

SYSTEMATIC REVIEW

Effect of exercise interventions on perceived fatigue in people with multiple sclerosis: synthesis of meta-analytic reviews

Reza Safari^{1,2}, Marietta L van der Linden¹ & Tom H Mercer^{*,1}

Practice points

- Information reviewed is largely obtained from studies of 30–50-year-old participants of moderate disability with nonprogressive multiple sclerosis (MS).
- Exercise interventions appear to elicit a moderately favorable effect on fatigue in people with MS when compared with usual care or a no-exercise comparison group.
- Mixed training and other types of exercise such as balance and yoga appear to show larger effects than other types of exercise.
- There is still a lack of convincing information regarding which type of exercise intervention/s and with what specific combination of exercise duration, frequency and intensity might be most effective in reducing MS-related fatigue.
- The results should be interpreted with caution because of high risk of bias, poor applicability of results, heterogeneity and small sample size in the randomized control trials on which this review of review evidence is ultimately based.

Although exercise training has been advocated as a nonpharmacological treatment for multiple sclerosis (MS) related fatigue, no consensus exists regarding its effectiveness. To address this, we collated meta-analytic reviews that explored the effectiveness of exercise training for the treatment of MS-related fatigue. We searched five online databases for relevant reviews, published since 2005, and identified 172 records. Five reviews were retained for systematic extraction of information and evidence quality analysis. Although our review synthesis indicated that exercise training interventions have a moderate effect on fatigue reduction in people with MS, no clear insight was obtained regarding the relative effectiveness of specific types or modes of exercise intervention. Moreover, Grading of Recommendation Assessment, Development and Evaluation revealed that the overall quality of evidence emanating from these five reviews was 'very low'.

First draft submitted: 10 February 2017; Accepted for publication: 24 April 2017; Published online: 20 June 2017

According to the WHO report 1.3 million people, in 112 countries, have multiple sclerosis (MS) [1]. The symptoms of MS include emotional lability, incontinence, pain, cognitive problems, oscillopsia, spasticity, mobility problems and fatigue. Fatigue, the most common symptom, has been reported as the most troubling symptom by more than 50% of people with MS (pwMS) [2,3]. The etiology of

KEYWORDS

- exercise • fatigue
- multiple sclerosis
- overview • review

¹Centre for Health & Social Care Research, College of Health & Social Care, University of Derby, Derby, DE22 1GB, UK

²Centre for Health, Activity & Rehabilitation Research, School of Health Sciences, Queen Margaret University, Musselburgh, EH21 6UU, UK

*Author for correspondence: Tel.: +44 131 474 0000; Fax: +44 131 474 0001; tmercer@qmu.ac.uk

fatigue is unknown, however, it is multifactorial, subjective in nature, can be physical and/or mental, secondary or primary, and interferes with desired and usual activities [4–6].

Recent reviews suggest that pharmacological intervention for treatment of fatigue has minimal effect and use of nonpharmacological intervention such as exercise has potential merit for the direct management of fatigue in pwMS rather than merely as an adjunct treatment [7,8].

Although there is a paucity of information on definite mechanisms behind the effect of exercise training on MS fatigue, it has been suggested that exercise may ameliorate primary MS fatigue through cardiovascular, immunologic, neuroendocrine and neurologic changes, while secondary fatigue could be affected by exercise as a result of symptomatic reduction in deconditioning, sleep disorders and depression [9,10].

A Cochrane review on the effectiveness of exercise-based rehabilitation programs for MS by Rietberg *et al.* [11] concluded that there was evidence to support the implementation of exercise training as a means to improve muscle power, physical function, exercise tolerance, mobility-related activities and mood. However, no evidence was found to support a role for exercise training on MS fatigue, possibly due to the limited number of studies available at the time of that particular review. Since then in excess of 60 randomized control trials (RCTs) have been published, reporting the effect of exercise therapy on patient reported fatigue in pwMS, mostly as secondary outcome. Subsequently, a number of similarly focused systematic/meta-analytic reviews have been published. Consequently, this paper aims to produce an overview synthesis of meta-analytic review-generated information about a potential role for exercise training interventions for MS-related fatigue. In doing so, we will appraise both the methodological quality of the included meta-analytic reviews as well as evaluating the overall quality of evidence generated by them with a view to contribute to the debate around the development of future exercise intervention guidelines and the identification of future research directions.

Methods

• Literature search

A search for relevant published reviews was conducted using online databases including Medline and Embase, both using Ovid,

Cochrane Library, Centre for Review and Dissemination, and Physiotherapy Evidence Database (PEDro), from 2005 to October 2016. Both Medical subject heading terms and standard keywords were used to build a search strategy using relevant terms including: ‘multiple sclerosis’, ‘Exercise’ and ‘Fatigue’. The search strategy was further refined for each specific database (**Supplementary Appendix 1**). No restriction was made on the search results except that the type of publication was filtered to ‘any type of review’. In addition, the reference lists of the reviews were handsearched for any additional potentially relevant articles.

• Inclusion criteria

Reviews evaluating the evidence regarding the effects of any type of exercise intervention on measure of perceived fatigue, which included pwMS and statistically combined the results of at least two RCTs, were included in this overview. Narrative reviews and systematic reviews with qualitative synthesis of the results of RCTs and/or studies with other research designs were identified and listed (**Supplementary Appendix 2**) but were not included in this metasynthesis.

• Data collection & analysis

Relevant information was extracted from the included reviews and entered into predefined data collection forms designed according to the guidance from the Cochrane Handbook of Systematic Reviews of Interventions. The information extracted included review questions or objectives, study inclusion/exclusion criteria, characteristics of participants, types and details of intervention and comparators, types of fatigue outcome measures, study quality assessment, statistical methods used, summary of results and limitations of the review and/or the included studies. Data were collected by one author and double checked by other authors.

• Quality of included reviews & evidence

The methodological quality of included reviews was assessed using A Measurement Tool to Assess Systematic Reviews (AMSTAR) [12]. The tool consisted of 11 items (overall rating score of 11) and has good face and content validity [12]. In a limited test setting, it was shown to have good agreement, reliability, construct validity and feasibility [13]. All three authors commented on the AMSTAR ratings of the included reviews. Systematic reviews with an AMSTAR score of

8–11 were ranked as high, 4–7 as medium and less than three as low methodological quality [14].

The quality of evidence presented in the reviews was assessed using the Grading of Recommendation Assessment, Development and Evaluation (GRADE) approach [15,16]. According to GRADE, the quality of the overall evidence is assessed based on the rating of five main domains:

limitation in design, publication bias, precision in results, indirectness and inconsistency of results.

Results

• Characteristics of included reviews

After adjusting for duplicates, 172 records were included for initial title and abstract screening. Records that clearly did not meet the inclusion

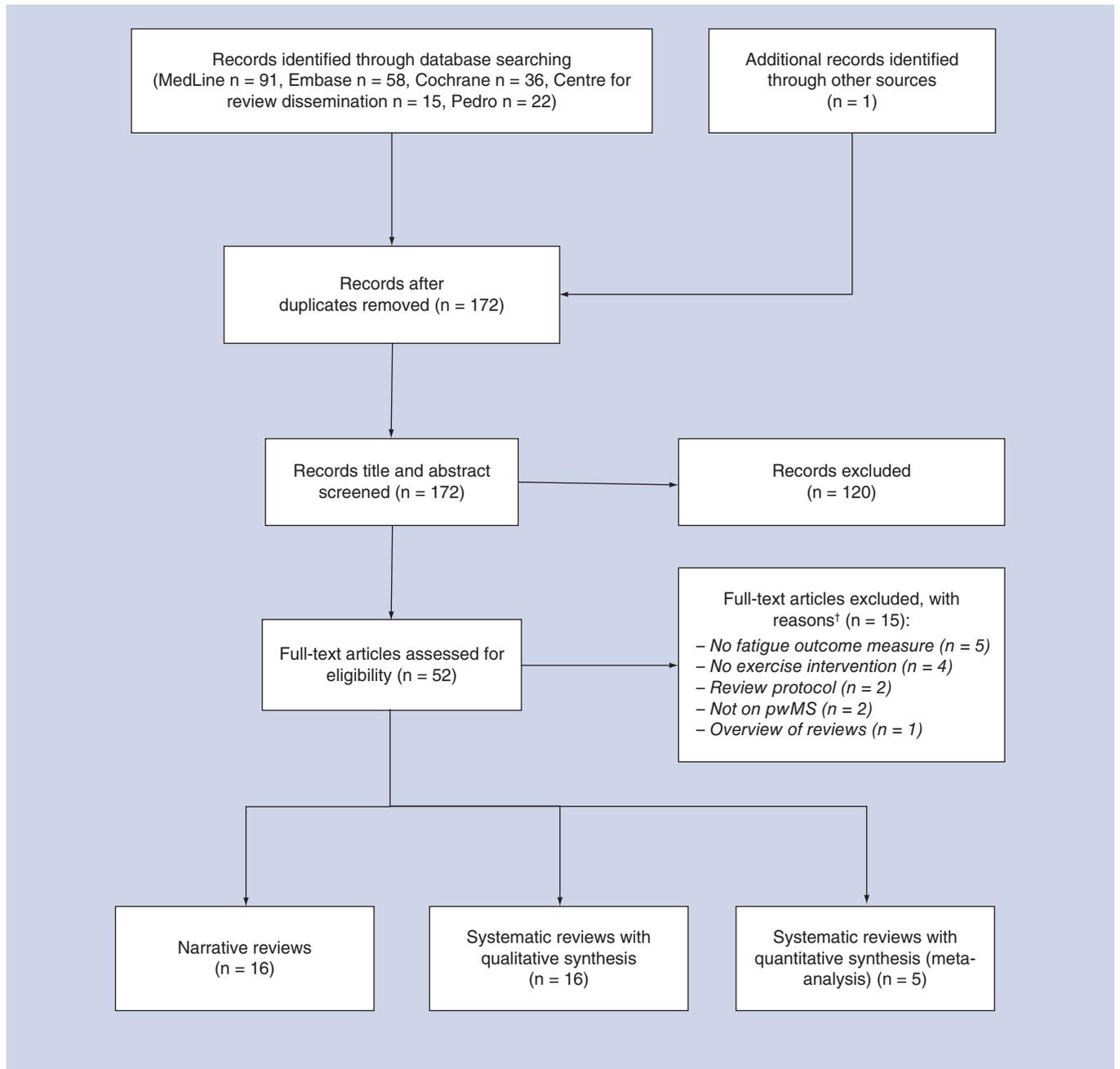


Figure 1. Study flow diagram.

†For a list of excluded reviews, see [Supplementary Appendix 3](#).

pwMS: People with multiple sclerosis.

Table 1. Characteristics of included reviews.					
Characteristics	Study				
	<i>Heine et al. (2015)</i>	<i>Cramer et al. (2014)</i>	<i>Asano & Finlayson (2014)</i>	<i>Pilutti et al. (2013)</i>	<i>Motl & Gosney (2008)</i>
Types of studies (n)	RCTs including Crossover designs (36)	RCT (5)	RCT (10) [†]	RCT (17)	Pre-post study (3), controlled trial (4), both pre-post and controlled (1)
Date assessed as up to date	November 2014	March 2014	August 2013	October 2012	November 2006
Population					
n (IG vs CG)	Exercise versus control: 1603 (969 vs 34), Exercise versus nonexercise control: 1325 (831 vs 494), exercise versus exercise: 278 (138 vs 140)	Yoga versus control: 203 (109 vs 94), yoga vs exercise: 369 (119 vs 250)	233 (112 vs 121)	568	347
Age (years)	18–65	32.2–54.4	34–50	NR	NR
Sex (female; %)	NR	NR	55–100%	NR	NR
Types of MS (n)	NR (9), RRMS (5), SPMS + PPMS (3), RRMS + SPMS (6), all types (13)	NR	Unclear (6), > 90% RRMS (3), 31% RRMS (1)	NR	NR
EDSS	EDDS < 6.5 in 35 studies, 6.5 < EDDS < 8.5 in one study	NR	Minimal to moderate (EDDS range: 3–7.5)	NR	NR
Baseline fatigue	Predefined in one study	NR	Predefined in one study	NR	NR
Intervention:					
– Type (n studies)	Endurance (11), muscle power training (4), task-oriented (2), mixed training (6), other training (9)	Hata yoga (2), Lyengar yoga (1), yoga (2)	Various exercise interventions	Various exercise interventions	Aerobic (6) Nonaerobic (1) Aerobic and nonaerobic(1)
– Duration of treatment (n studies)	Endurance: >3 weeks to <6 months (11), muscle power training: 8–12 weeks (4), task-oriented training: NA (2), mixed training: 2–26 weeks (6), other training: NR (9)	6–12 weeks (3), 24 weeks (2)	<6 weeks (2), 6–12 weeks (6), 24 weeks (2)	<6 weeks (3), 6–12 weeks (8), 12–26 weeks (4), NR (2)	3 months or more (6), <3 months (2)
– Timing and intensity (n studies)	Endurance: 1–5x/week, Muscle power training: 2–5x/week, task-oriented training: NR, mixed training 2–5x/week, other training: NR	1x/week for 60–90 min (4), 3 x/week for 60–70 min (1)	Range: 1–4x/week for <30–90 min	1x or 2x/week (6), 3x/week (7), 3–6x/week (4)	< 90 min/week (6), > 90 min/week (1), NR (1)
[†] Studies included a purpose statement saying that the intervention targeted fatigue/energy levels.					
CG: Control group; EDDS: Expanded Disability Status Scale; ES: Effect size; FSMC: Fatigue scale for motor and cognitive function; FSS: Fatigue Severity Scale; IG: Intervention group; MFI: Multidimensional fatigue inventory; MFIS: Modified Fatigue Impact Scale; MS: Multiple sclerosis; NR: Not reported; POMS: Profile of Mood State; PPMS: Primary progressive multiple sclerosis; QoL: Quality of life; RCT: Randomized control trial; RRMS: Relapsing-remitting multiple sclerosis; SF 36: Short Form (36); SPMS: Secondary progressive multiple sclerosis.					

Table 1 (cont.). Characteristics of included reviews.

Characteristics	Study				
	<i>Heine et al. (2015)</i>	<i>Cramer et al. (2014)</i>	<i>Asano & Finlayson (2014)</i>	<i>Pilutti et al. (2013)</i>	<i>Motl & Gosney (2008)</i>
Comparison (n studies)	Exercise versus control (36), exercise versus non-exercise control group (27), exercise versus exercise (9)	No intervention/ Usual care (4), aerobic (2), aerobic + resistance (2) and climbing (1)	No-intervention/usual care (5), other exercise – for example stretch, balance, yoga, home exercise (5)	No-intervention/ usual care	Either no control or no-exercise control group
Fatigue measure: – Type (n studies)	Endurance: FSS (7), MFIS (5), POMS fatigue subscale (2), FSMC (2) and MFI (1), muscle power training: FSS (4) and MFIS (4), task-oriented training: FSS (2), mixed training: FSS (7), MFIS (6)	MFIS, (3), MFI (1), FSS (1)	FSS (6), MFIS (3), MFI (1)	FSS (7), MFIS (5), MSQOL (4), SF 36 – vitality (4), POMS vigor/fatigue (2), MFI (2), Hamburg QoLMS (1), Sickness impact Profile-alertness (1)	FSS (3), MFIS (3), MFI (1), Visual Analog Fatigue Scale (1)
Review limitation	Lack of trials specifically targeted fatigue Only three studies measured fatigue as a primary outcome Possible effect of low methodological quality of studies on ES Presence of small study effect Underpowered studies with small sample sizes Only two studies included participants with fatigue at baseline	Limited overall study/sample size Insufficient reporting and/or low methodological quality of the included studies Studies in grey literature were not included Type of MS or level of disability was not reported.	Small sample size of the included studies Only one study included participants with fatigue at baseline Lack of reporting of quality of evidence Result for exercise subgroups not reported	Few studies with patients with preselected level of fatigue Result for exercise subgroups not reported	Combining controlled and noncontrolled effect sizes Effect on fatigue was analyzed as moderator of overall QoL outcome

¹Studies included a purpose statement saying that the intervention targeted fatigue/energy levels.

CG: Control group; EDDS: Expanded Disability Status Scale; ES: Effect size; FSMC: Fatigue scale for motor and cognitive function; FSS: Fatigue Severity Scale; IG: Intervention group; MFI: Multidimensional fatigue inventory; MFIS: Modified Fatigue Impact Scale; MS: Multiple sclerosis; NR: Not reported; POMS: Profile of Mood State; PPMS: Primary progressive multiple sclerosis; QoL: Quality of life; RCT: Randomized control trial; RRMS: Relapsing-remitting multiple sclerosis; SF 36: Short Form (36); SPMS: Secondary progressive multiple sclerosis.

criteria (n = 120) were excluded and the full text of the remaining 52 records were reviewed. Five of these included reviews statistically combined the results of the studies and were retained in this overview for further analysis. Fifteen records were further excluded due to the reasons stated in [Figure 1](#). The remaining records were either systematic reviews with qualitative synthesis (n = 16) or narrative reviews (n = 16; see [Supplementary Appendix 2](#) for a detailed list). No further records were found after handsearching the reference lists of the narrative and systematic reviews.

Included trials in the reviews

A summary description of included reviews is presented in [Table 1](#). *Cramer et al.*, *Pilutti et al.* and *Asano et al.* included five, 17 and 10 RCTs, respectively [8,17,18]. These reviews included RCTs only. In the other two reviews, *Heine et al.* included both RCTs and randomized crossover designs (n = 36 studies) [19], while *Motl and Gosney* included both RCTs and prepost trials (n = 8 studies) [20]. The earliest and the most recent reviews were assessed as up to date on November 2006 and November 2014, respectively. A total of 46 unique trial

reports were included in five reviews of which 19 reports appeared in two or more reviews.

Population

The included reviews reported on three different comparisons (see comparison group section below) and the total number of participants in each of these comparisons ranged from 203 to 1603 with a median of 358 (Inter-quartile range; IQR = 490). In the three reviews which reported the age of participants, one included studies with a wide range of sample age (18–65 years) [19], whereas, the other two reviews by Cramer *et al.*, Asano and Finlayson included studies with participants ranging from 32–54 to 34–50 years old, respectively [8,17]. Two reviews stated the types of MS in the included studies [8,19]. However, the types of MS in the majority of the trials included in these two reviews were unclear or not reported. The trials which did report the type of MS, mostly included participants with relapsing-remitting MS. The same two reviews [8,19] are also the only reviews which reported degree of disability of the pwMS in the included studies. Participants in the studies included in the review were minimally to moderately affected by MS (Expanded Disability Status Scale; EDSS levels ranging from 0 to 6.5). No reviews included studies with pwMS which were more severely affected, in other words, those unable to walk (EDSS of 7 and higher). Of all included trials in the reviews, only one trial [21] included participants with a predefined level of fatigue (score of 45 or above out of 84 on the Modified Fatigue Impact Scale questionnaire) and this study appeared in two reviews [8,19].

Types & mode of intervention

While the majority of the reviews included trials on a wide range of exercise interventions, one review included only studies on the effects of yoga [17]. The review by Motl and Gosney categorized the interventions in the included trials as either aerobic or nonaerobic [20]. Heine *et al.* grouped the exercise interventions as: endurance, muscle training, task-oriented training, mixed and 'other' (i.e., hippotherapy, balance training, yoga, inspiratory muscle training, motor learning, sports climbing and robot-assisted gait training) [19]. Aerobic exercise was the most common type of intervention in all reviews except the review on yoga

by Cramer *et al.* [17]. The majority of the trials implemented intervention durations of between 6 and 12 weeks. The intervention frequency and session duration varied among the trials, ranging from once a week to five-times a week with total session duration spanning of 30–90 min per week. There was a lack of information on long-term follow-up in all reviews and the included studies.

Comparison group

The reviews included in this metasynthesis reported a total of eight pooled effect sizes for exercise versus one or more of three different comparators. Four reviews calculated the pooled effect size (ES) for exercise interventions compared with nonexercise/usual care control group [17–20]. In addition, Heine *et al.* and Cramer *et al.* calculated pooled ES for the effect of exercise (yoga in the case of Cramer) compared with other exercise control conditions [17,19]. In two reviews the exercise interventions were compared with any type of control group (including either exercise or nonexercise group) [8,19].

Outcome measure

The most frequently reported fatigue outcome measure, reported in 37 of the 46 trials, was Fatigue Severity Scale, included in all eight comparisons made in the reviews followed by Modified Fatigue Impact Scale, reported in 25 trials. The remaining outcome measures included: multidimensional fatigue inventory (reported in 6 trials), Profile of Mood State Fatigue subscale (four trials), SF-36 vitality (four trials), MS Quality of life (QoL; four trials) and Hamburg QoLMS, Sickness Impact Profile-alertness and Visual Fatigue Analog Scale, each in one trial. All reviews reported the short-term effect of exercise on fatigue (i.e., pre- and postintervention).

• Methodological quality of included reviews

Methodological quality varied among the included reviews (see Table 2). Two reviews were rated as having high methodological quality; with AMSTAR ratings of 10 and 9, respectively [17,19]. The reviews conducted by Pilutti *et al.* and Asano *et al.* exhibited medium methodological quality (AMSTAR: 5) [8,18]. The lowest rating of methodological quality was assigned to Motl and Gosney's review (AMSTAR: 2) [20]. None of the reviews was scored as having a conflict

Table 2. AMSTAR rating.

Question	Heine <i>et al.</i> (2015)	Cramer <i>et al.</i> (2014)	Asano & Finlayson (2014)	Pilutti <i>et al.</i> (2013)	Motl & Gosney (2008)
1 Was an ' <i>a priori</i> ' design provided?	Y	Y	N	N	N
2 Was there duplicate study selection and data extraction?	Y	Y	Y	CA	CA
3 Was a comprehensive literature search performed?	Y	Y	CA	Y	Y
4 Was the status of publication (i.e., grey literature) used as an inclusion criterion?	Y	N	N	N	N
5 Was a list of studies (included and excluded) provided?	Y	Y	N	N	N
6 Were the characteristics of the included studies provided?	Y	Y	Y	Y	N
7 Was the scientific quality of the included studies assessed and documented?	Y	Y	N	Y	N
8 Was the scientific quality of the included studies used appropriately in formulating conclusions?	Y	Y	N	Y	N
9 Were the methods used to combine the findings of studies appropriate?	Y	Y	Y	Y	Y
10 Was the likelihood of publication bias assessed?	Y	Y	Y	CA	N
11 Was the conflict of interest included?	CA	N	CA	CA	N
12 Overall score	10	9	5	5	2

CA: Cannot Answer; N: No; Y: Yes.

of interest in either the review or the included trials. Four of the five reviews did not include grey literature and/or did not search secondary sources other than online databases. Three of the five reviews did not provide an *a priori* design or a list of excluded studies.

• Risk of bias assessment

One review assessed the risk of bias (RoB) among their included studies using both the Cochrane RoB assessment tool and PEDro checklist [19]. Two reviews reported the methodological quality of trials using Cochrane RoB assessment tool only [8,17] and one review used PEDro only [18]. Motl and Gosney did not report on the methodological quality of their included trials. Allocation concealment and blinding of participants, personnel or outcome assessment were reported as 'high risk' or 'unknown risk' of bias for the majority of the trials included in all four reviews who reported RoB (Table 1). Only Heine *et al.* commented on the quality of evidence. The authors employed the GRADE approach for evaluating the quality of body of evidence and reported, on this

basis, the evidence for the favorable effect of exercise therapy on fatigue was moderate.

• Effect of exercise on MS fatigue

In the five reviews in this synthesis, the pooled effect sizes for eight comparisons were calculated. These eight comparisons were grouped in this overview in three categories based on the type of control group(s), Table 3. The results are explained in the following sections.

Exercise versus nonexercise/usual care control/no control

The reviews reporting the effect of exercise on MS fatigue, when compared with no exercise or usual care conditions, showed a significant moderate random effect in favor of exercise. Heine *et al.*, Cramer *et al.* and Pilutti *et al.* calculated the pooled effect sizes (95% CI) as -0.58 (-0.8, -0.34), -0.52 (-1.02, -0.02) and -0.45 (-0.22, -0.68), respectively. All three reviews adopted a random effect model as there was significant heterogeneity among included studies (see Table 3). It is worth noting that Cramer *et al.* included only trials on yoga as

Table 3. Summary of findings with GRADE ratings.

Exercise for fatigue in multiple sclerosis								
Intervention/ Comparison	Study (year)	Effect size (95% CI)	p-value	Heterogeneity test	p-value	Number of participants (trials)	Quality of the evidence (GRADE) [†]	Comments
Exercise versus control	Heine <i>et al.</i> (2015)	-0.35 (-0.13, -0.57)	0.0001	I ² = 73%	<0.01	1603 (36)	⊕⊕⊕⊕ Very low ^{‡,§,¶}	Limitation –1 Indirectness –1 Inconsistency –1
	Asano and Finlayson (2014)	-0.57 (-0.1, -1.04)	0.02	Q = 26.39, I ² = 65%	0.003	233 (10)	⊕⊕⊕⊕ Very low ^{‡,¶,#}	Limitation –1 Imprecision –1 Inconsistency –1
Exercise versus nonexercise/usual care/no control	Heine <i>et al.</i> (2015)	-0.58 (-0.34, -0.81)	0.0001	I ² = 58%	<0.01	1325 (27)	⊕⊕⊕⊕ Very low ^{‡,§,¶}	Limitation –1 Indirectness –1 Inconsistency –1
	Pilutti <i>et al.</i> (2013)	-0.45 ^{**} (-0.22, -0.68)	<0.001	Q = 29.9	0.019	568 (17)	⊕⊕⊕⊕ Very low ^{‡,§,¶}	Limitation –1 Indirectness –1 Inconsistency –1
	Cramer <i>et al.</i> (2014)	-0.52 (-0.02, -1.02)	0.04	χ ² = 7.3, I ² = 60%	0.06	203 (4)	⊕⊕⊕⊕ Very low ^{‡,§,¶,#}	Limitation –1 Indirectness –1 Inconsistency –1 Imprecision –1
	Motle and Gosney (2008)	-0.19 ^{**,\$§} (-0.01, -0.39)	significant ^{**††}	Q = 20.53	0.0001	347 (8)	⊕⊕⊕⊕ Very low ^{‡,§,¶,#,††}	Limitation –1 Indirectness –1 Inconsistency –1 Imprecision –1
Exercise versus other exercise control	Heine <i>et al.</i> (2015)	0.28 (0.00, -0.56)	0.047	I ² = 23%	0.24	278 (9)	⊕⊕⊕⊕ Very low ^{‡,§,#}	Limitation –1 Imprecision –1 Indirectness –1
	Cramer <i>et al.</i> (2014)	0.03 (-0.24, 0.30)	0.83	χ ² = 4.59, I ² = 13%	0.33	369 (5)	⊕⊕⊕⊕ Very low ^{‡,§,#}	Limitation –1 Imprecision –1 Indirectness –1

[†]The argumentation for downgrading the grades of evidence is provided in the footnotes. GRADE Working Group grades of evidence.
 High quality: Further research is very unlikely to change our confidence in the estimate of effect.
 Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
 Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
 Very low quality: We are very uncertain about the estimate.
[‡]Crucial limitation on the blinding of the intervention to the participants, personnel and outcome assessors in most of studies. There is limitation in allocation concealment in most studies.
[§]Patients included did not report baseline fatigue. Fatigue was secondary outcome in majority of studies.
[¶]Statistical tests of heterogeneity have a low p-value and I² value is large or minimal or no overlap of confidence intervals.
[#]Total sample size is small.
^{††}High risk of bias due to inclusion of before and after trials.
^{§§}Hedges g
^{§§§}Authors tested the effect of exercise on QoL and tested for the effect of fatigue instrument as a moderator of the QoL effect size.
^{¶¶}p-value not reported.
 GRADE: Grading of Recommendation Assessment, Development and Evaluation.

an exercise intervention, whereas the other two reviews included trials on various types of exercise training. Besides, Pilutti *et al.*, Motle and Gosney adjusted for small sample size which calculated ES in accordance with Hedges g.

Motle and Gosney included both controlled and noncontrolled trials and stated that the effect of exercise intervention on QoL and fatigue as a measure of QoL was statistically significant. They further tested the effect of presence of fatigue outcome measure as

a moderator variable in order to examine whether these can explain the amount of the variability in the overall magnitude of change in QoL. According to the authors, the overall effect size was significantly influenced by the type of QoL outcome measure instruments used in the studies and concluded that there was a significant effect for fatigue ‘as a measure of QoL’. The pooled effect sizes (95% CI) for fatigue as a QoL moderator was calculated to be -0.19 (-0.01, -0.39).

Exercise versus other exercise control

Cramer *et al.* reported that effect of yoga on fatigue compared with other exercise types (i.e., climbing, treadmill, aerobic and resistance training) was not significant, ES (95% CI): 0.03 (-0.24, -0.30), $p = 0.83$. The same non-significant results were reported by Heine *et al.* for effect of exercise interventions compared with other types of exercise; ES (95% CI): 0.28 (0.00, -0.56), $p = 0.047$. The exercise intervention that was the primary focus of a trial was considered as an experimental intervention. The test of heterogeneity was not significant for either review. Additionally, both reviews were rated as having high methodological quality.

Exercise versus exercise or nonexercise control

Two reviews reported the pooled effect estimate for the effect of exercise interventions compared with control groups of exercise and/or nonexercise conditions [8,19]. Asano and Finlayson reported a significant moderate random effect of a wide range of exercise interventions (i.e., aerobic, aquatic and inspiratory muscle exercise; vestibular rehabilitation program; progressive resistance training; climbing; and yoga) on fatigue. The pooled ES (95% CI) was -0.57 (-0.1, -1.04) indicating a reduction on fatigue, with a significant heterogeneity ($Q = 26.39$, $I^2 = 65\%$, $p = 0.003$). This ES was larger than that of random pooled ES calculated by Heine *et al.* (ES [95% CI]: -0.35 [-0.57, -0.13], heterogeneity: $I^2 = 73\%$, $p < 0.01$). Based on AMSTAR rating, Heine *et al.*, Asano and Finlayson reviews were rated as having high and moderate methodological quality, respectively.

• Other results

Heine *et al.* [19] performed a sensitivity analysis to assess the effect of the methodological quality of their included trials on the effect size. For this analysis they included trials comparing exercise to nonexercise control with PEDro score >5 . Data from 14 studies with total number of 495 participants in the exercise group and 306 participants in the control group were used. The results showed a larger heterogeneous effect ($I^2 = 72\%$, p -value < 0.01), ES (95% CI): -0.64 (-0.95, -0.32), $p < 0.01$ than the overall effect size of -0.52, indicating that higher methodological quality produced a higher effect size. Cramer *et al.* also conducted a sensitivity analysis on the blinding of outcome measurement (detection) bias and found that

the intervention effect was not significant when only trials ($n = 3$) with low detection bias were entered for calculating the effect size (yoga vs usual care, ES [95% CI]: -0.32 [-0.72, -0.08], $p = 0.12$, heterogeneity: $I^2 = 36\%$, $p = 0.21$). The authors stated that no study had low risk of selection or attrition biases.

Furthermore, Heine *et al.* [19] in another set of sensitivity analysis, on the type of exercise training undertaken (i.e., endurance, muscle power, task oriented, mixed training and other training groups), reported that the mixed training (ES [95% CI]: -0.73 [-1.23, -0.23], $p < 0.01$, heterogeneity: $I^2 = 82\%$, $p < 0.01$) and 'other' training (ES [95% CI]: -0.54 [-0.79, -0.29], $p < 0.01$, heterogeneity: $I^2 = 45\%$, $p = 0.07$) showed a larger effect of exercise on fatigue than the endurance exercise (ES [95% CI]: -0.43 [-0.69, -0.17], $p = 0.004$, heterogeneity: $I^2 = 28\%$, $p > 0.18$), muscle power training (ES [95% CI]: 0.03 [-1.02, -0.33], $p = 0.93$, heterogeneity: $I^2 = 70\%$, $p > 0.02$) or task oriented training (ES [95% CI]: -0.34 [-1.02, -0.33], $p > 0.05$, heterogeneity: $I^2 = 0\%$, $p > 0.05$). However, the subgroup analysis did not show a statistically significant difference between these groups.

• Quality of evidence

Based on GRADE assessment, the quality of evidence was rated 'very low' for all three comparisons. We lowered the evidence quality rating due to the RoB (limitation), lack of applicability of results, in other words, 'indirectness', for all reported effect sizes. Heterogeneity of results was also a concern for exercise training effect compared with either any control or nonexercise/usual care control group. Furthermore, apart from two comparisons in Heine *et al.* review [19], the total sample size in the reviews was small ($n < 200$) and the majority of the trials were underpowered, therefore; these comparisons were rated down for imprecision, **Table 3**.

Discussion & conclusion

The results of this synthesis of reviews suggest that the exercise interventions have a moderate effect on ameliorating fatigue in pwMS when compared with usual care or a no-exercise comparison group [17–19]. This was concluded in two reviews with high methodological quality [17,19] and one moderate quality review [18]. However, the results should be interpreted with

caution because of high RoB, poor applicability of results, heterogeneity and small sample size in the RCTs on which this evidence is based. Blinding is not possible in exercise studies due to the nature of the intervention, and in the majority of the studies, authors failed to report whether the group allocation was concealed or whether the outcome assessors were blind to allocation. Furthermore, with one exception [21] none of the studies used the inclusion criterion that participants should be assessed as fatigued on study entry, and in the majority of trials, the intervention was not specifically designed to target fatigue, hence fatigue was measured as a secondary outcome.

The review by Asano and Finlayson included only those trials in which there was a 'clear statement' indicating that intervention was designed to target fatigue. Interestingly, the pooled effect size showed a moderate effect of exercise (Standardised mean difference; SMD: -0.57, 95% CI: -0.1 to -1.04) when compared with any type of comparison group (including exercise and nonexercise groups), which was larger than that of the reported effect size in Heine *et al.* review (SMD: -0.35, 95% CI: -0.13 to -0.57). This may plausibly indicate that the effect of exercise training on self-reported fatigue may be larger when the intervention is designed to target fatigue. However, after close inspection of the aim and rationale of the studies included by Asano and Finlayson [8], only two studies [21,22] reported an intervention which specifically targeted fatigue as the main outcome.

There was a significant heterogeneity in the results most likely due to the diversity in type/mode of exercise interventions, comparison groups and/or study population. Most studies included 30–50-year-old participants of moderate disability with nonprogressive MS. Heterogeneity was higher when reviews included both exercise and nonexercise control conditions into the comparison, compared with when only nonexercise/usual care control condition was used. In contrast, heterogeneity was not significant in two reviews, with high AMSTAR ratings, when exercise was compared with the other exercise control group. Despite a relatively large number of studies in the included reviews, there is still a paucity of information on which type of exercise intervention/s and with what specific combination of exercise duration, frequency and intensities might be most effective in reducing MS-related fatigue.

Asano and Finlayson [8] commented that those studies which reported larger effect sizes (ES >0.81) assessed a large variety of exercise training types and therefore no definitive conclusion could be made regarding type and prescription characteristics of exercise training that would be more effective to reduce fatigue. Pilutti *et al.* [18] concluded that aerobic training intervention elicited lower effect sizes compared with resistive training or combined mode, whereas; Heine *et al.* [19] reported that aerobic and mixed training (aerobic and resistive) and other training such as yoga, robotics and balance training showed significant effects, with mixed training showing the highest effect of all (ES: -0.73, 95% CI: -0.23 to -1.23). Cramer *et al.*'s review [17] on yoga suggested that yoga can be equally effective as other exercise training interventions on both patient reported and objective measured outcomes such as mobility and cognitive function.

In addition to the aforementioned limitations, there are few important issues with the available exercise studies that preclude meaningful interpretations regarding most effective type of exercise to treat fatigue. First, the vast majority of the trials failed to include a sufficiently detailed description of the exercise training intervention characteristics such as supervision of exercise, progression and overload training principles, and intervention fidelity and adherence. Second, there is lack of studies in which the intervention is guided by a theory regarding the possible pathophysiological or behavioral mechanisms through which the intervention is thought to influence either primary or secondary fatigue. Moreover, very few studies are designed to test these theories. In a recent scoping review, including 234 studies, Langeskov-Christensen *et al.* reported that only three studies explicitly investigated the potential pathophysiological pathways explaining the effect of exercise on MS-related fatigue [10]. These three studies all showed an effect of a variety of exercise intervention (aerobic and resistive) on the immunologic changes due to increase in anti-inflammatory factors. The authors also proposed other pathophysiological mechanistic pathways that could potentially explain the effect of the exercise on fatigue including cardiovascular, neuroendocrine and neurotrophic changes. Furthermore, it was hypothesized that secondary MS fatigue could be ameliorated through an increase in fitness

and motor efficiency, improved sleep quality and changes in depressive symptoms.

The other main challenge that impedes meaningful conclusions in fatigue studies is the lack of consistent measurement of perceived fatigue using psychometrically sound outcome measures. Fatigue is considered multidimensional that could result from changes in central neural system, psychological and physiological factors [23]. Often the current self-report fatigue scales and questionnaires suffer from construct contamination and measure one or only few dimensions of fatigue. In a systematic review into the evidence of psychometric properties of outcomes of fatigue in pwMS and Parkinson's disease, Elbers *et al.* reported that there is limited evidence for the psychometric properties of the Fatigue Severity Scale and Modified Fatigue Impact Scale, the most common self-reported fatigue measures in MS studies [24]. The authors concluded that the fatigue scale for motor and cognitive functions has moderate evidence for its internal consistency and structural validity, while the unidimensional fatigue impact scale has moderate evidence for its reliability and structural validity in pwMS. Apart from the relative lack of psychometric sound measure for self-reported fatigue in pwMS, studies on measuring motor fatigability show inconsistency in the concept being assessed, disparity of exercise protocols and fatigability outcome, limited information of psychometric properties of outcome measures and lack of cut-off value showing clinically significant motor fatigability [25]. Such outcomes may contribute to understanding of some of the mechanisms through which exercise interventions could potentially result in a decrease in self-reported fatigue in pwMS.

Considering the limitations of the current literature, it is suggested that future studies exploring the benefits of exercise interventions to reduce fatigue in pwMS are powered high quality randomized trials of exercise modalities including participants with a predefined level of fatigue, with detailed description of both participant characteristics and intervention

mediators, such as overload components, progression, duration and mode of delivery [26]. Such studies should also report not only on outcomes of perceived fatigue, but also motor fatigability and other MS symptoms associated with fatigue such as depression and assess these outcomes both short and long term. This will then allow future reviews to systematically characterize and explore the nuanced detail of exercise training program components used [26] to enable better categorization of exercise intervention types that may be hypothesized to favorably affect fatigue. This may not only result in a reduction of heterogeneity of the results and thus an increase in the quality of evidence generated, but may also help in understanding which component of the exercise intervention is more effective in reducing MS fatigue, and in which group of pwMS. Furthermore, it is suggested that future high quality reviews focus on exploring what component/s or mediators of exercise intervention and through which mechanistic pathway/s, may have the greatest effect in mitigating MS-related fatigue.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/nmt-2017-0009

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- 1 WHO. *Atlas: Multiple Sclerosis Resources in the World*. World Health Organisation, Geneva, Switzerland. www.who.int
- 2 Giovannoni G. Multiple sclerosis related fatigue. *J. Neurol. Neurosurg. Psych.* 77(1), 2–3 (2006).
- 3 Wynia K, Middel B, Van Dijk JP, De Keyser JH, Reijneveld SA. The impact of
- 4 Kluger BM, Krupp LB, Enoka RM. Fatigue and fatigability in neurologic illnesses: disabilities on quality of life in people with multiple sclerosis. *Mult. Scler.* 14(7), 972–980 (2008).

- proposal for a unified taxonomy. *Neurology* 80(4), 409–416 (2013).
- 5 Chaudhuri A, Behan PO. Fatigue in neurological disorders. *Lancet* 363(9413), 978–988 (2004).
 - 6 Multiple Sclerosis Society, UK. What is MS (2016).
www.msociety.org.uk
 - 7 Khan F, Amatya B, Galea M. Management of fatigue in persons with multiple sclerosis. *Front. Neurol.* 5, 177 (2014).
 - 8 Asano M, Finlayson ML. Meta-analysis of three different types of fatigue management interventions for people with multiple sclerosis: exercise, education, and medication. *Mult. Scler. Int.* 2014, 798285 (2014).
 - 9 Andreassen AK, Stenager E, Dalgas U. The effect of exercise therapy on fatigue in multiple sclerosis. *Mult. Scler.* 17(9), 1041–1054 (2011).
 - 10 Langeskov-Christensen M, Bisson EJ, Finlayson ML, Dalgas U. Potential pathophysiological pathways that can explain the positive effects of exercise on fatigue in multiple sclerosis: a scoping review. *J. Neurol. Sci.* 373, 307–320 (2017).
 - 11 Rietberg MB, Brooks D, Uitdehaag BM, Kwakkel G. Exercise therapy for multiple sclerosis. *Cochrane Database Syst. Rev.* (1), CD003980 (2005).
 - 12 Shea BJ, Grimshaw JM, Wells GA *et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med. Res. Methodol.* 7, 10 (2007).
 - 13 Shea BJ, Hamel C, Wells GA *et al.* AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J. Clin. Epidemiol.* 62(10), 1013–1020 (2009).
 - 14 Sharif MO, Janjua-Sharif FN, Ali H, Ahmed F. Systematic reviews explained: AMSTAR-how to tell the good from the bad and the ugly. *Oral Health Dent. Manag.* 12(1), 9–16 (2013).
 - 15 Atkins D, Best D, Briss PA *et al.* Grading quality of evidence and strength of recommendations. *BMJ* 328(7454), 1490 (2004).
 - 16 Guyatt GH, Oxman AD, Vist GE *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336(7650), 924–926 (2008).
 - 17 Cramer H, Lauche R, Azizi H, Dobos G, Langhorst J. Yoga for multiple sclerosis: a systematic review and meta-analysis. *PLoS ONE* 9(11), e112414 (2014).
 - 18 Pilutti LA, Greenlee TA, Motl RW, Nickrent MS, Petruzzello SJ. Effects of exercise training on fatigue in multiple sclerosis: a meta-analysis. *Psychosom Med.* 75(6), 575–580 (2013).
 - 19 Heine M, Van De Port I, Rietberg MB, Van Wegen EE, Kwakkel G. Exercise therapy for fatigue in multiple sclerosis. *Cochrane Database Syst. Rev.* (9), CD009956 (2015).
 - 20 Motl RW, Gosney JL. Effect of exercise training on quality of life in multiple sclerosis: a meta-analysis. *Mult. Scler.* 14(1), 129–135 (2008).
 - 21 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 [with consumer summary]. *Phys. Ther.* 91(8), 1166–1183 (2011).
 - 22 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. *Arch. Phys. Med. Rehabil.* 93(10), 1701–1708 (2012).
 - 23 Rudroff T, Kindred JH, Ketelhut NB. Fatigue in multiple sclerosis: misconceptions and future research directions. *Front. Neurol.* 7, 122 (2016).
 - 24 Elbers RG, Rietberg MB, Van Wegen EE *et al.* Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Qual. Life Res.* 21(6), 925–944 (2012).
 - 25 Severijns D, Zijdewind I, Dalgas U, Lamers I, Lismont C, Feys P. The assessment of motor fatigability in persons with multiple sclerosis. *Neurorehabil. Neural Repair* 31(5), 413–431 (2017).
 - 26 Hoffmann TC, Glasziou PP, Boutron I *et al.* Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 348, g1687 (2014).