no): yes – had to have moderate-severe fatigue (defined as score ≥4.0 on FSS) Age: mean age 52 years for whole population Type of MS: majority had relapsing-remitting MS (>80%). EDSS: not reported – on PDDS, required score of 1.0-5.0 for inclusion.	New study published since last version of guideline
	Physical activity + fatigue self-management As described above. vs. **Control** Delivered via phone for 12 weeks. Information on health topics relevant to MS. Designed as a contact control group.		
Rietberg 2014[93] N=48 randomised Conducted in The Netherlands	**Multidisciplinary outpatient rehabilitation programme** Individually tailored programme focused on optimising self-management behaviour in daily life on fitness, behaviours that worsen fatigue and energy conservation – this could include any combination of physical therapy (12 weeks aerobic training), occupational therapy (fatigue management skills) and social work (help with support and counselling). vs.	Multiple sclerosis Fatigue at baseline: median score on FFS 48.0-52.0 across the two groups; median score on MFIS total 36-43 across the two groups. Threshold for fatigue used for inclusion (yes/no): yes – had suffer from chronic fatigue (fatigue present	New study published since last guideline version.

Study	Intervention and comparison	Population	Comments
	Outpatient MS-nurse consultation Nurse discussed general principles of activity planning, prioritisation, energy conservation and accepting help from others. Physical activity was recommended. Advise on alcohol and nutrition was given.	for any length of time on at least 50% of days for more than 6 weeks) Age: mean age 45-47 years across the two groups Type of MS: majority had relapsing-remitting MS (>50%). EDSS: median 3-4 across the two groups.	
Fatigue management and energy conservation management programmes			
Abonie 2020[1] N=24 randomised Conducted in UK	**Tailored activity pacing intervention** Tailored programme based on accelerometer data for each person. Then provided with information relevant to their behaviours e.g., those avoiding activity in response to fatigue or to prevent it given information to develop consistent physical activity increase. Those overdoing activity when feeling better given information about balancing rest and exercise. vs. **Control** No definition provided – assume continue usual lifestyle without the intervention	Multiple sclerosis Fatigue at baseline: mean score on FFS 4.7-4.8 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 58.0-61.0 across the two groups Type of MS: majority were relapsing-remitting (48%) or secondary progressive (43%) EDSS: not reported – median PDDS score was 2.0-3.5 across the two groups.	New study published since previous guideline version
Blikman 2017[13] N=86 randomised	**Energy conservation management**	Multiple sclerosis	Part of TREFAMS-ACE study group

Study	Intervention and comparison	Population	Comments
Conducted in The Netherlands	Individual energy conservation management. Protocol adapted based on a group programme. 12 sessions over 4 months delivered by an occupational therapist. vs. **Control** Information only control. Three MS nurse consultations over 4 months. Standardised information about MS-related fatigue without providing treatment or advice.	Fatigue at baseline: mean score CIS20r-fatigue, MFIS total and FSS were 44 (for both groups), 43-45 and 5.1-5.3, respectively. Threshold for fatigue used for inclusion (yes/no): yes – severely fatigued according to CIS20r – fatigue subscale (score ≥35) Age: mean age 47.0-48.0 across the two groups Type of MS: majority were relapsing-remitting (>70%) EDSS: score ≤6.0 was an inclusion criterion. Median 1.8-2.5 across the two groups.	New study published since previous guideline version
Finlayson 2011[34] N=190 randomised Conducted in USA	**Fatigue management programme** 6-week group-based intervention delivered by teleconference. Weekly teleconferences for 6 weeks lasting 70 min each delivered by occupational therapist. Involved teaching sessions and homework. vs. **Control** Waitlist control group. Assume continued usual lifestyle for the duration of the intervention.	Multiple sclerosis Fatigue at baseline: mean score on FFS 5.0 for the whole population Threshold for fatigue used for inclusion (yes/no): yes – moderate to severe fatigue (score ≥4.0 on FSS) Age: mean age 56.0 years for the whole population Type of MS: majority were relapsing-remitting (52%) or secondary progressive (22%)	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
		EDSS: not reported – mean PDDS score for the whole population was 4.0	
García Jalón 2013[38] N=23 randomised Conducted in UK	**Energy conservation management** Group format with one weekly 2 h session for 5 weeks. Derived from previous publications and piloted, with modification subsequently made. vs. **Control** Peer support group. Education and discussion of common topics for people with MS as recommended by MS Society, MS Action and the MS Trust.	Multiple sclerosis Fatigue at baseline: mean score on FFS 5.9 for both groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥4.0 on FSS Age: mean age 46.0-52.0 across the two groups Type of MS: majority were secondary progressive (57%) EDSS: score ≤6.0 was an inclusion criterion.	Included in previous guideline version.
Hugos 2010[54] N=41 randomised Conducted in USA	**Fatigue management programme** Fatigue: Take Control programme, consisting of group-based 2 h sessions weekly for 6 weeks, including DVD viewing, topic-focused group discussion, goal setting and homework assignments. Focus on guiding environmental, behavioural and lifestyle changes needed to manage fatigue. vs. **Control**	Multiple sclerosis Fatigue at baseline: mean score on FFS and MFIS total was 52.0 and 45.0, respectively, for the whole population Threshold for fatigue used for inclusion (yes/no): no Age: mean age 57 years for the whole population	Included in previous guideline version

Study	Intervention and comparison	Population	Comments
	Waitlist control group. Met twice weekly for 20-30 min to complete study documents.	Type of MS: proportion with different types not reported. EDSS: score ≤6.0 was an inclusion criterion. Mean score 5.2 for the whole population	
Hugos 2019[53] Associated papers: Hugos 2019[52] N= 218 randomised Conducted in USA	**Fatigue management programme** Fatigue: Take Control programme. Face-to-face group programme. 6 weekly 2 h sessions. DVD viewing, topic-focused group discussion, individual goal setting, and homework assignments. Focus on guiding environmental, behavioural and lifestyle changes needed to manage fatigue. vs. **Control – general self-management programme** MS: Take Control programme. General education face-to-face group programme. 6 weekly 2-hour sessions. No DVD or goal setting processes. Structured based on educational pamphlets. Homework was to read the pamphlets to be discussed at the next session. If the topic of fatigue arose discussion was allowed to proceed until conversation led back to topic of that day's discussion.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 46.1-46.7 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – moderate to severe fatigue (score ≥25.0 on MFIS total) Age: mean age 54.0 in both groups Type of MS: majority were relapsing-remitting (59%) EDSS: score ≤6.5 was an inclusion criterion. Mean score 5.1-5.3 across the two groups.	New study published since previous guideline version
Kos 2007[64] N=51 randomised Conducted in Belgium	**Fatigue management programme** Four sessions lasting 2 h spread over 4 weeks. Structured as follows: information from instructor, followed by discussion with the group on current and planned strategies. People in this group were given information about possible strategies for fatigue management and energy conservation.	Multiple sclerosis Fatigue at baseline: median score on MFIS total 46.0 in both groups Threshold for fatigue used for inclusion (yes/no): yes – high	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
	vs. **Control** Four sessions lasting 2 h spread over 4 weeks. Structured as follows: information from instructor, followed by discussion with the group on current and planned strategies. This group received information that was interesting enough to avoid drop-outs but not directly related to fatigue.	fatigue impact (score ≥3.0 on fatigue subscale of Guy's Neurological Disability scale) Age: mean 43.0-45.0 years across the two groups Type of MS: majority were relapsing-remitting (67%) EDSS: not reported	
Kos 2016[65] N=31 randomised Conducted in Belgium	**Individual self-management occupational therapy intervention** Self-management occupational therapy (SMooTh) programme based on recommendations of MS Council. Covers strategies help take control of performance activities within the limits of their available energy and therefore raise their self-efficacy in managing fatigue. Includes techniques such as goal setting, self-monitoring and feedback. Consists of three individual sessions of 60-90 minutes for 3 consecutive weeks and is provided by an experienced occupational therapist. Also provided with a booklet containing information on fatigue, strategies to cope with fatigue and pace activities. vs. **Control – relaxation** Education about the role of stress (management) in MS	Multiple sclerosis Fatigue at baseline: mean MFIS total score 44.0-45.0 across the two groups. Threshold for fatigue used for inclusion (yes/no): yes – high impact of fatigue (fatigue VAS score ≥60 – scale unclear, possibly 0-100?) Age: mean 37.0-44.0 years across the two groups Type of MS: not reported EDSS: score ≤5.0 was an inclusion criterion. Median score 3.0-3.5 across the two groups.	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	and practicing relaxation techniques like Jacobson, Schultz and visualization. Information also assembled in an evidence-based information booklet, and participants completed a stress-reaction diary to register activities or events that evoke stress. This diary was used to coach clients in improving coping with similar future stress events. The mode, duration and frequency of the relaxation therapy were identical to the SMOoTh intervention.		
Mathiowetz 2005[70] N=169 randomised Conducted in USA	**Energy conservation management** 6-week energy conservation management course. Delivered by occupational therapists. Group-based intervention, consisting of 2 h sessions, including lectures, discussions, goal setting, practice activities and homework. vs. **Control** No intervention for 6 weeks.	Multiple sclerosis Fatigue at baseline: mean score on FSS 5.9 in both groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥4.0 on FSS Age: mean 48.0-49.0 years across the two groups Type of MS: majority were relapsing-remitting (>60% in both groups) EDSS: not reported	Included in previous guideline version
FACETS (Fatigue: Applying Cognitive behavioural and Energy Effectiveness Techniques to lifestyle)			
Thomas 2014[104] Associated papers: Thomas 2013[105]	**Fatigue management programme – FACETS** Elements of cognitive behavioural, social cognitive, energy effectiveness, self-management and self-efficacy theories. Aim to normalise fatigue experiences, learn helpful ways of thinking about	Multiple sclerosis Fatigue at baseline: mean score on Global Fatigue Severity subscale of	12-month follow-up of a trial already included in previous guideline version now available and extracted.

Study	Intervention and comparison	Population	Comments
N=164 randomised Conducted in UK	fatigue and use available energy more effectively. Consists of 6 group sessions (~90 min duration) held weekly by two health professionals such as occupational therapists, nurses or physiotherapists. Involves presentations, discussions, group activities and homework. vs. **Control** Current local practice. Ranged from advice and information provision about MS fatigue to more detailed individualised management advice from a variety of health professionals. Variation in the exact composition of what was usually provided, within and between centres, depending on local resources and patient need.	Fatigue Assessment Instrument 5.6 in both groups Threshold for fatigue used for inclusion (yes/no): yes – score >4.0 on FSS Age: mean 48.0-50.0 years across the two groups Type of MS: majority were relapsing-remitting (>40% in both groups) or secondary progressive (≥20% in both groups) EDSS: not reported - >70% in both groups had score ≥4 on adapted PDDS (MS interferes with walking)	
Diets			
Bohlouli 2021[14] N=180 randomised Conducted in Iran	**Modified Mediterranean diet** 17% protein, 51% carbohydrate and 32% fat based on higher consumption of fresh fruits and vegetables, whole grains, monounsaturated fatty acids, fish, and low to moderate consumption of dairy products, meat, and poultry. Prescribed diet was individualised based on cultural and personal preferences, and the elimination of any alcohol-containing foods and beverages. 6-month intervention. vs. **Traditional Iranian diet**	Multiple sclerosis Fatigue at baseline: mean score on MFIS total score was 72.4 and 69.5 in the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean age 39-40 years in the two groups	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	Low in low-fat dairy products, whole grains; high in red meats, solid oils, refined grains, and moderate intakes of legumes, fruits and vegetables). 13% protein, 58% carbohydrate and 29% fat. Did not continue normal eating pattern as diet was adjusted for energy intake to avoid unexpected body weight changes. Individualised plan for each person. 6-month intervention.	Type of MS: had to have relapsing-remitting MS to be included EDSS: score of up to 3.0 an inclusion criterion, mean scores were 1.7 and 2.0 in the two groups	
Irish 2017[55] N=34 randomised Conducted in USA	**Modified Paleo diet** Nine cups of vegetables and some fruits, meat protein including organ meat, and complete abstinence from products containing gluten, dairy, potatoes, and legumes. Continue concurrent MS therapy and/or medications. Intervention was for 3 months. vs. **Control** Usual care. Typical physician recommendations for MS. Continue concurrent MS therapy and/or medications.	Multiple sclerosis Fatigue at baseline: mean score on FSS 4.0-4.2 across the two groups (1-9 scale) Threshold for fatigue used for inclusion (yes/no): no Age: mean 35.0-37.0 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: not reported	New study published since previous guideline version.
Katz Sand 2019[59] N=36 randomised Conducted in USA	**Mediterranean diet** 6-month intervention. Encouraged intake of fresh vegetables and fruits, fish, nuts, legumes, whole grains, avocados and the use of olive oil in cooking. Advised against red and white meat, dairy, white grains and processed foods. Also advised to limit salt intake to <2 g/day and to refrain from eating for at least 12 h a day (e.g., window from 7 pm to 7 am). No specific advice given in terms of	Multiple sclerosis Fatigue at baseline: not reported Threshold for fatigue used for inclusion (yes/no): no Age: median 43.0 years for the whole population	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	calorie intake and weight loss. Group sessions with dietician attended at beginning of the dietary protocol. vs. **Control** Non-dietary group. Attended educational seminars on MS.	Type of MS: majority had relapsing-remitting MS (80% of whole population) EDSS: median score 2.0 for whole population	
Mousavi-Shirazi-Fard 2021[75] N=104 randomised Conducted in Iran	**Modified anti-inflammatory diet** Personalised diet for each participant based on an anti-inflammatory diet, lasting for 12 weeks. Aimed for 55% energy to come from carbohydrates, 15% from proteins and 30% from fat. Diet intended to maintain weight not to lose weight. Lots of fruits and vegetables included in the diet. Substitute white bread and rice with brown bread and rice, and high-fat dairy with low fat probiotics. Legumes and healthy oils for cooking were recommended. Nuts, spices, green and white tea, dark chocolate, lean poultry and fish also included. Advised to limit lean red meat and eggs to 1-2 times weekly. Refined carbohydrates and sugary products, as well as processed foods, were not recommended. vs. **Control** Received healthy diet recommendations based on World Health Organisation's healthy diet, rather than a personalised diet plan.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 48.0 in both groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 35.0-36.0 years across the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score <5.5 was an inclusion criterion. Majority had EDSS score 0-4 (87%).	New study published since previous guideline version.
Razeghi-Jahromi 2020[92]	**Mediterranean-based diet** Diets personalised to patients following interview with dietician based on dietary intake, habits and	Multiple sclerosis	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
N=80 randomised Conducted in Iran	preferences, as well as energy requirement calculation and nutritional needs. Prescribed diet adjusted according to new weight assessments. Energy needs and macronutrients proportional to age, sex and BMI. Generally, diet was modified in accordance with Mediterranean diet apart from wine and other unspecified foods. Advice focused on encouraging increased consumption of healthy oils (especially olive and olive oil), whole grains, vegetables, fruits and raw and unroasted nuts and seeds, legumes, and healthy plant-based foods. Consumption of fish and seafood (~2 times weekly), poultry, eggs, and low fat or skimmed dairy (daily to weekly) was recommended. Participants also instructed to limit the intake of red meat, fried foods, and refined grains and to minimise the consumption of simple sugar, sugary foods and beverages, processed meat, and animal-based fats to as low amounts as possible. Patients in both groups advised to have five meals daily and were not aware of whether they had received the intervention or control diet. 1-year intervention. vs. **Standard healthy diet** Nutritionist-aided diet in accordance with US Department of Agriculture dietary guidelines for Americans, 2010. Guidelines customised to be proportionate to age, sex and BMI. Propose food-based recommendations for promoting public health, aiming to ensure dietary requirements are met and to prevent development and progression of chronic	Fatigue at baseline: mean score on MFIS total was 40.05 and 38.19 in intervention and control groups, respectively Threshold for fatigue used for inclusion (yes/no): no Age: mean 34 years in both groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score <5.5 was an inclusion criterion. Mean EDSS score was 2.27 and 2.4 in intervention and control groups, respectively	

Study	Intervention and comparison	Population	Comments
	disease. Patients in both groups advised to have five meals daily and were not aware of whether they had received the intervention or control diet. 1-year intervention		
Wahls 2021[112] N=87 randomised Conducted in USA	**Wahls diet (modified Palaeolithic elimination diet)** 12-week run-in period for observation of usual diet and stability of pre-intervention outcomes. Followed by Wahls diet for 24 weeks. Two in-person and five telephone-based nutrition counselling sessions in first 12 weeks. Personalised emails with feedback on dietary checklists every 4 weeks. 6-9 servings of fruit and vegetables and provides 6-12 ounces meat per day according to gender. It excludes all grain, legumes, eggs, and dairy (except for clarified butter or ghee). Nightshade vegetables were also excluded in the Wahls group during the first 12-week period from baseline. Instructed to follow their assigned diet ad libitum and were given the following daily supplement regime: 1 teaspoon cod liver oil, 1,000 mg methyl-B12, 1,000 mg methylfolate, a multivitamin without iron, and 5,000 IU vitamin D3, the latter of which was adjusted based on serum levels with a target range of 40 to 80 ng/mL vs. **Swank diet (low-saturated fat diet)** 12-week run-in period for observation of usual diet and stability of pre-intervention outcomes. Followed by Wahls diet for 24 weeks. Two in-person and five telephone-based nutrition counselling sessions in first 12 weeks. Personalised emails with feedback on dietary checklists every 4 weeks. Restricts saturated fat to 15 g per day and provides 20-50 g (4-10 teaspoons) unsaturated fat per day and four servings each of grains, whole preferred, and fruits and	Multiple sclerosis Fatigue at baseline: mean score on FSS (scale 1-9 in this study) was 5.2-5.3 in the two groups Threshold for fatigue used for inclusion (yes/no): yes – moderate to severe fatigue (FSS at least 4.0) was an inclusion criterion Age: mean 46-47 years in the two groups Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: unclear, likely <6.0 as had to be able to walk unassisted or only with unilateral aid	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	vegetables. Instructed to follow their assigned diet ad libitum and were given the following daily supplement regimen: 1 teaspoon cod liver oil, 1,000 mg methyl-B12, 1,000 mg methylfolate, a multivitamin without iron, and 5,000 IU vitamin D3, the latter of which was adjusted based on serum levels with a target range of 40 to 80 ng/mL.		
Mindfulness training			
Grossman 2010[43] N=150 randomised Conducted in Iran	**Mindfulness-based intervention** Involved an interview to set realistic goals, 8 weekly 2.5 h group classes in mindfulness practices, 1 7-hour session at week 6, homework assignments (~ 40 min/day). Each class covered specific exercises and topics within the context of mindfulness training. Conducted by 2 experienced teachers with >9 years teaching experience. Also received usual care as described below. vs. **Control** Usual care – currently optimal medical care during the duration of the study, as provided by the neurology department of the hospital. This included one medical examination at preintervention and another at 6 months post-intervention, with additional measures as individually required.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 30.0-35.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 47.3 years for the whole population Type of MS: relapsing-remitting MS or secondary progressive MS was an inclusion criterion. Majority had relapsing-remitting MS (82% of whole population) EDSS: score ≤6.0 was an inclusion criterion. Mean score 3.01 for the whole population.	Included in previous guideline version.
Self-management programme			
Afrasiabifar 2016[2] N=63 randomised	**Orem's self-care model** Orem's self-care model was applied during six sessions of 45 - 60 min duration (3 weeks). After the sessions were over, the self-care model was applied	Multiple sclerosis Fatigue at baseline: mean score on FSS 6.0-6.2 across the two groups	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
Conducted in Iran	for 4 weeks at home. Methods of helping included acting, guiding, teaching, supporting and providing an environment. Covered wide range of areas depending on participant need including nutrition, energy management, bladder training, management of pain, social interaction and accepting the disease, among others. vs. **Control** No intervention was conducted, and the participants received only care and training routines.	Threshold for fatigue used for inclusion (yes/no): no Age: mean 29.0-31.0 years across the two groups Type of MS: Majority had relapsing-remitting MS (>90% in both groups) EDSS: not reported	
Barlow 2009[11] N=142 randomised Conducted in UK	**Lay-led self-management for MS** Chronic Disease Self-Management Course. Not MS specific and used for different types of chronic diseases. Weekly sessions for 6 weeks each lasting ~2 h delivered by tutors trained in course delivery. Covers generic topics such as self-management principles, management of pain, fatigue, exercise, relaxation techniques, dealing with depression, nutrition, communication with family and health professionals, problem solving and goal setting. Format consisted of short lectures, group discussion, role plays and trying out techniques. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean score on fatigue VAS (0-10 scale) 4.8-5.7 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 48.0-51.0 years across the two groups Type of MS: not reported EDSS: not reported	Study not included previously but upon review considered to be relevant.
Ehde 2015[32] N=163 randomised	**Telephone-delivered self-management programme** 8 weekly 45-60 min telephone sessions. Cognitive behavioural and positive psychology strategies to help	Multiple sclerosis	

Study	Intervention and comparison	Population	Comments
Conducted in USA	self-manage pain, depression and fatigue in daily lives. At final session a comprehensive self-management plan was created integrating preferred skills and goals for use post-treatment. vs. **Control** 8 weekly 45-60 min telephone sessions. Telephone education programme covering topics such as fatigue, pain and nutrition.	Fatigue at baseline: mean score on 5-item MFIS 48.0-51.0 across the two groups. 81% met the criteria for fatigue (score ≥10 on 5-item MFIS short form) Threshold for fatigue used for inclusion (yes/no): partially yes – had to have at least one of following: score 10-14 on PHQ-9 (depression), score ≥3 for pain intensity (scale 0-10) or significant fatigue (score ≥10 on 5-item MFIS short form) Age: mean 51.0-53.0 years across the two groups Type of MS: Majority had relapsing-remitting MS (>50% in both groups) EDSS: majority (>60% in both groups) had EDSS score 4.5-6.0	
Self-management programme + exercise			
Lutz 2017[69] N=18 randomised Conducted in Germany	**Exercise-based patient education programme** 6-week exercise patient education programme. Provide with knowledge to work out independently. Participants were Various types of exercise training (cardiorespiratory, strength, coordination/reflex-based, and flexibility) were offered based on individual performance abilities. Psychological determinants for adoption and maintenance of health-related behaviour, such as self-efficacy, problem-solving, and patient-generated goal setting	Multiple sclerosis Fatigue at baseline: mean WEIMuS fatigue score 21.0-26.0 across the two groups. Threshold for fatigue used for inclusion (yes/no): no Age: mean 52.0-56.0 years across the two groups	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	were taught to enhance motivation and self-management skills. Delivered over 6 weeks, twice a week for 60 to 90 min per session. Information booklets and homework also provided. After sessions, exercise programme was continued at home for 12 weeks and beyond until last follow-up. vs. **Control** Waitlist control group.	Type of MS: majority relapsing-remitting or primary progressive MS (79% of whole population) EDSS: Median score 3.5 in both groups.	
Functional electrical stimulation + exercise			
Backus 2020[10] N=24 randomised Conducted in USA	**Functional electrical stimulation cycling** 12-week functional electrical stimulation cycling training (aim was three times weekly for 12 weeks). Performed while seated in wheelchair. Electrodes placed over muscles of gluteus maximus, hamstrings and quadriceps. Cycled volitionally with assistance from electrical stimulation as needed – 2 min warm-up, 30 min active phase (cycling with/without stimulation) and 2 min cool down. Goal was cycling speed of 35-50 rpm. Stimulation parameters were 200 microseconds for pulse width and frequency of 50 Hz. Resistance added in 0.14 Nm increments once participants could achieve the target exercise duration and speed for three consecutive sessions. vs. **Control** Waitlist control group. Encouraged to keep current activities and medications constant.	Multiple sclerosis Fatigue at baseline: mean score on FSS 4.4 for the whole population Threshold for fatigue used for inclusion (yes/no): yes – score >2.3 on FSS Age: mean age 55.0 years for the whole population Type of MS: includes some with relapsing-remitting and some with secondary progressive MS, with some not specified EDSS: score >6.5 was an inclusion criterion. Median score 7.2 for the whole population.	New study published since previous guideline version.
Yoga			

Study	Intervention and comparison	Population	Comments
Ahmadi 2010[5] N=21 randomised Conducted in Iran	**Yoga** 8-week programme of Hatha yoga classes (three sessions per week lasting 60-70 min each). Stretching techniques followed by standing, supine, prone-lying and sitting postures. Poses held for 10-30 seconds with rest periods between poses of 30 seconds to 1 min. Patients supported for most poses by a chair, Swiss ball or wall. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean FSS score 4.1 for the whole population Threshold for fatigue used for inclusion (yes/no): no Age: mean 34.0 years for the whole population Type of MS: not reported EDSS: score between 1.0 and 4.0 was an inclusion criterion. Mean score 2.1 for the whole population.	Included in previous guideline version.
Ahmadi 2013[4] N=31 randomised across the three groups Conducted in Iran	**Yoga** Three sessions of Hatha yoga (60-70 min each) for 8-weeks. Includes postures, breathing techniques and meditation. The postures started with stretching techniques followed by standing, supine and prone-lying and sitting postures. Yoga teacher was familiar with problems common to people with MS. Each pose was held for 10-30 seconds (or 8 seconds for subjects who were unable to maintain some techniques) with resting periods between poses lasting 30 seconds to 1 min. Patients were supported for the majority of poses with a chair, Swiss ball or wall. Performed in physiotherapist clinic. vs.	Multiple sclerosis Fatigue at baseline: mean FSS score 3.9 for the whole population Threshold for fatigue used for inclusion (yes/no): no Age: mean 35.0 years for the whole population Type of MS: not reported EDSS: score between 1.0 and 4.0 was an inclusion criterion. Mean score 2.2 for the whole population	Included in previous guideline version. Additional comparison of exercise vs. control included above under the ‘exercise’ section

Study	Intervention and comparison	Population	Comments
	Treadmill training Supervised treadmill training exercises three times a week for 8 consecutive weeks. Each training session consisted of 30 min of treadmill exercise. The exercise class began and ended with 10 min stretching of muscles and flexion and rotation movements of the trunk and the lower limb. Training intensity was 40-75% age predicted maximal heart rate.		
	Yoga As described above. vs. **Control** Waitlist control group.		
Garrett 2013[39] Associated studies: Garrett 2013[40] N=148 randomised across the two interventions Conducted in Ireland	**Yoga** Not pre-defined and differed depending on which yoga instructor gave the class vs. **Control** Asked not to change their exercise habits during the 10-week treatment period	Multiple sclerosis Fatigue at baseline: mean score on MFIS 36-40 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 49-50 across the two groups Type of MS: majority relapsing remitting (>50%) EDSS: not reported	Additional comparisons from this study involved resistance training + aerobic exercise and are included under a separate section.
	Yoga	Multiple sclerosis	

Study	Intervention and comparison	Population	Comments
Hasanpour Dehkordi 2016[45] N=90 randomised across the three groups Conducted in Iran	Three sessions weekly for 12 weeks. Hatha yoga classes 60-70 min in duration. Included postures, breathing techniques and meditation. Postures started with stretching techniques followed by, standing, supine and prone-lying and sitting procedures. Each pose held for 10-30 seconds with rest periods in between of 30 seconds to 1 min. Each session ended with 10 min deep relaxation. Practice at home advised and given a booklet explaining the poses. vs. **Aerobic exercise** Three sessions weekly for 12 weeks. Each session lasted 40 min, with 5-10 min warm-up, 25-30 min exercise (walking) and 5 min cooling down. Exercise aimed to reach 60% of heart rate reserve. After 6 sessions duration of exercise increased to 30-35 min at a heart rate of 70% heart rate reserve.	Fatigue at baseline: mean 3.4-4.9 across the three groups on 'Rhoten Fatigue Test' Threshold for fatigue used for inclusion (yes/no): no Age: mean 31.9 years for the whole population Type of MS: not reported EDSS: score not reported	New study published since previous guideline version Additional comparisons of yoga vs. aerobic exercise and control groups included under a separate section.
	Yoga As described above. vs. **Control** Educational support with no exercise protocol. Asked to continue medications and usual lifestyle.		
Oken 2004[80] N=69 randomised across the three groups	**Yoga** Once weekly classes (90 min). Lyengar yoga classes. Modified to take into account fatigue and cerebellar dysfunction. All 19 included poses were supported (e.g., using chair or performing on floor or against the wall). Each pose was held for 10-30 seconds with rest	Multiple sclerosis Fatigue at baseline: mean score on Multidimensional Fatigue Inventory –	Excluded previously but on review deemed relevant.

Study	Intervention and comparison	Population	Comments
Conducted in USA	periods between poses lasting 30-60 seconds. Emphasis on breathing for concentration and relaxation. Each class ended with a 10-minute-deep relaxation. Daily home practice was strongly encouraged. vs. **Aerobic exercise** One session per week along with home exercise. Cycling on recumbent or dual-action stationary bicycles. The weekly exercise class began and ended with 5 min stretching of cycling muscles. Intensity was very light to moderate. Sometimes option of using Swiss ball and arm, trunk and balance work, though cycling main form of exercise. Encouraged to exercise regularly at home in addition to in-person sessions.	Physical subscale was 13.0-14.0 across the three groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 48.0-50.0 years across the three groups Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Mean score 2.9-3.2 across the three groups.	Additional comparison of exercise vs. control group included under a separate section.
	Yoga As described above. vs. **Control** Waitlist control group.		
Razazian 2016[91] N=54 randomised across the three groups Conducted in Iran	**Yoga** Three times weekly Hatha yoga sessions (60 min per session) for 8 weeks under supervision of a certified yoga instructor. Yoga sequences for beginners were taught. vs. **Exercise – aerobic + resistance**	Multiple sclerosis Fatigue at baseline: mean score on FSS 38.9-48.7 across the three groups Threshold for fatigue used for inclusion (yes/no): no	Additional comparison of exercise vs. control group is included under a separate section. New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	Three weekly sessions of aquatic exercise for 8 weeks, including a series of water activities (1 h per session). Included warming up, 10 min walking, stretching and gymnastics, 40 min power endurance activities, strength training and 10 min cooling down, relaxing, stretching and breathing exercises. **Yoga** As described above. vs. **Control** Non-exercise control group. Met 2-3 times weekly for 60-90 min. Able to talk to physicians and hospital staff, to complete everyday duties, to participate in occupational therapy and to meet and to talk to other patients.	Age: mean age 33.0-35.0 years across the three groups Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Mean score 3.3-3.9 across the three groups.	
Velikonja 2010[111] N=20 randomised Conducted in Slovenia	**Yoga** Once weekly sessions for 10 weeks. Hatha yoga supervised by a yoga instructor. vs. **Sports climbing** Once weekly sessions for 10 weeks. Supervised by qualified sports climbing instructors.	Multiple sclerosis Fatigue at baseline: median score 32.0-40.0 on MFIS total across the two groups. Threshold for fatigue used for inclusion (yes/no): no Age: median 41.0-41.0 years across the two groups Type of MS: relapsing-remitting, primary progressive or secondary progressive MS. Proportion with each not reported.	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
		EDSS: score ≤6.0 was an inclusion criterion. Median score 4.0-4.2 across the two groups.	
Pilates			
Bulguroglu 2017[19] N=45 randomised Conducted in Turkey	**Pilates** 60-90 min sessions two times weekly for 8 weeks. Movements controlled by physiotherapist. 10 repetitions for the first 2 weeks followed by 20 repetitions after 2 weeks. Therabands used in mat Pilates to increase difficult level for certain exercises, while resistance of springs adjusted in Reformer Pilates. vs. **Control** Home programme consisting of relaxation and respiration exercises for 8 weeks, with twice weekly sessions.	Multiple sclerosis Fatigue at baseline: median score 44.0-49.0 on FSS across the three original groups Threshold for fatigue used for inclusion (yes/no): no Age: median 37.0-45.0 across the three original groups Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Median score 1.0-2.0 across the three original groups.	New study published since previous guideline version. Two separate groups of reformer and mat Pilates combined into a single group to compare with the control group.
Eftekhari 2018[31] N=30 randomised Conducted in Iran	**Mat Pilates** 8 weeks of mat Pilates (three weekly sessions, with 48 h rest between each session). Exercises based on core stability of low-moderate intensity. Main exercise in each session was 30-40 min in duration. vs. **Control** Continued routine life.	Multiple sclerosis Fatigue at baseline: mean MFIS score 8.5-10.0 across the two groups (unclear which subscale) Threshold for fatigue used for inclusion (yes/no): no Age: mean 31.0-35.0 years across the two groups	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
		Type of MS: all had relapsing-remitting MS EDSS: score 2.0-6.0 was an inclusion criterion. Mean/median score not reported.	
Fleming 2019[36] N=18 randomised Conducted in Ireland	**Pilates** Home-based or supervised. Two sessions (1 h each) per week for 8 weeks, with 48 h between sessions. Mat-based beginner level exercises. Repetitions gradually progressed at 2-week intervals. vs. **Control** Waitlist control group. Asked to maintain pre-trial activity levels.	Multiple sclerosis Fatigue at baseline: mean MFIS total score 41.0 for the whole population Threshold for fatigue used for inclusion (yes/no): no Age: mean 50.2 years for the whole population Type of MS: not reported EDSS: not reported – PDDS score <3.0 was an inclusion criterion.	New study published since previous guideline version. Two separate groups of home-based and supervised Pilates combined and compared to the control group for the purpose of this review.
Fleming 2021[37] N=80 randomised Conducted in Ireland	**Pilates** Home-based Pilates. Twice weekly sessions 48 h apart for 8 weeks at home. Supported by DVD that had been developed and evaluated in a feasibility trial. Supported by weekly phone calls about sessions and any adverse events or relapses. vs. **Control**	Multiple sclerosis Fatigue at baseline: 69.2% and 68.3% in intervention and control groups, respectively, said to be fatigued at baseline (MFIS total score >38). Mean scores at baseline were 43.6 in both groups. Threshold for fatigue used for inclusion (yes/no): no	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	Waitlist control group. Asked to maintain pre-trial activity levels.	Age: mean 46.7 and 47.4 years in intervention and control groups, respectively Type of MS: not reported EDSS: not reported – PDDS score <3.0 was an inclusion criterion.	
Kucuk 2016[66] N=20 randomised Conducted in Turkey	**Pilates** Two days per week training for 8 weeks. Taught elements of Pilates exercises prior to starting them. Exercises checked and corrections made where applicable by physical therapist. Exercises repeated 8-10 times. Difficulty increased when participants could complete them correctly. Group sessions, with sessions 45-60 min in duration. vs. **Control – traditional exercise programme** Two days per week training for 8 weeks. Strength, balance and coordination training exercises.	Multiple sclerosis Fatigue at baseline: mean score on MFIS Physical subscale was 10.0-12.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 47.0-50.0 years across the two groups Type of MS: not reported EDSS: score ≤6.0 was an inclusion criterion. Mean score 2.8-3.2 across the two groups.	New study published since previous guideline version.
Relaxation – including relaxation, reflexology, massage and acupressure			
Arab 2019[8] N=80 randomised Conducted in Iran	**Massage** Three techniques used for massage therapy (four techniques for feet massage, three techniques for back, two techniques for neck and four techniques for hand). Family member taking responsibility for delivering the home massage were completely trained	Multiple sclerosis Fatigue at baseline: mean score on FSS was 48.3 and 47.7 in intervention and control groups, respectively	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	by physiotherapist at a one-hour session. Each patient in the intervention group received the massage therapy programme three days per week for 4 weeks and 20 min per session. The massage time was planned with consent of the patient before bedtime. The minimum number of massage therapy sessions to enter the information in the data analysis stage included 10 sessions. Moreover, an SMS was sent to patients and a weekly massage table was provided to them as a reminder of planned sessions. vs. **Control** Routine medical care only for 4 weeks	Threshold for fatigue used for inclusion (yes/no): yes (FSS at least 36) Age: mean 33.88 and 32.88 years in intervention and control groups, respectively Type of MS: not reported EDSS: not reported	
Atashi 2014[9] N=62 randomised Conducted in Iran	**Slow stroke back massage** Massage for seven 10-min sessions delivered by the researcher and a co-researcher. Unclear whether sessions were delivered weekly or twice weekly for example. Massage therapy was administrated by the researcher with the patient sat on massage chair with his/her head on a pillow. Small circular massage was conducted on patients' neck by researcher's thumb. Slow stroke back massage was administrated from neck area to sacrum by the researcher's palm and repetition of the action by her other palm on the other side of spine in a reverse direction simultaneously (toward neck). It also included slow stroke with thumb in both sides of spine from shoulder to waist and sweep stroke from neck nearly down to sacrum by two palms vs.	Multiple sclerosis Fatigue at baseline: mean score on FSS was 48.31 and 48.86 in the intervention and control groups, respectively Threshold for fatigue used for inclusion (yes/no): no Age: not reported Type of MS: not reported	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	Control Not defined, assume no intervention		
Bastani 2015[12] N=100 randomised Conducted in Iran	**Acupressure** Pressure on acupoints performed for 3 min on each of the points and repeated for opposite side of the body – total time was 18 min daily. Taught to do intervention in the first session and then performed themselves twice daily for two weeks at home. Also given booklets explaining the procedure. vs. **Control** Touching only at the same points as in the acupressure group. Taught to do intervention in the first session and then performed themselves twice daily for two weeks at home. Also given booklets explaining the procedure.	Multiple sclerosis Fatigue at baseline: mean score on FSS 83.0-89.0 Threshold for fatigue used for inclusion (yes/no): yes – score ≥5.0 on FSS Age: mean 32.0 years in both groups Type of MS: not reported EDSS: not reported.	New study published since previous guideline version.
Dilek Dogan 2021[29] N=66 randomised Conducted in Turkey	**Reflexology** 12-week reflexology intervention. Applied in ergonomic and adjustable therapy chair in a neurology clinic. Performed by considering sympathetic and parasympathetic nervous systems with more intense focus on certain points in line with expert opinion. Three sessions weekly using pure olive oil. Process involved warm up movements for 1 min using rotation, stretching of Achilles tendon, wrist release, running the toe on the soles of the feet and laundry ringing methods. Warm up methods completed by applying pressure to solar plexus. Brain area then massaged for 4 min. Epiphyseal, hypothalamus	Multiple sclerosis Fatigue at baseline: mean score on FSS 4.33 and 4.91 (1-7 scale) Threshold for fatigue used for inclusion (yes/no): no Age: mean 36.4 and 39.5 years in intervention and control groups, respectively	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	and pituitary gland points in the toes massaged. Reflexology also applied to spinal region, lymphatic system, shoulder, elbow, hip and knee regions, intestinal regions, reproductive organs, bladder region, mouth and jaw muscles. Foot loosening movements performed also. Session completed in 15-20 min by applying pressure to solar plexus. Repeated for each foot. Also received routine treatment. vs. **Control** No intervention was performed for the 12-week trial period and patients continued their routine clinical treatment	Type of MS: majority with relapsing-remitting (80% and 76.7% in two groups) EDSS: score ≤5.5 was an inclusion criterion. Mean scores 2.33 and 2.25 in intervention and control groups, respectively.	
Nazari 2015[77] N=75 randomised Conducted in Iran	**Reflexology or relaxation** Twice weekly sessions for 4 weeks (40 min per session). Performed in bright, silent, warm room. Combination of Jacobson and Benson methods for those receiving relaxation. For those that had reflexology, general reflex therapy was performed by massaging all plantar reflexology points followed by special reflex therapy. Major reflex points in feet under pressure using thumb and index finger. Ended with massage of solar plexus. vs. **Control** Routine treatment and care recommended by attending physician.	Multiple sclerosis Fatigue at baseline: mean score on FSS 4.9-5.0 across the three original groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥4.0 on FSS Age: mean 34.0 years for all three original groups Type of MS: not reported	New study published since previous guideline version. Combined reflexology and relaxation groups into a single group compared with the control group for purpose of this review.

Study	Intervention and comparison	Population	Comments
		EDSS: score between 0.0 and 5.5 was an inclusion criterion.	
Negahban 2013[79] N=24 randomised across the two groups Conducted in Iran	**Massage alone** Three 30 min supervised intervention sessions per week for 5 weeks Swedish massage by trained massage therapist. vs. **Control** Continue standard medical care and asked to avoid participation in any new exercise programme or change usual activities for 5 weeks	Multiple sclerosis Fatigue at baseline: mean score on FSS 41.3-42.3 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 36.7-36.8 years across the two groups Type of MS: relapsing-remitting MS or secondary progressive MS was an inclusion criterion. Proportion with each not reported. EDSS: mean 3.8 for each of the two groups	Included in previous guideline version. Additional comparisons from this study involved aerobic, strengthening and balance exercises, and a combination of massage + these exercises, and are included under a separate section.
Rahimi 2020[89] N=106 randomised Conducted in Iran	**Self-acupressure** Three training sessions of 30-40 min for participants. Psychological and physical complications of MS discussed, and intervention explained. Participants taught location of acupoints and method and amount of pressure on the acupoints explained, with pressure to be applied using pulp of the thumb. Asked to press each acupoint for 30 seconds and gradually increase pressure to feel warmth and tingling in target areas. Then asked to hold the weight for 4 minutes and release hand pressure for 30 seconds. Each acupoint pressed individually and then this was repeated on another acupoint. Intervention to be conducted at	Multiple sclerosis Fatigue at baseline: mean score on FSS 4.26 and 4.06 (scale 1-7) across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean not reported, most between 26 and 45 years	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	home every day between 9.00 and 10.00 am for 15 min (5 min per acupoint). In the third session a CD containing acupressure video was presented to participants. Intervention lasted for 1 month, during which researchers reminded participants to perform between 9 and 10 am by auto SMS reminder vs. **Sham** Taught to use the pulp of the thumb to press 2.5 cm below Shenmen point (to the forearm) and 3 cm above the Yin Tang acupoint. Length and frequency of the intervention was the same as the self-acupressure group. 1 month duration.	Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score ≤5.5 was an inclusion criterion.	
Sajadi 2020[98] N=70 randomised Conducted in Iran	**Reflexology** Rwo Shur method. Twice weekly sessions (30-40 min) for 4 weeks. Individual sessions in a private room. General foot massage followed by specialised massage to pituitary gland, hypothalamus, pineal gland and solar plexus reflex points. vs. **Control** Twice weekly sessions of non-specialised foot massage for 4 weeks. Sham massage without applying pressure on any particular reflex points.	Multiple sclerosis Fatigue at baseline: mean score on Fatigue Impact Scale 75.0-77.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: 35% 20-29 years, 35% 30-39 years and 30% 40-49 years for the whole population Type of MS: relapsing-remitting MS was an inclusion criterion EDSS: score ≤4.0 was an inclusion criterion.	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
Sgoifo 2017[101] N=48 randomised Conducted in Italy	**Relaxation** Integrated Imaginative Distention Therapy. Once weekly training group sessions (60 min) for 8 weeks. Includes Jacobsen relaxation exercises with breath awareness, motor imaging, body imaginative scan and imaginative experience. Following practice participants could take part in group discussion. Encouraged to repeat process at home. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 39.0-40.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: not reported – 18-75 years was an inclusion criterion Type of MS: majority relapsing-remitting (>85% of whole population) EDSS: mean 3.3 for the whole population	New study published since previous guideline version.
Cognitive behavioural therapy and motivational interviewing			
Bombardier 2008[16] N=130 randomised Conducted in USA	**Motivational interviewing** Initial 60-90 min motivational interview and goal setting. Followed by 5 follow-up telephone counselling sessions (30 min each) over 12 weeks. vs. **Control** Waitlist control group.	Multiple sclerosis Fatigue at baseline: mean score on MFIS total 32.0-40.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 45.0-48.0 years across the two groups Type of MS: majority relapsing-remitting (≥70% in both groups)	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
		EDSS: score ≤5.5 was an inclusion criterion. Mean/median score not reported.	
Borji 2018[17] N=60 randomised Conducted in Iran	**Motivational interviewing** According to Miller and Rollnick model. Group-based programme. 45-60 min per sessions, with each participant receiving 5 sessions over five weeks (1 session per week). vs. **Control** Not defined. Presumably received no intervention and continued usual lifestyle.	Multiple sclerosis Fatigue at baseline: mean score on Fatigue Impact Scale 63.0-66.0 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 33.0-35.0 years across the two groups Type of MS: not reported EDSS: not reported	New study published since previous guideline version.
Khayeri 2016[61] N=140 randomised Conducted in Iran	**Cognitive behavioural therapy** Fordyce Happiness Model. Twice weekly, with 8 sessions overall (60-90 min per session). Involved lectures, group discussions, question and answers. Asked to go through certain drills outside of the research environment. Consisted of various elements including defining depression, stress and anxiety, defining happiness, reviewing results of studies on happiness, increasing physical activity, being productive and doing useful and meaningful things, planning better and social relationships. vs. **Control**	Multiple sclerosis Fatigue at baseline: mean score on Piper scale was 6.3-6.6 across the two groups Threshold for fatigue used for inclusion (yes/no): no Age: mean 49.0-50.0 years across the two groups Type of MS: not reported EDSS: not reported	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	Not defined – presumably no intervention and continued usual lifestyle.		
Moss-Morris 2012[73] N=40 randomised Conducted in UK	**Cognitive behavioural-based self-management programme** MS Invigor8: Breaking the Cycle of Fatigue. Once weekly sessions for 8 weeks (25-50 min per session). Interactive and could be tailored to individual. Includes homework tasks. Automated emails encouraging completion of the online sessions. Three telephone support sessions in addition between 30-50 min, included goal setting and measuring progress, and challenging unhelpful thoughts. Performed by assistant psychologist. vs. **Control** Waitlist control group. Standard care received.	Multiple sclerosis Fatigue at baseline: mean score on MFIS 13.2-12.7 (unclear whether total or subscale) across the two groups Threshold for fatigue used for inclusion (yes/no): yes – score >4.0 on a fatigue scale (scale used unclear) Age: mean 40.0-42.0 years across the two groups Type of MS: majority relapsing-remitting (43.5 and 70.6% in the two groups) or secondary progressive (30.4% and 11.8% in the two groups) EDSS: not reported	Included in previous guideline version.
Pottgen 2018[88] N=275 randomised Conducted in Germany	**Cognitive behavioural intervention** 12-week internet-based intervention (ELEVIDA). Based on cognitive behavioural therapy strategies primarily through technique of 'simulated dialogue'. Includes introduction, summary and homework tasks. Advised to access the programme 1-2 times weekly. Offers tailored information based on individual needs following responses to statements in multiple choice format.	Multiple sclerosis Fatigue at baseline: mean score on Fatigue Scale of Motor and Cognition total score 76.0-77.0 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥43.0	New study published since previous guideline version.

Study	Intervention and comparison	Population	Comments
	vs. **Control** Waitlist control group. Standard care.	on Fatigue Scale of Motor and Cognition total score Age: mean 41.0-42.0 years across the two groups Type of MS: majority relapsing-remitting (>70% in both groups) EDSS: not reported – majority had at least mild impairment on PDDS	
van den Akker 2017[108] N=91 randomised Conducted in The Netherlands	**Cognitive behavioural therapy** 12 sessions of cognitive behavioural therapy delivered face-to-face over a 4-month period. Consists of 10 modules covering topics such as setting goals, changing beliefs, reducing focus on fatigue and the role of the environment. Patient-tailored based on baseline questionnaires. vs. **Control** Three MS nurse consultations (45 min) over the 4-month period. Delivered information about MS fatigue but did not allow advice to be given about treatment or referral to a psychologist or other healthcare professionals for fatigue treatment.	Multiple sclerosis Fatigue at baseline: mean score on FSS 5.4-5.5 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – severe fatigue (CIS20r fatigue subscale score ≥35.0) Age: mean 46.0-51.0 years across the two groups Type of MS: majority relapsing-remitting (>70% in both groups) EDSS: median 2.5-3.0 across the two groups	New study published since previous guideline version. TREFAMS-CBT study
van Kessel, 2008[110] N=72 randomised	**Cognitive behavioural therapy** Once weekly session for 8 weeks (up to 50 min each). Three sessions were face-to-face and other five by telephone. Included workbook with homework tasks.	Multiple sclerosis	Included in previous guideline version.

Study	Intervention and comparison	Population	Comments
Conducted in New Zealand	Developed with fatigue in mind. Challenge behavioural, cognitive, emotional and external factors contributing to MS fatigue. Individually tailored. vs. **Control – relaxation** Once weekly session for 8 weeks (up to 50 min each). Three sessions were face-to-face and other five by telephone. Included workbook with homework tasks. Taught series of relaxation techniques during 8 sessions, including diaphragmatic breathing, progressive muscle relaxation, visualisation, cue-controlled relaxation and rapid relaxation. No advice given about scheduling, rest, managing sleep or cognitive strategies.	Fatigue at baseline: mean score on Chalder fatigue scale 20.0-21.0 across the two groups Threshold for fatigue used for inclusion (yes/no): yes – score ≥4.0 on Chalder fatigue scale Age: mean 43.0-47.0 years across the two groups Type of MS: majority relapsing-remitting (66% and 49% in the two groups) or secondary progressive (31% and 30% in the two groups) EDSS: score ≤6.0 was an inclusion criterion. Mean score 3.0-3.9 across the two groups.	
Motivational interviewing + exercise			
Flachenecker 2020[35] N=84 randomised Conducted in Germany	**Internet-based physical activity promotion** Based on physical activity-related health competence and the self-determination theory and integrated various behavioural change techniques and motivational interviewing. Programme involved web- and telephone-based, behaviour-oriented physical activity coaching with one individual and four group sessions, and an individual exercise prescription in a one-to-one approach using a browser-based software. Participants used the software to document their exercises and to plan their activities and sessions in a physical activity diary. Exercise therapists used patient feedback and exercise parameters to supervise and manage exercises and activities.	Multiple sclerosis Fatigue at baseline: median score WEIMuS fatigue score 39.0-45.0 across the two groups. Threshold for fatigue used for inclusion (yes/no): yes – WEIMuS score ≥32 Age: mean 46.0-48.0 years across the two groups	New study published since previous guideline version

Study	Intervention and comparison	Population	Comments
	Individual exercise prescription was based on general recommendations for strength training and endurance training. The recommendation for exercise intensity was light to moderate. Training performed for 3 months. vs. **Control** Usual care – did not receive any study intervention and were told not to change any of their habits, including physical activity.	Type of MS: majority with relapsing-remitting MS (>50% in both groups) EDSS: score ≤6.0 was an inclusion criterion. Median score 4.0-4.3 across the two groups.	

1 See appendices for full evidence tables.

2

3
4

1.1.6 Summary of the effectiveness evidence

Results for each comparison are given below in the form of GRADE summary tables. See appendices for full GRADE tables.

Aerobic exercise vs. control

Table 3: Clinical evidence summary: Aerobic exercise vs. control – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
Fatigue Severity Scale (1-7) Scale from: 1 to 7 follow up: range 4 weeks to 26 weeks	129 (4 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (1-7) was 4.59	MD 0.71 lower (1.87 lower to 0.45 higher)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: range 7 weeks to 12 weeks	183 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue Severity Scale (9-63) was 40.82	MD 7.59 lower (17.64 lower to 2.47 higher)
Modified Fatigue Impact Scale - total (0-84) Scale from: 0 to 84 follow up: range 8 weeks to 26 weeks	125 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean modified Fatigue Impact Scale - total (0-84) was 37.63	MD 3.21 lower (12.34 lower to 5.92 higher)
Modified Fatigue Impact Scale - physical (0-36) Scale from: 0 to 36 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean modified Fatigue Impact Scale - physical (0-36) was 14.5	MD 4.8 lower (9.69 lower to 0.09 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
Modified Fatigue Impact Scale - cognitive (0-40) Scale from: 0 to 40 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean modified Fatigue Impact Scale - cognitive (0-40) was 14.0	MD 4.3 lower (9.38 lower to 0.78 higher)
Modified Fatigue Impact Scale - psychosocial (0-8) Scale from: 0 to 8 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean modified Fatigue Impact Scale - psychosocial (0-8) was 1.8	MD 0.1 lower (1.3 lower to 1.1 higher)
Fatigue subscale of Checklist Individual Strength-20 (8-56) Scale from: 8 to 56 follow up: 26 weeks	71 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean fatigue subscale of Checklist Individual Strength-20 (8-56) was 40.6	MD 0.4 lower (4.82 lower to 4.02 higher)
Fatigue Scale for Motor and Cognitive Challenge (FSMC) - physical (10-50) Scale from: 10 to 50 follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean fatigue Scale for Motor and Cognitive Challenge (FSMC) - physical (10-50) was 29.6	MD 3.4 lower (9 lower to 2.2 higher)
Fatigue Scale for Motor and Cognitive Challenge (FSMC) - cognitive (10-50) Scale from: 10 to 50 follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean fatigue Scale for Motor and Cognitive Challenge (FSMC) - cognitive (10-50) was 28.9	MD 0.9 lower (7.81 lower to 6.01 higher)
Rhoten Fatigue Scale (0-10) Scale from: 0 to 10 follow up: 12 weeks	41 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean rhoten Fatigue Scale (0-10) was 3.55	MD 1 lower (1.67 lower to 0.33 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
Fatigue Impact Scale (0-160) Scale from: 0 to 160 follow up: 24 weeks	138 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean fatigue Impact Scale (0-160) was 62.63	MD 8.21 lower (19.44 lower to 3.02 higher)
Multidimensional Fatigue Inventory - general fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - general fatigue (4-20) was 14.9	MD 2.8 lower (4.73 lower to 0.87 lower)
Multidimensional Fatigue Inventory - physical fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - physical fatigue (4-20) was 13.9	MD 3.1 lower (5.93 lower to 0.27 lower)
Multidimensional Fatigue Inventory - reduced activity (4-20) Scale from: 4 to 20 follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - reduced activity (4-20) was 11.5	MD 1.6 lower (4.39 lower to 1.19 higher)
Multidimensional Fatigue Inventory - reduced motivation (4-20) Scale from: 4 to 20 follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - reduced motivation (4-20) was 9.8	MD 2.1 lower (4.27 lower to 0.07 higher)
Multidimensional Fatigue Inventory - mental fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - mental fatigue (4-20) was 11.2	MD 3.4 lower (6.21 lower to 0.59 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
MSQOL-54 physical composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean MSQOL-54 physical composite (0-100) was 66.64	MD 5.15 higher (4.71 lower to 15.01 higher)
MSQOL-54 mental composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean MSQOL-54 mental composite (0-100) was 66.54	MD 1.92 lower (15.07 lower to 11.23 higher)
MSQOL-54 change in health domain (0-100) Scale from: 0 to 100 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean MSQOL-54 change in health domain (0-100) was 52.5	MD 0 (24.11 lower to 24.11 higher)
MSIS-29 - physical (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 24 weeks	180 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean MSIS-29 - physical (0-100) was 34.19	MD 5.75 lower (11.5 lower to 0.01 lower)
MSIS-29 - psychological (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 24 weeks	180 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 - psychological (0-100) was 32.8	MD 3.36 lower (9.18 lower to 2.47 higher)
SF-36 physical functioning (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	76 (2 RCTs)	⊕◯◯◯ VERY LOW a, c,e	-	The mean SF-36 physical functioning (0-100) was 47.87	MD 10.89 higher (0.53 higher to 21.25 higher)
SF-36 emotional limitations (0-100) Scale from: 0 to 100	76 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-36 emotional limitations (0-100) was 59.37	MD 0.85 higher (25.92 lower to 27.62 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
follow up: range 12 weeks to 6 months					
SF-36 physical role limitations (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	76 (2 RCTs)	⊕◯◯◯ VERY LOW a ,c,e	-	The mean SF-36 physical role limitations (0-100) was 52.46	MD 4.91 lower (12.54 lower to 2.72 higher)
SF-36 energy/vitality (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	76 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 energy/vitality (0-100) was 40.09	MD 12.76 higher (7.21 higher to 18.32 higher)
SF-36 mental health (0-100) Scale from: 0 to 100 follow up: 12 weeks	41 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 mental health (0-100) was 50.44	MD 11.34 higher (3.54 higher to 19.14 higher)
SF-36 social functioning (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	76 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 social functioning (0-100) was 55.38	MD 6.95 higher (1.94 higher to 11.96 higher)
SF-36 body pain (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	76 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-36 body pain (0-100) was 62.14	MD 8.24 lower (25.69 lower to 9.21 higher)
SF-36 general health (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	76 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 general health (0-100) was 48.87	MD 10.85 higher (5.45 higher to 16.25 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
SF-36 health transition (0-100) Scale from: 0 to 100 follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 health transition (0-100) was 48.6	MD 11.9 lower (28.63 lower to 4.83 higher)
EDSS scale (0-10) Scale from: 0 to 10 follow up: 8 weeks	47 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean EDSS scale (0-10) was 1.98	MD 0.29 higher (0.67 lower to 1.25 higher)
Guy's neurological disability scale (0-60) Scale from: 0 to 60 follow up: 7 weeks	16 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean guy's neurological disability scale (0-60) was 0.13	MD 0.62 higher (1.24 lower to 2.48 higher)
HAQUAMS - fatigue/thinking (1-5) Scale from: 1 to 5 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean HAQUAMS - fatigue/thinking (1-5) was 2.7	MD 0.8 lower (1.51 lower to 0.09 lower)
HAQUAMS - total (1-5) Scale from: 1 to 5 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean HAQUAMS - total (1-5) was 2.0	MD 0.4 lower (0.71 lower to 0.09 lower)
HAQUAMS - mood (1-5) Scale from: 1 to 5 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean HAQUAMS - mood (1-5) was 2.1	MD 0.4 lower (0.86 lower to 0.06 higher)
HAQUAMS - social function (1-5) Scale from: 1 to 5 follow up: 8 weeks	28 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean HAQUAMS - social function (1-5) was 1.9	MD 0.1 lower (0.58 lower to 0.38 higher)
Cognitive - Digit Symbol Substitution Test follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - Digit Symbol Substitution Test was 85.5	MD 8.8 higher (0.23 higher to 17.37 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
Cognitive - Word List Generation follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - Word List Generation was 31.4	MD 1.1 higher (3.5 lower to 5.7 higher)
Cognitive - Selective reminding test (long-term storage) follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - Selective reminding test (long-term storage) was 50.8	MD 3.6 lower (9.23 lower to 2.03 higher)
Cognitive - Selective reminding test (consistent long-term retrieval) follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean cognitive - Selective reminding test (consistent long-term retrieval) was 62.0	MD 8.8 lower (14.64 lower to 2.96 lower)
Cognitive - Spatial Recall Test follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean cognitive - Spatial Recall Test was 44.4	MD 3.6 higher (0.09 lower to 7.29 higher)
Cognitive - Paced Auditory Serial Attention Test (PASAT) follow up: 12 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean cognitive - Paced Auditory Serial Attention Test (PASAT) was 48.6	MD 2.1 higher (2.6 lower to 6.8 higher)
Cognitive - checklist individual strength concentration (5-35) Scale from: 5 to 35 follow up: 26 weeks	71 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - checklist individual strength concentration (5-35) was 18.8	MD 0.9 higher (2.43 lower to 4.23 higher)
Cognitive - Stroop Colour Word Interference (attention/concentration) follow up: 6 months	35 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - Stroop Colour Word Interference (attention/concentration) was 8.1	MD 1.8 higher (1.88 lower to 5.48 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: range 8 weeks to 10 weeks	46 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean beck Depression Inventory (0-63) was 14.82	MD 5.65 lower (9.9 lower to 1.39 lower)
Beck Depression Inventory - fast screen (0-21) Scale from: 0 to 21 follow up: 8 weeks	47 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean beck Depression Inventory - fast screen (0-21) was 6.52	MD 1.4 lower (4.16 lower to 1.36 higher)
Beck Anxiety Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean beck Anxiety Inventory (0-63) was 8.2	MD 2.1 lower (7.61 lower to 3.41 higher)
Incidence of adverse events - only MS exacerbations reported follow up: 24 weeks	138 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RR 0.71 (0.37 to 1.36)	Moderate	
				246 per 1,000	71 fewer per 1,000 (155 fewer to 89 more)
Incidence of adverse events - mixed follow up: range 6 weeks to 6 months	141 (5 RCTs)	⊕◯◯◯ VERY LOW a,f	RD 0.14 (0.04 to 0.24)	0 per 1,000	140 more per 1,000 (40 more to 240 more)
Incidence of adverse events - orthopaedic problems reported separately follow up: 24 weeks	138 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RR 0.67 (0.39 to 1.14)	Moderate	
				348 per 1,000	115 fewer per 1,000 (212 fewer to 49 more)
Incidence of adverse events - at least one	138 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RR 0.57 (0.31 to 1.07)	Moderate	
				304 per 1,000	131 fewer per 1,000 (210 fewer to 21 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - up to 6-month outcomes	Risk difference with Aerobic exercise
fall reported separately follow up: 24 weeks					
Adverse events leading to withdrawal follow up: 6 months	26 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 6.92 (0.41 to 118.14)	0 per 1,000	143 more per 1,000 (73 fewer to 359 more)
Acceptability - Completed all 1-1 phone calls	138 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 0.64 (0.30 to 1.37)	Moderate	
				768 per 1,000	89 fewer per 1,000 (270 fewer to 51 more)
Acceptability - Completed all teleconference calls with or without at least one makeup session	138 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 1.12 (0.44 to 2.84)	Moderate	
				841 per 1,000	15 more per 1,000 (142 fewer to 97 more)

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

b. Heterogeneity present that could not be explained by prespecified subgrouping strategies and I2 >75%

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Downgraded by 1 increment as the follow-up time was less than the minimum of 3 months specified in the protocol for the majority of the evidence

e. Downgraded by 1 increment as point estimates differ widely despite I2 being below 50%

f. Imprecision assessed using OIS due to zero events in both arms of at least one study. Downgraded by 1 increment if power 80-90% and 2 increments if power <80%.

Table 4: Clinical evidence summary: Aerobic exercise vs. control – outcomes >6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control, education only) - >6 months outcomes	Risk difference with Aerobic exercise
Fatigue Severity Scale (1-7) Scale from: 1 to 7 follow up: 52 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Scale (1-7) was 5.1	MD 0.1 higher (0.44 lower to 0.64 higher)
Modified Fatigue Impact Scale - total (0-84) Scale from: 0 to 84 follow up: 52 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - total (0-84) was 39.9	MD 0.9 lower (7.15 lower to 5.35 higher)
Fatigue subscale of Checklist Individual Strength-20 (8-56) Scale from: 8 to 56 follow up: 52 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue subscale of Checklist Individual Strength-20 (8-56) was 41.2	MD 0.5 higher (4.52 lower to 5.52 higher)
Cognitive - checklist individual strength concentration (5-35) Scale from: 5 to 35 follow up: 52 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cognitive - checklist individual strength concentration (5-35) was 19.5	MD 1.2 higher (2.4 lower to 4.8 higher)
Incidence of adverse events - MS relapse follow up: 52 weeks	65 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.28 (0.10 to 0.81)	Could not be calculated as no control group risk given [c]	

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Control group risk could not be calculated as number of events not reported - therefore absolute effect could not be calculated.

Aerobic exercise vs. neurological rehabilitation (respiratory, postural and stretching)

Table 5: Clinical evidence summary: Aerobic exercise vs. neurological rehabilitation – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with neurological rehabilitation (respiratory, postural and stretching)	Risk difference with Aerobic exercise
Average adherence rate	22 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean average adherence rate was 90%	MD 3 lower (8.91 lower to 2.91 higher)

Only other available evidence from this study was reported as median values making it difficult to analyse.

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Functional electrical stimulation + aerobic exercise vs. control (waitlist)

Table 6: Clinical evidence summary: Functional electrical stimulation + aerobic exercise vs. control (waitlist) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist)	Risk difference with Functional electrical stimulation + aerobic exercise
5-item MFIS score (0-20) Scale from: 0 to 20 follow up: 12 weeks	12 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean 5-item MFIS score (0-20) was 0.17	MD 2.57 lower (7.61 lower to 2.47 higher)
Decrease in fatigue on MFIS 5-item (any decrease) follow up: 12 weeks	12 (1 RCT)	⨁◯◯◯ VERY LOW a,b	OR 2.00 (0.19 to 20.61)	Moderate	
				500 per 1,000	167 more per 1,000 (340 fewer to 454 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist)	Risk difference with Functional electrical stimulation + aerobic exercise
Fatigue Scale of Motor and Cognitive Functions - Total score (20-100) Scale from: 20 to 100 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Scale of Motor and Cognitive Functions - Total score (20-100) was -2.17	MD 2.5 lower (10.09 lower to 5.09 higher)
Fatigue Scale of Motor and Cognitive Functions - Cognitive score (10-50) Scale from: 10 to 50 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Scale of Motor and Cognitive Functions - Cognitive score (10-50) was -1.5	MD 1 lower (4.84 lower to 2.84 higher)
Fatigue Scale of Motor and Cognitive Functions - Motor score (10-50) Scale from: 10 to 50 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Scale of Motor and Cognitive Functions - Motor score (10-50) was -0.67	MD 1.5 lower (6.95 lower to 3.95 higher)
Decrease in fatigue on FSMC total score (any decrease) follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 2.50 (0.16 to 38.60)	Moderate	
				667 per 1,000	167 more per 1,000 (424 fewer to 321 more)
MSQOL-54 (0-100 for all) - Mental health composite Scale from: 0 to 100 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSQOL-54 (0-100 for all) - Mental health composite was 1.05	MD 0.72 higher (12.95 lower to 14.39 higher)
MSQOL-54 (0-100 for all) - Physical health composite Scale from: 0 to 100 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSQOL-54 (0-100 for all) - Physical health composite was -2.18	MD 8.95 higher (2.1 higher to 15.8 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist)	Risk difference with Functional electrical stimulation + aerobic exercise
MSQOL-54 (0-100 for all) - Change in health domain Scale from: 0 to 100 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSQOL-54 (0-100 for all) - Change in health domain was 0.0	MD 4.17 lower (19.23 lower to 10.89 higher)
PHQ-9 (depression; 0-27) Scale from: 0 to 27 follow up: 12 weeks	12 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean PHQ-9 (depression; 0-27) was -2.5	MD 2.83 higher (1.96 lower to 7.62 higher)
Adverse events leading to withdrawal follow up: 12 weeks	18 (1 RCT)	⊕⊕◯◯ LOW a,b	RR 3.18 (0.46 to 21.85)	Moderate	
				143 per 1,000	312 more per 1,000 (77 fewer to 2,979 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Resistance training vs. control (waitlist control, no intervention, usual care or education only)

Table 7: Clinical evidence summary: Resistance training vs. control (waitlist control, no intervention, usual care or education only) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, no intevention, usual care or education only)	Risk difference with Resistance training
Modified Fatigue Impact Scale - total (0-84)	133 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean modified Fatigue Impact Scale - total (0-84) was 1.86	MD 4.85 lower (14.33 lower to 4.64 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, no intevention, usual care or education only)	Risk difference with Resistance training
Scale from: 0 to 84 follow up: range 4 weeks to 22 weeks					
Modified Fatigue Impact Scale - physical (0-36) Scale from: 0 to 36 follow up: range 4 weeks to 22 weeks	90 (2 RCTs)	⊕⊕⊕◯ MODERATE a	-	The mean modified Fatigue Impact Scale - physical (0-36) was 1.73	MD 0.81 lower (3.5 lower to 1.88 higher)
Modified Fatigue Impact Scale - cognitive (0-40) Scale from: 0 to 40 follow up: range 4 weeks to 22 weeks	90 (2 RCTs)	⊕⊕◯◯ LOW a,c	-	The mean modified Fatigue Impact Scale - cognitive (0-40) was 0.70	MD 1.3 higher (1.49 lower to 4.1 higher)
Modified Fatigue Impact Scale - psychosocial (0-8) Scale from: 0 to 8 follow up: range 4 weeks to 22 weeks	80 (2 RCTs)	⊕◯◯◯ VERY LOW a,d,e	-	The mean modified Fatigue Impact Scale - psychosocial (0-8) was 0.59	MD 0.32 lower (2.05 lower to 1.41 higher)
Fatigue Severity Scale (1-7) Scale from: 1 to 7 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean fatigue Severity Scale (1-7) was 5.1	MD 0.2 lower (1.2 lower to 0.8 higher)
Multidimensional Fatigue Inventory (4-20) - General fatigue Scale from: 4 to 20 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean multidimensional Fatigue Inventory (4-20) - General fatigue was 11.8	MD 0.9 higher (2.37 lower to 4.17 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, no intevention, usual care or education only)	Risk difference with Resistance training
Multidimensional Fatigue Inventory (4-20) - Physical fatigue Scale from: 4 to 20 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean multidimensional Fatigue Inventory (4-20) - Physical fatigue was 12.6	MD 1.6 lower (4.48 lower to 1.28 higher)
Multidimensional Fatigue Inventory (4-20) - Reduced activity Scale from: 4 to 20 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean multidimensional Fatigue Inventory (4-20) - Reduced activity was 10.9	MD 0.6 lower (3.54 lower to 2.34 higher)
Multidimensional Fatigue Inventory (4-20) - Reduced motivation Scale from: 4 to 20 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean multidimensional Fatigue Inventory (4-20) - Reduced motivation was 6.7	MD 0.5 lower (2.2 lower to 1.2 higher)
Multidimensional Fatigue Inventory (4-20) - Mental fatigue Scale from: 4 to 20 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean multidimensional Fatigue Inventory (4-20) - Mental fatigue was 10.6	MD 0 (3.79 lower to 3.79 higher)
SF-36 quality of life (0-100) - Physical summary Scale from: 0 to 100 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean SF-36 quality of life (0-100) - Physical summary was 41.5	MD 3.8 higher (0.85 lower to 8.45 higher)
SF-36 quality of life (0-100) - Mental summary Scale from: 0 to 100 follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean SF-36 quality of life (0-100) - Mental summary was 57.8	MD 2.4 lower (9.28 lower to 4.48 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, no intevention, usual care or education only)	Risk difference with Resistance training
SF-36 quality of life (0-100) - General health domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - General health domain was 41.1	MD 8.4 higher (8.96 lower to 25.76 higher)
SF-36 quality of life (0-100) - Physical functioning domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Physical functioning domain was 43.9	MD 5.4 lower (41.29 lower to 30.49 higher)
SF-36 quality of life (0-100) - Physical limitation domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Physical limitation domain was 44.4	MD 5.6 higher (28.3 lower to 39.5 higher)
SF-36 quality of life (0-100) - Emotional limitation domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Emotional limitation domain was 59.1	MD 27.6 higher (7.32 lower to 62.52 higher)
SF-36 quality of life (0-100) - Emotional wellbeing domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Emotional wellbeing domain was 64.0	MD 11.6 higher (4.01 lower to 27.21 higher)
SF-36 quality of life (0-100) - Pain domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Pain domain was 64.2	MD 12.1 higher (17.41 lower to 41.61 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, no intevention, usual care or education only)	Risk difference with Resistance training
SF-36 quality of life (0-100) - Energy/fatigue domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Energy/fatigue domain was 49.1	MD 11.4 higher (6.55 lower to 29.35 higher)
SF-36 quality of life (0-100) - Social functioning domain Scale from: 0 to 100 follow up: 4 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 quality of life (0-100) - Social functioning domain was 58.6	MD 14.9 higher (11.14 lower to 40.94 higher)
WHOQOL-BREF (0-100) - Overall score Scale from: 0 to 100 follow up: 22 weeks	71 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean WHOQOL-BREF (0-100) - Overall score was 0.1	MD 0 (0.51 lower to 0.51 higher)
WHOQOL-BREF (0-100) - Overall health change Scale from: 0 to 100 follow up: 22 weeks	71 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean WHOQOL-BREF (0-100) - Overall health change was 0.9	MD 0.6 lower (2.11 lower to 0.91 higher)
WHOQOL-BREF (0-100) - Overall physical health change Scale from: 0 to 100 follow up: 22 weeks	71 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean WHOQOL-BREF (0-100) - Overall physical health change was 0.1	MD 0.2 lower (0.65 lower to 0.25 higher)
Functional capacity (% - baseline set at 100%) follow up: 12 weeks	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean functional capacity (% - baseline set at 100%) was 108.9	MD 12.1 higher (4.35 higher to 19.85 higher)
Major Depression Inventory (scale	34 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean major Depression Inventory (scale unclear) was 8.9	MD 0.2 lower (4.5 lower to 4.1 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, no intevention, usual care or education only)	Risk difference with Resistance training
unclear) follow up: 12 weeks					
Incidence of adverse events (harm) follow up: 4 weeks	19 (1 RCT)	⨁◯◯◯ VERY LOW a, c,f	RD 0.00 (-0.18 to 0.18)	0 per 1,000	0 fewer per 1,000 (180 fewer to 180 more)
Adverse events leading to withdrawal follow up: 10 weeks	43 (1 RCT)	⨁◯◯◯ VERY LOW a,c	OR 7.90 (1.24 to 50.09)	0 per 1,000	217 more per 1,000 (37 more to 398 more)

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

b. Heterogeneity that cannot be explained by prespecified subgrouping strategies and I2 >75%

c. Downgraded by 1 increment as the follow-up duration for the majority of the evidence is less than the 3-month minimum specified in the protocol

d. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

e. Heterogeneity that cannot be explained by prespecified subgrouping strategies

f. Imprecision assessed based on sample size as zero events in both arms of a single study. Downgraded by 2 increments as sample size <70.

Vestibular/balance training vs. control (waitlist control, routine care, information only)

Table 8: Clinical evidence summary: Vestibular/balance training vs. control (waitlist control, routine care, information only) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, routine care, information only)	Risk difference with Vestibular/balance training
Modified Fatigue Impact Scale - total (0-84) Scale from: 0 to 84 follow up: range 10 weeks to 14 weeks	149 (3 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - total (0-84) was 32.46	MD 11.13 lower (15.43 lower to 6.84 lower)
Modified Fatigue Impact Scale - physical (0-36) Scale from: 0 to 36 follow up: 14 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact Scale - physical (0-36) was 20.7	MD 4.7 lower (7.89 lower to 1.51 lower)
Modified Fatigue Impact Scale - cognitive (0-40) Scale from: 0 to 40 follow up: 14 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact Scale - cognitive (0-40) was 19.3	MD 5.1 lower (8.43 lower to 1.77 lower)
Modified Fatigue Impact Scale - psychosocial (0-8) Scale from: 0 to 8 follow up: 14 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact Scale - psychosocial (0-8) was 3.61	MD 1.17 lower (2.02 lower to 0.32 lower)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 8 weeks	87 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was 42.55	MD 8.51 lower (14.75 lower to 2.27 lower)
Fatigue Impact Scale - total score (0-160) Scale from: 0 to 160 follow up: 12 weeks	72 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fatigue Impact Scale - total score (0-160) was 96.5	MD 25.7 lower (34.3 lower to 17.1 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, routine care, information only)	Risk difference with Vestibular/balance training
Fatigue Impact Scale - physical subscale (0-40) Scale from: 0 to 40 follow up: 12 weeks	72 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fatigue Impact Scale - physical subscale (0-40) was 28.8	MD 9.8 lower (12.92 lower to 6.68 lower)
Fatigue Impact Scale - cognitive subscale (0-40) Scale from: 0 to 40 follow up: 12 weeks	72 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fatigue Impact Scale - cognitive subscale (0-40) was 22.0	MD 4.9 lower (6.65 lower to 3.15 lower)
Fatigue Impact Scale - psychosocial subscale (0-80) Scale from: 0 to 80 follow up: 12 weeks	72 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fatigue Impact Scale - psychosocial subscale (0-80) was 45.8	MD 13.5 lower (18.87 lower to 8.13 lower)
SF-36 physical summary (0-100) Scale from: 0 to 100 follow up: 14 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 physical summary (0-100) was 37.3	MD 3.7 higher (0.18 lower to 7.58 higher)
SF-36 mental summary (0-100) Scale from: 0 to 100 follow up: 14 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 mental summary (0-100) was 44.6	MD 3.6 higher (0.22 higher to 6.98 higher)
MusiQoL (0-100) Scale from: 0 to 100 follow up: 8 weeks	42 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean musiQoL (0-100) was 63.08	MD 10 higher (2.02 higher to 17.98 higher)
EDSS (0-10) Scale from: 0 to 10 follow up: 8 weeks	45 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean EDSS (0-10) was 1.98	MD 1.12 higher (0.08 higher to 2.16 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist control, routine care, information only)	Risk difference with Vestibular/balance training
Cognitive - perceived deficits questionnaire (0-80) Scale from: 0 to 80 follow up: 14 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - perceived deficits questionnaire (0-80) was 35.3	MD 6.3 lower (12.54 lower to 0.06 lower)
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 10 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (0-63) was 16.6	MD 5 lower (13.7 lower to 3.7 higher)
Beck Depression Inventory - fast screen (0-21) Scale from: 0 to 21 follow up: 8 weeks	45 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory - fast screen (0-21) was 6.52	MD 1.23 lower (4.34 lower to 1.88 higher)
Adverse events follow up: range 6 weeks to 10 weeks	66 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,d	RD 0.00 (-0.09 to 0.09)	0 per 1,000	0 fewer per 1,000 (90 fewer to 0 more)
Adverse events leading to withdrawal follow up: range 10 weeks to 14 weeks	227 (3 RCTs)	⊕◯◯◯ VERY LOW a,e	RD 0.03 (-0.03 to 0.08)	27 per 1,000	30 more per 1,000 (30 fewer to 80 more)

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

b. Downgraded by 1 increment as the follow-up was less than the minimum of 3 months specified in the protocol for the majority of the evidence

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Imprecision assessed using sample size as zero events in both arms of all studies. Downgraded by 2 increments as sample size <70.

e. Imprecision assessed based on OIS as zero events in both arms of at least one study. Downgraded by 1 increment if power 80-90% and 2 increments if power <80%.

Vestibular/balance training vs. standard neurorehabilitation

Table 9: Clinical evidence summary: Vestibular/balance training vs. standard neurorehabilitation – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard neurorehabilitation	Risk difference with Vestibular/balance training
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 4 weeks	23 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was 47.1	MD 2.1 higher (6.35 lower to 10.55 higher)
Functional - Barthel Index (0-100) Scale from: 0 to 100 follow up: 4 weeks	23 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean functional - Barthel Index (0-100) was 81.3	MD 3.2 higher (6.41 lower to 12.81 higher)

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

b. Downgraded by 1 increment as the majority of the evidence was at a follow-up less than the 3 months minimum specified in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Resistance training vs. aerobic exercise

Table 10: Clinical evidence summary: Resistance training vs. aerobic exercise – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Resistance training
Modified Fatigue Impact Scale - physical (0-36)	32 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact Scale - physical (0-36) was -2.7	MD 1.1 higher (1.96 lower to 4.16 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Resistance training
Scale from: 0 to 36 follow up: 8 weeks					
Modified Fatigue Impact Scale - cognitive (0-40) Scale from: 0 to 40 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact Scale - cognitive (0-40) was -2.3	MD 1 lower (5.82 lower to 3.82 higher)
Modified Fatigue Impact Scale - psychosocial (0-8) Scale from: 0 to 8 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact Scale - psychosocial (0-8) was -0.8	MD 0.8 lower (6.53 lower to 4.93 higher)
SF-36 physical composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-36 physical composite (0-100) was -0.2	MD 3.9 higher (0.88 lower to 8.68 higher)
SF-36 mental composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-36 mental composite (0-100) was 2.3	MD 4.2 lower (11.24 lower to 2.84 higher)
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (0-63) was 0.6	MD 2.9 lower (6.16 lower to 0.36 higher)
Incidence of adverse events follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b,d	RD 0.00 (-0.11 to 0.11)	0 per 1,000	0 fewer per 1,000 (110 fewer to 110 more)

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

b. Downgraded by 1 increment as follow-up for the majority of the evidence was less than the 3 months minimum specified in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Imprecision assessed using sample size as zero events in both arms of at least one study. Downgraded by 2 increments as sample size <70.

Vestibular/balance training vs. aerobic exercise

Table 11: Clinical evidence summary: Vestibular/balance training vs. aerobic exercise – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Vestibular/balance training
Modified Fatigue Impact Scale - total (0-84) Scale from: 0 to 84 follow up: 10 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact Scale - total (0-84) was 44.7	MD 14.4 lower (29.13 lower to 0.33 higher)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 8 weeks	50 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was 39.31	MD 5.23 lower (14.21 lower to 3.75 higher)
Improvement in MFIS from baseline follow up: 3 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 4.50 (0.37 to 54.16)	Moderate	
				667 per 1,000	233 more per 1,000 (241 fewer to 324 more)
Improvement in MFIS (motor) from baseline follow up: 3 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 1.13 (0.06 to 21.09)	Moderate	
				889 per 1,000	12 more per 1,000 (565 fewer to 105 more)
Improvement in HAQUAMS (motor) from baseline follow up: 3 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 1.87 (0.28 to 12.31)	Moderate	
				556 per 1,000	145 more per 1,000 (296 fewer to 383 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Vestibular/balance training
EDSS (0-10) Scale from: 0 to 10 follow up: 8 weeks	50 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean EDSS (0-10) was 2.27	MD 0.83 higher (0.15 lower to 1.81 higher)
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 10 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (0-63) was 12.9	MD 1.3 lower (9.51 lower to 6.91 higher)
Beck Depression Inventory - fast screen (0-21) Scale from: 0 to 21 follow up: 8 weeks	50 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory - fast screen (0-21) was 5.12	MD 0.17 higher (2.74 lower to 3.08 higher)
Improvement in Beck Depression Inventory from baseline follow up: 3 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 4.50 (0.37 to 54.16)	Moderate	
				667 per 1,000	233 more per 1,000 (241 fewer to 324 more)
Adverse events follow up: 6 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	OR 0.15 (0.00 to 7.39)	Moderate	
				77 per 1,000	77 fewer per 1,000 (270 fewer to 116 more)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up of less than the 3 months minimum specified in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Vestibular/balance training vs. resistance training

Table 12: Clinical evidence summary: Vestibular/balance training vs. resistance training – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with resistance training	Risk difference with Vestibular/balance training
Modified Fatigue Impact Scale - total (0-84) Scale from: 0 to 84 follow up: 10 weeks	51 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c,	-	The mean modified Fatigue Impact Scale - total (0-84) was -12.8	MD 1.7 higher (4.43 lower to 7.83 higher)
Adverse events leading to withdrawal follow up: 10 weeks	51 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.09 (0.01 to 0.56)	Moderate	
				217 per 1,000	217 fewer per 1,000 (43 fewer to 392 fewer)

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

b. Downgraded by 1 increment as the follow-up for the majority of the evidence was less than the minimum of 3 months specified in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Resistance training + aerobic exercise vs. control (waitlist control, no intervention, information only)

Table 13: Clinical evidence summary: Resistance training + aerobic exercise vs. control (waitlist control, no intervention, information only) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only)	Risk difference with Resistance + aerobic
Modified Fatigue Impact scale - Total score (0-84)	312 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean modified Fatigue Impact scale - Total score (0-84) ranged from -1.1 to -4.5	MD 5.43 lower (9.93 lower to 0.92 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only)	Risk difference with Resistance + aerobic
Scale from: 0 to 84 follow up: range 12 weeks to 6 months					
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 12 weeks	112 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was 0.4	MD 4.3 lower (6.42 lower to 2.18 lower)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 12 weeks	112 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was -0.51	MD 1.59 lower (3.15 lower to 0.03 lower)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 8 weeks	36 (1 RCT)	⊕⊕◯◯ LOW a,d	-	The mean fatigue Severity Scale (9-63) was 41.22	MD 15.94 lower (24.2 lower to 7.68 lower)
WEIMuS Fatigue score (0-68) Scale from: 0 to 68 follow up: 6 months	177 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean wEIMuS Fatigue score (0-68) was -0.89	MD 2.05 lower (5.26 lower to 1.16 higher)
MSIS-29 physical (0-100) Scale from: 0 to 100 follow up: 12 weeks	112 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSIS-29 physical (0-100) was 0.3	MD 7.2 lower (12.87 lower to 1.53 lower)
MSQoL-54 mental composite Scale from: 0 to 100 follow up: 12 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean mSQoL-54 mental composite was -5.2	MD 16.3 higher (2.78 higher to 29.82 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only)	Risk difference with Resistance + aerobic
MSQoL-54 physical composite Scale from: 0 to 100 follow up: 12 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean mSQoL-54 physical composite was 3.3	MD 6.7 higher (13.13 lower to 26.53 higher)
Beck Depression Inventory (0-63) - Maurer 18 - e-training individualised exercise protocol Scale from: 0 to 63 follow up: 6 months	177 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean beck Depression Inventory (0-63) - Maurer 18 - e-training individualised exercise protocol was -0.65	MD 0.65 lower (2.94 lower to 1.64 higher)
Beck Depression Inventory (0-63) - Razazian 2016 - aquatic exercises at rehab centre Scale from: 0 to 63 follow up: 8 weeks	36 (1 RCT)	⊕⊕◯◯ LOW a,d	-	The mean beck Depression Inventory (0-63) - Razazian 2016 - aquatic exercises at rehab centre was 21.33	MD 16.55 lower (20.1 lower to 13 lower)
Beck Depression Inventory (0-63) - Correale 2021 - training sessions at centre Scale from: 0 to 63 follow up: 12 weeks	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean beck Depression Inventory (0-63) - Correale 2021 - training sessions at centre was -2.3	MD 4.7 lower (11.39 lower to 1.99 higher)
Adverse events leading to withdrawal follow up: range 12 weeks to 6 months	288 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,e	RR 0.57 (0.12 to 2.81)	Moderate	
				77 per 1,000	33 fewer per 1,000 (67 fewer to 138 more)
				Moderate	

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only)	Risk difference with Resistance + aerobic
Any adverse event follow up: 6 months	178 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 0.91 (0.50 to 1.66)	607 per 1,000	23 fewer per 1,000 (171 fewer to 112 more)

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

b. Heterogeneity that cannot be explained by subgroup analysis exists, based on point estimates varying between studies and I2 >50%

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Downgraded by 1 increment as the follow-up for the majority of the evidence is less than the minimum 3 months specified in the protocol

e. Heterogeneity that cannot be explained by subgroup analysis exists, based on point estimates differing widely between the two studies

Resistance training + balance exercises vs. control (no intervention, waitlist control)

Table 14: Clinical evidence summary: Resistance training + balance exercises vs. control (no intervention, waitlist control) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control)	Risk difference with Resistance training + balance
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: range 8 weeks to 12 weeks	132 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue Severity Scale (9-63) was 1.95	MD 5.7 lower (16.5 lower to 5.1 higher)
SF-36 (0-100) - Physical functioning	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 (0-100) - Physical functioning was 7.7	MD 9.71 higher (2.75 higher to 16.66 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control)	Risk difference with Resistance training + balance
Scale from: 0 to 100 follow up: 8 weeks					
SF-36 (0-100) - Role-physical functioning Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 (0-100) - Role-physical functioning was 5	MD 12.75 higher (19.28 lower to 44.78 higher)
SF-36 (0-100) - Bodily pain Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 (0-100) - Bodily pain was 4	MD 1.97 higher (1.51 lower to 5.44 higher)
SF-36 (0-100) - General health Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 (0-100) - General health was 3.2	MD 0.31 higher (8.29 lower to 8.91 higher)
SF-36 (0-100) - vitality Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 (0-100) - vitality was 11	MD 0.75 lower (16.45 lower to 14.95 higher)
SF-36 (0-100) - Social functioning Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 (0-100) - Social functioning was 5	MD 1.15 higher (12.37 lower to 14.67 higher)
SF-36 (0-100) - Role-emotional functioning Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 (0-100) - Role-emotional functioning was 19.9	MD 8.57 lower (46.08 lower to 28.93 higher)
SF-36 (0-100) - Mental health Scale from: 0 to 100 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 (0-100) - Mental health was 7	MD 1.55 lower (7.84 lower to 4.74 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention, waitlist control)	Risk difference with Resistance training + balance
MusiQoL (0-100) Scale from: 0 to 100 follow up: 12 weeks	99 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean musiQoL (0-100) was -0.4	MD 2.38 higher (0.41 higher to 4.35 higher)
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	33 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean beck Depression Inventory (0-63) was -1.6	MD 0.94 lower (5.5 lower to 3.62 higher)
Adverse events leading to withdrawal follow up: range 8 weeks to 12 weeks	142 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	RR 0.39 (0.11 to 1.36)	Moderate	
				154 per 1,000	94 fewer per 1,000 (137 fewer to 56 more)

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

b. Heterogeneity that cannot be explained by subgrouping analyses is present and I2 >75%

c. Downgraded by 1 increment as the majority of the evidence has a follow-up of less than the 3 months specified in the protocol

d. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Vestibular/balance training + aerobic exercise vs. control (education only)

Table 15: Clinical evidence summary: Vestibular/balance training + aerobic exercise vs. control (education only) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (education only)	Risk difference with Balance + aerobic exercise
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Total score (0-84) was 61	MD 28.2 lower (33.21 lower to 23.19 lower)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was 29.4	MD 15.3 lower (18.45 lower to 12.15 lower)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was 24.9	MD 10.4 lower (13.19 lower to 7.61 lower)
Modified Fatigue Impact scale - Psychosocial scale (0-8) Scale from: 0 to 8 follow up: 8 weeks	32 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Psychosocial scale (0-8) was 6.7	MD 2.5 lower (3.54 lower to 1.46 lower)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum specified in the protocol

Resistance training + balance exercise + aerobic exercise vs. control (usual care, no intervention)

Table 16: Clinical evidence summary: Resistance training + balance exercise + aerobic exercise vs. control (usual care, no intervention) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care, no intervention), up to 6 months	Risk difference with Resistance + balance + aerobic exercise
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 8 weeks	58 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean modified Fatigue Impact scale - Total score (0-84) was 48.89	MD 19.25 lower (37.92 lower to 0.58 lower)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was 29.5	MD 15.5 lower (19.49 lower to 11.51 lower)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was 24.5	MD 10.1 lower (13.95 lower to 6.25 lower)
Modified Fatigue Impact scale - Psychosocial scale (0-8) Scale from: 0 to 8 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean modified Fatigue Impact scale - Psychosocial scale (0-8) was 6.7	MD 2.8 lower (4.18 lower to 1.42 lower)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: range 5 weeks to 12 weeks	37 (3 RCTs)	⊕◯◯◯ VERY LOW a,c,d,e	-	The mean fatigue Severity Scale (9-63) was 41.15	MD 8.59 lower (14.44 lower to 2.74 lower)
Fatigue Severity Scale (1-7)	49 (2 RCTs)	⊕◯◯◯ VERY LOW a,d	-	The mean fatigue Severity Scale (1-7) was 6.11	MD 0.64 lower (1.2 lower to 0.07 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care, no intervention), up to 6 months	Risk difference with Resistance + balance + aerobic exercise
Scale from: 1 to 7 follow up: 3 months					
MSQOL-54 - physical summary (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSQOL-54 - physical summary (0-100) was 44.2	MD 21.2 higher (16.35 higher to 26.05 higher)
MSQOL-54 - mental summary (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSQOL-54 - mental summary (0-100) was 43.6	MD 26.6 higher (20.26 higher to 32.94 higher)
MSIS-29 - physical score (0-100) Scale from: 0 to 100 follow up: 3 months	24 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean MSIS-29 - physical score (0-100) was 53	MD 3.84 lower (17.9 lower to 10.22 higher)
MSIS-29 - psychological score (0-100) Scale from: 0 to 100 follow up: 3 months	24 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean MSIS-29 - psychological score (0-100) was 53.7	MD 10.74 lower (23.79 lower to 2.31 higher)
Multicultural quality of life index (MQLIM; scale 0-100) Scale from: 0 to 100 follow up: 8 weeks	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multicultural quality of life index (MQLIM; scale 0-100) was 66.52	MD 13.54 higher (7.52 higher to 19.56 higher)
MS-specific quality of life - mental domain (name and range of scale unclear) - MS-specific quality of life - mental domain (name and range of scale	61 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean mS-specific quality of life - mental domain (name and range of scale unclear) - MS-specific quality of life - mental domain (name and range of scale unclear) was not reported	MD 16.36 higher (7.1 higher to 25.62 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care, no intervention), up to 6 months	Risk difference with Resistance + balance + aerobic exercise
unclear) follow up: 11 weeks					
MS-specific quality of life - physical domain (name and range of scale unclear) follow up: 11 weeks	61 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean mS-specific quality of life - physical domain (name and range of scale unclear) was not reported	MD 12.17 higher (5.28 higher to 19.06 higher)
EDSS (0-10) Scale from: 0 to 10 follow up: 11 weeks	61 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean EDSS (0-10) was not reported	MD 0.13 lower (0.61 lower to 0.35 higher)
Hospital Anxiety and Depression Scale (0-63) Scale from: 0 to 63 follow up: 12 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean hospital Anxiety and Depression Scale (0-63) was 13.8	MD 2.1 lower (7.16 lower to 2.96 higher)
Leeds MS quality of life (0-24) Scale from: 0 to 24 follow up: 12 weeks	25 (1 RCT)	⊕◯◯◯ VERY LOW a,d	-	The mean leeds MS quality of life (0-24) was 12.4	MD 1.5 lower (4.25 lower to 1.25 higher)
Adverse events leading to withdrawal follow up: 11 weeks	64 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	RR 1.12 (0.11 to 11.71)	Moderate	
				44 per 1,000	5 more per 1,000 (39 fewer to 466 more)

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

b. Heterogeneity that could not be explained by subgrouping strategies and I2 >75%

c. Downgraded by 1 increment as the majority of the evidence has a follow-up less than the minimum 3 months specified in the protocol

d. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

e. Heterogeneity present that could not be explained by subgrouping analyses

Table 17: Clinical evidence summary: Resistance training + balance exercise + aerobic exercise vs. control (usual care, no intervention) – outcomes >6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care, no intervention), >6 months	Risk difference with Resistance + balance + aerobic exercise
Fatigue Severity Scale (9-63) - Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 1 years	55 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fatigue Severity Scale (9-63) - Fatigue Severity Scale (9-63) was not reported	MD 10.2 lower (16.84 lower to 3.56 lower)
MS-specific quality of life - mental domain (name and range of scale unclear) - MS-specific quality of life - mental domain (name and range of scale unclear) follow up: 1 years	55 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean mS-specific quality of life - mental domain (name and range of scale unclear) - MS-specific quality of life - mental domain (name and range of scale unclear) was not reported	MD 13.54 higher (2.48 higher to 24.6 higher)
MS-specific quality of life - physical domain (name and range of scale unclear) - MS-specific quality of life - physical domain (name and range of scale unclear) follow up: 1 years	55 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean mS-specific quality of life - physical domain (name and range of scale unclear) - MS-specific quality of life - physical domain (name and range of scale unclear) was not reported	MD 10.9 higher (1.99 higher to 19.81 higher)
EDSS (0-10) Scale from: 0 to 10 follow up: 1 years	55 (1 RCT)	⊕◯◯◯ VERY LOW a	-	The mean EDSS (0-10) was not reported	MD 0.28 lower (0.86 lower to 0.3 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Standard exercises (resistance + balance + aerobic) + high-intensity lower limb resistance training vs. standard exercises alone

Table 18: Clinical evidence summary: Standard exercises (resistance + balance + aerobic) + high-intensity lower limb resistance training vs. standard exercises alone – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard exercises alone	Risk difference with Standard exercises (resistance + balance + aerobic) + high-intensity lower limb resistance training
Fatigue Severity Scale (10 max score) follow up: 12 weeks	19 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean fatigue Severity Scale (10 max score) was -1.38	MD 0.44 higher (0.5 lower to 1.38 higher)
Adverse events follow up: 12 weeks	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.12 (0.00 to 6.14)	Moderate	
				111 per 1,000	96 fewer per 1,000 (111 fewer to 323 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Resistance + balance + aerobic exercise vs. massage

Table 19: Clinical evidence summary: Resistance + balance + aerobic exercise vs. massage – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with massage	Risk difference with Resistance + balance + aerobic exercise
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 5 weeks	24 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was -8.08	MD 2.67 lower (8.61 lower to 3.27 higher)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum specified in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Massage + exercise (resistance, balance + aerobic) vs. control (no intervention)

Table 20: Clinical evidence summary: Massage + exercise (resistance, balance + aerobic) vs. control (no intervention) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (no intervention)	Risk difference with Massage + exercise (resistance, balance, aerobic)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 5 weeks	24 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was 3	MD 12.42 lower (18.87 lower to 5.97 lower)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Massage + exercise (resistance, balance + aerobic) vs. exercise only

Table 21: Clinical evidence summary: Massage + exercise (resistance, balance + aerobic) vs. exercise only – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with exercise alone	Risk difference with Massage + exercise (resistance, balance, aerobic)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 5 weeks	24 (1 RCT)	⨁◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was -10.75	MD 1.33 higher (5.96 lower to 8.62 higher)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Massage + exercise (resistance, balance + aerobic) vs. massage only

Table 22: Clinical evidence summary: Massage + exercise (resistance, balance + aerobic) vs. massage only – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with massage alone	Risk difference with Massage + exercise (resistance, balance, aerobic)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 5 weeks	24 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was -8.08	MD 1.34 lower (8.73 lower to 6.05 higher)

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

b. Downgraded by 1 increment as the majority of the evidence has a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Resistance + aerobic exercise vs. yoga

Table 23: Clinical evidence summary: Resistance + aerobic exercise vs. yoga – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with yoga	Risk difference with Resistance + aerobic
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 24 weeks	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Total score (0-84) was 33.9	MD 1 lower (8.63 lower to 6.63 higher)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 12 weeks	126 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was -2.1	MD 1.8 lower (4.09 lower to 0.49 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with yoga	Risk difference with Resistance + aerobic
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 12 weeks	126 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was -0.96	MD 1.14 lower (2.5 lower to 0.22 higher)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 8 weeks	36 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was 38.94	MD 13.66 lower (21.96 lower to 5.36 lower)
MSIS-29 (0-100) - Physical domain Scale from: 0 to 100 follow up: 24 weeks	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) - Physical domain was 34	MD 6.3 lower (14.9 lower to 2.3 higher)
MSIS-29 (0-100) - Psychological domain Scale from: 0 to 100 follow up: 24 weeks	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) - Psychological domain was 30.19	MD 6.7 lower (14.82 lower to 1.42 higher)
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	36 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean beck Depression Inventory (0-63) was 5.06	MD 0.28 lower (2.36 lower to 1.8 higher)
Adherence - classes attended out of possible 10 Scale from: 0 to 10 follow up: 12 weeks	126 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean adherence - classes attended out of possible 10 was 7.8	MD 0.3 higher (0.53 lower to 1.13 higher)
Adverse events leading to withdrawal follow up: 24 weeks	90 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 2.23 (0.63 to 7.87)	Moderate	
				73 per 1,000	90 more per 1,000 (27 fewer to 503 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

Fatigue/energy management programme vs. control (waitlist, no intervention, information only)

Table 24: Clinical evidence summary: Fatigue/energy management programme vs. control (waitlist, no intervention, information only) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), up to 6 months	Risk difference with Fatigue/energy management programme
Fatigue Severity Scale (1-7) Scale from: 1 to 7 follow up: range 4 weeks to 4.25 months	296 (4 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Scale (1-7) was 5.01	MD 0.07 lower (0.29 lower to 0.15 higher)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 6 weeks	30 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (9-63) was 45.82	MD 2.78 higher (1.43 lower to 6.99 higher)
MFIS - total (0-84) Scale from: 0 to 84 follow up: range 6 weeks to 26 weeks	101 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean MFIS - total (0-84) was 40.19	MD 2.6 lower (8.84 lower to 3.64 higher)
MFIS - physical (0-36) Scale from: 0 to 36 follow up: range 6 weeks to 26 weeks	101 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean MFIS - physical (0-36) was 19.49	MD 0.78 lower (3.29 lower to 1.73 higher)
MFIS - cognitive (0-40) Scale from: 0 to 40 follow up: range 6 weeks to 26 weeks	101 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean MFIS - cognitive (0-40) was 17.04	MD 1.63 lower (4.43 lower to 1.16 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), up to 6 months	Risk difference with Fatigue/energy management programme
MFIS - psychosocial (0-8) Scale from: 0 to 8 follow up: range 6 weeks to 26 weeks	101 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,e	-	The mean MFIS - psychosocial (0-8) was 3.68	MD 0.23 lower (1.06 lower to 0.61 higher)
Fatigue Impact Scale - total (0-160) Scale from: 0 to 160 follow up: 4.25 months	23 (1 RCT)	⊕⊕◯◯ LOW a,c	-	The mean fatigue Impact Scale - total (0-160) was 79.4	MD 20.7 lower (43.1 lower to 1.7 higher)
Fatigue Impact Scale - cognitive (0-40) Scale from: 0 to 40 follow up: range 6 weeks to 4.25 months	377 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Impact Scale - cognitive (0-40) was 21.1	MD 3.14 lower (4.55 lower to 1.73 lower)
Fatigue Impact Scale - physical (0-40) Scale from: 0 to 40 follow up: range 6 weeks to 4.25 months	377 (3 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Impact Scale - physical (0-40) was 23.6	MD 3.05 lower (4.53 lower to 1.56 lower)
Fatigue Impact Scale - psychosocial (0-80) Scale from: 0 to 80 follow up: range 6 weeks to 4.25 months	377 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Impact Scale - psychosocial (0-80) was 34.7	MD 6.1 lower (8.79 lower to 3.41 lower)
CIS20r - fatigue (8-56) Scale from: 8 to 56 follow up: 26 weeks	71 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cIS20r - fatigue (8-56) was 40.1	MD 3.55 lower (7.52 lower to 0.42 higher)
Clinically significant improvement in fatigue - 0.5-point reduction on	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	RR 1.64 (0.18 to 15.26)	Moderate	
				111 per 1,000	71 more per 1,000 (91 fewer to 1,584 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), up to 6 months	Risk difference with Fatigue/energy management programme
FSS follow up: 4 weeks					
Clinically significant improvement in fatigue - 10-point improvement on MFIS follow up: 4 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	RR 0.38 (0.13 to 1.09)	Moderate	
				438 per 1,000	271 fewer per 1,000 (381 fewer to 39 more)
SF-36 physical function (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 physical function (0-100) was 59.2	MD 1.68 higher (1.21 lower to 4.56 higher)
SF-36 role physical (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean SF-36 role physical (0-100) was 51.4	MD 9.45 higher (5.45 lower to 24.34 higher)
SF-36 body pain (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 body pain (0-100) was 65.1	MD 3.34 higher (0.93 lower to 7.62 higher)
SF-36 general health (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 general health (0-100) was 47.9	MD 2.71 higher (0.33 lower to 5.75 higher)
SF-36 vitality (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean SF-36 vitality (0-100) was 43.3	MD 6.04 higher (1.48 lower to 13.57 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), up to 6 months	Risk difference with Fatigue/energy management programme
SF-36 social function (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 social function (0-100) was 67.6	MD 4.43 higher (0.29 lower to 9.15 higher)
SF-36 role emotional (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,d	-	The mean SF-36 role emotional (0-100) was 81.1	MD 4.67 higher (7.15 lower to 16.49 higher)
SF-36 mental health (0-100) Scale from: 0 to 100 follow up: range 6 weeks to 26 weeks	425 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean SF-36 mental health (0-100) was 71.6	MD 4.74 higher (1.73 higher to 7.76 higher)
MSIS-29 - total (0-100) Scale from: 0 to 100 follow up: 4.25 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSIS-29 - total (0-100) was 42.67	MD 4.65 lower (17.97 lower to 8.67 higher)
MSIS-29 - physical (0-100) Scale from: 0 to 100 follow up: 4.25 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSIS-29 - physical (0-100) was 45.12	MD 6.66 lower (21.22 lower to 7.9 higher)
MSIS-29 - psychological (0-100) Scale from: 0 to 100 follow up: 4.25 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSIS-29 - psychological (0-100) was 37.49	MD 1.17 lower (16.95 lower to 14.61 higher)
CIS20r - concentration (5-35) Scale from: 5 to 35 follow up: 26 weeks	71 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean cIS20r - concentration (5-35) was 19.1	MD 0.4 higher (2.54 lower to 3.34 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), up to 6 months	Risk difference with Fatigue/energy management programme
Adverse events follow up: 6 weeks	181 (1 RCT)	⊕◯◯◯ VERY LOW a,b,f	RD 0.00 (-0.02 to 0.02)	0 per 1,000	0 fewer per 1,000 (20 fewer to 20 more)
BDI fast screen (0-21) Scale from: 0 to 21 follow up: 4.25 months	23 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean BDI fast screen (0-21) was 2.2	MD 0.11 higher (2.02 lower to 2.24 higher)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Heterogeneity that cannot be explained by subgrouping strategies

e. Heterogeneity that cannot be explained by subgrouping strategies and I2 >75%

f. Imprecision assessed using sample size as zero events in both arms of a single study. Downgraded by 1 increment as sample size <350 and >70

Table 25: Clinical evidence summary: Fatigue/energy management programme vs. control (waitlist, no intervention, information only) – outcomes >6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), >6 months	Risk difference with Fatigue/energy management programme
Fatigue Severity Scale (1-7)	69 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean fatigue Severity Scale (1-7) was 5.3	MD 0.02 lower (0.37 lower to 0.33 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), >6 months	Risk difference with Fatigue/energy management programme
Scale from: 1 to 7 follow up: 52 weeks					
MFIS - total (0-84) Scale from: 0 to 84 follow up: 52 weeks	69 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS - total (0-84) was 40.6	MD 0.1 higher (5.46 lower to 5.66 higher)
MFIS - physical (0-36) Scale from: 0 to 36 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - physical (0-36) was 20.0	MD 0.07 higher (2.56 lower to 2.7 higher)
MFIS - cognitive (0-40) Scale from: 0 to 40 follow up: 52 weeks	69 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MFIS - cognitive (0-40) was 16.9	MD 0.2 higher (2.66 lower to 3.06 higher)
MFIS - psychosocial (0-8) Scale from: 0 to 8 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - psychosocial (0-8) was 3.6	MD 0.22 higher (0.48 lower to 0.92 higher)
CIS20r - fatigue (8-56) Scale from: 8 to 56 follow up: 52 weeks	73 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r - fatigue (8-56) was 42.1	MD 1.45 lower (5.46 lower to 2.56 higher)
SF-36 physical function (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-36 physical function (0-100) was 54.0	MD 2.91 higher (3.45 lower to 9.27 higher)
SF-36 role physical (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 role physical (0-100) was 37.1	MD 3.88 higher (13.53 lower to 21.29 higher)
SF-36 body pain (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 body pain (0-100) was 68.2	MD 5.37 lower (13.62 lower to 2.88 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), >6 months	Risk difference with Fatigue/energy management programme
SF-36 general health (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 general health (0-100) was 49.6	MD 1.88 higher (3.52 lower to 7.28 higher)
SF-36 vitality (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 vitality (0-100) was 42.2	MD 2.87 higher (3.98 lower to 9.72 higher)
SF-36 social function (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-36 social function (0-100) was 65.7	MD 1.14 lower (9.48 lower to 7.2 higher)
SF-36 role emotional (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 role emotional (0-100) was 68.6	MD 7.3 higher (9.98 lower to 24.58 higher)
SF-36 mental health (0-100) Scale from: 0 to 100 follow up: 52 weeks	69 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-36 mental health (0-100) was 69.5	MD 0.56 higher (5.92 lower to 7.04 higher)
CIS20r - concentration (5-35) Scale from: 5 to 35 follow up: 52 weeks	69 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean cIS20r - concentration (5-35) was 20.7	MD 0.26 lower (3.23 lower to 2.71 higher)
Adverse events (serious) follow up: 52 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.11 (0.30 to 4.12)	Moderate	
				100 per 1,000	11 more per 1,000 (70 fewer to 312 more)
Adverse events leading to withdrawal follow up: 52 weeks	76 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RD 0.00 (-0.05 to 0.05)	0 per 1,000	0 fewer per 1,000 (50 fewer to 50 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention, information only), >6 months	Risk difference with Fatigue/energy management programme
Adherence to programme	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.79 (0.24 to 2.58)	Moderate	
				864 per 1,000	30 fewer per 1,000 (260 fewer to 79 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Imprecision assessed using sample size as zero events in both arms of a single study. Downgraded by 1 increment as sample size <350 and >70.

Fatigue/energy management programme vs. general self-management programme

Table 26: Clinical evidence summary: Fatigue/energy management programme vs. general self-management programme – outcomes up to 6 months and >6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with general self-management programme	Risk difference with Fatigue/energy management programme
MFIS - total (0-84) - 6 months Scale from: 0 to 84 follow up: 6 months	203 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean MFIS - total (0-84) - 6 months was 41.9	MD 1 lower (5.33 lower to 3.33 higher)
MFIS - total (0-84) - 12 months Scale from: 0 to 84 follow up: 12 months	78 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - total (0-84) - 12 months was 43.7	MD 5.1 lower (12.17 lower to 1.97 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with general self-management programme	Risk difference with Fatigue/energy management programme
BDI (0-63) - 6 weeks Scale from: 0 to 63 follow up: 6 weeks	204 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean BDI (0-63) - 6 weeks was 10.7	MD 1.2 lower (3.31 lower to 0.91 higher)
Adverse events (all relapses) - 6 weeks follow up: 6 weeks	204 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	RR 1.04 (0.27 to 4.05)	Moderate	
				39 per 1,000	2 more per 1,000 (28 fewer to 117 more)
Adherence - completed at least 4 sessions	218 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 1.00 (0.46 to 2.16)	Moderate	
				862 per 1,000	0 fewer per 1,000 (120 fewer to 69 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months specified in the protocol

Fatigue/energy management programme vs. relaxation

Table 27: Clinical evidence summary: Fatigue/energy management programme vs. relaxation – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with Fatigue/energy management programme
MFIS - Total (0-84) Scale from: 0 to 84 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - Total (0-84) was 41.9	MD 9.6 lower (20.4 lower to 1.2 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with Fatigue/energy management programme
MFIS - Physical (0-36) Scale from: 0 to 36 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - Physical (0-36) was 20.4	MD 3.8 lower (9.06 lower to 1.46 higher)
MFIS - Cognitive (0-40) Scale from: 0 to 40 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - Cognitive (0-40) was 17.7	MD 4.9 lower (10.93 lower to 1.13 higher)
MFIS - Psychosocial (0-8) Scale from: 0 to 8 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MFIS - Psychosocial (0-8) was 3.8	MD 0.9 lower (2.41 lower to 0.61 higher)
Checklist individual strength - Total (20-140) Scale from: 20 to 140 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean checklist individual strength - Total (20-140) was 74.8	MD 2.2 higher (18.58 lower to 22.98 higher)
Checklist individual strength - Concentration (5-35) Scale from: 5 to 35 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean checklist individual strength - Concentration (5-35) was 17.1	MD 1.5 higher (5.35 lower to 8.35 higher)
Checklist individual strength - Physical activity (3-21) Scale from: 3 to 21 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean checklist individual strength - Physical activity (3-21) was 9.4	MD 1.2 higher (3.14 lower to 5.54 higher)
Checklist individual strength - Motivation (4-28) Scale from: 4 to 28 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean checklist individual strength - Motivation (4-28) was 9.4	MD 1.2 higher (3.14 lower to 5.54 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with Fatigue/energy management programme
Checklist individual strength - Subjective fatigue (8-56) Scale from: 8 to 56 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean checklist individual strength - Subjective fatigue (8-56) was 36.6	MD 1.3 higher (9.04 lower to 11.64 higher)
SF-36 (0-100 for all) - Physical functioning Scale from: 0 to 100 follow up: 3 months	225 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Physical functioning was 58.3	MD 8.6 higher (8.17 lower to 25.37 higher)
SF-36 (0-100 for all) - Role physical function Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Role physical function was 66.7	MD 7.3 lower (36.91 lower to 22.31 higher)
SF-36 (0-100 for all) - Physical pain Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-36 (0-100 for all) - Physical pain was 59.2	MD 24.1 higher (12.31 higher to 35.89 higher)
SF-36 (0-100 for all) - General health Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - General health was 47.6	MD 1.2 higher (11 lower to 13.4 higher)
SF-36 (0-100 for all) - Vitality Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Vitality was 48.9	MD 5.5 higher (7.59 lower to 18.59 higher)
SF-36 (0-100 for all) - Social functioning Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Social functioning was 68.1	MD 3.8 higher (9.63 lower to 17.23 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with Fatigue/energy management programme
SF-36 (0-100 for all) - Role emotional function Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Role emotional function was 85.2	MD 6 lower (33.25 lower to 21.25 higher)
SF-36 (0-100 for all) - Mental health Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Mental health was 70.7	MD 6.7 lower (18.87 lower to 5.47 higher)
SF-36 (0-100 for all) - Health change Scale from: 0 to 100 follow up: 3 months	25 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 (0-100 for all) - Health change was 58.3	MD 14.5 lower (31.63 lower to 2.63 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Aerobic exercise + fatigue self-management vs. control (information only)

Table 28: Clinical evidence summary: Aerobic exercise + fatigue self-management vs. control (information only) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (information only)	Risk difference with Aerobic exercise + fatigue self-management
Fatigue Impact scale - total (0-160) Scale from: 0 to 160 follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Impact scale - total (0-160) was 62.63	MD 8.68 lower (19.33 lower to 1.97 higher)
MSIS-29 (0-100) - Physical function Scale from: 0 to 100 follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) - Physical function was 37.81	MD 6.7 lower (13.43 lower to 0.03 higher)
MSIS-29 (0-100) - Mental function Scale from: 0 to 100 follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) - Mental function was 35.77	MD 6.21 lower (12.93 lower to 0.51 higher)
Adverse events (exacerbations) follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 0.81 (0.43 to 1.52)	Moderate	
				246 per 1,000	47 fewer per 1,000 (140 fewer to 128 more)
Adverse events (orthopaedic problems) follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.15 (0.75 to 1.77)	Moderate	
				348 per 1,000	52 more per 1,000 (87 fewer to 268 more)
Adverse events (at least 1 fall) follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.03 (0.63 to 1.70)	Moderate	
				304 per 1,000	9 more per 1,000 (113 fewer to 213 more)
Adherence - completed all 1-1 calls	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 1.21 (0.54 to 2.71)	Moderate	
				768 per 1,000	32 more per 1,000 (127 fewer to 132 more)
Adherence - completed all group calls with or without at least 1 makeup session	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 1.71 (0.62 to 4.70)	Moderate	
				841 per 1,000	60 more per 1,000 (75 fewer to 121 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Aerobic exercise + fatigue self-management vs. aerobic exercise only

Table 29: Clinical evidence summary: Aerobic exercise + fatigue self-management vs. aerobic exercise only – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise only	Risk difference with Aerobic exercise + fatigue self-management
Fatigue Impact scale - total (0-160) Scale from: 0 to 160 follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Impact scale - total (0-160) was 68.03	MD 14.08 lower (24.07 lower to 4.09 lower)
MSIS-29 (0-100) - Physical function Scale from: 0 to 100 follow up: 24 weeks	139 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 (0-100) - Physical function was 32.19	MD 1.08 lower (7.5 lower to 5.34 higher)
MSIS-29 (0-100) - Mental function Scale from: 0 to 100 follow up: 24 weeks	139 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 (0-100) - Mental function was 31.08	MD 1.52 lower (8.09 lower to 5.05 higher)
Adverse events (exacerbations) follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.15 (0.57 to 2.31)	Moderate	
				174 per 1,000	26 more per 1,000 (75 fewer to 228 more)
Adverse events (orthopaedic problems) follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.73 (1.03 to 2.89)	Moderate	
				232 per 1,000	169 more per 1,000 (7 more to 438 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise only	Risk difference with Aerobic exercise + fatigue self-management
Adverse events (at least 1 fall) follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.81 (0.97 to 3.36)	Moderate 174 per 1,000	141 more per 1,000 (5 fewer to 410 more)
Adherence - completed all 1-1 calls follow up: 24 weeks	139 (1 RCT)	⊕⊕◯◯ LOW a,b	OR 1.87 (0.86 to 4.06)	Moderate 681 per 1,000	119 more per 1,000 (34 fewer to 215 more)
Adherence - completed all group calls with or without at least 1 makeup session follow up: 24 weeks	139 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 1.53 (0.55 to 4.27)	Moderate 855 per 1,000	45 more per 1,000 (91 fewer to 107 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Fatigue management + CBT vs. control (local/standard care)

Table 30: Clinical evidence summary: Fatigue management + CBT vs. control (local/standard care) – outcomes up to 6 months and >6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (local/standard care)	Risk difference with Fatigue management + CBT
Global fatigue severity (1-7) - 5.5 months Scale from: 1 to 7 follow up: 5.5 months	146 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean global fatigue severity (1-7) - 5.5 months was 5.66	MD 0.36 lower (0.63 lower to 0.09 lower)
Global fatigue severity (1-7) - 12 months Scale from: 1 to 7 follow up: 12 months	131 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean global fatigue severity (1-7) - 12 months was 5.7	MD 0.38 lower (0.72 lower to 0.04 lower)
MFIS total (0-84) Scale from: 0 to 84 follow up: 10 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MFIS total (0-84) was 12.88	MD 3.88 lower (6.28 lower to 1.48 lower)
Chalder fatigue scale (0-33) Scale from: 0 to 33 follow up: range 10 weeks to 12 weeks	315 (2 RCTs)	⊕◯◯◯ VERY LOW b,d	-	The mean chalder fatigue scale (0-33) was 19.57	MD 4.39 lower (9.25 lower to 0.46 higher)
MS fatigue self-efficacy scale (10-100) - 5.5 months Scale from: 10 to 100 follow up: 5.5 months	146 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MS fatigue self-efficacy scale (10-100) - 5.5 months was 43.0	MD 6 higher (0 to 12 higher)
MS fatigue self-efficacy scale (10-100) - 12 months Scale from: 10 to 100 follow up: 12 months	131 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MS fatigue self-efficacy scale (10-100) - 12 months was 53.0	MD 4 higher (1.65 lower to 9.65 higher)
Fatigue Scale of Motor and Cognition - Total (20-100) Scale from: 20 to 100 follow up: 12 weeks	275 (1 RCT)	⊕⊕⊕⊕ HIGH	-	The mean fatigue Scale of Motor and Cognition - Total (20-100) was not reported	MD 3.47 lower (5.89 lower to 1.05 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (local/standard care)	Risk difference with Fatigue management + CBT
Fatigue Scale of Motor and Cognition - Motor (0-50) Scale from: 0 to 50 follow up: 12 weeks	275 (1 RCT)	⊕⊕⊕⊕ HIGH	-	The mean fatigue Scale of Motor and Cognition - Motor (0-50) was not reported	MD 1.49 lower (2.74 lower to 0.24 lower)
Fatigue Scale of Motor and Cognition - Cognition (0-50) Scale from: 0 to 50 follow up: 12 weeks	275 (1 RCT)	⊕⊕⊕⊕ HIGH	-	The mean fatigue Scale of Motor and Cognition - Cognition (0-50) was not reported	MD 2.01 lower (3.38 lower to 0.64 lower)
SF-36 vitality - 12 months Scale from: 0 to 100 follow up: 12 months	131 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 vitality - 12 months was 32.43	MD 5.27 higher (0.99 lower to 11.53 higher)
MSIS-29 total (0-100) - 5.5 months Scale from: 0 to 100 follow up: 5.5 months	146 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean MSIS-29 total (0-100) - 5.5 months was 43.0	MD 1.56 lower (6.45 lower to 3.33 higher)
MSIS-29 total (0-100) - 12 months Scale from: 0 to 100 follow up: 12 months	131 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 total (0-100) - 12 months was 47.2	MD 1 lower (7.28 lower to 5.28 higher)
MSIS-29 physical (0-100) - 12 months Scale from: 0 to 100 follow up: 12 months	131 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 physical (0-100) - 12 months was 50.5	MD 3.1 lower (10.16 lower to 3.96 higher)
MS neuropsychological screening questionnaire (0-60?) Scale from: 0 to 60 follow up: 12 weeks	275 (1 RCT)	⊕⊕⊕⊕ HIGH	-	The mean MS neuropsychological screening questionnaire (0-60?) was not reported	MD 0.27 lower (2.21 lower to 1.67 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (local/standard care)	Risk difference with Fatigue management + CBT
HADS anxiety (0-21) Scale from: 0 to 21 follow up: range 10 weeks to 12 weeks	315 (2 RCTs)	⊕◯◯◯ VERY LOW b,d	-	The mean HADS anxiety (0-21) was 11.65	MD 2.72 lower (7.11 lower to 1.66 higher)
HADS depression (0-21) Scale from: 0 to 21 follow up: range 10 weeks to 12 weeks	315 (2 RCTs)	⊕◯◯◯ VERY LOW b,d	-	The mean HADS depression (0-21) was 8.73	MD 0.76 lower (1.41 lower to 0.11 lower)
Withdrawal due to adverse events (relapse) - 5.5 months follow up: 5.5 months	133 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 9.00 (0.55 to 146.78)	0 per 1,000	33 more per 1,000 (20 fewer to 85 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Downgraded by 1 increment as the majority of the evidence had a follow-up of less than the 3 months minimum specified in the protocol

d. Heterogeneity present that could not be explained by subgrouping strategies and I2 >75%

Multidisciplinary rehabilitation + fatigue self-management vs. control (consultation only)

Table 31: Clinical evidence summary: Multidisciplinary rehabilitation + fatigue self-management vs. control (consultation only)– outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (consultation only)	Risk difference with Multidisciplinary rehabilitation + fatigue self-management
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Total score (0-84) was -0.6	MD 1.8 higher (5 lower to 8.6 higher)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was -0.6	MD 1.7 higher (1.42 lower to 4.82 higher)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was 0.1	MD 0.2 lower (4.16 lower to 3.76 higher)
Modified Fatigue Impact scale - Psychosocial scale (0-8) Scale from: 0 to 8 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Psychosocial scale (0-8) was -0.1	MD 0.2 higher (0.79 lower to 1.19 higher)
Fatigue Severity Scale (1-7) Scale from: 1 to 7 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Scale (1-7) was 0.3	MD 1.9 lower (6.41 lower to 2.61 higher)
MSIS-29 (0-100) - Physical function Scale from: 0 to 100 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) - Physical function was 2.0	MD 1 lower (4.67 lower to 2.67 higher)
MSIS-29 (0-100) - Mental function	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) - Mental function was 1.0	MD 1 lower (4.21 lower to 2.21 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (consultation only)	Risk difference with Multidisciplinary rehabilitation + fatigue self-management
Scale from: 0 to 100 follow up: 3 months					
Functional independence measure (1-7) Scale from: 1 to 7 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean functional independence measure (1-7) was -1.0	MD 3 higher (0.39 higher to 5.61 higher)
CIS20r - Total (0-140) Scale from: 0 to 140 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r - Total (0-140) was 2.2	MD 3 lower (8.08 lower to 2.08 higher)
CIS20r - Subjective fatigue (8-56) Scale from: 8 to 56 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r - Subjective fatigue (8-56) was 1.7	MD 1.1 lower (3.51 lower to 1.31 higher)
CIS20r - Concentration (5-35) Scale from: 5 to 35 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r - Concentration (5-35) was -0.3	MD 0.8 lower (2.87 lower to 1.27 higher)
CIS20r - Motivation (4-28) Scale from: 4 to 28 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r - Motivation (4-28) was 0.3	MD 0.9 lower (2.75 lower to 0.95 higher)
CIS20r - Physical activity (3-21) Scale from: 3 to 21 follow up: 3 months	46 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r - Physical activity (3-21) was 0.6	MD 0.3 lower (1.75 lower to 1.15 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Multidisciplinary rehabilitation + fatigue self-management vs. relaxation

Table 32: Clinical evidence summary: Multidisciplinary rehabilitation + fatigue self-management vs. relaxation – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with Multidisciplinary rehabilitation + fatigue self-management
Modified Fatigue Impact scale - total (0-84) Scale from: 0 to 84 follow up: 4 months	29 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - total (0-84) was 34.5	MD 0 (10.3 lower to 10.3 higher)
SF-36 physical functioning (0-100) Scale from: 0 to 100 follow up: 4 months	29 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 physical functioning (0-100) was 30.0	MD 14.8 higher (0.6 lower to 30.2 higher)
SF-36 fatigue/vitality (0-100) Scale from: 0 to 100 follow up: 4 months	29 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean SF-36 fatigue/vitality (0-100) was 43.5	MD 3 higher (9.7 lower to 15.7 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Multidisciplinary rehabilitation (medical, exercise, counselling and fatigue self-management) vs. no rehabilitation in those treated with methylprednisolone for a relapse

Table 33: Clinical evidence summary: Multidisciplinary rehabilitation (medical, exercise, counselling and fatigue self-management) vs. no rehabilitation in those treated with methylprednisolone for a relapse – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with no rehab in those treated with methylprednisolone for relapse	Risk difference with Multidisciplinary rehab (medical, exercise, counselling + fatigue SM)
Fatigue Severity Scale (9-63) Scale from: 9 to 63 follow up: 3 months	39 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Scale (9-63) was 40.6	MD 4 lower (15.77 lower to 7.77 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Self-management programme vs. control

Table 34: Clinical evidence summary: Self-management programme vs. control – outcomes up to 6 months and >6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Self-management programme
Fatigue severity scale (1-7) Scale from: 1 to 7 follow up: 11 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue severity scale (1-7) was 0.41	MD 5.86 lower (6.08 lower to 5.64 lower)
Fatigue VAS (0-10) Scale from: 0 to 10 follow up: 4 months	142 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean fatigue VAS (0-10) was -0.8	MD 0.5 higher (0.54 lower to 1.54 higher)
MFIS - total (0-84) - 6 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MFIS - total (0-84) - 6 months was 41.7	MD 4.4 lower (9.67 lower to 0.87 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Self-management programme
Scale from: 0 to 84 follow up: 6 months					
MFIS - total (0-84) - 12 month Scale from: 0 to 84 follow up: 12 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MFIS - total (0-84) - 12 month was 43.3	MD 3.1 lower (8.41 lower to 2.21 higher)
MFIS - at least 10-point reduction vs. baseline - 6 months follow up: 6 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 1.74 (0.78 to 3.88)	395 per 1,000	137 more per 1,000 (58 fewer to 322 more)
MFIS - at least 10-point reduction vs. baseline - 12 months follow up: 12 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 1.74 (0.79 to 3.83)	358 per 1,000	134 more per 1,000 (52 fewer to 323 more)
SF-8 physical domain (0-100) - 6 months Scale from: 0 to 100 follow up: 6 months	145 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-8 physical domain (0-100) - 6 months was 40.4	MD 0.1 lower (3.17 lower to 2.97 higher)
SF-8 physical domain (0-100) - 12 month Scale from: 0 to 100 follow up: 12 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a	-	The mean SF-8 physical domain (0-100) - 12 month was 40.3	MD 1.7 lower (4.59 lower to 1.19 higher)
SF-8 mental health domain (0-100) - 6 months Scale from: 0 to 100 follow up: 6 months	145 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-8 mental health domain (0-100) - 6 months was 47.0	MD 1.2 higher (1.97 lower to 4.37 higher)
SF-8 mental health domain (0-100) - 12 month Scale from: 0 to 100 follow up: 12 months	145 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-8 mental health domain (0-100) - 12 month was 47.2	MD 0.5 higher (2.63 lower to 3.63 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Self-management programme
MSIS-29 (0-100) - Physical Scale from: 0 to 100 follow up: 4 months	142 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 (0-100) - Physical was 3.3	MD 6.6 lower (12.44 lower to 0.76 lower)
MSIS-29 (0-100) - Psychological Scale from: 0 to 100 follow up: 4 months	142 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MSIS-29 (0-100) - Psychological was -2.3	MD 3.6 lower (12.64 lower to 5.44 higher)
HADS - anxiety (0-21) Scale from: 0 to 21 follow up: 4 months	142 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean HADS - anxiety (0-21) was -0.2	MD 0.5 lower (1.82 lower to 0.82 higher)
HADS - depression (0-21) Scale from: 0 to 21 follow up: 4 months	142 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean HADS - depression (0-21) was 0.0	MD 0.9 lower (1.85 lower to 0.05 higher)
PHQ-9 (depression; 0-27) - 6 months Scale from: 0 to 27 follow up: 6 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean PHQ-9 (depression; 0-27) - 6 months was 6.7	MD 1 lower (2.47 lower to 0.47 higher)
PHQ-9 (depression; 0-27) - 12 months Scale from: 0 to 27 follow up: 12 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean PHQ-9 (depression; 0-27) - 12 months was 7.3	MD 1 lower (2.5 lower to 0.5 higher)
PHQ-9 (depression) - at least 50% reduction vs. baseline - 6 months follow up: 6 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 1.41 (0.45 to 4.42)	123 per 1,000	42 more per 1,000 (64 fewer to 260 more)
PHQ-9 (depression) - at least 50% reduction vs. baseline - 12 months follow up: 12 months	145 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 1.00 (0.31 to 3.23)	173 per 1,000	0 fewer per 1,000 (112 fewer to 230 more)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Self-management programme
Adverse events leading to withdrawal follow up: 11 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,d	RD 0.00 (-0.06 to 0.06)	0 per 1,000	0 fewer per 1,000 (0 fewer to 0 fewer)
Serious adverse events - 6 months follow up: 6 months	141 (1 RCT)	⊕◯◯◯ VERY LOW a,nd	RD 0.00 (-0.03 to 0.03)	0 per 1,000	0 fewer per 1,000 (0 fewer to 0 fewer)
Serious adverse events - 12 months follow up: 12 months	140 (1 RCT)	⊕◯◯◯ VERY LOW a,d	RD 0.00 (-0.03 to 0.03)	0 per 1,000	0 fewer per 1,000 (0 fewer to 0 fewer)
Treatment adherence - attending all 8 sessions	163 (1 RCT)	⊕⊕⊕◯ MODERATE c	OR 0.49 (0.21 to 1.12)	Moderate	
				875 per 1,000	101 fewer per 1,000 (280 fewer to 12 more)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up of less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Imprecision assessed based on sample size as zero events in both arms of a single study. Downgraded by 2 increments if sample size was <70 and 1 increment if sample size was >70 and <350

Self-management programme + exercise vs. control (waitlist)

Table 35: Clinical evidence summary: Self-management programme + exercise vs. control (waitlist) – outcomes up to 6 months

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist)	Risk difference with Self-management + exercise
WEIMuS fatigue scale - Total (0-68) Scale from: 0 to 68 follow up: 6 weeks	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean wEIMuS fatigue scale - Total (0-68) was 18.8	MD 3.3 higher (9.72 lower to 16.32 higher)
WEIMuS fatigue scale - Mental (0-36) Scale from: 0 to 36 follow up: 6 weeks	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean wEIMuS fatigue scale - Mental (0-36) was 7.5	MD 2 higher (4.1 lower to 8.1 higher)
WEIMuS fatigue scale - Physical (0-32) Scale from: 0 to 32 follow up: 6 weeks	14 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean wEIMuS fatigue scale - Physical (0-32) was 11.3	MD 1.3 higher (7.55 lower to 10.15 higher)
MusiQoL score (0-100) Scale from: 0 to 100 follow up: 6 weeks	14 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean musiQoL score (0-100) was 74.6	MD 2.6 higher (9.53 lower to 14.73 higher
Adverse events follow up: 6 weeks	14 (1 RCT)	⊕◯◯◯ VERY LOW a,d	RD 0.00 (-0.24 to 0.24)	0 per 1,000	0 fewer per 1,000 (240 fewer 240 more)

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

b. Downgraded by 1 increment as the majority of the evidence has a follow-up less than the 3 months specified in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Imprecision assessed using sample size as zero events in both arms of a single study. Downgraded by 2 increments as sample size <70.

CBT vs. control

Table 36: Clinical evidence summary: CBT vs. control – up to 6 months and >6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with CBT
CIS20r fatigue (8-56) - 16 weeks Scale from: 8 to 56 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean cIS20r fatigue (8-56) - 16 weeks was 40.3	MD 6.3 lower (10.74 lower to 1.86 lower)
CIS20r fatigue (8-56) - 52 weeks Scale from: 8 to 56 follow up: 52 weeks	74 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cIS20r fatigue (8-56) - 52 weeks was 39.5	MD 0.6 lower (4.86 lower to 3.66 higher)
CIS20r fatigue - at least 8-point improvement - 16 weeks follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	RR 2.19 (1.17 to 4.11)	Moderate	
				257 per 1,000	306 more per 1,000 (44 more to 800 more)
FSS score (1-7) - 16 weeks Scale from: 1 to 7 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean FSS score (1-7) - 16 weeks was 5.2	MD 0.7 lower (1.12 lower to 0.28 lower)
FSS score (1-7) - 52 weeks Scale from: 1 to 7 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean FSS score (1-7) - 52 weeks was 5.1	MD 0.1 lower (0.51 lower to 0.31 higher)
MFIS total (0-84) - 16 weeks Scale from: 0 to 84 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS total (0-84) - 16 weeks was 41.2	MD 2.5 lower (8.98 lower to 3.98 higher)
MFIS total (0-84) - 52 weeks Scale from: 0 to 84 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS total (0-84) - 52 weeks was 39.1	MD 3.4 higher (2.56 lower to 9.36 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with CBT
MFIS physical subscale (0-36) - 16 weeks Scale from: 0 to 36 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS physical subscale (0-36) - 16 weeks was 19.6	MD 1.8 lower (4.9 lower to 1.3 higher)
MFIS physical subscale (0-36) - 52 weeks Scale from: 0 to 36 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS physical subscale (0-36) - 52 weeks was 18.1	MD 2.2 higher (0.76 lower to 5.16 higher)
MFIS cognitive subscale (0-40) - 16 weeks Scale from: 0 to 40 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS cognitive subscale (0-40) - 16 weeks was 18.1	MD 0.7 lower (4.37 lower to 2.97 higher)
MFIS cognitive subscale (0-40) - 52 weeks Scale from: 0 to 40 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,	-	The mean MFIS cognitive subscale (0-40) - 52 weeks was 17.6	MD 1 higher (2.28 lower to 4.28 higher)
MFIS psychosocial (0-8) - 16 weeks Scale from: 0 to 8 follow up: 16 weeks	74 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean MFIS psychosocial (0-8) - 16 weeks was 3.4	MD 0 (0.71 lower to 0.71 higher)
MFIS psychosocial (0-8) - 52 weeks Scale from: 0 to 8 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MFIS psychosocial (0-8) - 52 weeks was 3.4	MD 0.2 higher (0.53 lower to 0.93 higher)
Piper Fatigue Scale (0-10?) Scale from: 0 to 10 follow up: 4 months	140 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean piper Fatigue Scale (0-10?) was 6.6	MD 2.27 lower (3.9 lower to 0.64 lower)
DASS-21 - anxiety subscale (0-21)	140 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean DASS-21 - anxiety subscale (0-21) was 16.08	MD 1.15 lower (2.04 lower to 0.26 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with CBT
Scale from: 0 to 21 follow up: 4 months					
DASS-21 - depression subscale (0-21) Scale from: 0 to 21 follow up: 4 months	140 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean DASS-21 - depression subscale (0-21) was 14.06	MD 1.4 lower (2.16 lower to 0.64 lower)
SF-36 vitality (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 vitality (0-100) - 16 weeks was 45.4	MD 7.8 higher (1.04 higher to 14.56 higher)
SF-36 vitality (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SF-36 vitality (0-100) - 52 weeks was 46.2	MD 0.7 higher (7 lower to 8.4 higher)
SF-36 physical functioning (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 physical functioning (0-100) - 16 weeks was 61.3	MD 3.1 lower (13.39 lower to 7.19 higher)
SF-36 physical functioning (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 physical functioning (0-100) - 52 weeks was 60.3	MD 4.4 lower (14.5 lower to 5.7 higher)
SF-36 physical role functioning (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 physical role functioning (0-100) - 16 weeks was 32.4	MD 15.6 higher (1.63 lower to 32.83 higher)
SF-36 physical role functioning (0-100) - 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 physical role functioning (0-100) - 52 weeks was 38.5	MD 9.7 lower (27.25 lower to 7.85 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with CBT
Scale from: 0 to 100 follow up: 52 weeks					
SF-36 emotional role functioning (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean SF-36 emotional role functioning (0-100) - 16 weeks was 72.2	MD 2.6 higher (14.73 lower to 19.93 higher)
SF-36 emotional role functioning (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean SF-36 emotional role functioning (0-100) - 52 weeks was 71.2	MD 0.6 higher (17.49 lower to 18.69 higher)
SF-36 social functioning (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 social functioning (0-100) - 16 weeks was 61.7	MD 7.2 higher (1.89 lower to 16.29 higher)
SF-36 social functioning (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 social functioning (0-100) - 52 weeks was 73.6	MD 5.9 lower (14.96 lower to 3.16 higher)
SF-36 mental health (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean SF-36 mental health (0-100) - 16 weeks was 71.7	MD 0 (6.03 lower to 6.03 higher)
SF-36 mental health (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,	-	The mean SF-36 mental health (0-100) - 52 weeks was 71.1	MD 2.8 lower (10 lower to 4.4 higher)
SF-36 general health (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean SF-36 general health (0-100) - 16 weeks was 48.2	MD 1.7 lower (8.45 lower to 5.05 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with CBT
SF-36 general health (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⨁⨁◯◯ LOW a,	-	The mean SF-36 general health (0-100) - 52 weeks was 50.3	MD 1.7 lower (8.68 lower to 5.28 higher)
SF-36 bodily pain (0-100) - 16 weeks Scale from: 0 to 100 follow up: 16 weeks	74 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean SF-36 bodily pain (0-100) - 16 weeks was 68.6	MD 4.7 higher (4.68 lower to 14.08 higher)
SF-36 bodily pain (0-100) - 52 weeks Scale from: 0 to 100 follow up: 52 weeks	74 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean SF-36 bodily pain (0-100) - 52 weeks was 70.5	MD 0.1 lower (10.78 lower to 10.58 higher)
CIS20r concentration (5-35) - 16 weeks Scale from: 5 to 35 follow up: 16 weeks	74 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean cIS20r concentration (5-35) - 16 weeks was 21.3	MD 1.2 lower (4.6 lower to 2.2 higher)
CIS20r concentration (5-35) - 52 weeks Scale from: 5 to 35 follow up: 52 weeks	74 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean cIS20r concentration (5-35) - 52 weeks was 20.4	MD 0.4 higher (3.04 lower to 3.84 higher)
Serious adverse events - 16 weeks follow up: 16 weeks	74 (1 RCT)	⨁◯◯◯ VERY LOW a,b	RR 0.45 (0.04 to 4.74)	Moderate	
				57 per 1,000	31 fewer per 1,000 (55 fewer to 214 more)
Serious adverse events - 52 weeks follow up: 52 weeks	74 (1 RCT)	⨁◯◯◯ VERY LOW a,b	RR 1.20 (0.29 to 4.98)	Moderate	
				86 per 1,000	17 more per 1,000 (61 fewer to 341 more)

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

2 b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of
3 this evidence review.

CBT vs. relaxation – up to 6 months and >6 months outcomes

Table 37: Clinical evidence summary: CBT vs. relaxation – up to 6 months and >6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with CBT
Chalder fatigue scale (0-33) - 5 months Scale from: 0 to 33 follow up: 5 months	72 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean chalder fatigue scale (0-33) - 5 months was 11.11	MD 2.12 lower (4.41 lower to 0.17 higher)
Chalder fatigue scale (0-33) - 8 months Scale from: 0 to 33 follow up: 8 months	72 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean chalder fatigue scale (0-33) - 8 months was 12.49	MD 2.12 lower (4.82 lower to 0.58 higher)
Fatigue-related impairment (work and social adjustment scale; 0-40) - 5 months Scale from: 0 to 40 follow up: 5 months	72 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean fatigue-related impairment (work and social adjustment scale; 0-40) - 5 months was 19.24	MD 5.86 lower (9.99 lower to 1.73 lower)
Fatigue-related impairment (work and social adjustment scale; 0-40) - 8 months Scale from: 0 to 40 follow up: 8 months	72 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean fatigue-related impairment (work and social adjustment scale; 0-40) - 8 months was 20.16	MD 5.19 lower (9.9 lower to 0.48 lower)
HADS - depression (0-21) - 5 months Scale from: 0 to 21 follow up: 5 months	72 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean HADS - depression (0-21) - 5 months was 5.13	MD 1.51 lower (2.87 lower to 0.15 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with CBT
HADS - depression (0-21) - 8 months Scale from: 0 to 21 follow up: 8 months	72 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean HADS - depression (0-21) - 8 months was 5.05	MD 1.08 lower (2.56 lower to 0.4 higher)
HADS - anxiety (0-21) - 5 months Scale from: 0 to 21 follow up: 5 months	72 (1 RCT)	⨁⨁⨁◯ MODERATE a	-	The mean HADS - anxiety (0-21) - 5 months was 5.81	MD 0.21 lower (1.71 lower to 1.29 higher)
HADS - anxiety (0-21) - 8 months Scale from: 0 to 21 follow up: 8 months	72 (1 RCT)	⨁⨁⨁◯ MODERATE a	-	The mean HADS - anxiety (0-21) - 8 months was 5.81	MD 0.19 higher (1.48 lower to 1.86 higher)
Acceptability - usefulness end of treatment (0-4)	72 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean acceptability - usefulness end of treatment (0-4) was 0.97	MD 0.21 lower (0.63 lower to 0.21 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Motivational interviewing vs. control

Table 38: Clinical evidence summary: Motivational interviewing vs. control – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Motivational interviewing
MFIS - total (0-84) Scale from: 0 to 84 follow up: 9 weeks	60 (1 RCT)	⨁◯◯◯ VERY LOW a,b	-	The mean MFIS - total (0-84) was 62.13	MD 20.38 lower (26.11 lower to 14.65 lower)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum specified in the protocol

Resistance + aerobic exercise + CBT vs. control (waitlist)

Table 39: Clinical evidence summary: Resistance + aerobic + CBT vs. control (waitlist) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist), up to 6 months	Risk difference with Resistance + aerobic exercise + CBT
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 3 months	107 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean modified Fatigue Impact scale - Total score (0-84) was 43.2	MD 7.4 lower (14.13 lower to 0.67 lower)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 3 months	107 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was 21.2	MD 3.3 lower (6.56 lower to 0.04 lower)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 3 months	107 (1 RCT)	⨁⨁◯◯ LOW a,b	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was 17.7	MD 2.8 lower (6.19 lower to 0.59 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist), up to 6 months	Risk difference with Resistance + aerobic exercise + CBT
Modified Fatigue Impact scale - Psychosocial scale (0-8) Scale from: 0 to 8 follow up: 3 months	107 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean modified Fatigue Impact scale - Psychosocial scale (0-8) was 4.2	MD 1.3 lower (2.12 lower to 0.48 lower)
MSQOL-54 score (0-100) Scale from: 0 to 100 follow up: 3 months	107 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MSQOL-54 score (0-100) was 60.6	MD 7.5 higher (0.01 higher to 14.99 higher)
EQ-5D follow up: 3 months	107 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean EQ-5D was 0.68	MD 0.06 higher (0.03 lower to 0.15 higher)
EDSS (0-10) Scale from: 0 to 10 follow up: 3 months	107 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean EDSS (0-10) was 3.7	MD 0.2 lower (0.73 lower to 0.33 higher)
Cognitive - PASAT follow up: 3 months	107 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean cognitive - PASAT was 46.0	MD 4.1 lower (9.55 lower to 1.35 higher)
Adverse events (MS relapse) leading to withdrawal follow up: 3 months	109 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 0.98 (0.06 to 15.30)	Moderate	
				19 per 1,000	0 fewer per 1,000 (17 fewer to 265 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Table 40: Clinical evidence summary: Resistance + aerobic + CBT vs. control (waitlist) – >6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist), >6 months	Risk difference with Resistance + aerobic exercise + CBT
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 9 months	99 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean modified Fatigue Impact scale - Total score (0-84) was 41.3	MD 1.7 lower (8.69 lower to 5.29 higher)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 84 follow up: 9 months	99 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was 20.7	MD 0.6 lower (3.82 lower to 2.62 higher)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 9 months	99 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was 16.7	MD 0.7 lower (4.33 lower to 2.93 higher)
Modified Fatigue Impact scale - Psychosocial scale (0-8) Scale from: 0 to 8 follow up: 9 months	99 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean modified Fatigue Impact scale - Psychosocial scale (0-8) was 4.0	MD 0.5 lower (1.35 lower to 0.35 higher)
MSQOL-54 score (0-100) Scale from: 0 to 100 follow up: 9 months	99 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean MSQOL-54 score (0-100) was 60.4	MD 5.5 higher (2.62 lower to 13.62 higher)
EQ-5D follow up: 9 months	99 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean EQ-5D was 0.73	MD 0.01 higher (0.09 lower to 0.1 higher)
EDSS (0-10) Scale from: 0 to 10 follow up: 9 months	99 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean EDSS (0-10) was 3.9	MD 0.2 lower (0.83 lower to 0.43 higher)
Cognitive - PASAT follow up: 9 months	99 (1 RCT)	⊕⊕⊕◯ MODERATE a	-	The mean cognitive - PASAT was 46.9	MD 0.5 higher (4.26 lower to 5.26 higher)
				Moderate	

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist), >6 months	Risk difference with Resistance + aerobic exercise + CBT
Adverse events (relapse) follow up: 9 months	120 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 0.64 (0.30 to 1.37)	233 per 1,000	84 fewer per 1,000 (163 fewer to 86 more)
Adverse events (MS relapse) leading to withdrawal follow up: 9 months	102 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 2.00 (0.19 to 21.37)	Moderate	
				20 per 1,000	20 more per 1,000 (16 fewer to 399 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Diet vs. control

Table 41: Clinical evidence summary: Diet vs. control – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care/no dietary intervention)	Risk difference with Diet
Fatigue Severity Scale (1-9) Scale from: 1 to 9 follow up: 3 months	17 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Scale (1-7) was 0.2	MD 1.6 lower (3.07 lower to 0.13 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care/no dietary intervention)	Risk difference with Diet
>1-point reduction on FSS follow up: 3 months	17 (1 RCT)	⊕⊕◯◯ LOW a	OR 13.67 (1.55 to 120.73)	0 per 1,000	500 more per 1,000 (147 more to 854 more)
Modified Fatigue Impact Scale - total score Scale from: 0 to 84 follow up: 6 months	147 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - total score was 75.9	MD 12 lower (16.77 lower to 7.23 lower)
Modified Fatigue Impact Scale - physical subscale Scale from: 0 to 36 follow up: 6 months	147 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - physical subscale was 33.7	MD 5.2 lower (8.27 lower to 2.13 lower)
Modified Fatigue Impact Scale - cognitive sub score Scale from: 0 to 40 follow up: 6 months	147 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - cognitive sub score was 36.1	MD 5.9 lower (8.46 lower to 3.34 lower)
Modified Fatigue Impact Scale - psychosocial sub score Scale from: 0 to 8 follow up: 6 months	147 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - psychosocial sub score was 6.1	MD 0.9 lower (1.87 lower to 0.07 higher)
Neurological fatigue index - MS (scale unclear but likely 0-30)	36 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean neurological fatigue index - MS (scale unclear but likely 0-30) was not reported	MD 4.55 lower (7.65 lower to 1.45 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care/no dietary intervention)	Risk difference with Diet
Scale from: 0 to 30 follow up: 6 months					
At least 5-point reduction on MSQOL-54 mental health composite follow up: 3 months	17 (1 RCT)	⊕⊕◯◯ LOW a	OR 31.57 (1.37 to 725.23)	Moderate	
				333 per 1,000	607 more per 1,000 (73 more to 664 more)
Improvement (no threshold) on MSQOL-54 physical health composite follow up: 3 months	17 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 14.00 (1.14 to 172.64)	Moderate	
				333 per 1,000	542 more per 1,000 (30 more to 655 more)
MSIS-29 (0-100) Scale from: 0 to 100 follow up: 6 months	36 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSIS-29 (0-100) was not reported	MD 7.36 lower (16.32 lower to 1.6 higher)
EDSS score (0-10) Scale from: 0 to 10 follow up: 6 months	183 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c	-	The mean EDSS score (0-10) ranged from 2.1 – unclear	MD 0.59 lower (1.12 lower to 0.06 lower)
Adverse events follow up: 3-6 months	167 (2 RCTs)	⊕◯◯◯ VERY LOW a,d	RD -0.01 (-0.05 to 0.04)	Moderate	
				91 per 1,000	10 fewer per 1,000 (50 fewer to 40 more)
Adverse events leading to withdrawal follow up: 3 months	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 0.61 (0.07 to 5.70)	Moderate	
				182 per 1,000	71 fewer per 1,000 (169 fewer to 854 more)
Adherence to intervention or control	19 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 0.81 (0.57 to 1.15)[e]	Moderate	
				1,000 per 1,000	190 fewer per 1,000 (430 fewer to 150 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Downgraded by 1 increment as heterogeneity is present that cannot be explained by subgroup analyses, based on I2 value >50%

d Imprecision assessed by calculating OIS and assessing power, as zero events in both arms of some but not all studies. Downgraded by 2 increments as power <80%.

e. Presented as RR despite event rate >50%, as using OR would not allow absolute effect to be calculated given the risk in the control group is 100%

Diet (individualised) vs. standard healthy diet recommendations

Table 42: Clinical evidence summary: Diet (individualised) vs. standard healthy diet recommendations – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard healthy diet recommendations	Risk difference with Diet (individualised)
Modified Fatigue Impact scale - Total score (0-84) Scale from: 0 to 84 follow up: 12 weeks	100 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean modified Fatigue Impact scale - Total score (0-84) was 47.92	MD 0.7 lower (5.34 lower to 3.94 higher)
Modified Fatigue Impact scale - Physical subscale (0-36) Scale from: 0 to 36 follow up: 12 weeks	100 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Physical subscale (0-36) was 22.98	MD 0.8 lower (2.92 lower to 1.32 higher)
Modified Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 12 weeks	100 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean modified Fatigue Impact scale - Cognitive subscale (0-40) was 22.72	MD 0.48 lower (3.62 lower to 2.66 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard healthy diet recommendations	Risk difference with Diet (individualised)
Modified Fatigue Impact scale - Psychosocial scale (0-8) Scale from: 0 to 8 follow up: 12 weeks	100 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - Psychosocial scale (0-8) was 2.28	MD 0.38 higher (0.25 lower to 1.01 higher) c
MSQOL-54 (0-100) - Physical composite Scale from: 0 to 100 follow up: 12 weeks	100 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSQOL-54 (0-100) - Physical composite was 46.57	MD 2.93 higher (6.32 lower to 12.18 higher) d
MSQOL-54 (0-100) - Mental health composite Scale from: 0 to 100 follow up: 12 weeks	100 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean MSQOL-54 (0-100) - Mental health composite was 64.43	MD 5.91 lower (16.21 lower to 4.39 higher) e
Adverse events leading to withdrawal (relapse) follow up: 12 weeks	103 (1 RCT)	⊕◯◯◯ VERY LOW a,b	RR 1.96 (0.18 to 20.97)	Moderate	
				20 per 1,000	19 more per 1,000 (16 fewer to 391 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Note there is a larger baseline difference between groups for this outcome - scores improved from baseline in the intervention group and worsened slightly in the control group.

d. Note differences at baseline may mislead interpretation - results changed very little in both groups from baseline but were higher at baseline in the intervention group

e. Note differences at baseline may mislead interpretation - results changed very little in both groups from baseline but were lower at baseline in the intervention group

Table 43: Clinical evidence summary: Diet (individualised) vs. standard healthy diet recommendations – >6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard healthy diet recommendations > 6 months	Risk difference with Diet (individualised)
Modified Fatigue Impact scale Scale from: 0 to 84 follow up: 1 years	72 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean modified Fatigue Impact scale was 37.98	MD 4.05 lower (5.38 lower to 2.72 lower)
PASAT - cognitive follow up: 1 years	56 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean PASAT - cognitive was 42.37	MD 0.31 higher (3.36 lower to 3.98 higher)
SDMT - cognitive follow up: 1 years	56 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean SDMT - cognitive was 45.89	MD 2.52 lower (6.03 lower to 0.99 higher)
California Verbal Learning Test II - delayed recall - cognitive follow up: 1 years	56 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test II - delayed recall - cognitive was 10.12	MD 1.38 higher (0.21 lower to 2.97 higher)
California Verbal Learning Test II - total learning - cognitive follow up: 1 years	56 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean california Verbal Learning Test II - total learning - cognitive was 50.94	MD 0.15 lower (5.15 lower to 4.85 higher)
Judgement of line orientation test - cognitive follow up: 1 years	56 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean judgement of line orientation test - cognitive was 19.57	MD 0.95 lower (2.72 lower to 0.82 higher)
Brief Visuospatial Memory Test-Revised - cognitive follow up: 1 years	56 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean brief Visuospatial Memory Test-Revised - cognitive was 23.73	MD 3.17 lower (5.74 lower to 0.6 lower)
North American Adult Reading Test - cognitive follow up: 1 years	56 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean north American Adult Reading Test - cognitive was 40.95	MD 0.57 higher (1.15 lower to 2.29 higher)
Controlled Oral Word Association Test -	56 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean controlled Oral Word Association Test - cognitive was 8.63	MD 0.19 higher (0.85 lower to 1.23 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with standard healthy diet recommendations > 6 months	Risk difference with Diet (individualised)
cognitive follow up: 1 years					
Delis-Kaplan Executive Function System description- cognitive follow up: 1 years	56 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean delis-Kaplan Executive Function System description- cognitive was 11.69	MD 0.72 lower (2.72 lower to 1.28 higher)
Delis-Kaplan Executive Function System total scoring - cognitive follow up: 1 years	56 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean delis-Kaplan Executive Function System total scoring - cognitive was 3.39	MD 0.47 lower (1.04 lower to 0.1 higher)
Adherence to intervention (scale 0-14) Scale from: 0 to 14 follow up: 1 years	72 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean adherence to intervention (scale 0-14) was 7.0	MD 2.45 higher (1.29 higher to 3.61 higher)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Wahls diet (modified Palaeolithic elimination diet) vs. Swank diet (low-saturated fat diet)

Table 44: Clinical evidence summary: Mindfulness vs. control (usual care) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Swank diet (low-saturated fat diet), up to 6 months	Risk difference with Wahls diet (modified Palaeolithic elimination diet)
Fatigue Severity Score (scale 1-9) Scale from: 1 to 9 follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Score (scale 1-9) was 4.32	MD 0.45 lower (1.17 lower to 0.27 higher)
Modified Fatigue Impact Scale - Total score (0-84) Scale from: 0 to 84 follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - Total score (0-84) was 30.2	MD 3.7 lower (11.52 lower to 4.12 higher)
Modified Fatigue Impact Scale - Physical sub score (0-36) Scale from: 0 to 36 follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - Physical sub score (0-36) was 14.7	MD 3.4 lower (6.98 lower to 0.18 higher)
Modified Fatigue Impact Scale - Cognitive sub score (0-40) Scale from: 0 to 40 follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - Cognitive sub score (0-40) was 13.5	MD 0.7 lower (5.03 lower to 3.63 higher)
Modified Fatigue Impact Scale - Psychosocial sub score (0-8) Scale from: 0 to 8 follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact Scale - Psychosocial sub score (0-8) was 3.03	MD 0.66 lower (1.62 lower to 0.3 higher)
MSQoL-54 (0-100) - Physical composite Scale from: 0 to 100 follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean mSQoL-54 (0-100) - Physical composite was 64.9	MD 6.1 higher (2.7 lower to 14.9 higher)
MSQoL-54 (0-100) - Mental composite	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean mSQoL-54 (0-100) - Mental composite was 73.6	MD 2.7 higher (6.24 lower to 11.64 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with Swank diet (low-saturated fat diet), up to 6 months	Risk difference with Wahls diet (modified Palaeolithic elimination diet)
Scale from: 0 to 100 follow up: 6 months					
Serious adverse events follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RD 0.00 (-0.05 to 0.05)	Moderate	
				0 per 1,000	0 fewer per 1,000 (50 fewer to 50 more)
Adherence to diet follow up: 6 months	72 (1 RCT)	⊕◯◯◯ VERY LOW a,b	OR 0.67 (0.22 to 2.06)	Moderate	
				811 per 1,000	69 fewer per 1,000 (326 fewer to 87 more)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

c. Imprecision assessed based on sample size as zero events in both arms of a single study. Downgraded by 1 increment as sample size >70 and <350

Mindfulness vs. control (usual care)

Table 45: Clinical evidence summary: Mindfulness vs. control (usual care) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care)	Risk difference with Mindfulness
Modified Fatigue Impact scale - total (0-84) Scale from: 0 to 84 follow up: 6 months	150 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean modified Fatigue Impact scale - total (0-84) was 0.09	MD 6.03 lower (10.08 lower to 1.98 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care)	Risk difference with Mindfulness
HAQUAMS (1-5) Scale from: 1 to 5 follow up: 6 months	150 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean HAQUAMS (1-5) was 0.05	MD 0.18 lower (0.35 lower to 0.01 lower)
CES-D depression (0-60) Scale from: 0 to 60 follow up: 6 months	150 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean CES-D depression (0-60) was -0.86	MD 3.77 lower (6.63 lower to 0.91 lower)
STAI anxiety (20-80) Scale from: 20 to 80 follow up: 6 months	150 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean STAI anxiety (20-80) was -0.13	MD 3.55 lower (6.09 lower to 1.01 lower)

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

b. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Yoga vs. control

Table 46: Clinical evidence summary: yoga vs. control – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Yoga
Fatigue severity scale (1-7) Scale from: 1 to 7 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue severity scale (1-7) was 4.23	MD 1.79 lower (2.89 lower to 0.69 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Yoga
Fatigue Severity Scale (9-63) follow up: 8 weeks	36 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean fatigue Severity Scale (9-63) was 41.22	MD 25 lower (32.66 lower to 17.34 lower)
MFIS - total (0-84) Scale from: 0 to 84 follow up: 12 weeks	112 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MFIS - total (0-84) was -1.1	MD 4.7 lower (9.4 lower to 0)
MFIS - physical (0-36) Scale from: 0 to 36 follow up: 12 weeks	112 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean MFIS - physical (0-36) was 0.4	MD 2.5 lower (4.55 lower to 0.45 lower)
MFIS - cognitive (0-40) Scale from: 0 to 40 follow up: 12 weeks	112 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MFIS - cognitive (0-40) was -0.51	MD 0.45 lower (1.92 lower to 1.02 higher)
Multidimensional Fatigue Inventory - general fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - general fatigue (4-20) was 14.9	MD 1.9 lower (3.69 lower to 0.11 lower)
Multidimensional Fatigue Inventory - physical fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - physical fatigue (4-20) was 13.9	MD 1.8 lower (4.5 lower to 0.9 higher)
Multidimensional Fatigue Inventory - reduced activity (4-20) Scale from: 4 to 20 follow up: 6 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - reduced activity (4-20) was 11.5	MD 0.3 lower (2.91 lower to 2.31 higher)
Multidimensional Fatigue Inventory - reduced motivation (4-20)	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - reduced motivation (4-20) was 9.8	MD 0.6 lower (2.42 lower to 1.22 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Yoga
Scale from: 4 to 20 follow up: 6 months					
Multidimensional Fatigue Inventory - mental fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - mental fatigue (4-20) was 11.2	MD 0.5 lower (2.89 lower to 1.89 higher)
Rhoten Fatigue Scale (0-10) Scale from: 0 to 10 follow up: 12 weeks	41 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean rhoten Fatigue Scale (0-10) was 3.55	MD 0.2 lower (0.83 lower to 0.43 higher)
MSQOL-54 physical health composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MSQOL-54 physical health composite (0-100) was 66.64	MD 0.94 lower (11.15 lower to 9.27 higher)
MSQOL-54 mental health composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MSQOL-54 mental health composite (0-100) was 65.54	MD 8.76 higher (4.18 lower to 21.7 higher)
MSQOL-54 change in health domain (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MSQOL-54 change in health domain (0-100) was 52.5	MD 0.23 lower (22.25 lower to 21.79 higher)
MSIS-29 physical component (0-100) Scale from: 0 to 100 follow up: 12 weeks	112 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean MSIS-29 physical component (0-100) was 0.3	MD 4.3 lower (9.72 lower to 1.12 higher) Clinically important benefit/No difference?
SF-36 physical functioning (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	83 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean SF-36 physical functioning (0-100) was 48.59	MD 11 higher (5.4 higher to 16.59 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Yoga
SF-36 emotional limitations (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	83 (2 RCTs)	⊕◯◯◯ VERY LOW ac,d	-	The mean SF-36 emotional limitations (0-100) was 59.97	MD 0.88 higher (25.13 lower to 26.88 higher)
SF-36 physical role limitations (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	83 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 physical role limitations (0-100) was 52.48	MD 6.5 lower (13.21 lower to 0.22 higher)
SF-36 energy/vitality (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	83 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 energy/vitality (0-100) was 39.93	MD 10.7 higher (5.26 higher to 16.13 higher)
SF-36 mental health (0-100) Scale from: 0 to 100 follow up: 12 weeks	41 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 mental health (0-100) was 50.44	MD 10.1 higher (1.25 higher to 18.95 higher)
SF-36 social functioning (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	83 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 social functioning (0-100) was 55.38	MD 3.5 higher (12.79 lower to 19.78 higher)
SF-36 body pain (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	83 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 body pain (0-100) was 62.14	MD 9.27 lower (26.67 lower to 8.12 higher)
SF-36 general health (0-100) Scale from: 0 to 100	83 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 general health (0-100) was 48.87	MD 7.79 higher (2.93 higher to 12.65 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control	Risk difference with Yoga
follow up: range 12 weeks to 6 months					
SF-36 health transition (0-100) Scale from: 0 to 100 follow up: 6 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 health transition (0-100) was 48.6	MD 12.9 lower (25.28 lower to 0.52 lower)
Cognitive - Stroop colour word interference (attention/concentration) follow up: 6 months	42 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean cognitive - Stroop colour word interference (attention/concentration) was 8.1	MD 0.4 higher (2.29 lower to 3.09 higher)
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	57 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean beck Depression Inventory (0-63) was 18.18	MD 9.43 lower (23.95 lower to 5.08 higher)
Beck Anxiety Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Anxiety Inventory (0-63) was 8.2	MD 1.75 lower (6.8 lower to 3.3 higher)
Adverse events leading to withdrawal follow up: 12 weeks	122 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RR 0.22 (0.05 to 0.99)	Moderate	
				140 per 1,000	110 fewer per 1,000 (133 fewer to 1 fewer)
Adverse events (MS exacerbation) follow up: 6 months	43 (1 RCT)	⊕◯◯◯ VERY LOW a,c	OR 6.49 (0.13 to 329.99)	0 per 1,000	44 more per 1,000 (73 fewer to 160 more)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Heterogeneity that cannot be explained by subgrouping analyses and I2 >75%

Yoga vs. aerobic exercise

Table 47: Clinical evidence summary: yoga vs. aerobic exercise – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Yoga
Fatigue severity scale (1-7) Scale from: 1 to 7 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue severity scale (1-7) was 1.9	MD 0.54 higher (0.46 lower to 1.54 higher)
Multidimensional Fatigue Inventory - general fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - general fatigue (4-20) was 12.1	MD 0.9 higher (0.96 lower to 2.76 higher)
Multidimensional Fatigue Inventory - physical fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - physical fatigue (4-20) was 10.8	MD 1.3 higher (1.43 lower to 4.03 higher)
Multidimensional Fatigue Inventory - reduced activity (4-20) Scale from: 4 to 20 follow up: 6 months	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - reduced activity (4-20) was 9.9	MD 1.3 higher (1.31 lower to 3.91 higher)
Multidimensional Fatigue Inventory - reduced motivation (4-20) Scale from: 4 to 20 follow up: 6 months	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - reduced motivation (4-20) was 7.7	MD 1.5 higher (0.63 lower to 3.63 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Yoga
Multidimensional Fatigue Inventory - mental fatigue (4-20) Scale from: 4 to 20 follow up: 6 months	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean multidimensional Fatigue Inventory - mental fatigue (4-20) was 7.8	MD 2.9 higher (0.12 higher to 5.68 higher)
Rhoten Fatigue Scale (0-10) Scale from: 0 to 10 follow up: 12 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean rhoten Fatigue Scale (0-10) was 2.55	MD 0.8 higher (0.26 higher to 1.34 higher)
MSQOL-54 physical health composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MSQOL-54 physical health composite (0-100) was 71.19	MD 5.49 lower (14.73 lower to 3.75 higher)
MSQOL-54 mental health composite (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MSQOL-54 mental health composite (0-100) was 64.62	MD 9.68 higher (3.36 lower to 22.72 higher)
MSQOL-54 change in health domain (0-100) Scale from: 0 to 100 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MSQOL-54 change in health domain (0-100) was 52.5	MD 0.23 lower (22.25 lower to 21.79 higher)
SF-36 physical functioning (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean SF-36 physical functioning (0-100) was 55.50	MD 1.68 lower (7.86 lower to 4.51 higher)
SF-36 emotional limitations (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕⊕◯◯ LOW a	-	The mean SF-36 emotional limitations (0-100) was 58.12	MD 0.73 lower (7.86 lower to 6.39 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Yoga
SF-36 physical role limitations (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕◯◯◯ VERY LOW a	-	The mean SF-36 physical role limitations (0-100) was 52.81	MD 1.59 lower (8.74 lower to 5.57 higher)
SF-36 energy/vitality (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 energy/vitality (0-100) was 55.48	MD 2.32 lower (8.5 lower to 3.86 higher)
SF-36 mental health (0-100) Scale from: 0 to 100 follow up: 12 weeks	40 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 mental health (0-100) was 61.78	MD 1.24 lower (9.16 lower to 6.68 higher)
SF-36 social functioning (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean SF-36 social functioning (0-100) was 62.00	MD 5.18 lower (25.78 lower to 15.41 higher)
SF-36 body pain (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 body pain (0-100) was 53.0	MD 1.13 lower (6.69 lower to 4.42 higher)
SF-36 general health (0-100) Scale from: 0 to 100 follow up: range 12 weeks to 6 months	77 (2 RCTs)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 general health (0-100) was 57.70	MD 3.25 lower (8.61 lower to 2.12 higher)
SF-36 health transition (0-100) Scale from: 0 to 100 follow up: 6 months	37 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean SF-36 health transition (0-100) was 36.7	MD 1 lower (17.67 lower to 15.67 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with aerobic exercise	Risk difference with Yoga
Beck Depression Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Depression Inventory (0-63) was 5.6	MD 5.49 lower (2.17 lower to 13.15 higher)
Beck Anxiety Inventory (0-63) Scale from: 0 to 63 follow up: 8 weeks	21 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean beck Anxiety Inventory (0-63) was 6.1	MD 0.35 higher (3.39 lower to 4.09 higher)
Cognitive - Stroop colour word interference (attention/concentration) follow up: 6 months	42 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean cognitive - Stroop colour word interference (attention/concentration) was 9.9	MD 1.4 lower (4.7 lower to 1.9 higher)
Adverse events (MS exacerbation) follow up: 6 months	39 (1 RCT)	⊕◯◯◯ VERY LOW a,c	RR 0.70 (0.05 to 10.32)	Moderate	
				63 per 1,000	19 fewer per 1,000 (59 fewer to 583 more)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

d. Heterogeneity that cannot be explained by subgrouping analyses and I2 >75%

Pilates vs. control (waitlist, no intervention)

Table 48: Clinical evidence summary: Pilates vs. control (waitlist, no intervention) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention)	Risk difference with Pilates
MFIS total (0-84) Scale from: 0 to 84 follow up: 8 weeks	120 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean MFIS total (0-84) was 10.5-48.3	MD 10.4 lower (18.98 lower to 1.82 lower)
MFIS physical (0-36) Scale from: 0 to 36 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean MFIS physical (0-36) was 21.3-22.8	MD 6.14 lower (8.9 lower to 3.39 lower)
MFIS cognitive (0-40) Scale from: 0 to 40 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d,e	-	The mean MFIS cognitive (0-40) was 15.3-20.8	MD 6.73 lower (14.62 lower to 1.15 higher)
MFIS psychosocial (0-8) Scale from: 0 to 8 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d,f	-	The mean MFIS psychosocial (0-8) was 4.0-4.7	MD 1.57 lower (3.14 lower to 0 lower)
STAY-Y1 - anxiety (20-80) Scale from: 20 to 80 follow up: 8 weeks	15 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean STAY-Y1 - anxiety (20-80) was 43.0	MD 18.5 lower (24.85 lower to 12.15 lower)
STAY-Y2 - anxiety (20-80) Scale from: 20 to 80 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean STAY-Y2 - anxiety (20-80) was 38.7-48.5	MD 7.44 lower (21.22 lower to 6.33 higher)
HADS - anxiety (0-21) Scale from: 0 to 21 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d,g	-	The mean HADS - anxiety (0-21) was 5.8-10.7	MD 0.64 higher (2.29 lower to 3.56 higher)
HADS - depression (0-21) Scale from: 0 to 21 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d,e	-	The mean HADS - depression (0-21) was 5.3-9.3	MD 2.72 lower (6.48 lower to 1.03 higher)
QIDS - depression (0-27) Scale from: 0 to 27 follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,d	-	The mean QIDS - depression (0-27) was 7.4-9.5	MD 2.45 lower (3.83 lower to 1.07 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist, no intervention)	Risk difference with Pilates
POMS-B total mood (scale unclear) follow up: 8 weeks	15 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean POMS-B total mood (scale unclear) was 26.0	MD 24.4 lower (41.28 lower to 7.52 lower)
POMS-B depression subscale (scale unclear) follow up: 8 weeks	15 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean POMS-B depression subscale (scale unclear) was 4.3	MD 4.2 lower (7.33 lower to 1.07 lower)
POMS-B fatigue subscale (scale unclear) follow up: 8 weeks	15 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean POMS-B fatigue subscale (scale unclear) was 9.3	MD 7.6 lower (13.07 lower to 2.13 lower)
Adverse events follow up: 8 weeks	95 (2 RCTs)	⊕◯◯◯ VERY LOW a,c,h	RD 0.00 (-0.06 to 0.06)	Moderate	
				0 per 1,000	0 fewer per 1,000 (60 fewer to 60 more)
Discontinuation possibly related to intervention follow up: 8 weeks	80 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	RR 0.88 (0.29 to 2.64)	Moderate	
				146 per 1,000	18 fewer per 1,000 (104 fewer to 240 more)

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

b. Downgraded by 2 increments as there was heterogeneity present that could not be explained by subgrouping strategies. Point estimates vary widely across studies and I2 >75%

c. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

d. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

e. Downgraded by 2 increments as there was heterogeneity present that could not be explained by subgrouping strategies, with I2 >60%

f. Downgraded by 2 increments as there was heterogeneity present that could not be explained by subgrouping strategies, with I2 >80%

g. Downgraded by 2 increments as there was heterogeneity present that could not be explained by subgrouping strategies. Point estimates vary widely across studies and I2 >70%

h. Imprecision assessed by sample size as zero events in both arms. Downgraded by 1 increment as sample size >70 and <350

Pilates vs. resistance + balance exercises

Table 49: Clinical evidence summary: Pilates vs. resistance + balance exercises – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with resistance + balance exercises	Risk difference with Pilates
MFIS physical (0-36) Scale from: 0 to 36 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MFIS physical (0-36) was 7.44	MD 0.26 lower (4.32 lower to 3.8 higher)
MFIS cognitive (0-40) Scale from: 0 to 40 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MFIS cognitive (0-40) was 7.33	MD 1.51 lower (6.75 lower to 3.73 higher)
MFIS psychosocial (0-8) Scale from: 0 to 8 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MFIS psychosocial (0-8) was 13.11	MD 5.47 lower (14.24 lower to 3.3 higher)
MusiQoL (0-100) Scale from: 0 to 100 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean musiQoL (0-100) was 40.05	MD 16.23 lower (28.78 lower to 3.68 lower)
Cognitive - PASAT follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b	-	The mean cognitive - PASAT was 27.89	MD 19.93 higher (9.07 higher to 30.79 higher)
BDI (0-63) Scale from: 0 to 63 follow up: 8 weeks	20 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean BDI (0-63) was 9.78	MD 1.87 lower (7.18 lower to 3.44 higher)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Pilates + balance training vs. relaxation

Table 50: Clinical evidence summary: Pilates + balance training vs. relaxation – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with relaxation	Risk difference with Pilates + balance training
Adverse or harmful events follow up: 8 weeks	39 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	RD 0.00 (-0.11 to 0.11)	0 per 1,000	0 fewer per 1,000 (110 fewer to 110 more)
Adherence - discontinuation due to work intensity	47 (1 RCT)	⊕◯◯◯ VERY LOW a,d	OR 5.11 (0.95 to 27.46)	0 per 1,000	235 more per 1,000 (63 more to 407 more)

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

b. Downgraded by 1 increment as the majority of evidence had a follow-up less than the minimum 3 months in the protocol

c. Imprecision assessed using sample size as zero events in both arms of a single study. Downgraded by 2 increments as sample size <70.

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

Relaxation vs. control (waitlist)

Table 51: Clinical evidence summary: Relaxation vs. control (waitlist) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (waitlist)	Risk difference with Relaxation
MFIS - total (0-84) Scale from: 0 to 84 follow up: 8 weeks	45 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean MFIS - total (0-84) was 38.1	MD 3.8 lower (12.93 lower to 5.33 higher)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the minimum 3 months in the protocol

88

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Acupressure vs. control (touching/sham only)

Table 52: Clinical evidence summary: Acupressure vs. control (touching only) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (touching only/sham)	Risk difference with Acupressure
Fatigue Severity Scale (scale unclear) follow up: 4 weeks	100 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (scale unclear) was 95.5	MD 30 lower (58.233 lower to 1.77 lower)
Fatigue Severity Scale (scale 1-7) Scale from: 1 to 7 follow up: 4 weeks	86 (1 RCT)	⊕⊕◯◯ LOW a,b	-	The mean fatigue Severity Scale (scale 1-7) was 4.01	MD 0.16 lower (0.81 lower to 0.49 higher)
Depression - DASS-42 (scale 0-42) Scale from: 0 to 42 follow up: 4 weeks	86 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean depression - DASS-42 (scale 0-42) was 11.36	MD 1.7 lower (3.01 lower to 0.39 lower)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up that was less than the minimum 3 months in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Reflexology/relaxation vs. control (usual care)

Table 53: Clinical evidence summary: Reflexology/relaxation vs. control (usual care) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care)	Risk difference with Reflexology/relaxation
Fatigue Severity Scale (1-7) - Foot reflexology vs. control Scale from: 1 to 7 follow up: 8-12 weeks	110 (2 RCTs)	⊕◯◯◯ VERY LOW a,b	-	The mean fatigue Severity Scale (1-7) - Foot reflexology vs. control was 4.74-4.97	MD 1.99 lower (2.41 lower to 1.56 lower)
Fatigue Severity Scale (1-7) - Relaxation vs. control Scale from: 1 to 7 follow up: 8 weeks	50 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Severity Scale (1-7) - Relaxation vs. control was 4.74	MD 0.47 lower (0.93 lower to 0.01 lower)
MSQoL-54 physical composite (0-100 usually) - Foot reflexology vs. control Scale from: 0 to 100 follow up: 12 weeks	60 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean mSQoL-54 physical composite (0-100 usually) - Foot reflexology vs. control was 41.12	MD 24.43 higher (15.66 higher to 33.2 higher)
MSQoL-54 mental composite (0-100 usually) - Foot reflexology vs. control Scale from: 0 to 100 follow up: 12 weeks	60 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean mSQoL-54 mental composite (0-100 usually) - Foot reflexology vs. control was 44.48	MD 28.83 higher (18.85 higher to 37.81 higher)
MSQoL-54 health change (0-100 usually) - Foot reflexology vs.	60 (1 RCT)	⊕⊕◯◯ LOW a	-	The mean mSQoL-54 health change (0-100	MD 39.17 higher (28.82 higher to 49.52 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care)	Risk difference with Reflexology/relaxation
control Scale from: 0 to 100 follow up: 12 weeks				usually) - Foot reflexology vs. control was 34.16	

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the 3 months minimum in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Massage vs. control (usual care/no intervention)

Table 54: Clinical evidence summary: Massage vs. control (usual care) – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care/no intervention)	Risk difference with Massage
Fatigue Severity Scale (9-63) mix of change from BL and final values Scale from: 9 to 63 follow up: 4-7 weeks	164 (3 RCTs)	⊕◯◯◯ VERY LOW a,b,c,d	-	The mean fatigue Severity Scale (9-63) mix of change from BL and final values was 3.0 (change value), 46.91-53.2 (final values)	MD 11.38 lower (22.08 lower to 0.68 lower)
Fatigue relief and effectiveness of fatigue reduction VAS (scale 0-10) Scale from: 0 to 10 follow up: 4 weeks	80 (1 RCT)	⊕◯◯◯ VERY LOW a,c,d	-	The mean fatigue relief and effectiveness of fatigue reduction VAS (scale 0-10) was 5.55	MD 1.3 higher (0.11 higher to 2.49 higher)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with control (usual care/no intervention)	Risk difference with Massage
Spielberger Overt Anxiety Questionnaire (scale 20-80) Scale from: 20 to 80 follow up: 7 weeks	60 (1 RCT)	⊕◯◯◯ VERY LOW a,c	-	The mean spielberger Overt Anxiety Questionnaire (scale 20-80) was 52.13	MD 13.48 lower (15.97 lower to 10.99 lower)

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

b. Downgraded by 2 increments as heterogeneity present that could not be explained by subgroup analyses, based on wide variation in point estimates across studies and I2 >90%

c. Downgraded by 1 increment as the majority of the evidence had a follow-up of less than the 3 months minimum in the protocol

d. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Reflexology vs. non-specialised foot massage

Table 55: Clinical evidence summary: Reflexology vs. non-specialised foot massage – up to 6 months outcomes

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specialised foot massage	Risk difference with Reflexology
Fatigue Impact scale - Total score (0-160) Scale from: 0 to 160 follow up: 4 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Impact scale - Total score (0-160) was 81.33	MD 13.57 lower (31.22 lower to 4.08 higher)
Fatigue Impact scale - Physical subscale (0-40) Scale from: 0 to 40 follow up: 4 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Impact scale - Physical subscale (0-40) was 22.3	MD 5.06 lower (9.89 lower to 0.23 lower)

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with non-specialised foot massage	Risk difference with Reflexology
Fatigue Impact scale - Cognitive subscale (0-40) Scale from: 0 to 40 follow up: 4 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Impact scale - Cognitive subscale (0-40) was 19.53	MD 1.98 lower (7.05 lower to 3.09 higher)
Fatigue Impact scale - Psychosocial scale (0-80) Scale from: 0 to 80 follow up: 4 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean fatigue Impact scale - Psychosocial scale (0-80) was 40.1	MD 6.83 lower (16.22 lower to 2.56 higher)
State trait anxiety inventory (20-80) Scale from: 20 to 80 follow up: 4 weeks	63 (1 RCT)	⊕◯◯◯ VERY LOW a,b,c	-	The mean state trait anxiety inventory (20-80) was 49.5	MD 6.2 lower (7.3 lower to 5.1 lower)

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

b. Downgraded by 1 increment as the majority of the evidence had a follow-up less than the minimum of 3 months in the protocol

c. Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs. For specific MIDs used for continuous outcomes, see footnotes in GRADE tables in Appendix F of this evidence review.

Evidence that could not be analysed using GRADE

Narrative summary for studies included in previous guideline version

Aerobic exercise versus control

McCullagh 2008[72] reported their results as medians and interquartile ranges (IQR), and so these could not be analysed in Revman. The results, which showed a clear advantage to aerobic exercise in reducing fatigue and improving function, are shown in the table below.

Table 56: Results from McCullagh 2008 for aerobic exercise versus control

Outcome	Exercise [median(IQR)]	Control [median(IQR)]	P (based on Mann-Whitney U test)
MFIS change from baseline to 3 months (lower better)	-13 (-20.5, -3)	1(-4, +4.5)	0.02
MSIS-29 change from baseline to 3 months (lower better)	-6.5(-10, +1)	-1(-4.5, +4.5)	0.13
FAMS change from baseline to 3 months (higher better)	23(+9.5, +42.5)	-3.5(-16, +5)	0.006
MFIS change from baseline to 6 months (lower better)	-8.5(-19.5, -1)	0.5(-2.5, +6.5)	0.02
MSIS-29 change from baseline to 6 months (lower better)	-6(-9, +0.5)	0(-1, +1)	0.10
FAMS change from baseline to 6 months (higher better)	19(+14, +31)	-4.5(-25, +8)	0.002

Gervasoni 2014[42] reported their results as medians and range so these could not be analysed in Revman. The median (range) FSS at 2 weeks was 5.5 (2.4-7) in the treadmill group and 5.3(1.6-7) in the control group. There was thus no clear difference between the groups.

Aerobic training versus neurorehabilitation

Rampello 2007[90] reported their results for fatigue and quality of life as medians and ranges, and so these could not be analysed in Revman. The results, which showed no difference between aerobic exercise and neurorehabilitation in reducing fatigue and quality of life, are shown the table below.

Table 57: Results from Rampello 2007 for aerobic exercise versus control

	Aerobic training N=11 [median (range)]	Neurological rehab N=11 [median (range)]	p
MFIS total median range	29 (4-56)	26 (3-67)	0.86
MFIS physical median range	14 (4-23)	13 (3-26)	0.89

MFIS cognitive median range	8 (0-36)	10 (0-40)	0.71
MFIS psychosocial median range	3 (0-7)	2 (0-6)	0.92
MSQOL-54 Overall quality of life median range	28 (10-82)	36 (20-82)	
MSQOL-54 physical median range	59 (44-81)	57 (41-81)	
MSQOL-54 mental health median range	66 (24-90)	66 (32-87)	

Motivational interviewing versus control

Bombardier 2008[16] reported their results as medians and interquartile ranges (IQR), and so these could not be analysed in Revman. The results, which showed a clear advantage to motivational interviewing in reducing fatigue and mental quality of life, but a possible disadvantage in terms of physical quality of life and no clear effect in improving function, are shown in the table below.

Table 58: Results from Bombardier 2008 for aerobic exercise versus control

	Motivational interviewing [median(IQR)]	Control [median(IQR)]	P
MS Fatigue Impact Scale	-1 (-9.5 to 0.5)	0 (-7 to 5)	0.02
SF-36 mental component	3.6 (0.3 to 8.0)	0.7 (-2.7 to 6.3)	0.02
SF-36 Physical component	-0.3 (-3.4 to 2.1)	1.0 (-2.8 to 5.1)	0.11
Bicycle ergometer time s	0 (-45 to 23)	0 (-34 to 31)	0.62
Self-selected walking speed	-0.4 (-2.0 to 0.5)	0.0 (-1.7 to 1.0)	0.28

Wii balance versus resistance training

Brichetto 2013[18] compared wii balance board training to static and dynamic exercises carried out with or without a balance board. After 12 sessions over 2 weeks, the wii group had improved by 10.1 points on the MFIS total scale, compared to 2.2 points in the control group. This was described as non-significant with a $p>0.05$.

Post-test values with standard deviations were reported but because of the baseline inequivalence it was deemed inappropriate to use them in this review. Hence change values were used, but no standard deviations for these change scores were available. Because of the imprecise p value it was not possible to estimate the standard deviations of these change scores.

Resistance training versus Yoga

Velikonja 2010[111] used non-parametric analyses for analysis, presenting their data as medians (IQR). Only within –group analyses were carried out, and so the imprecision of between-group comparisons is not possible to ascertain. Nevertheless, climbing appeared to lead to greater improvements in fatigue than yoga, but this may partly be explained by the climbing group starting off at a worse level. EDSS also improved more in the climbing group but again the climbing group were worse at baseline. Neither group seemed to change much in spasticity, though climbing was numerically more improved.

Table 59: Results from Velikonja 2010 for resistance training versus yoga

Variable	Climbing (n=10)			Yoga (n=10)		
	baseline	10 weeks	p	baseline	10 weeks	p
MFIS total	40(36.5-53)	27(21.5-45.5)	0.015	32(22-42)	23(20.5-36)	0.057
MFIS cog	17(8.5-21.5)	8(6-19.5)	0.024	12(4.5-14.3)	7(3.8-12.5)	0.282
MFIS ps	3(1.5-6)	3(1-5.5)	0.334	4(1-4.5)	3(0.8-4)	0.234
MFISphys	25(21.5-28.5)	19(9-26.5)	0.021	17.5(14.3-24.5)	18(9.8-19)	0.064
Spasticity MSA	10(8.5-18.3)	12.5(10-17.3)	0.574	9.3(3.5-18.4)	8.8(5.5-17.1)	0.673
EDSSpyr	4(3-4)	3(2.5-4)	0.046	2.5(2-4)	2(2-3.3)	0.317

Individualised rehabilitation versus group wellness intervention

Plow 2009[87] did not provide data for between group analyses except effect sizes. However, the paper reported that the modified fatigue impact scale and SF-36 did not differ significantly between groups at post-test.

Summary for new studies included in the current update of the review (see table below for summary)

Motivational interviewing + exercise (as well as inpatient rehabilitation) vs. control (inpatient rehabilitation only)

One study[35] with n=34 and n=30 in intervention and control groups, respectively, reported scores on the WEIMuS Fatigue Scale at 6 months, with results indicating reduction in score in the intervention and control groups compared to baseline. The results indicated a lower score (better outcome) in the control group compared to intervention, with P<0.001. Risk of bias was graded high.

Pilates vs. control (relaxation and respiration exercises)

One study[19] reported outcomes as median (IQR) for three groups (mat Pilates, reformer Pilates and control groups, n=12, n=13 and n=13, respectively) at the end of an 8-week intervention period. The results for Fatigue Severity Scale and the physical and mental health composite of the MS-Quality of Life-54 questionnaire indicated that scores significantly (P<0.05) improved in the mat Pilates and reformer Pilates groups compared to baseline. However, values in the control group also improved, though the P-value compared to baseline was only <0.05 for the control group for the physical health composite of the quality-of-life scale. At the end of the intervention, values were better in the control group for all outcomes, however there were some differences in the scores at baseline with control group values being slightly better before intervention. Risk of bias was graded high, with indirectness also an issue as the time-point was <3 months.

Balance training + Pilates vs. control (relaxation exercises)

One study[82] reported Fatigue Severity Scale in three groups (virtual reality balance training + Pilates, balance training + Pilates without virtual reality and relaxation control group, n=13 in each group) at the end of an 8-week intervention period as median (IQR). The results demonstrated a decrease in fatigue at 8 weeks compared to baseline in the two intervention groups, with the score increasing at 8 weeks in the control group compared to baseline. Risk of bias was graded high. Risk of bias was graded high, with indirectness also an issue as the time-point was <3 months.

Table 60: Further clinical outcomes reported incompletely by studies or as median values – new studies

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
Motivational interviewing + exercise (as well as inpatient rehabilitation) vs. control (inpatient rehabilitation only)					
Flachenecker 2020[35]	WEIMuS Fatigue Scale (scale	Baseline • Intervention: 45 (38-52) • Control: 39 (36-46)	N=34	N=30	High

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
	0-68) at 6 months Lower better.	6 months (3 months after last intervention session) • Intervention: 22.5 (8-30), P<0.001 vs control • Control: 5.5 (1-11) Values reported as median (IQR), with P-value vs. control given for 6-month time-point			
Pilates vs. control (relaxation and respiration exercises)					
Bulgurogl u 2017[19]	Fatigue Severity Scale (scale usually 9-63) at 8 weeks. Lower better.	Baseline • Mat Pilates: 49.0 (33.25-54.25) • Reformer Pilates: 48.0 (30.5-51.0) • Control: 44.0 (18.0-53.5) End of intervention (8 weeks) • Mat Pilates: 43.5 (26.75-50.50), P=0.034 vs. baseline • Reformer Pilates: 39.0 (32.5-48.0), P=0.008 vs. baseline • Control: 32.0 (19.5-47.0), P=0.221 vs. baseline Values reported as median (IQR) and P-value vs. baseline	N=25 (N=12 mat Pilates and N=13 reformer Pilates)	N=13	High, also indirectness as time-point <3 months
	MS Quality of Life-54 – physical health composite (scale usually 0-100) at 8 weeks.	Baseline • Mat Pilates: 74.54 (65.43-83.41) • Reformer Pilates: 71.14 (67.26-74.35) • Control: 77.35 (68.17-88.31) End of intervention (8 weeks) • Mat Pilates: 75.8 (70.83-86.42), P=0.005 vs. baseline • Reformer Pilates: 76.3 (74.39-83.37), P=0.002 vs. baseline • Control: 82.64 (66.77-91.27), P=0.023 vs. baseline			

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
	Higher better.	Values reported as median (IQR) and P-value vs. baseline			
	MS Quality of Life-54 – mental health composite (scale usually 0-100) at 8 weeks. Higher better.	<u>Baseline</u> • Mat Pilates: 74.54 (65.43-83.41) • Reformer Pilates: 69.2 (65.86-71.41) • Control: 75.65 (68.08-86.38) <u>End of intervention (8 weeks)</u> • Mat Pilates: 77.23 (70.2-84.54), P=0.006 vs. baseline • Reformer Pilates: 74.58 (70.39-80.58), P=0.002 vs. baseline • Control: 78.52 (64.77-89.21), P=0.249 vs. baseline Values reported as median (IQR) and P-value vs. baseline			
Balance training + Pilates vs. control (relaxation exercises)					
Ozkul 2020[82]	Fatigue Severity Scale (scale usually 9-63) at 8 weeks	<u>Baseline</u> • Virtual reality balance training + Pilates group: 48 (41.5-52.5) • Balance training + Pilates with no virtual reality group: 49 (34.5-54.5) • Relaxation control group: 46.0 (32.5-53.5) <u>8 weeks</u> • Virtual reality balance training + Pilates group: 37 (30.5-44.0) • Balance training + Pilates with no virtual reality group: 29 (26.0-46.5) • Relaxation control group: 52.0 (35.5-58.0)	N=26	N=13	High, also indirectness as time-point <3 months

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
		Values reported as median (IQR)			

1

2 **Table 61: Adherence and satisfaction outcomes reported by studies (those not suitable for GRADE analysis)**

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
Aerobic exercise vs. control (no intervention, or usual care with nurse consultations)					
Geddes 2009[41]	Adherence to programme at 12 weeks	Adherence reported to be 75% in the intervention group (walking) Not applicable to control group as no intervention involved.	N=8	NA	High
Heine 2017[50]	Acceptability of intervention (adherence %) at 26 weeks	Mean (SD) % of completed sessions reported for the aerobic exercise group: 74 (25)%. The intervention consisted of 12 sessions. For the control group, 87% reported to have completed all three sessions.	N=37	N=34	High
McCullagh 2008[72]	Adherence at 12 weeks	In the intervention group, all completed at least 20/24 hospital-based classes (only 2 completed all 24) but none completed >50% of prescribed home sessions. Not applicable in the control group as no intervention involved	N=12	NA	High
Functional electrical stimulation cycling vs. control					
Backus 2020[10]	Adherence – completion of all 36 training sessions at 12 weeks	Reported that in the intervention group, all but one (5/6 analysed) completed all of the 36 sessions. Not applicable for the control group as no intervention was completed as a waitlist control group.	N=6	NA	High

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
Exercise – resistance training vs. control (waitlist control, usual care or education only)					
Grubic Kezele 2019[44]	Compliance - % of exercise sessions attended at 4 weeks	Mean (SD) % reported to be 98.0 (4.2) in the exercise intervention group. Not applicable for the control group as no intervention was completed as a waitlist control group.	N=10	NA	High, also indirectness as time-point <3 months
Dalgas 2010[27]	Adherence at 12 weeks	Intervention group reported to have completed a total of 23.9 (95% CI 23.7-24.0) out of 24 planned sessions. Not applicable for control group as no intervention involved.	N=19	NA	High
Dodd 2011[30]	Adherence –number of scheduled sessions (out of 20 in intervention group and 10 in control group) attended at 10 weeks	Progressive resistance training group Mean (SD) of 18.4 (2.9), range 6-20 (20 possible sessions) Usual care group with social intervention Mean (SD) of 6.2 (3.1), range 0-10 (10 possible sessions)	N=36	N=35	Some concerns
Exercise – vestibular balance rehabilitation vs. control (waitlist control)					
Hebert 2018[48]	Compliance % at 14 weeks	Compliance reported to be 92% and 88%, respectively, in phase 1 and 2 supervised training. 81% reported to have returned the home-based exercise log. Not applicable for the control group as no intervention was completed as a waitlist control group.	N=38	NA	High
Yazgan 2019[113]	Compliance at 8 weeks	Statement that all in the intervention group completed 16 sessions of exercise with excellent adherence to exergaming systems.	N=27	NA	High, also indirectness as

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
		Not applicable for the control group as no intervention was completed as a waitlist control group.			time-point <3 months
Exercise – vestibular balance rehabilitation vs. Exercise – aerobic training					
Dettmers 2009[28]	Acceptance at 3 weeks	Acceptance stated to be high, with one participant dropping out as they found it too demanding Not reported for the aerobic group, though no dropouts in that group.	N=15	N=15	High, also indirectness as time-point <3 months
Exercise – progressive resistance training + aerobic exercise vs. control (no intervention)					
Maurer 2018[71]	Compliance % at 6 months (completing at least 70% of scheduled exercise sessions during months 1-6)	In the intervention group, % sessions completed was variable (0-442.0%). Mean compliance was 82.4 (64.1)%. 39.8% were described as non-compliant. Not applicable for the control group as no intervention involved.	N=94	NA	High
	Feasibility and acceptance questionnaire at 6 months	• Usability in general (software): 2.34 (0.94), n=129 analysed (lower better) • Usability – graphical appeal (software): 4.12 (0.98), n=126 analysed (higher better) • Usability – problems with software: 2.31 (0.93), n=127 analysed (lower better) • Therapeutic support – satisfaction with therapist and support at introductory session: 1.4 (0.64), n=128 analysed (lower better)	See previous column for each domain	NA	

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
		• Therapeutic support – satisfaction with the training support: 1.4 (0.66), n=128 analysed (lower better) • Therapeutic support – satisfaction with the support at the central assessment centre: 1.4 (0.56), n=128 analysed (lower better) • Satisfaction about the quality of the information about the internet-based training and to independently conduct the training at home at the introductory group session: 4.4 (0.72), n=128 analysed (higher better) • Usefulness and meaningfulness of an internet-supported training: 4.4 (0.89), n=126 analysed (higher better) • Interest in the continuation of the training: 3.9 (1.1), n=127 analysed (higher better) Scores for each domain were graded on a scale of 1-5 and results given as mean (SD). Results include anyone that eventually had the intervention, including some that were originally in the waitlist group but had the intervention after this period.			
Exercise – progressive resistance training + aerobic exercise vs. yoga vs. control (no intervention)					
Garrett 2013[39] and Garrett 2013[40]	Adherence –classes attended (out of possible 10 classes)	<u>Progressive resistance + aerobic exercise group (led by physiotherapist group)</u> Mean (95% CI): 8.1 (7.5-8.5) <u>Yoga group</u> Mean (95% CI): 7.8 (7.2-8.3) Not applicable for control group as no intervention involved	N=63 in both groups	NA	High
Exercise – progressive resistance training + balance exercises vs. control (no intervention)					

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
Cakit 2010[20]	Adherence to training protocol at 8 weeks	In the supervised training group, 209/224 prescribed sessions were completed – average adherence rate of 93%. In the home-based training group, 136/224 prescribed sessions were completed – average adherence rate of 60%. Not applicable for the control group as no intervention involved.	N=224 in both groups	NA	High
Exercise - progressive resistance training + aerobic + balance vs. control (usual care)					
Learmonth 2012[68]	Adherence at 12 weeks	Adherence at classes was 69% in the intervention group. Not applicable for the control group as no intervention involved.	N=15	NA	High
Standard exercises (progressive resistance training + aerobic + balance) + high-intensity lower limb resistance training vs. standard exercises alone					
Hayes 2011[46]	Participation - % only at 12 weeks	<u>Resistance + standard exercise</u> Average of 30/36 days of exercise (82% participation) <u>Standard exercise</u> Average of 30/36 days of exercise (82% participation)	N=10	N=9	
FACETS (Fatigue: Applying Cognitive behavioural and Energy Effectiveness Techniques to lifestyle) vs. control (local/standard care)					
Thomas 2014[104] and Thomas 2013[105]	Adherence – attended at least 4 sessions (out of possible 6) at 6 weeks	In the intervention group, 72/84 (85.7%) attended at least four of the six sessions Not applicable for the control group as no sessions to attend.	N=84	NA	Some concerns
	Satisfaction – content/format/usefulness/pace/le	Mean (SD) for various domains: • Content: 4.6 (0.6) • Format: 4.5 (0.7) • Usefulness: 4.6 (0.7)	N=84	NA	

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
	ngth. Scale 1-5 (5=ideal) at 6 weeks	• Pace: 3.1 (0.6) • Length: 3.1 (0.6) Not applicable for the control group as no intervention.			
Multidisciplinary rehabilitation + fatigue self-management programme vs. control (nurse consultation only)					
Rietberg 2014[93]	Adherence to homework tasks at 6 months	Adherence to homework tasks as part of the 6-month intervention reported to be: • 96% in the multidisciplinary rehabilitation + fatigue self-management group • 89% in the MS nurse consultation group	N=21	N=23	Some concerns
Self-management programme vs. control (education intervention)					
Ehde 2015[32]	Treatment satisfaction (unclear how this was measured and scale unclear) at 12 months	Median (IQR): 9 (8-10) vs. 8 (5-9) in self-management and control groups, respectively	Unclear, N=64 for other outcomes	Unclear, N=81 for other outcomes	High
Self-management programme + exercise vs. control (waitlist control)					
Lutz 2017[69]	Compliance % at 6 weeks	Stated that in the intervention group, all participants had at least 80% compliance, with none missing >2 sessions. Not applicable to control group as was a waitlist control with no intervention.	N=8	NA	High, also indirectness as time-point <3 months
CBT vs. control (no intervention, or MS nurse consultations)					
Moss-Morris 2012[73]	Adherence at 10 weeks	In the intervention group, mean (SD) sessions completed: 4.91 (2.10) of 8 sessions. Only one finished all 8 sessions. 60.8% finished >5 sessions.	N=23	NA	Some concerns

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
		Not applicable to control group as was a waitlist control with no intervention.			
van den Akker 2017[108]	Compliance with protocol at 16 weeks	CBT: 64% completed at least 10 sessions. Median (IQR) was 10.5 (8.8-11.0) sessions. Control: 79% completed all three consultations. Median (IQR) was 3 (3-3).	N=39	N=35	Some concerns
Exercise (aerobic + resistance training) + cognitive behavioural therapy vs. control (waitlist control)					
Carter 2014[22]	Adherence to intervention at 3 months	In the intervention group, participants attended an average of 16.2 of 18 supervised sessions (90%, range 7-18 sessions) and completed an average of 14.6 of 18 home exercise sessions (81%, range 2-18 sessions). Not applicable to the control group as no intervention involved.	N=60	NA	Some concerns
Diet vs. control (education sessions)					
Katz Sand 2019[59]	Engagement and adherence at 6 months	Attendance at monthly sessions or by phone was 90.6% overall. Mean self-reported adherence was 90.3%. Outcome not reported in a way that could also apply to the control group.	N=18	NA	High
Mindfulness training vs. control					
Grossman 2010[43]	Adherence at 8 weeks (average adherence rate)	Average adherence rate in the intervention group reported to be 92% of all sessions. Not applicable in the control group as no intervention involved.	N=76	NA	High
Yoga vs. aerobic exercise vs. control (waitlist control)					

Study	Outcome definition	Results	Intervention group (n analysed)	Comparator group (n analysed)	Risk of bias
Oken 2004[80]	Adherence - % attendance at sessions at 6 months	Yoga: Attendance reported to be 68% - home practice reported for an average of 51% of non-class days for an average of 39 min (14-80 min) Aerobic exercise: Attendance reported to be 65% - home practice reported for an average of 45% of non-class days for an average of 32 min (15-57 min) Not applicable for the control group as no intervention was completed as a waitlist control group.	N=21 in yoga group and N=15 in aerobic exercise group	NA	High
Balance training + Pilates vs. control (relaxation exercises)					
Ozkul 2020[82]	Adherence – participation rate at 8 weeks	Median (IQR) participation was reported to be 80.8% (68.8-100.0) for virtual reality and 82.7% (68.8-100) for balance training. Values for the control group were not reported.	N=26	NA	High, also indirectness as time-point <3 months

1.1.7 Economic evidence

1.1.7.1 Included studies

Four health economic studies with the relevant comparison were included in this review[73, 76, 105, 106].

These are summarised in the health economic evidence profile below (**Table 62**) and the health economic evidence tables in Appendix H.

1.1.7.2 Excluded studies

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

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

2 ### 1.1.8 Summary of included economic evidence

3 **Table 62: Health economic evidence profile: Non-pharmacological management of fatigue**

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Moss-Morris 2012[73] (UK)	Partially applicable [(a)]	Very serious limitations [(b)]	• Within trial analysis (pilot RCT: Moss-Morris 2012 [73]) • Cost-utility analysis (QALYs) • Population: Adults with MS • Comparators: 1. Waitlist 2. Online CBT programme • Analysis of individual level data for health outcomes, EQ-5D and resource use, with unit costs applied. • Follow-up: 10 weeks	2-1: Saves £4 [(c)]	2-1: 0.015 QALYs gained	Online CBT programme dominates Waitlist Mean costs were similar between groups with a small improvement in quality of life.	Probability online CBT program cost effective (£20/30k threshold): NR Uncertainty: Results retained their significance levels for all outcomes when the analysis was rerun controlling for gender, ambulation status and completion.
Thomas 2013[105](UK)	Partially applicable [(d)]	Potentially serious limitations [(e)]	• Within trial analysis (RCT: Thomas 2013[105]) • Population: Adults with clinical definite MS diagnosis (FSS total score >4; ambulant) receiving either • Comparators:	2-1: £488[(f)]	2-1: 0.02 fewer QALYs	Current local practice dominates FACETS (less costly and more effective)	A probabilistic sensitivity analysis was undertaken to analyse the impact of the uncertainty in the level of staff input for FACETS programme delivery on costs. The mean cost of the intervention was £453 with 95% of estimates in

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			1. Current local practice 2. Group based fatigue management programme (FACETS) and current local practice. • Analysis of individual level data for health outcomes, EQ-5D and resource use, with unit costs applied. • Follow-up: 5.5 months (4 months after final session)				the range of £331 to £585 per participant.
Tosh 2014[106] (UK)	Partially applicable [(g)]	Potentially serious limitations[(h)]	• Within trial analysis (RCT: Carter 2014[22]) • Population: adults with clinically definite MS diagnosis; EDSS score 1.0–6.5; able to walk a 10-metre distance and physically able to participate in exercise three times per week • Comparators: 1. Current local practice 2. Programme incorporating	2-1: £466[(i)]	2-1: 0.046 QALYs	2 vs. 1: £10,137 per QALY gained	Probability cost-effective (£20k): 75% Scenario analyses conducted: • Scenario 1 (EDSS score): <4.0 = dominated; ≥4.0 = £5,092 per QALY gained • Scenario 2 (GLTEQ score): >14 = £9,558 per QALY; <14 = £11,470 per QALY gained

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			aerobic and resistance exercise and CBT (EXIMS) and current local practice. • Analysis of individual level data for health outcomes, EQ-5D and resource use, with unit costs applied. • Follow-up: 9 months (6 months after final session)				• Scenario 3 (private provision of intervention): £11,938 per QALY gained • Scenario 4 (SF-6D utility score): £19,783 per QALY gained
National Institute for Health and Care Excellence, P.421, 2014[76] (UK)	Directly applicable[(j)]	Potentially serious limitations [(k)]	• de novo health economic analysis conducted as part of NICE 2014 guideline based on an RCT included in the clinical review (Cakit 2010 [20]) • Population: people with multiple sclerosis. • Comparators: 1. Control 2. Homebased resistance and balance 3. Supervised resistance and balance • Time horizon: one year.	2-1: £52 3-2: £398 [(l)]	2-1: 0.011QALY 3-2: 0.052 QALY	2 vs. 1: £4,867 per QALY 3 vs. 2: £7,619 per QALY	Sensitivity analysis was conducted with a shorter time horizon of 8 weeks. Assuming the improvement in quality of life is not maintained beyond the 8-week intervention duration, the ICER increased to £31,633 per QALY and £49,526 per QALY for comparison 1 and 2 respectively

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty

Abbreviations: CBT = cognitive behavioural therapy; ICER = incremental cost-effectiveness ratio; MS = multiple sclerosis; NA = not applicable; NR = not reported; QALYs= quality-adjusted life years; RCT = randomised controlled trial; EDSS = Expanded Disability Status Scale; EQ5D = Euroqol 5 dimensions (scale: 0.0 [death] to 1.0 [full health]; EXIMS = EXercise Intervention for people with MS; GLTEQ = Godin Leisure Time Exercise Questionnaire; QALYs = quality-adjusted life years; SF-6D = Short form 6 dimension; FACETS: Fatigue Applying Cognitive behavioural and Energy effectiveness Techniques to lifeStyle; FSS = Fatigue Severity Scale.

(a) EQ-5D scoring tariff was not reported.

(b) Cost utility model based on a single pilot RCT. Sample size was small (n=40) with a high non-completion rate. The study was a small feasibility trial with no long-term follow-up data; Cost-effectiveness would be heavily influenced by the maintenance of treatment gains. 10 weeks may be too short to show much change in healthcare resource use between groups. Intervention effects were obtained from the current trial, which was a pilot trial and not designed to evaluate intervention effects with certainty nor long enough to estimate the duration of treatment effect. Costs did not include development or administration of the intervention, which would depend on how many people used it. Medication costs were not included. Resource use was self-reported by trial participants at 10 weeks, which may be unreliable. The only reference for unit costs was Personal Social Services Research Unit. However, for some unit costs the NHS Tariff may have been a more appropriate source. The analysis was rerun controlling for gender, ambulation status and completion but detailed results of these analyses were not reported. No probabilistic sensitivity analysis.

(c) 2008 UK pounds. Cost components incorporated: Outpatient appointments (neurology and other), inpatient care (urology, intensive care unit, other), residential care, general practitioner, specialists (neurologist, other), physiotherapist, social worker, nurse, home help, other.

(d) *QALYs derived from EQ5D (from patients, tariffs not stated) with maximum QALY equalling 0.46, assuming full health over 24 weeks.*

(e) Analysis based on a single RCT (Thomas 2013[105]). No probabilistic sensitivity analysis for ICER and short follow-up

(f)2010 UK pounds. Costs incorporated are FACETS programme including training, equipment, session facilitators (two Band 7 therapists), venue hire, refreshments, printing, administrative support. Cost for NHS and social care (over a 3-month period) assessed at 4 months follow up for both interventions.

(g) Analysis was based on a single RCT (Carter 2014[22]). QALYs derived from EQ-5D (from patients, tariff used not stated).

(h) Analysis based on a single RCT (Carter 2014[22]).Short follow-up

(i) 2011 UK pounds. Costs incorporated are: EXIMS programme including staff, equipment, and overheads. Costs for NHS and social care services over 9-month period (intervention start to end of follow-up) assessed for both interventions.

(j) Direct EQ-5D data was not available. QALYs were estimated through the mapping of changes in SF-36 scores obtained from the RCT using algorithm by Ara and Brazier (2008). The improvement in EQ-5D was assumed to be maintained, beyond the 8-week intervention period, over 1 year.

(k) Analysis based on a single RCT (Cakit 2010[20]); utilities were estimated through a mapping function which is associated with limitations. The results were sensitive to the assumption of a continued treatment effect beyond the trial follow-up Cost of a cycling machine and downstream costs were excluded from the analysis. No probabilistic sensitivity analysis.

(l) Cost of staff time only.

1.1.9 Unit costs

Relevant unit costs are provided below to aid consideration of cost effectiveness.

Resource	Unit cost per working hour [(a)]
Hospital-based staff	
Consultant: Medical	£148
Consultant: psychiatric	£146
Clinical psychologist (band 8a)	£72
Hospital physiotherapist (band 7)	£62
Hospital occupational therapist (band 7)	£62
Clinical Nurse specialist (band 7)	£62
Community-based staff	
Physiotherapy (band 7)	£60
Occupational therapy (band 7)	£60
Clinical psychologist, Counsellor (specialist) (band 7)	£60
Nurse (GP practice)	£41
Interventions	
Cognitive behavioural therapy (CBT) per session	£106 [(b)]
Mindfulness-based cognitive therapy – group-based intervention	£91 per hour of direct contact £181 per session, £16 per service user [(c)]

Source: PSSRU 2020[26]

(a) Qualification costs included (excluding individual and productivity costs)

(b) Taken from PSSRU (2017)[25] and inflated to 2018/19 prices using OECD purchasing power parities (PPPs)[81]

(c) Taken from PSSRU (2013)[24] and inflated to 2018/2019 prices using OECD purchasing power parities (PPPs)[81]

1.1.10 Evidence statements

Effectiveness

For results that could be assessed using GRADE, see summary of evidence in Tables 3-53. A narrative summary of studies that could not be analysed using GRADE in the previous version of this review is provided under the 'evidence that could not be analysed using GRADE' section of the results above.

Clinical outcome data from new studies that could not be analysed is provided in Table 59. A narrative summary of this evidence is provided alongside the table.

Data for adherence outcomes and satisfaction could often not be analysed using GRADE due to the fact that the outcome only applied to the intervention group (for example, adherence or satisfaction could not be assessed in waitlist control groups as there was no intervention to adhere to or rate satisfaction for). Where it was possible to analyse using GRADE, this data is provided in the main GRADE summary tables for each comparison. Additional data from studies where

comparative data was not available is presented in Table 60. This data was of limited use due to its non-comparative nature.

Economic

One cost–utility analysis found that an online CBT program was dominant (less costly and more effective) than waitlist for the management of fatigue in adults with multiple sclerosis. This analysis was assessed as partially applicable with very serious limitations.

One cost–utility analysis found that current local practice was dominant (less costly and more effective) than a group-based fatigue management programme (FACETS) for the management of fatigue in adults with multiple sclerosis. This analysis was assessed as partially applicable with potentially serious limitations.

One cost–utility analysis found that a programme incorporating aerobic and resistance exercise was cost effective compared to current local for the management of fatigue in people with multiple sclerosis (ICER: £10,137 per QALY gained). This analysis was assessed as partially applicable with potentially serious limitations.

One cost–utility analysis found that a supervised resistance and balance intervention was the most cost-effective intervention when compared to control and a homebased resistance and balance intervention for the management of fatigue in people with multiple sclerosis (ICER: £7,619 per QALY gained compared to homebased resistance and balance). This analysis was assessed as partially applicable with potentially serious limitations.

1.1.11 The committee's discussion and interpretation of the evidence

1.1.11.1. The outcomes that matter most

All outcomes listed in the protocol were considered to be equally important for decision-making. Outcomes included in the protocol were patient-reported outcome measures assessing MS fatigue, health-related quality of life measures, impact on carers, functional scales quantifying level of disability, cognitive functions such as memory and concentration, psychological symptoms and adverse effects including incidence of different adverse events and those leading to withdrawal. Outcomes measuring how acceptable an intervention was, for example through satisfaction or adherence, were also reported where available.

Fatigue outcomes were well-reported across studies, though the specific fatigue scale used differed across studies, though most reported either Modified Fatigue Impact Scale or Fatigue Severity Scale. Adverse events were also well-reported across studies, though some gave the number of specific adverse events separately and others reported total number of adverse events or those requiring withdrawal only. Quality of life and psychological outcomes, such as anxiety and depression, were reported for most comparisons, though not all studies reported these outcomes and there was variation in the scale used for those that did. Other outcomes were less well-reported across studies and comparisons but for some comparisons there was data available, such as functional scales quantifying disability (for example, EDSS) and measures of cognitive function. Impact on carers was the only outcome where no data was available for any intervention or comparison.

The preferred format of continuous outcomes (as a continuous or dichotomous measure) was not specified in the protocol and any format these outcomes were reported in were therefore extracted. In the vast majority of cases studies reported outcomes in a continuous format. However, some studies reported a continuous outcome in both a continuous and dichotomous format, and in this case both versions of the same outcome were extracted (for example, continuous final value for Modified Fatigue Impact Scale and also a dichotomous version of the outcome where the study reports the proportion that achieved any improvement on this scale compared to baseline). A few studies only reported certain outcomes in a dichotomous format. Caution was noted when interpreting continuous outcomes that had been reported in a dichotomous format as there are various limitations associated with this.

Two different time-points were prespecified in the protocol and some evidence was found for both of these time points (3-6 months and >6 months – 12 months), though fewer studies reported data for the later time-point. Among studies included in the 3–6-month time-point, many of these were indirect, as they reported outcomes at a time-point <3 months (for example, 8 weeks) but were included and downgraded for indirectness as specified in the protocol.

No relevant randomised controlled trials including the following interventions were identified:

- 'Getting to Grips' programme
- Gym prescription
- 'FatiMa' patient education programme
- Tai Chi
- Hyperbaric oxygen

1.1.11.2 The quality of the evidence

A total of 89 randomised controlled trials (RCTs) were included in this review; most of these were parallel RCTs but did also include five crossover trials and one cluster-randomised trial. This included 29 studies that had already been included in the previous version of this review and an additional 60 studies identified as relevant during the update. Studies covered a wide range of interventions and different comparators. Pooling was performed where possible but even then, the total sample size of the meta-analyses remained small, with most being <200 people, as the majority of individual studies were small. Despite the largest study having over 300 people included, very few studies had a sample size >100 and many of these had sample sizes <50. The small size of included studies, and small sample sizes when even one two or three studies were pooled, meant that the majority of reported outcomes across comparisons were based on data from very small populations, often <100 or <50 people if only a single study reported the outcome. This contributed to a lot of uncertainty in the size and direction of the effect, meaning the committee could not be confident in most of the results that were reported based on confidence intervals for the absolute effect.

The quality of the evidence as assessed by GRADE ranged from very low to high, with the majority being of low or very low quality. Across all outcomes and comparisons, downgrading was primarily due to imprecision and/or risk of bias. Within risk of bias ratings, the most common reasons contributing to a rating of 'some concerns' or 'high' risk of bias for an outcome were concerns about bias arising from the randomisation process, concerns about the degree of missing data and a lack of blinding for subjective outcome measures. Many outcomes were also downgraded for indirectness if the majority of the evidence for that outcome came from studies where the outcome was reported a time-point less than the 3-month minimum specified in the protocol (for example, at 8 weeks).

A number of outcomes were also downgraded for inconsistency as there was heterogeneity present in the meta-analyses that could not be explained by prespecified subgrouping strategies due to there being three or fewer studies included or most or all studies falling into the same subgroup categories and heterogeneity therefore not being explained by these subgrouping strategies. A random effects analysis was used for these outcomes and downgrading for inconsistency performed as part of the GRADE quality rating.

1.1.11.3 Benefits and harms

These initial paragraphs cover a summary of the decisions that were made and the factors contributing to these decisions. Because there were a wide range of interventions and comparisons included in this review, a description of the benefits and harms identified for specific comparisons is included below under individual headings for type of intervention and comparator.

Overall, the committee agreed that despite there being a large number of new studies since the previous update, based on the limitations described above, they could not make existing recommendations any stronger based on the evidence alone, but used the additional evidence identified within this update as further support for existing recommendations on which interventions may be beneficial in MS-related fatigue.

In terms of which non-pharmacological interventions could be used in MS-related fatigue, the committee agreed that there was evidence of some benefit from fatigue/energy management interventions and well-being techniques such as CBT

and mindfulness, meaning they should be mentioned as per the previous guideline. However, based on the limitations of the evidence as described above, recommendations for formal programmes were not made and the recommendation instead suggested that some elements of them could be factors to include in fatigue management discussions. The wording of the recommendation was altered to provide improved clarity and highlight how various factors should be considered and included as appropriate as part of a tailored discussion about fatigue management with each individual, and the recommendation strength was 'offer' rather than 'consider', as the committee agreed that in practice people would routinely be provided with this by occupational therapists and sometimes MS nurses or physiotherapists. A non-exhaustive list of factors that could be included in the discussion about fatigue management was included in the recommendation, with the list consisting of identification of goals and priorities for each person, advice on energy conservation, review of lifestyle factors such as diet and exercise and the use of well-being techniques such as cognitive behavioural principles for managing day-to-day activities and mindfulness-based techniques, all of which were agreed to be used in fatigue management discussions in current practice.

Previous recommendations on exercise-based interventions for MS-related fatigue were edited for clarity in line with current practice and clinical experience as well as based on the evidence included in the review. The previous recommendation to advise people that aerobic, balance and stretching exercises, including yoga, may be helpful was retained, as the committee agreed that the new evidence combined with that previously included did suggest possible benefits of these types of exercise. The committee further edited this recommendation to also include resistive exercises, as the evidence review demonstrated some possible benefits for this type of exercise as well as aerobic and balance exercises, and they also included Pilates as an example of a form of exercise that might be beneficial for the same reason. Although many of the studies involved supervised programmes, this recommendation covers self-directed exercise in the form of advice to people with MS, as the evidence included in the review was too limited to make recommendations for structured programmes to be provided, which included a lack of cost-effectiveness evidence to support providing structured programmes to all people with MS-related fatigue. An exception to this, where there was cost-effectiveness evidence to allow a supervised programme to be recommended, is described in the following paragraph.

The previous recommendation about a comprehensive programme of aerobic and moderate progressive resistance activity combined with cognitive behavioural techniques for those with fatigue and EDSS score of at least 4 was also retained. This was based on the clinical evidence and modest economic evidence from Tosh (2014), covering the EXIMS study, and the original economic analysis from the last guideline supporting the cost-effectiveness of combined exercise programmes. The population was limited to those with an EDSS score of at least 4 based on scenario analyses reported by the study, which indicated the intervention was cost-effective in those with more severely impaired mobility (EDSS >4) but dominated (more costly and less effective) in those with EDSS scores <4. However, the wording was edited to 'consider providing' to differentiate this from advice and make it clear that this refers to actively providing a supervised programme rather than self-directed exercise, as the committee noted that there was evidence of benefits of this type of exercise and that a supervised programme may be beneficial in terms of avoiding injury and improving adherence, as well as the fact that the EXIMS study was mostly a supervised programme. The word 'comprehensive' was also removed as it may allow for a more tailored programme to be provided according to the needs and abilities of each individual.

The committee noted the absence of RCT evidence for hyperbaric oxygen in MS-related fatigue and they were concerned that people with MS may be spending their own money on this treatment or accessing it through charities where resources would be better diverted elsewhere. Based on a lack of evidence and clinical experience and the fact that it is quite an expensive treatment, the committee made a recommendation that hyperbaric oxygen should not be used to treat fatigue in people with MS.

The committee noted that although some studies on diet were identified, the evidence was weak with only a few small studies, each looking at different interventions or comparisons. This meant it was not possible to specify a specific diet that should be followed, but as there was some evidence to support dietary interventions, the committee agreed to include diet as a factor within the tailored fatigue management discussion recommendation discussed above. A separate recommendation to highlight the lack of evidence for specific diets but to follow healthy diet principles was also made. It was noted that the effects of a healthy diet may not be specific to fatigue but health in general.

Further edits to recommendations were made to improve clarity and were not based on evidence identified in the review but on clinical experience and practice. The committee restructured the initial recommendation about assessing and offering treatment for other conditions to those with MS-related fatigue to improve clarity. The original recommendation was split into two separate recommendations, with the first being a clear statement to ask people with MS about the presence of fatigue to ensure that it is discussed as needed. The second recommendation made was a statement that it should not be assumed that fatigue is caused by MS, as other factors may be contributing to fatigue, which should be considered and managed appropriately. The list of examples provided included anxiety, depression, difficulty sleeping, which were included in the previous version of the recommendation, but in line with clinical experience the updated recommendation also highlighted the role other symptoms of MS (such as pain, spasticity and bladder dysfunction), side effects of medicines and illness such as infections (as well as anaemia and thyroid dysfunction already mentioned in the previous guideline version) can have in either causing or exacerbating fatigue. References to existing NICE guidance were made where appropriate.

An existing recommendation about explaining how MS-related fatigue may be precipitated by heat, overexertion and stress and that it may be related to the time of day was edited based on clinical experience to improve clarity. The recommendation was edited to explain that MS-related fatigue may be brought on by heat and biological, physical and emotional stress, wording that was considered to be a more accurate reflection currently based on clinical experience. Specific mention of overexertion was removed from the recommendation as there is no direct link between overexertion and MS-related fatigue and time of day was also removed from the recommendation as it was noted that MS-related fatigue can occur at any time of day and the link between time of day and fatigue is unclear.

Exercise-based interventions vs. control

Aerobic exercise vs. control

Depending on the outcome, up to five studies (up to 141 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most specific measures were only reported by one study. Most outcomes were low to very low quality based on GRADE. A range of fatigue scales were used

across studies, with most point estimates suggesting a possible benefit for aerobic exercise in terms of fatigue; however, in all cases there was uncertainty in the direction and/or size of effect. Similarly, some results suggested possible benefits for quality of life, but this was not consistent across studies and subscales, and uncertainty existed as described above for fatigue scales. Two studies reported different disability scales (EDSS and Guy's neurological disability scale), with the point estimates for both suggesting possible harms of aerobic exercise; however, uncertainty in the direction and/or size of the effect based on confidence intervals for the absolute effect existed. Data for various cognitive tests was available, though only from one study for each measure, with the results of most suggesting no difference between the groups and there being uncertainty in the direction and/or size of effect for those where point estimates suggested a possible harm or benefit. Similar to fatigue measures, studies reporting anxiety and/or depression outcomes suggested possible benefits of aerobic exercise based on point estimates, but uncertainty existed based on confidence intervals. One study suggested fewer adverse events with the intervention, including MS exacerbations, orthopaedic problems and falls reported individually, but a pooled analysis of 5 studies for mixed adverse events suggested increased events in the aerobic exercise group compared to control and one study reporting those leading to withdrawal also suggested increased events in the intervention group, though there was uncertainty in the direction of effect for the latter.

Only one study reported data for some outcomes at a time-point >6 months (12 months), with all outcomes assessed as very low quality. The results for this study suggested no difference for all four of five outcomes that were reported: three different fatigue scale measures and a measure of cognitive function. The odds ratio reported for the incidence of adverse events, specifically MS relapse in this case, suggested fewer events in the aerobic exercise group compared to control.

<u>Aerobic exercise vs. neurological rehabilitation (respiratory, postural and stretching)</u>

One study that included 22 people covered this comparison at 8 weeks; however, all clinical outcomes were reported as median values with their range, meaning GRADE analysis could not be performed and limiting the interpretation of these results. Risk of bias assessment for this study led to downgrading of two increments. Results for total fatigue score and subscales within the fatigue score suggested better scores for most in the neurological rehabilitation group, apart from the cognitive subscale; however, P-values were all >0.05. Similarly, quality of life data demonstrated very little difference between the two groups and average adherence rate was also similar.

<u>Resistance training vs. control (waitlist control, no intervention, usual care or education only)</u>

Depending on the outcome, up to three studies (up to 133 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. Most outcomes were low to very low quality based on GRADE. A range of fatigue scales were used across studies, point estimates for some suggested a possible benefit for resistance training in terms of fatigue, while others suggested no difference; however, in all cases there was uncertainty in the direction and/or size of effect. Quality of life was also reported using different scales across studies; results for some subscales indicated no difference, some a benefit and some a harm of resistance training based on the point

estimate, but uncertainty in the results existed based on confidence intervals. A possible benefit of resistance training was identified for functional capacity from one study, but there was uncertainty in the size of this effect and whether it was clinically important. Results suggested no difference based on the point estimate for depression and incidence of adverse events (defined as 'harm') from one study each, while another study reporting adverse events leading to withdrawal suggested a harm of resistance training, though the size of the difference was uncertain based on confidence intervals.

Vestibular/balance training vs. control (waitlist control, routine care, information only)

Depending on the outcome, up to three studies (up to 227 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. All outcomes were low to very low quality based on GRADE. Across all fatigue scales reported, including total and subscales for three different scales, point estimates suggested a benefit of vestibular/balance training compared to control. In some cases the confidence intervals were also consistent with this conclusion, but for others there was uncertainty in the size of the effect and whether it was clinically important. Two studies reported quality of life data using different scales; although all suggested increased (better) scores in the intervention group, only one of these was considered to be a clinically important difference and even for this result, confidence intervals meant there was uncertainty in the size of the effect and whether it was clinically important. One study each reported data for disability (EDSS) and cognitive function, with point estimates suggesting a harm and benefit for these outcomes, respectively; however, confidence intervals demonstrated uncertainty in the size of the effects for both. Two studies reported data for depression using different scales, with both suggesting a benefit of intervention based on the point estimate but there being uncertainty in the size and direction of effect. Data for adverse events in general and those leading to withdrawal suggested no clinically important difference between groups.

Vestibular/balance training vs. standard neurorehabilitation

One study including 23 people covered this comparison at 4 weeks and reported a measure of fatigue severity and functional ability. In both cases the point estimate suggested no clinically important difference between the two groups and evidence was very low quality based on GRADE.

Resistance training + aerobic exercise vs. control (waitlist control, no intervention, information only)

Depending on the outcome, up to three studies (up to 312 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. All outcomes were low to very low quality based on GRADE. Across all fatigue scales reported, including subscales and/or total scores for three different scales, point estimates suggested a benefit of resistance + aerobic exercise training compared to control. In some cases the confidence intervals were also consistent with this conclusion, but for others there was uncertainty in the direction and/or size of the effect and whether it was clinically important. Point estimates also suggested possible benefits for the intervention in

terms of quality of life and depression, but confidence intervals indicated uncertainty in the direction and/or size of the effect. Data for adverse events and adverse events leading to withdrawal suggested no clinically important difference between groups based on point estimates.

Resistance training + balance exercises vs. control (no intervention, waitlist control)

Depending on the outcome, up to two studies (up to 142 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. All outcomes were low to very low quality based on GRADE. Both studies assessed fatigue using the Fatigue Severity Scale and the pooled point estimate suggested a possible benefit of the intervention compared to control, though confidence intervals indicated uncertainty in the direction and size of the effect. Quality of life data was available from both studies, but different scales were used; many subscales suggested a benefit of intervention based on point estimates, some suggested a harm and some no difference. However, as with other outcomes, uncertainty existed based on confidence intervals. Similarly, results from one and two studies for depression and adverse events leading to withdrawal, respectively, suggested a possible benefit of the intervention, though confidence intervals highlighted uncertainty in these conclusions.

Vestibular/balance training + aerobic exercise vs. control (education only)

One study including 32 people covered this comparison at 8 weeks and reported results for a fatigue measure, both in terms of the total score and individual subscales. For the total and three individual subscales, the point estimate suggested a clinically important benefit of the intervention compared to the control, with confidence intervals also being consistent with this conclusion in all cases. Quality was very low for all four outcomes based on GRADE.

Resistance training + balance exercise + aerobic exercise vs. control (usual care, no intervention)

Depending on the outcome, up to three studies (up to 64 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. All outcomes were very low quality based on GRADE. Across all fatigue scales reported, including subscales and/or total scores for two different scales, point estimates suggested a benefit of resistance + balance + aerobic exercise training compared to control. In some cases the confidence intervals were also consistent with this conclusion, but for others there was uncertainty in the size of the effect and whether it was clinically important. Quality of life data was available from five studies, but all reported different scales; across the eight different scales or subscales that results were available for, the point estimate suggested a benefit of the intervention in seven cases while for the other no difference was indicated. For five of the seven suggesting a possible benefit based on the point estimate, the confidence intervals were also consistent with this conclusion and for the other two uncertainty existed based on confidence intervals. Adverse events leading to withdrawal were reported by one study, with the point estimate suggesting no difference between the two groups and uncertainty being present based on confidence intervals.

Resistance + balance + aerobic exercise vs. massage

One study including 24 people covered this comparison at 5 weeks and reported results for a fatigue measure, Fatigue Severity Scale, with the quality of the evidence very low based on GRADE. The point estimate suggested a clinically important benefit of the combined exercise intervention compared to the massage group, though confidence intervals indicated uncertainty in the size and direction of the effect.

Yoga vs. control

Depending on the outcome, up to two studies (up to 83 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. All outcomes were low to very low quality based on GRADE. Across studies, four different fatigue scales were reported which included eleven different scales and subscales. For seven of these eleven subscales, point estimates suggested a benefit of yoga compared to control, though confidence intervals were only consistent with this conclusion for three scales or subscales and the confidence intervals for the remaining four suggested uncertainty in the direction and/or size of the effect. The point estimates for the other four fatigue scales or subscales suggested no difference between the two groups. Quality of life was reporting across studies using three different scales, including thirteen different subscales reported. Results varied, with point estimates suggesting a benefit, harm or no difference for yoga compared to control depending on the specific subscale. In most cases there was uncertainty in the direction and/or size of the effect based on confidence intervals but there were two subscales of the SF-36 where a benefit of yoga was demonstrated and confidence intervals were also consistent with this conclusion. Data for cognitive functions was only available from one study, with the point estimate suggesting no difference between groups. Results for depression and anxiety suggested a benefit of yoga based on the point estimates, but again there was uncertainty in this conclusion based on confidence intervals. Adverse events (MS exacerbation specifically) and adverse events leading to withdrawal were reported by a single study each, with results suggesting no difference and a possible benefit of yoga, respectively, though the for the latter confidence intervals indicated uncertainty in the size of the effect.

Pilates vs. control (waitlist, no intervention)

Depending on the outcome, up to three studies (up to 120 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by two of these three studies. All outcomes were very low quality based on GRADE. All three studies assessed fatigue using the Modified Fatigue Impact Scale, with two reporting total scale score and individual subscales and the other only reporting the total score. Results for the four scores all indicated a possible benefit of Pilates vs. control based on the point estimate; there was uncertainty in the size of effect for the total score, cognitive sub score and psychosocial sub score, but for the physical subscale confidence intervals were also consistent with a benefit of Pilates. One study also suggested a benefit of Pilates in terms of the fatigue subscale of Profile of Mood States (POMS)-B, with confidence intervals and the point estimate consistent with this conclusion. Various measures of mood, including anxiety and depression, were reported by two of the

studies. For measures of anxiety, the conclusion differed depending on the scale; point estimates for the two subscales of the STAY scale suggested a benefit of Pilates, with there being uncertainty based on confidence intervals for one but not the other, while for the HADS anxiety scale the point estimate suggested a possible harm of Pilates with uncertainty in the direction and size of effect based on confidence intervals. All three depression scales reported by the study suggested a benefit of Pilates based on point estimates; in two cases confidence intervals were consistent with this conclusion. Results for the total mood subscale of POMS-B also suggested a benefit of Pilates vs. control, based on the point estimate and confidence intervals. Information on adverse events was available from one two studies and suggested no difference between the two groups, with one of these studies also reporting discontinuations that may have been related to intervention and indicating no clinically important difference between the two groups, though there was uncertainty based on confidence intervals.

Pilates + balance training vs. relaxation

One study including 39 people covered this comparison at 8 weeks; however, all clinical outcomes other than adverse events were reported as median values with their interquartile range, meaning GRADE analysis could not be performed and limiting the interpretation of these results. Risk of bias assessment for this study led to downgrading of two increments for all outcomes, as risk of bias was graded 'high'. Results for fatigue as measured by the Fatigue Severity Scale indicated lower (better) scores in the two Pilates + balance training groups vs. relaxation; with the P-value between the three groups being reported as $P<0.001$. Data for adverse or harmful events suggested no difference between the two groups and lack of adherence, as measured by discontinuation due to work intensity, suggested increased events in the Pilates + balance training group compared to relaxation, though the size of this effect varied based on confidence intervals.

Exercise-based interventions vs. other exercise interventions

Resistance training vs. aerobic exercise

One study including 32 people covered this comparison at 8 weeks, with evidence for all outcomes being very low quality based on GRADE. Results demonstrated either a possible harm (physical subscale) or benefit (cognitive and psychosocial subscale) of resistance training on Modified Fatigue Impact Scale subscales based on point estimates, though there was uncertainty in the size and direction of the effect. Similarly, the results indicated a possible benefit and possible harm of resistance training on SF-36 physical and mental composite scores, respectively, with confidence intervals again indicating uncertainty in the size and direction of effect. A possible benefit of resistance training as measured on a depression scale was also indicated based on the point estimate alone and results for adverse events suggested no difference between the two groups, though there was uncertainty present for both of these outcomes as well.

Vestibular/balance training vs. aerobic exercise

Three studies reported outcomes for this comparison but due to no overlap in outcome reporting pooling was not possible. All reported data for the up to 6-month time-point. All outcomes were very low quality based on GRADE. Results for fatigue

measures, which were Modified Fatigue Impact Scale (one study reporting as a continuous value and another reporting dichotomously as the proportion with any improvement on this scale compared to baseline) and Fatigue Severity Scale, suggested a possible benefit of vestibular/balance training based on the point estimate, with uncertainty in the direction and size of the effect present based on confidence intervals. However, one study reporting those with any improvement on Modified Fatigue Impact Scale (motor) demonstrated no difference based on the point estimate, with uncertainty also present. Other outcomes where the point estimate suggested a benefit but where there was uncertainty present were those with improvement on a quality-of-life measure (motor) compared to baseline, depression scale on a 0 to63 scale and those with improvement on a depression scale compared to baseline, and adverse events. There was one outcome where the point estimate suggested a possible harm of vestibular/balance training and another where the point estimate suggested no difference, with uncertainty based on confidence intervals for both, which were EDSS scale and a depression scale measured on a 0 to 21 scale, respectively.

Vestibular/balance training vs. resistance training

One study including 51 people covered this comparison at 10 weeks, reporting two outcomes with both being very low quality based on GRADE. Results demonstrated either a possible harm (adverse events leading to withdrawal) or benefit (total score on Modified Fatigue Impact Scale) of vestibular/balance training based on point estimates, though there was uncertainty in the direction and/or size of the effect.

Standard exercises (resistance + balance + aerobic) + high-intensity lower limb resistance training vs. standard exercises alone

One study including 19 people covered this comparison at 12 weeks, reporting two outcomes with both being assessed as low or very low quality based on GRADE. Results demonstrated either a possible harm (Fatigue Severity Scale score) or benefit (adverse events) of standard exercises + high-intensity lower limb resistance training based on point estimates, though there was uncertainty in the size and direction of the effect.

Resistance + aerobic exercise vs. yoga

Three studies reported outcomes for this comparison but due to no overlap in outcome reporting pooling was not possible. All reported data for the up to 6-month time-point. All outcomes were low to very low quality based on GRADE. Results for the following outcomes, based on point estimate, suggested a possible benefit of resistance + aerobic exercise: Fatigue as measured on physical and cognitive subscales of the Modified Fatigue Impact Scale and the Fatigue Severity Scale; and physical and psychological subscales of the MSIS-29. Of these outcomes, there was uncertainty for all based on confidence intervals apart from the Fatigue Severity Scale, where confidence intervals were also consistent with this conclusion. No difference (total score on Modified Fatigue Impact Scale, a depression scale and adherence) and a possible harm (adverse events leading to withdrawal) were also identified based on point estimates, though for all of these outcomes there was uncertainty based on confidence intervals in the size and direction of effect.

Yoga vs. aerobic exercise

Depending on the outcome, up to two studies (up to 77 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. All outcomes were low to very low quality based on GRADE. Based on point estimates, results suggested a possible harm for various outcomes (Fatigue Severity Scale, physical fatigue, reduced activity, reduced motivation and mental fatigue subscales of the Multidimensional Fatigue Inventory, Rhoten Fatigue Scale and cognitive function as measured by Stroop colour word interference test), though confidence intervals were only consistent with this conclusion for one outcome (Rhoten Fatigue Scale) and there was uncertainty in the direction and/or size of the effect for the others. No difference was suggested for the remaining subscale of the Multidimensional Fatigue Inventory (general fatigue), eleven of the twelve reported quality of life subscales, an anxiety scale and MS exacerbations (adverse events) based on point estimates, with uncertainty present based on confidence intervals. Possible benefits of yoga identified were the mental health composite on the MSQoL-54 scale and a depression scale, with uncertainty in the results based on confidence intervals.

Pilates vs. resistance + balance exercises

One study including 20 people covered this comparison at 8 weeks, reporting subscale scores for a fatigue scale, a quality-of-life scale and a measure of cognitive function and depression, with all outcomes being very low quality based on GRADE. Results demonstrated either a possible harm (quality of life measured by The Multiple Sclerosis International Quality of Life questionnaire) or benefit (cognitive and psychosocial subscales of Modified Fatigue Impact Scale, cognitive function measured by Paced Auditory Serial Addition Test and depression as measured by Beck Depression Inventory) of Pilates based on point estimates, though there was uncertainty in the direction and/or size of the effect for all outcomes apart from the cognitive function test.

Fatigue or self-management interventions vs. control

Fatigue/energy management programme vs. control (waitlist, no intervention, information only)

Depending on the outcome, up to four studies (up to 425 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though some outcomes were only reported by one or two studies. All outcomes were low to very low quality based on GRADE. Depending on the fatigue scale, point estimates suggested no difference (Fatigue Severity Scale on 1 to 7 and 9 to 63 scales, including continuous measures and one study reporting a dichotomous measure of those with 0.5-point reduction in fatigue compared to baseline, Modified Fatigue Impact Scale, including total score and subdomain scores for physical, cognitive and psychosocial fatigue, and the fatigue subscale of Checklist Individual Strength scale), a possible benefit (Fatigue Impact Scale, including total score and subdomain scores for physical, cognitive and psychosocial fatigue) or a possible harm (those with 10-point improvement on Modified Fatigue Impact Scale as reported by one study) of the fatigue or energy management programme. However, based on confidence intervals there was uncertainty in the direction and size of effect

for all of these outcomes. Results for most quality-of-life subdomains (SF-36 physical function, body pain, general health, social function, role-emotional and mental health, as well as the psychological subdomain of MSIS-29) suggested no difference between groups based on point estimates and results for four suggested a possible benefit (SF-36 role-physical, SF-36 vitality and MSIS-29 total score and physical subdomain score) of the intervention, with uncertainty in all of these results being present. No difference, with uncertainty based on confidence intervals, was also suggested for concentration measured on Checklist Individual Strength, adverse events and depression.

For the >6-month time-point, one study of 69 to 86 people covered this comparison at 12 months, with outcomes assessed as moderate to very low quality based on GRADE. Based on point estimates, results for all but two outcomes (SF-36 role-physical and role-emotional, with point estimate suggesting a benefit of intervention) suggested no difference (fatigue measured by Fatigue Severity Scale, total and subdomain scores on Modified Fatigue Impact Scale, and Checklist Individual Strength, remaining SF-36 subscales, concentration measured by Checklist Individual Strength, serious adverse events and adverse events leading to withdrawal, and adherence to the programme) between the two groups. However, there was uncertainty in the direction and size of effect for all outcomes.

Fatigue/energy management programme vs. relaxation

One study including 25 people covered this comparison at 3 months, with all outcomes being assessed as very low quality based on GRADE. Results demonstrated either a possible benefit (total, physical, cognitive and psychosocial scores on the Modified Fatigue Impact Scale, and physical functioning, physical pain and vitality subdomains of the SF-36 scale), harm (physical activity and motivation subdomains on the Checklist Individual Strength scale and role-physical function and health change subdomains of SF-36) or no difference (total, concentration and subjective fatigue scores on Checklist Individual Strength, and general health, social functioning, mental health and role-emotional function subdomains of SF-36) for the intervention vs. relaxation based on point estimates, though there was uncertainty in the direction and/or size of the effect for all but one outcome (SF-36 physical pain subdomain).

Self-management programme vs. control

Three studies reported outcomes for this comparison with all three-reporting data for the <6-month time-point, but due to no overlap in outcome reporting pooling was not possible, with 63 to163 people analysed. One of the studies also reported data for outcomes as 12 months, which was included in the >6-month time-point. Outcomes were moderate to very low quality based on GRADE, with all but one being low or very low quality. Results for the following outcomes, based on point estimate, suggested a possible benefit of the self-management programme at the <6-month time-point: Fatigue Severity Scale (scale 1 to 7), fatigue measured by visual analogue scale, total score and at least 10-point reduction on Modified Fatigue Impact Scale compared to baseline, quality of life measured by MS Impact Scale-29 (physical and psychological domains) and anxiety and depression measured by the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 (as a continuous outcome). Of these outcomes, there was uncertainty in the direction and/or size of effect for all but two (Fatigue Severity Scale and MS Impact Scale-29 Physical subdomain) based on confidence intervals. No difference (quality of life

measured by SF-8 physical and mental subdomains, depression measured by proportion with 50% reduction compared to baseline on Patient Health Questionnaire-9, adverse events leading to withdrawal and serious adverse events) was also identified based on point estimates, though for the depression outcome there was uncertainty based on confidence intervals in the size of the effect. Results for treatment adherence suggested an important difference between the two groups, with fewer adhering in the intervention group; however, there was uncertainty based on confidence intervals in the size and direction of the effect.

One outcome reported data at the >6-month time-point for this comparison, with outcomes reported at 12 months, 140 to 163 people analysed, and evidence assessed as moderate to very low quality based on GRADE. Based on point estimates, results suggested no difference (total score on the Modified Fatigue Impact Scale, quality of life measured on physical and mental subdomains of SF-8, depression measured by proportion with 50% reduction compared to baseline on Patient Health Questionnaire-9 and serious adverse events) or a possible benefit (at least 10-point reduction on total Modified Fatigue Impact Scale score compared to baseline and Patient Health Questionnaire-9 score for depression as a continuous measure) of the self-management programme. For all but two outcomes (SF-8 mental health domain and serious adverse events) there was uncertainty based on confidence intervals in the direction and/or size of effect.

Fatigue or self-management interventions vs. each other or other interventions

Fatigue/energy management programme vs. general self-management programme

Two studies reported outcomes for this comparison at the up to 6-month time-point but due to no overlap in outcome reporting pooling was not possible, with up to 218 people analysed depending on the outcome and outcomes assessed as moderate to very low quality based on GRADE. Results demonstrated either a possible benefit (depression scale) or no difference (Modified Fatigue Impact Scale total score, relapses and adherence measured in terms of those completing at least four sessions) based on point estimates, but there was uncertainty in the direction and size of the effect for all outcomes.

One study reported one of the outcomes (total score on Modified Fatigue Impact Scale) at 12 months, with evidence assessed as very low quality based on GRADE and the point estimate suggesting a possible benefit of the fatigue management programme compared to general self-management programme, with uncertainty in the size and direction of effect present.

CBT/motivational interviewing/mindfulness interventions

CBT vs. control

Two studies reported outcomes for this comparison at the up to 6-month time-point but due to no overlap in outcome reporting pooling was not possible, with 74 to140 people analysed. Outcomes were assessed as moderate to very low quality based on GRADE. Results for the following outcomes, based on point estimate, suggested a possible benefit of CBT: Fatigue as measured by Checklist Individual Strength, Fatigue Severity Scale and Piper Fatigue Scale, a depression scale, and vitality, physical role functioning and social functioning subdomains of SF-36. Of these

outcomes, there was uncertainty in the direction and/or size of effect for all based on confidence intervals. No difference (total score and physical, cognitive and psychosocial subdomain scores on Modified Fatigue Impact Scale, an anxiety scale, physical functioning, emotional role functioning, mental health, general health and bodily pain SF-36 subdomains, concentration subdomain on Checklist Individual Strength and serious adverse events) was also identified based on point estimates, though for all of these outcomes there was uncertainty based on confidence intervals in the direction and/or size of effect.

One outcome reported data at the >6-month time-point for this comparison, with outcomes reported at 12 months, 74 people analysed and evidence assessed as moderate to very low quality based on GRADE. Based on point estimates, results suggested no difference (Fatigue as measured by Checklist Individual Strength, Fatigue Severity Scale and total score and cognitive and psychosocial subdomain scores on Modified Fatigue Impact Scale, all but one SF-36 subdomain, concentration measured on Checklist Individual Strength and serious adverse events) or a possible harm (physical subdomain on Modified Fatigue Impact Scale and physical role functioning subdomain of SF-36) of CBT. For all outcomes there was uncertainty based on confidence intervals in the direction and/or size of effect.

CBT vs. relaxation

One study including 72 people covered this comparison at 5 months, with all outcomes being moderate to low quality based on GRADE. Results demonstrated either a possible benefit (Chalder Fatigue Scale, fatigue-related impairment measured by Work and Social Adjustment scale, a scale measuring depression and acceptability of treatment measured by rating usefulness of treatment on 0 to 4 scale) or no difference (scale measuring anxiety) for CBT vs. relaxation based on point estimates, though there was uncertainty in the direction and/or size of the effect for all outcomes.

The same study reported outcomes at the >6-month time-point, reporting outcomes at 8 months, with 72 people being analysed and evidence of moderate to low quality based on GRADE. Results demonstrated either a possible benefit (Chalder Fatigue Scale, fatigue-related impairment measured by Work and Social Adjustment scale and a scale measuring depression) or no difference (scale measuring anxiety) for CBT vs. relaxation based on point estimates, though there was uncertainty in the direction and/or size of the effect for all outcomes.

Motivational interviewing vs. control

One study including 60 people covered this comparison at 9 weeks, reporting a single outcome measuring fatigue (Modified Fatigue Impact Scale total score) which was assessed as very low quality based on GRADE. Results demonstrated a clinically important benefit of motivational interviewing for this outcome, with the point estimate and confidence intervals being consistent with this conclusion.

Mindfulness vs. control (usual care)

One study including 150 people covered this comparison at 6 months, with all outcomes very low quality based on GRADE. Results demonstrated a clinically important benefit of mindfulness for all four outcomes reporting (Modified Fatigue

Impact Scale total score, Hamburg Quality of Life Questionnaire in Multiple Sclerosis and measures of depression and anxiety), with point estimates and confidence intervals being consistent with the same conclusion in all cases.

Dietary interventions

Diet vs. control

Three studies reported outcomes for this comparison at the up to 6-month time-point but due to limited overlap in outcome reporting pooling was only possible for the outcomes of EDSS score and adverse events. Up to 183 people were analysed across outcomes, though many had <40 analysed. Diets also differed across studies (modified Mediterranean in two studies and Palaeolithic in one study). Outcomes were of low to very low quality based on GRADE. For some outcomes (Fatigue Severity Scale on a 1 to 9 scale as continuous measure and proportion achieving 1-point reduction on this scale at follow-up, and Neurological Fatigue Index in MS), point estimates and confidence intervals were consistent with a clinically important benefit of the dietary intervention. For most of the remaining outcomes (total and sub scores of the Modified Fatigue Impact Scale, proportion with reductions on quality-of-life scales, continuous measure of quality of life through MSIS-29, EDSS score, and adverse events leading to withdrawal), point estimates suggested a possible benefit of the intervention but there was uncertainty in the result based on point estimates. The results for adverse events overall suggested no difference between groups based on the point estimate, and for adherence suggested worse adherence in the intervention group compared to control, though there was uncertainty based on confidence intervals for adherence. There was concern about the selection of thresholds for improvement for some continuous outcomes in one study (Fatigue Severity Scale and physical and mental domains of a quality of life measure) reported in the form of dichotomous outcomes, as these thresholds did not appear to be pre-specified in the methods section and differed without explanation (for example, threshold of 5 for improvement on mental health domain of quality of life but ‘any improvement’ for the physical domain of the same quality of life scale).

Diet (individualised) vs. standard healthy diet recommendations

One study including 100 people covered this comparison at 12 weeks, with all outcomes of low to very low quality based on GRADE. Results demonstrated a possible harm of individualised diet for one outcome (psychosocial subdomain of Modified Fatigue Impact Scale) based on the point estimate. However, differences in baseline values for this outcome make the results misleading, as there was actually an improvement in the intervention group and a slight worsening of the score in the control group. Uncertainty was also present based on confidence intervals. For all other outcomes (total score and physical and cognitive subdomains of the Modified Fatigue Impact Scale, physical and mental subdomains on a quality-of-life scale and relapse events leading to withdrawal) the point estimate suggested no difference between groups, with uncertainty in the direction and size of effect. For the physical and mental health subdomains of quality of life, baseline differences further supported the conclusion of no difference between groups, as although scores were higher/lower in the intervention group at the end of the treatment period, these differences also existed at baseline and very little change was observed for both arms at follow-up.

One study covered a similar comparison, though the dietary intervention differed slightly, and this intervention was compared to a different set of standard healthy diet recommendations. These results were not combined with those mentioned above as the study reported the results at >6 months, which was listed as a separate time-point in the protocol for this review. This study consisted of 56 to 72 people that were analysed at 1 year, with all outcomes of low to very low quality based on GRADE. Results demonstrated a possible benefit of the intervention in terms of fatigue measured on the Modified Fatigue Impact Scale total score, though there was uncertainty in the size of this effect based on confidence intervals. The study reported various measures of cognitive function. For most of these the results suggested no difference between the two groups. A possible benefit was suggested for the delayed recall component of the California Verbal Learning Test II and a possible harm or worse outcome in the intervention was suggested for the Brief Visuospatial Memory Test-Revised and the total score on Delis-Kaplan Executive Function System, though there was also uncertainty in the direction and size of effect for all cognitive outcomes. The study suggested that adherence was higher in the intervention compared to control group, with confidence intervals consistent with this conclusion as well.

Wahls diet (modified Palaeolithic elimination diet) vs. Swank diet (low-saturated fat diet)

A single study compared outcomes between two different diets, the Wahls diet and the Swank diet, which was a 6-month intervention. Across two different fatigue scales (Fatigue Severity Scale and Modified Fatigue Impact Scale), the point estimates for most (Fatigue Severity Scale and total Modified Fatigue Impact Scale score as well as physical and psychosocial sub scores of this scale) results indicated a possible benefit of the Wahls diet compared to the Swank diet, with uncertainty in the direction and size of the effect based on confidence intervals. The point estimate for the cognitive sub score of Modified Fatigue Impact Scale suggested no difference between the two groups. For quality-of-life measures, based on point estimates, no difference between the groups was indicated for the mental composite of the MSQoL-54 scale and a possible benefit for the physical composite of this scale, though for both the confidence intervals suggested uncertainty in the direction and size of the effect. No events were reported for serious adverse events in either group and despite increased adherence in the Swank diet compared to Wahls, the absolute effect was <100 per 1000 meaning it did not reach the threshold for an important difference. Evidence for all outcomes was very low quality based on GRADE and the study was small with 72 people analysed at the end of the intervention.

Relaxation interventions

Relaxation vs. control (waitlist)

One study including45 people covered this comparison at 8 weeks, with a single outcome measure of fatigue (total score on Modified Fatigue Impact Scale) reported and of very low quality based on GRADE. Results demonstrated a possible benefit of the intervention based on the point estimate, but there was uncertainty in this result based on confidence intervals.

Acupressure vs. control (touching only/sham)

Two studies covered this comparison (100 and 86 people, respectively) covered this comparison at 4 weeks. Although both reported fatigue using the Fatigue Severity Scale, the scale used was unclear in one study and the numbers did not match scales that are usually used for this outcome scale, so the two studies were not pooled. Results for one study on this scale suggested a possible benefit of acupressure based on the point estimate, though confidence intervals indicated uncertainty in the size and direction of effect. The second study reporting this outcome suggested no difference between the two groups according to the point estimate. Depression was also reported by one study, with the point estimate suggesting a possible benefit in the intervention group, though there was uncertainty in the size of the effect. Quality of the evidence was low to very low quality based on GRADE.

Reflexology/relaxation vs. control (usual care)

Depending on the outcome, one or two studies of 50 to110 people covered this comparison at 8-12 weeks. Two studies reported results for the Fatigue Severity Scale when comparing foot reflexology with control. The results suggested a possible benefit of foot reflexology compared to control, with confidence intervals consistent with this conclusion. One of the studies reported quality of life outcomes for this comparison (MS Quality of Life-54 score, including physical and mental composites and health change scores); the results for all three sub scores suggested a possible benefit of the intervention based on point estimates and confidence intervals. In addition to results comparing foot reflexology vs. control, one of the studies also compared a relaxation group vs. control. For relaxation vs. control, the only outcome reported was Fatigue Severity Scale; the point estimate also suggested a possible benefit of the intervention, but there was uncertainty in the size of the effect based on confidence intervals. All outcomes were low to very low quality based on GRADE.

Massage vs. control (usual care/no intervention)

Up to three studies (including 60 to164 people) covered this comparison at 4-7 weeks. All three studies reported a measure of fatigue (Fatigue Severity Scale), which was very low quality based on GRADE. Results demonstrated a clinically important benefit of the intervention based on the point estimate, but there was uncertainty in the size of the effect based on confidence intervals. One study (80 people) also reported a VAS scale for fatigue relief and effectiveness of fatigue reduction, and another study (60 people) reported results for an anxiety scale. For both of these outcomes, quality was very low, and results indicated a possible benefit of intervention based on the point estimate, though confidence intervals indicated uncertainty in the size of the effect.

Reflexology vs. non-specialised foot massage

One study including 63 people covered this comparison at 4 weeks, with total and subdomain scores for a fatigue scale reported as well as a measure of anxiety. All outcomes were of very low quality based on GRADE. Results demonstrated a possible benefit of the reflexology vs. non-specialised foot massage based on the point estimates for all outcomes (Fatigue Impact Scale, including total score and physical, cognitive and psychosocial subdomain scores, and anxiety as measured by

State-Trait Anxiety Inventory). However, for all outcomes confidence intervals indicated uncertainty in the direction and/or size of the effect.

Multi-component interventions vs. control

Functional electrical stimulation + aerobic exercise vs. control (waitlist)

One study of 12 people covered this comparison at 12 weeks, reporting multiple fatigue scales as both a continuous and dichotomous measure, a quality-of-life scale, a measure of depression and adverse events leading to withdrawal, with all outcomes being of low to very low quality based on GRADE. Results demonstrated either a possible harm (adverse events leading to withdrawal) or benefit (fatigue measured on Modified Fatigue Impact Scale and Fatigue Scale of Motor and Cognitive Functions, both as continuous and dichotomous measures, mental and physical composite scores of MSQoL-54 and a depression scale) of the intervention based on point estimates, though there was uncertainty in the direction and size of the effect for all outcomes apart from the physical health composite of MSQoL-54.

Massage + exercise (resistance, balance + aerobic) vs. control (no intervention)

One study of 24 people covered this comparison at 5 weeks and reported a measure of fatigue, with evidence being very low quality based on GRADE. Based on the point estimate, results suggested a clinically important benefit of the intervention for the Fatigue Severity Scale, with confidence intervals also being consistent with this conclusion.

Aerobic exercise + fatigue self-management vs. control (information only)

One study of 139 people covered this comparison at 24 weeks and reported a measure of fatigue and quality of life as well as adverse event and adherence outcomes, with evidence being of moderate or very low quality for all outcomes based on GRADE. Based on the point estimate, results for clinical outcomes suggested a possible harm (orthopaedic problems), benefit (fatigue measured by Fatigue Impact Scale and physical and mental subdomains of the MSIS-29 scale) or no difference (exacerbations and falls) for the intervention, with confidence intervals indicating uncertainty in the direction and size of the effect. For adherence outcomes, point estimates suggested no difference for completion of all 1-1 calls but more people in the intervention completed all group calls with or without at least one make-up session, although for both, there was uncertainty in the result based on confidence intervals.

Fatigue management + CBT vs. control (local/standard care)

Depending on the outcome, up to two studies (up to 315 people analysed) reported data that could be pooled for this comparison for the up to 6-month time-point, though most outcomes were only reported by a single study. Outcomes were of high to very low quality based on GRADE. Based on point estimates, results suggested a possible benefit for various outcomes (fatigue measured by total score on Modified Fatigue Impact Scale and Chalder fatigue scale, self-efficacy measured on MS Fatigue Self-Efficacy scale and a measure of anxiety), though confidence intervals

were only consistent with this conclusion for one outcome (Modified Fatigue Impact Scale total score) and there was uncertainty in the direction and size of the effect for the others. No difference was suggested other measures of fatigue (Global Fatigue Severity on a 1 to 7 scale and total, motor and cognition scores on the Fatigue Scale of Motor and Cognition scale), quality of life measured by total MSIS-29 score, cognitive function based on the MS neuropsychological screening questionnaire, a measure of depression and withdrawal due to adverse events (relapse) based on point estimates, with uncertainty present based on confidence intervals.

One study also reported data at >6 months, with quality being low to very low quality for all outcomes based on GRADE. Most outcomes (global fatigue severity based on a 1 to 7 scale, self-efficacy measured on MS Fatigue Self-Efficacy scale, and MSIS-29 total score and physical subdomain score) demonstrated no difference based on the point estimate, with results for only one outcome (SF-36 vitality score) suggesting a possible benefit of the intervention, though for all outcomes there was uncertainty present based on confidence intervals.

Multidisciplinary rehabilitation + fatigue self-management vs. control (consultation only)

One study of 46 people covered this comparison at 3 months and reported three measures of fatigue, a measure of quality of life and a measure of functional independence, with evidence being of very low quality for all outcomes based on GRADE. Based on the point estimate, results suggested a possible harm (total score and physical and psychosocial subdomain scores on Modified Fatigue Impact Scale) or benefit (cognitive subdomain score on Modified Fatigue Impact Scale, Fatigue Severity Scale, physical and mental function subdomain scores on MSIS-29, functional independence measure on 1 to 7 scale and the Checklist of Individual Strength fatigue scale, including total score, subjective fatigue, concentration, motivation and physical activity subdomains) of the intervention, with confidence intervals indicating uncertainty in the direction and size of the effect for all but one outcome (functional independence measure).

Multidisciplinary rehabilitation (medical, exercise, counselling and fatigue self-management) vs. no rehabilitation in those treated with methylprednisolone for a relapse

One study of 39 people covered this comparison at 3 months and reported a fatigue scale, with evidence being of very low quality based on GRADE. Based on the point estimate, results for the outcome reported (Fatigue Severity Scale) suggested a possible benefit of the intervention, with confidence intervals indicating uncertainty in the direction and size of the effect.

Self-management programme + exercise vs. control (waitlist)

One study of 14 people covered this comparison at 6 weeks and reported a fatigue scale, a measure of quality of life and adverse events, with evidence being of very low quality for all outcomes based on GRADE. Based on the point estimate, results for the fatigue outcomes (WEIMuS fatigue scale, including total score and mental and physical subdomain scores) suggested a possible harm of the intervention and results for the quality-of-life scale and adverse events suggested no difference

between the two groups, with confidence intervals indicating uncertainty in the direction and size of the effect for all outcomes.

Resistance + aerobic exercise + CBT vs. control (waitlist)

One study of 107 to 109 people covered this comparison at 3 months and reported a fatigue scale, two quality of life measures, a measure of disability, a cognitive function measure and adverse events (relapse) leading to withdrawal, with evidence being moderate to very low quality based on GRADE. Based on the point estimate, results for the fatigue outcomes (Modified Fatigue Impact Scale, including total score and physical, cognitive and psychosocial subdomain scores) and quality of life measured by MSQoL-54 suggested a possible benefit of the intervention and results for EQ-5D quality of life, EDSS scale, cognitive function measured by Paced Auditory Serial Addition Test and MS relapse leading to withdrawal suggested no difference between the two groups, with confidence intervals indicating uncertainty in the direction and/or size of the effect for all but one outcome (psychosocial subdomain of Modified Fatigue Impact Scale).

Outcomes were also reported for this study at the >6-month time-point, with 99 to 120 people analysed and evidence being moderate to very low quality based on GRADE. Based on the point estimate, results for the psychosocial subdomain of the Modified Fatigue Impact Scale and MS relapse suggested a possible benefit of the intervention and results for all other outcomes (Modified Fatigue Impact Scale, including total score and physical and cognitive subdomains, quality of life measured by MSQoL-54 and EQ-5D, EDSS scale, cognitive function measured by Paced Auditory Serial Addition Test and MS relapse leading to withdrawal) suggested no difference between the two groups, with confidence intervals indicating uncertainty in the direction and size of the effect for all outcomes.

Multi-component interventions vs. other intervention

Massage + exercise (resistance, balance + aerobic) vs. exercise only

One study of 24 people covered this comparison at 5 weeks and reported a measure of fatigue, with evidence being very low quality based on GRADE. Based on the point estimate, results suggested a possible harm of the intervention for the Fatigue Severity Scale, with confidence intervals indicating uncertainty in the direction and size of the effect.

Massage + exercise (resistance, balance + aerobic) vs. massage only

One study of 24 people covered this comparison at 5 weeks and reported a measure of fatigue, with evidence being very low quality based on GRADE. Based on the point estimate, results suggested a possible benefit of the intervention for the Fatigue Severity Scale, with confidence intervals indicating uncertainty in the direction and size of the effect.

Aerobic exercise + fatigue self-management vs. aerobic exercise only

One study of 139 people covered this comparison at 24 weeks and reported a measure of fatigue and quality of life as well as adverse event and adherence outcomes, with evidence being moderate to very low quality for all outcomes based on GRADE. Based on the point estimate, results for clinical outcomes suggested a possible harm (falls and orthopaedic problems), benefit (fatigue measured by Fatigue Impact Scale) or no difference (physical and mental subdomains of the MSIS-29 scale, and exacerbations) for the intervention, with confidence intervals indicating uncertainty in the direction and/or size of the effect. For adherence outcomes, point estimates suggested no difference for completion of all group calls with or without at least one make-up session but more people in the intervention group completed all 1-1 calls, although for both, there was uncertainty in the result based on confidence intervals.

Multidisciplinary rehabilitation + fatigue self-management vs. relaxation

One study of 29 people covered this comparison at 4 months and reported a fatigue scale and two subdomains of a quality-of-life scale, with evidence being very low quality for all outcomes based on GRADE. Based on the point estimate, results for one outcome (SF-36 physical functioning) suggested a possible benefit of the intervention and results for two outcomes (Modified Fatigue Impact Scale total score and SF-36 fatigue/vitality subdomain) suggested no difference between the two groups, with confidence intervals indicating uncertainty in the direction and size of the effect for all outcomes.

1.1.11.4 Cost effectiveness and resource use

Four economic studies were identified for this review comparing non-pharmacological interventions for the management of fatigue. Unit costs of the staff and treatment programmes included in the clinical and economic evidence were also presented to aid committee consideration of cost-effectiveness.

The first study was by Tosh (2014), which was a cost-utility analysis of an RCT where people with MS were randomised to either usual care or an exercise program called EXIMS that incorporated aerobic and resistance exercise along with CBT techniques for 12 weeks to encourage improvements to exercise behaviour. The results found that aerobic and resistance exercise in combination with CBT and usual care was cost effective compared to usual care for treating fatigue (incremental cost-effectiveness ratio (ICER): £10,137 per QALY gained). In terms of current practice, the committee agreed that it was typical for physiotherapists and occupational therapists to apply CBT principles like goal setting and that it doesn't need to be a formal CBT intervention delivered by a psychologist.

The second economic evaluation by Thomas (2013) was based on an RCT that compared a six-session group-based fatigue management intervention called FACETS, which was delivered by health professionals, with current local practice for adults with a confirmed diagnosis of MS and significant fatigue levels. This analysis found current local practice alone to be dominant (less costly and more effective) compared to the FACETS combined with current local practice for treating MS-related fatigue. Probabilistic sensitivity analysis (PSA) was not performed on the ICER which limited the robustness of the study findings. The committee noted that it was not nationwide practice that people are automatically referred for fatigue management.

The original economic analysis from the previous MS guideline was also included as part of this review. The analysis was based on a study by Cakit (2010) which included people who had clinically definite relapsing-remitting or secondary progressive MS with an EDSS of less than 6.0. There were three comparators which were control (no intervention), home based resistance and balance and supervised resistance and balance. This study reported SF-36 data that could be mapped to EQ-5D allowing quality-adjusted life years (QALYs) to be estimated and cost-effectiveness to be explored. The results found that in adults with either RRMS or SPMS, 'supervised resistance and balance' was found to be the most cost-effective option compared to home-based resistance and balance and a control group.

The studies were assessed as partially applicable and thought to contain potentially serious limitations due the lack of all relevant comparators for this review and the fact that analyses were based on single RCTs.

The final study included was by Moss-Morris (2012), which was a cost-utility model based on a pilot RCT that assessed the effectiveness of an internet-based CBT self-management programme (MSInvigor8). The study was a small feasibility trial with no long-term follow-up data and no active control. The results suggested that in adults with a MS score >4 on the Fatigue scale, an online CBT program (MSInvigor8) may be cost-effective compared to a control group. Some committee members were surprised that the results showed a significant benefit and was not particularly costly, despite analysis occurring on an intention-to-treat basis with a short time horizon. Of note, this study was assessed as partially applicable with very serious limitations. The main limitation being the study size and high non-completion rate. Some committee members were wary of the effectiveness of online CBT compared to in-person and raised concerns that worsening symptoms of MS may inhibit ease of use of online technology. Variation in current practice for this intervention was also noted; some committee members found that general practice often offers CBT clinical online modules and has done so in the past few years while others had experience with either telephone or face-to-face CBT, making the resource impact of implementing such programmes difficult to estimate. Concerns were also raised over the time taken to complete the online sessions, which ranged between 25-50 minutes, as in-person CBT would have less variation and cautioned that this could affect the benefits of the intervention.

When discussing the costs of exercise programmes, the committee also noted that aerobic and resistance exercise interventions tend not to be costly in terms of equipment. They also noted that while the interventions used in the studies are often provided within clinical practice, the difference is that in a research setting, people are supervised for the duration of the intervention period. Clinical practice, however, faces resource restrictions which often means that people are given home-based programmes where they are not supervised for any length of time. The committee felt that a key benefit from these programmes was the period under supervision which provided the knowledge and confidence to allow people to embed these exercise programmes into their daily lives. Providing additional supervision in clinical practice would have a resource impact on the NHS; some suggested that this supervision would be reflected by the cost per working hour of a physiotherapist and that resistance training programmes typically schedule a couple sessions a week for a minimum period of 8 weeks. This was not universal across the UK, however, as some committee members experienced physiotherapy services that provided as little as six sessions over a two-year period. It was also stated that there would be considerable variation in terms of who provides the intervention and felt that there should be some form of follow-up to monitor progress and adherence. This disparity of treatment provided

across the UK creates uncertainty on the cost of providing these interventions and the potential resource impact.

A recommendation was made to offer people with MS and fatigue a personalised discussion on how they can manage fatigue. This includes discussing mindfulness and CBT techniques as well as other approaches to self-manage fatigue. This type of discussion is current practice and therefore this discussion with not result in a significant resource impact.

Given the clinical evidence and modest economic evidence from Tosh (2014) and the original economic analysis from the last guideline supporting the cost-effectiveness of combined exercise programmes, the committee made a consider recommendation for supervised aerobic and moderate progressive resistance activity combined with CBT for treating fatigue in people with MS. Due to the lack of evidence, the weaker recommendations of advising the use of aerobic, resistive and balance exercises including Yoga and Pilates was given.

1.1.11.5 Other factors the committee took into account

The committee made a research recommendation for future studies to be conduct which are adequately powered to detect a difference in outcomes. They also supported the development of a core outcome set for multiple sclerosis to facilitate the pooling of studies.

1.1.12 Recommendations supported by this evidence review

This evidence review supports recommendations 1.5.2 to 1.5.11 and the research recommendation on non-pharmacological management of fatigue.

1.1.13 References

1. Abonie US, Hettinga FJ. Effect of a tailored activity pacing intervention on fatigue and physical activity behaviours in adults with multiple sclerosis. International Journal of Environmental Research and Public Health. 2020; 18(1):17

2. Afrasiabifar A, Mehri Z, Javad Sadat S, Ghaffarian Shirazi HR. The effect of Orem's self-care model on fatigue in patients with multiple sclerosis: A single blind randomized clinical trial study. Iranian Red Crescent Medical Journal. 2016; 18(8):e31955

3. Ahmadi A, Arastoo AA, Nikbakht M. The effects of a treadmill training programme on balance, speed and endurance walking, fatigue and quality of life in people with multiple sclerosis. International SportMed Journal. 2010; 11(4):389-397

4. Ahmadi A, Arastoo AA, Nikbakht M, Zahednejad S, Rajabpour M. Comparison of the effect of 8 weeks aerobic and yoga training on ambulatory function, fatigue and mood status in MS patients. Iranian Red Crescent Medical Journal. 2013; 15(6):449-454

5. Ahmadi A, Nikbakh M, Arastoo A, Habibi AH. The effects of a yoga intervention on balance, speed and endurance of walking, fatigue and quality of life in people with multiple sclerosis. Journal of Human Kinetics. 2010; 23(1):71-78

6. Amatya B, Khan F, Galea M. Rehabilitation for people with multiple sclerosis: an overview of Cochrane Reviews. Cochrane Database of Systematic Reviews 2019, Issue 1. Art. No.: CD012732. DOI: 10.1002/14651858.cd012732.pub2.

7. Amatya B, Young J, Khan F. Non-pharmacological interventions for chronic pain in multiple sclerosis. Cochrane Database of Systematic Reviews 2018, Issue 12. Art. No.: CD012622. DOI: 10.1002/14651858.cd012622.pub2.

8. Arab M, Radfar A, Madadizadeh N, Pour ZSA, Karzari Z. The effect of massage therapy on fatigue of patients with multiple sclerosis. Journal of Advanced Pharmacy Education and Research. 2019; 9:45

9. Atashi V. The effect of SSBM massage on anxiety and fatigue of patients with multiple sclerosis. Journal of Applied Environmental and Biological Sciences. 2014; 4(8):217-223

10. Backus D, Moldavskiy M, Sweatman WM. Effects of functional electrical stimulation cycling on fatigue and quality of life in people with multiple sclerosis who are nonambulatory. International Journal of MS Care. 2020; 22(4):193-200

11. Barlow J, Turner A, Edwards R, Gilchrist M. A randomised controlled trial of lay-led self-management for people with multiple sclerosis. Patient Education and Counseling. 2009; 77(1):81-89

12. Bastani F, Sobhani M, Emamzadeh Ghasemi HS. Effect of acupressure on fatigue in women with multiple sclerosis. Global Journal of Health Science. 2015; 7(4):375-381

13. Blikman LJM, van Meeteren J, Twisk JWR, de Laat FAJ, de Groot V, Beckerman H et al. Energy conservation management for people with multiple sclerosis-related fatigue: Who benefits? American Journal of Occupational Therapy. 2019; 73(4):7304205040p7304205041-7304205040p7304205049

14. Bohlouli J, Namjoo I, Borzoo-Isfahani M, Poorbaferani F, Moravejolahkami AR, Clark CCT et al. Modified Mediterranean diet vs. traditional Iranian diet: Efficacy of dietary interventions on dietary inflammatory index score, fatigue severity and disability in multiple sclerosis patients. British Journal of Nutrition. 2021:1-11

15. Boldt I, Eriks-Hoogland I, Brinkhof MWG, de Bie R, Joggi D, von Elm E. Non-pharmacological interventions for chronic pain in people with spinal cord injury. Cochrane Database of Systematic Reviews 2014, Issue 11. Art. No.: CD009177. DOI: 10.1002/14651858.cd009177.pub2.

16. Bombardier CH, Cunniffe M, Wadhwani R, Gibbons LE, Blake KD, Kraft GH. The efficacy of telephone counseling for health promotion in people with multiple sclerosis: a randomized controlled trial. Archives of Physical Medicine and Rehabilitation. 2008; 89(10):1849-1856

17. Borji M, Taghinejad H, Salimi AH. The effect of motivational interviewing on fatigue in patients with multiple sclerosis. Archives of Neuroscience. 2018; 5(3):e63436

18. Brichetto G, Spallarossa P, De Carvalho MLL, Battaglia MA. The effect of Nintendo Wii on balance in people with multiple sclerosis: A pilot randomized control study. Multiple Sclerosis. 2013; 19(9):1219-1221

19. Bulguroglu I, Guclu-Gunduz A, Yazici G, Ozkul C, Irkec C, Nazliel B et al. The effects of mat pilates and reformer pilates in patients with multiple sclerosis: A randomized controlled study. NeuroRehabilitation. 2017; 41(2):413-422

20. Cakit BD, Nacir B, Genc H, Saracoglu M, Karagoz A, Erdem HR et al. Cycling progressive resistance training for people with multiple sclerosis: a randomized controlled study. American Journal of Physical Medicine and Rehabilitation. 2010; 89(6):446-457

21. Callesen J, Cattaneo D, Brincks J, Kjeldgaard Jorgensen ML, Dalgas U. How do resistance training and balance and motor control training affect gait performance and fatigue impact in people with multiple sclerosis? A randomized controlled multi-center study. Multiple Sclerosis. 2020; 26(11):1420-1432

22. Carter A, Daley A, Humphreys L, Snowdon N, Woodroofe N, Petty J et al. Pragmatic intervention for increasing self-directed exercise behaviour and improving important health outcomes in people with multiple sclerosis: a randomised controlled trial. Multiple Sclerosis. 2014; 20(8):1112-1122

23. Correale L, Buzzachera CF, Liberali G, Codrons E, Mallucci G, Vandoni M et al. Effects of combined endurance and resistance training in women with multiple sclerosis: A randomized controlled study. Frontiers in Neurology. 2021; 12:698460

24. Curtis L. Unit costs of health and social care 2013. Canterbury. Personal Social Services Research Unit University of Kent, 2013. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2013/index.php

25. Curtis l, Burns A. Unit Costs of Health and Social Care 2017. Canterbury, England. Personal Social Services Research Unit University of Kent, 2017. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2017/

26. Curtis L, Burns A. Unit Costs of Health and Social Care 2020. Canterbury, England. Personal Social Services Research Unit University of Kent, 2020. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2020/

27. Dalgas U, Stenager E, Jakobsen J, Petersen T, Hansen HJ, Knudsen C et al. Fatigue, mood and quality of life improve in MS patients after progressive resistance training. Multiple Sclerosis. 2010; 16(4):480-490

28. Dettmers C, Sulzmann M, Ruchay-Plossl A, Gutler R, Vieten M. Endurance exercise improves walking distance in MS patients with fatigue. Acta Neurologica Scandinavica. 2009; 120(4):251-257

29. Dilek Dogan H, Tan M. Effects of reflexology on pain, fatigue, and quality of life in multiple sclerosis patients: A clinical study. Alternative Therapies in Health and Medicine. 2021; 27(12):AT6133

30. Dodd KJ, Taylor NF, Shields N, Prasad D, McDonald E, Gillon A. Progressive resistance training did not improve walking but can improve muscle performance, quality of life and fatigue in adults with multiple sclerosis: a randomized controlled trial. Multiple Sclerosis. 2011; 17(11):1362-1374

31. Eftekhari E, Etemadifar M. Impact of clinical mat pilates on body composition and functional indices in female patients with multiple sclerosis. Crescent Journal of Medical and Biological Sciences. 2018; 5(4):297-305

32. Ehde DM, Elzea JL, Verrall AM, Gibbons LE, Smith AE, Amtmann D. Efficacy of a telephone-delivered self-management intervention for persons with multiple sclerosis: A randomized controlled trial with a one-year follow-up. Archives of Physical Medicine and Rehabilitation. 2015; 96(11):1945-1958.e1942

33. Feys P, Moumdjian L, Van Halewyck F, Wens I, Eijnde BO, Van Wijmeersch B et al. Effects of an individual 12-week community-located "start-to-run" program on physical capacity, walking, fatigue, cognitive function, brain volumes, and structures in persons with multiple sclerosis. Multiple Sclerosis. 2019; 25(1):92-103

34. Finlayson M, Preissner K, Cho C, Plow M. Randomized trial of a teleconference-delivered fatigue management program for people with multiple sclerosis. Multiple Sclerosis. 2011; 17(9):1130-1140

35. Flachenecker P, Bures AK, Gawlik A, Weiland AC, Kuld S, Gusowski K et al. Efficacy of an internet-based program to promote physical activity and exercise after inpatient rehabilitation in persons with multiple sclerosis: A randomized, single-blind, controlled study. International Journal of Environmental Research and Public Health. 2020; 17(12):4544

36. Fleming KM, Coote SB, Herring MP. The feasibility of pilates to improve symptoms of anxiety, depression, and fatigue among people with multiple sclerosis: an eight-week randomized controlled pilot trial. Psychology of Sport and Exercise. 2019; 45:101573

37. Fleming KM, Coote SB, Herring MP. Home-based pilates for symptoms of anxiety, depression and fatigue among persons with multiple sclerosis: An 8-week randomized controlled trial. Multiple Sclerosis. 2021:13524585211009216

38. García Jalón EG, Lennon S, Peoples L, Murphy S, Lowe-Strong A. Energy conservation for fatigue management in multiple sclerosis: a pilot randomized controlled trial. Clinical Rehabilitation. 2013; 27(1):63-74

39. Garrett M, Hogan N, Larkin A, Saunders J, Jakeman P, Coote S. Exercise in the community for people with minimal gait impairment due to MS: an assessor-blind randomized controlled trial. Multiple Sclerosis. 2013; 19(6):782-789

40. Garrett M, Hogan N, Larkin A, Saunders J, Jakeman P, Coote S. Exercise in the community for people with multiple sclerosis - a follow-up of people with minimal gait impairment. Multiple Sclerosis. 2013; 19(6):790-798

41. Geddes EL, Costello E, Raivel K, Wilson R. The effects of a twelve-week home walking program on cardiovascular parameters and fatigue perception of individuals with multiple sclerosis: a pilot study. Cardiopulmonary Physical Therapy Journal. 2009; 20(1):5-12

42. Gervasoni E, Cattaneo D, Jonsdottir J. Effect of treadmill training on fatigue in multiple sclerosis: a pilot study. International Journal of Rehabilitation Research. 2014; 37(1):54-60

43. Grossman P, Kappos L, Gensicke H, D'Souza M, Mohr DC, Penner IK et al. MS quality of life, depression, and fatigue improve after mindfulness training: a randomized trial. Neurology. 2010; 75(13):1141-1149

44. Grubic Kezele T, Babic M, Stimac D. Exploring the feasibility of a mild and short 4-week combined upper limb and breathing exercise program as a possible home base program to decrease fatigue and improve quality of life in ambulatory and non-ambulatory multiple sclerosis individuals. Neurological Sciences. 2019; 40(4):733-743

45. Hasanpour Dehkordi A. Influence of yoga and aerobics exercise on fatigue, pain and psychosocial status in patients with multiple sclerosis: a randomized trial. Journal of Sports Medicine and Physical Fitness. 2016; 56(11):1417-1422

46. Hayes HA, Gappmaier E, LaStayo PC. Effects of high-intensity resistance training on strength, mobility, balance, and fatigue in individuals with multiple sclerosis: a randomized controlled trial. Journal of Neurologic Physical Therapy. 2011; 35(1):2-10

47. Hebert JR, Corboy JR, Manago MM, Schenkman M. Effects of vestibular rehabilitation on multiple sclerosis-related fatigue and upright postural control: a randomized controlled trial. Physical Therapy. 2011; 91(8):1166-1183

48. Hebert JR, Corboy JR, Vollmer T, Forster JE, Schenkman M. Efficacy of balance and eye-movement exercises for persons with multiple sclerosis (BEEMS). Neurology. 2018; 90(9):e797-e807

49. Heine M, van de Port I, Rietberg MB, van Wegen EEH, Kwakkel G. Exercise therapy for fatigue in multiple sclerosis. Cochrane Database of Systematic Reviews 2015, Issue 9. Art. No.: CD009956. DOI: 10.1002/14651858.cd009956.pub2.

50. Heine M, Verschuren O, Hoogervorst EL, van Munster E, Hacking HG, Visser-Meily A et al. Does aerobic training alleviate fatigue and improve societal participation in patients with multiple sclerosis? A randomized controlled trial. Multiple Sclerosis. 2017; 23(11):1517-1526

51. Hersche R, Weise A, Michel G, Kesselring J, Bella SD, Barbero M et al. Three-week inpatient energy management education (IEME) for persons with multiple sclerosis-related fatigue: Feasibility of a randomized clinical trial. Multiple Sclerosis and Related Disorders. 2019; 35:26-33

52. Hugos CL, Cameron MH, Chen Z, Chen Y, Bourdette D. A multicenter randomized controlled trial of two group education programs for fatigue in multiple sclerosis: Long-term (12-month) follow-up at one site. Multiple Sclerosis. 2019; 25(6):871-875

53. Hugos CL, Chen Z, Chen Y, Turner AP, Haselkorn J, Chiara T et al. A multicenter randomized controlled trial of two group education programs for fatigue in multiple sclerosis: Short- and medium-term benefits. Multiple Sclerosis. 2019; 25(2):275-285

54. Hugos CL, Copperman LF, Fuller BE, Yadav V, Lovera J, Bourdette DN. Clinical trial of a formal group fatigue program in multiple sclerosis. Multiple Sclerosis. 2010; 16(6):724-732

55. Irish AK, Erickson CM, Wahls TL, Snetselaar LG, Darling WG. Randomized control trial evaluation of a modified Paleolithic dietary intervention in the treatment of relapsing-remitting multiple sclerosis: a pilot study. Degenerative Neurological & Neuromuscular Disease. 2017; 7:1-18

56. Karami F, Afrasiabifar A, Doulatabad SN. Comparing the effectiveness of vestibular rehabilitation and frenkel exercise on fatigue reduction in patients with multiple sclerosis: A randomized controlled trial. Iranian Red Crescent Medical Journal. 2018; 20(12):e68913

57. Kargarfard M, Etemadifar M, Baker P, Mehrabi M, Hayatbakhsh R. Effect of aquatic exercise training on fatigue and health-related quality of life in patients with multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 2012; 93(10):1701-1708

58. Kargarfard M, Shariat A, Ingle L, Cleland JA, Kargarfard M. Randomized controlled trial to examine the impact of aquatic exercise training on functional capacity, balance, and perceptions of fatigue in female patients with multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 2018; 99(2):234-241

59. Karpatkin HI, Napolione D, Siminovich-Blok B. Acupuncture and multiple sclerosis: a review of the evidence. Evidence-Based Complementary and Alternative Medicine. 2014; 2014:972935

60. Khan F, Amatya B, Kesselring J, Galea M. Telerehabilitation for persons with multiple sclerosis. Cochrane Database of Systematic Reviews 2015, Issue 4. Art. No.: CD010508. DOI: 10.1002/14651858.cd010508.pub2.

61. Khayeri F, Rabiei L, Shamsalinia A, Masoudi R. Effect of Fordyce Happiness Model on depression, stress, anxiety, and fatigue in patients with multiple sclerosis. Complementary Therapies in Clinical Practice. 2016; 25:130-135

62. Kooshiar H, Moshtagh M, Sardar MA, Foroughipour M, Shakeri MT, Vahdatinia B. Fatigue and quality of life of women with multiple sclerosis: a randomized controlled clinical trial. Journal of Sports Medicine and Physical Fitness. 2015; 55(6):668-674

63. Köpke S, Solari A, Rahn A, Khan F, Heesen C, Giordano A. Information provision for people with multiple sclerosis. Cochrane Database of Systematic Reviews 2018, Issue 10. Art. No.: CD008757. DOI: 10.1002/14651858.cd008757.pub3.

64. Kos D, Duportail M, D'Hooghe M, Nagels G, Kerckhofs E. Multidisciplinary fatigue management programme in multiple sclerosis: a randomized clinical trial. Multiple Sclerosis. 2007; 13(8):996-1003

65. Kos D, Duportail M, Meirte J, Meeus M, D'Hooghe M B, Nagels G et al. The effectiveness of a self-management occupational therapy intervention on activity

performance in individuals with multiple sclerosis-related fatigue: a randomized-controlled trial. International Journal of Rehabilitation Research. 2016; 39(3):255-262

66. Kucuk F, Kara B, Poyraz EC, Idiman E. Improvements in cognition, quality of life, and physical performance with clinical Pilates in multiple sclerosis: a randomized controlled trial. Journal of Physical Therapy Science. 2016; 28(3):761-768

67. Latorraca COC, Martimbianco ALC, Pachito DV, Torloni MR, Pacheco RL, Pereira JG et al. Palliative care interventions for people with multiple sclerosis. Cochrane Database of Systematic Reviews 2019, Issue 10. Art. No.: CD012936. DOI: 10.1002/14651858.cd012936.pub2.

68. Learmonth YC, Paul L, Miller L, Mattison P, McFadyen AK. The effects of a 12-week leisure centre-based, group exercise intervention for people moderately affected with multiple sclerosis: a randomized controlled pilot study. Clinical Rehabilitation. 2012; 26(7):579-593

69. Lutz C, Kersten S, Haas CT. Short-term and long-term effects of an exercise-based patient education programme in people with multiple sclerosis: A pilot study. Multiple Sclerosis International. 2017; 2017:2826532

70. Mathiowetz VG, Finlayson ML, Matuska KM, Chen HY, Luo P. Randomized controlled trial of an energy conservation course for persons with multiple sclerosis. Multiple Sclerosis. 2005; 11(5):592-601

71. Maurer M, Schuh K, Seibert S, Baier M, Hentschke C, Streber R et al. A randomized study to evaluate the effect of exercise on fatigue in people with relapsing-remitting multiple sclerosis treated with fingolimod. Multiple Sclerosis Journal Experimental Translational & Clinical. 2018; 4(1):2055217318756688

72. McCullagh R, Fitzgerald AP, Murphy RP, Cooke G. Long-term benefits of exercising on quality of life and fatigue in multiple sclerosis patients with mild disability: a pilot study. Clinical Rehabilitation. 2008; 22(3):206-214

73. Moss-Morris R, McCrone P, Yardley L, van Kessel K, Wills G, Dennison L. A pilot randomised controlled trial of an internet-based cognitive behavioural therapy self-management programme (MS Invigor8) for multiple sclerosis fatigue. Behaviour Research and Therapy. 2012; 50(6):415-421

74. Mostert S, Kesselring J. Effects of a short-term exercise training program on aerobic fitness, fatigue, health perception and activity level of subjects with multiple sclerosis. Multiple Sclerosis. 2002; 8(2):161-168

75. Mousavi-Shirazi-Fard Z, Mazloom Z, Izadi S, Fararouei M. The effects of modified anti-inflammatory diet on fatigue, quality of life, and inflammatory biomarkers in relapsing-remitting multiple sclerosis patients: a randomized clinical trial. International Journal of Neuroscience. 2021; 131(7):657-665

76. National Institute for Health and Care Excellence. Multiple sclerosis: management of multiple sclerosis in primary and secondary care. NICE guideline 186. London. National Institute for Health and Care Excellence, 2014. Available from: https://www.nice.org.uk/guidance/cg186/evidence/full-guideline-pdf-193254305#page=187&zoom=100,129,212

77. Nazari F, Shahreza MS, Shaygannejad V, Valiani M. Comparing the effects of reflexology and relaxation on fatigue in women with multiple sclerosis. Iranian Journal of Nursing and Midwifery Research. 2015; 20(2):200-204

78. Nedeljkovic U, Raspopovic ED, Ilic N, Vujadinovic ST, Soldatovic I, Drulovic J. Effectiveness of rehabilitation in multiple sclerosis relapse on fatigue, self-efficacy and physical activity. Acta Neurologica Belgica. 2016; 116(3):309-315

79. Negahban H, Rezaie S, Goharpey S. Massage therapy and exercise therapy in patients with multiple sclerosis: a randomized controlled pilot study. Clinical Rehabilitation. 2013; 27(12):1126-1136

80. Oken BS, Kishiyama S, Zajdel D, Bourdette D, Carlsen J, Haas M et al. Randomized controlled trial of yoga and exercise in multiple sclerosis. Neurology. 2004; 62(11):2058-2064

81. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: http://www.oecd.org/std/ppp Last accessed: July 2021.

82. Ozkul C, Guclu-Gunduz A, Yazici G, Atalay Guzel N, Irkec C. Effect of immersive virtual reality on balance, mobility, and fatigue in patients with multiple sclerosis: A single-blinded randomized controlled trial. European Journal of Integrative Medicine. 2020; 35:101092

83. Parks NE, Jackson-Tarlton CS, Vacchi L, Merdad R, Johnston BC. Dietary interventions for multiple sclerosis-related outcomes. Cochrane Database of Systematic Reviews 2020, Issue 5. Art. No.: CD004192. DOI: 10.1002/14651858.cd004192.pub4.

84. Payne C, Wiffen PJ, Martin S. Interventions for fatigue and weight loss in adults with advanced progressive illness. Cochrane Database of Systematic Reviews 2017, Issue 4. Art. No.: CD008427. DOI: 10.1002/14651858.cd008427.pub3.

85. Pazokian M, Shaban M, Zakermoghdam M, Mehran A, Sangelagi B. A comparison between the effect of stretching with aerobic and aerobic exercises on fatigue level in multiple sclerosis patients. Qom University of Medical Sciences Journal. 2013; 7(1):50-56

86. Plow M, Finlayson M, Liu J, Motl RW, Bethoux F, Sattar A. Randomized controlled trial of a telephone-delivered physical activity and fatigue self-management interventions in adults with multiple sclerosis. Archives of Physical Medicine and Rehabilitation. 2019; 100(11):2006-2014

87. Plow MA, Mathiowetz V, Lowe DA. Comparing individualized rehabilitation to a group wellness intervention for persons with multiple sclerosis. American Journal of Health Promotion. 2009; 24(1):23-26

88. Pottgen J, Moss-Morris R, Wendebourg JM, Feddersen L, Lau S, Kopke S et al. Randomised controlled trial of a self-guided online fatigue intervention in multiple sclerosis. Journal of Neurology, Neurosurgery and Psychiatry. 2018; 89(9):970-976

89. Rahimi H, Mehrpooya N, Nahayati MA, Vagharseyyedin S, Izadpanahi AM, Rezaee Z. Self-acupressure for multiple sclerosis-related depression and fatigue: A feasibility randomized controlled trial. Journal of Advances in Medical and Biomedical Research. 2020; 28(130):276-283

90. Rampello A, Franceschini M, Piepoli M, Antenucci R, Lenti G, Olivieri D et al. Effect of aerobic training on walking capacity and maximal exercise tolerance in patients with multiple sclerosis: a randomized crossover controlled study. Physical Therapy. 2007; 87(5):545-555

91. Razazian N, Yavari Z, Farnia V, Azizi A, Kordavani L, Bahmani DS et al. Exercising impacts on fatigue, depression, and paresthesia in female patients with multiple sclerosis. Medicine and Science in Sports and Exercise. 2016; 48(5):796-803

92. Razeghi-Jahromi S, Doosti R, Ghorbani Z, Saeedi R, Abolhasani M, Akbari N et al. A randomized controlled trial investigating the effects of a mediterranean-like diet in patients with multiple sclerosis-associated cognitive impairments and fatigue. Current journal of neurology. 2020; 19(3):112-121

93. Rietberg MB, van Wegen EE, Eyssen IC, Kwakkel G, MS Study Group. Effects of multidisciplinary rehabilitation on chronic fatigue in multiple sclerosis: a randomized controlled trial. PloS One. 2014; 9(9):e107710

94. Rietberg MB, Veerbeek JM, Gosselink R, Kwakkel G, van Wegen EEH. Respiratory muscle training for multiple sclerosis. Cochrane Database of Systematic Reviews 2017, Issue 12. Art. No.: CD009424. DOI: 10.1002/14651858.cd009424.pub2.

95. Rosti-Otajärvi EM, Hämäläinen PI. Neuropsychological rehabilitation for multiple sclerosis. Cochrane Database of Systematic Reviews 2014, Issue 2. Art. No.: CD009131. DOI: 10.1002/14651858.cd009131.pub3.

96. Sabapathy NM, Minahan CL, Turner GT, Broadley SA. Comparing endurance- and resistance-exercise training in people with multiple sclerosis: a randomized pilot study. Clinical Rehabilitation. 2011; 25(1):14-24

97. Sadeghi Bahmani D, Razazian N, Farnia V, Alikhani M, Tatari F, Brand S. Compared to an active control condition, in persons with multiple sclerosis two different types of exercise training improved sleep and depression, but not fatigue, paresthesia, and intolerance of uncertainty. Multiple Sclerosis and Related Disorders. 2019; 36:101356

98. Sajadi M, Davodabady F, Ebrahimi-Monfared M. The effect of foot reflexology on fatigue, sleep quality and anxiety in patients with multiple sclerosis: A randomized controlled trial. Archives of Neuroscience. 2020; 7(3):e102591

99. Sangelaji B, Nabavi SM, Estebsari F, Banshi MR, Rashidian H, Jamshidi E et al. Effect of combination exercise therapy on walking distance, postural balance, fatigue and quality of life in multiple sclerosis patients: a clinical trial study. Iranian Red Crescent Medical Journal. 2014; 16(6):e17173

100. Schulz KH, Gold SM, Witte J, Bartsch K, Lang UE, Hellweg R et al. Impact of aerobic training on immune-endocrine parameters, neurotrophic factors, quality of life and coordinative function in multiple sclerosis. Journal of the Neurological Sciences. 2004; 225(1-2):11-18

101. Sgoifo A, Bignamini A, La Mantia L, Celani MG, Parietti P, Ceriani MA et al. Integrated imaginative distention therapy to cope with fatigue. DIMMI SI study: The first randomized controlled trial in multiple sclerosis. Neurology & Therapy. 2017; 6(2):213-223

102. Straudi S, Martinuzzi C, Pavarelli C, Sabbagh Charabati A, Benedetti MG, Foti C et al. A task-oriented circuit training in multiple sclerosis: a feasibility study. BMC Neurology. 2014; 14:124

103. Tarakci E, Yeldan I, Huseyinsinoglu BE, Zenginler Y, Eraksoy M. Group exercise training for balance, functional status, spasticity, fatigue and quality of life in

multiple sclerosis: a randomized controlled trial. Clinical Rehabilitation. 2013; 27(9):813-822

104. Thomas PW, Thomas S, Kersten P, Jones R, Slingsby V, Nock A et al. One year follow-up of a pragmatic multi-centre randomised controlled trial of a group-based fatigue management programme (FACETS) for people with multiple sclerosis. BMC Neurology. 2014; 14:109

105. Thomas S, Thomas PW, Kersten P, Jones R, Green C, Nock A et al. A pragmatic parallel arm multi-centre randomised controlled trial to assess the effectiveness and cost-effectiveness of a group-based fatigue management programme (FACETS) for people with multiple sclerosis. Journal of neurology, neurosurgery, and psychiatry. 2013; 84(10):1092-1099

106. Tosh J, Dixon S, Carter A, Daley A, Petty J, Roalfe A et al. Cost effectiveness of a pragmatic exercise intervention (EXIMS) for people with multiple sclerosis: economic evaluation of a randomised controlled trial. Multiple Sclerosis. 2014; 20(8):1123-1130

107. Tramontano M, Martino Cinnera A, Manzari L, Tozzi FF, Caltagirone C, Morone G et al. Vestibular rehabilitation has positive effects on balance, fatigue and activities of daily living in highly disabled multiple sclerosis people: A preliminary randomized controlled trial. Restorative Neurology and Neuroscience. 2018; 36(6):709-718

108. van den Akker LE, Beckerman H, Collette EH, Twisk JW, Bleijenberg G, Dekker J et al. Cognitive behavioral therapy positively affects fatigue in patients with multiple sclerosis: Results of a randomized controlled trial. Multiple Sclerosis. 2017; 23(11):1542-1553

109. van den Berg M, Dawes H, Wade DT, Newman M, Burridge J, Izadi H et al. Treadmill training for individuals with multiple sclerosis: a pilot randomised trial. Journal of neurology, neurosurgery, and psychiatry. 2006; 77(4):531-533

110. van Kessel K, Moss-Morris R, Willoughby E, Chalder T, Johnson MH, Robinson E. A randomized controlled trial of cognitive behavior therapy for multiple sclerosis fatigue. Psychosomatic Medicine. 2008; 70(2):205-213

111. Velikonja O, Curic K, Ozura A, Jazbec SS. Influence of sports climbing and yoga on spasticity, cognitive function, mood and fatigue in patients with multiple sclerosis. Clinical Neurology and Neurosurgery. 2010; 112(7):597-601

112. Wahls TL, Titcomb TJ, Bisht B, Eyck PT, Rubenstein LM, Carr LJ et al. Impact of the Swank and Wahls elimination dietary interventions on fatigue and quality of life in relapsing-remitting multiple sclerosis: The WAVES randomized parallel-arm clinical trial. Multiple Sclerosis Journal Experimental Translational & Clinical. 2021; 7(3):DOI: 10.1177/20552173211035399

113. Yazgan YZ, Tarakci E, Tarakci D, Ozdincler AR, Kurtuncu M. Comparison of the effects of two different exergaming systems on balance, functionality, fatigue, and quality of life in people with multiple sclerosis: A randomized controlled trial. Multiple Sclerosis and Related Disorders. 2019; 39:101902

1

2