Systematic Review

Which exercise and behavioural interventions show most promise for treating fatigue in multiple sclerosis? A network meta-analysis

Anthony M Harrison*, Reza Safari*, Tom Mercer, Federica Picariello, Marietta L van der Linden, Claire White, Rona Moss-Morris** and Sam Norton**

Abstract

Background: Fatigue is a common, debilitating symptom of multiple sclerosis (MS) without a current standardised treatment.

Objective: The aim of this systematic review with network meta-analyses was to estimate the relative effectiveness of both fatigue-targeted and non-targeted exercise, behavioural and combined (behavioural and exercise) interventions.

Methods: Nine electronic databases up to August 2018 were searched, and 113 trials (n=6909) were included: 34 were fatigue-targeted and 79 non-fatigue-targeted trials. Intervention characteristics were extracted using the Template for Intervention Description and Replication guidelines. Certainty of evidence was assessed using GRADE.

Results: Pairwise meta-analyses showed that exercise interventions demonstrated moderate to large effects across subtypes regardless of treatment target, with the largest effect for balance exercise (SMD=0.84). Cognitive behavioural therapies (CBTs) showed moderate to large effects (SMD=0.60), with fatigue-targeted treatments showing larger effects than those targeting distress. Network meta-analysis showed that balance exercise performed significantly better compared to other exercise and behavioural intervention subtypes, except CBT. CBT was estimated to be superior to energy conservation and other behavioural interventions. Combined exercise also had a moderate to large effect.

Conclusion: Treatment recommendations for balance and combined exercise are tentative as the certainty of the evidence was moderate. The certainty of the evidence for CBT was high.

Keywords: Fatigue, multiple sclerosis, network meta-analysis, randomised controlled trials, quasi-randomised controlled trials, behavioural interventions, exercise interventions, TIDieR

Date received: 2 October 2020; revised: 19 January 2021; accepted: 29 January 2021

Introduction

Fatigue affects around two-thirds of people with multiple sclerosis (pwMS).1,2 MS-related fatigue is a complex and subjective symptom characterised by a lack of energy or overwhelming sense of physical and/or mental tiredness.3-5 Fatigue is associated with poorer quality of life (QoL) even when controlling for disease severity and is a major reason for stopping work in pwMS.6

Treating fatigue has been identified as priority by pwMS.7 In routine clinical care, pharmacological treatments tend to be the treatment of choice, with behavioural and exercise interventions considered as alternative or adjunctive treatment options.8 In many cases, patients are never offered these adjunctive treatments. This is concerning as the current evidence base suggests pharmacological interventions to date are largely ineffective, while exercise and behavioural interventions have larger effects.9

So, why are behavioural treatments for fatigue not part of standard treatment? One reason may be there are many randomised controlled trials (RCTs) of very different behavioural interventions for fatigue in MS. There is little standardisation of these interventions across the studies, making it unclear which of these should be part of the routine offer. Standard pairwise
meta-analytic systematic reviews of exercise and/or behavioural interventions have attempted to summarise the findings across studies.9–17 However, control conditions of included trials within these reviews differ considerably, making it challenging to directly compare effect sizes across intervention types. In contrast to pairwise meta-analysis which has been used in all previous reviews, network meta-analysis (NMA) allows for trial arm data to be compared across trials (e.g. aerobic exercise to energy conservation (EC) methods) even when these direct head-to-head comparisons do not exist.18 This allows a ranking of the most effective treatments against each other19 and provides greater clarity of which may be the best to take forward.

In most previous systematic reviews, interventions designed specifically for fatigue have not been distinguished from those where fatigue is one of the secondary outcomes.10,11,12,16 Combining these in one analysis may dilute the effect. The overarching aim of the current meta-analysis was to elucidate which non-pharmacological interventions are likely to be the most effective for the management of fatigue in MS by:

1. Thoroughly reviewing the interventions included in the trials so that subgroups of exercise, behavioural and combined interventions could be clearly defined including whether they are targeting fatigue as a primary or secondary outcome;
2. Directly comparing these defined intervention groupings using NMA at the end of treatment and at follow-up;
3. Exploring if fatigue-targeted interventions (intervention exclusively designed with an aim to reduce fatigue) have larger effects than non-fatigue-targeted interventions, including some predefined within category comparisons (1) cognitive behavioural therapy (CBT) for depression, where fatigue is measured as a secondary outcome, versus CBT for fatigue which is a different protocol where fatigue is the primary target, and (2) whether systematic differences in treatment effects exist depending on setting of general exercise delivery, that is, aquatic versus land-based;
4. Defining the quality of the evidence using GRADE criteria;
5. Exploring how treatment effects vary according to type of MS, duration of health care professional (HCP) contact and study quality (i.e. risk-of-bias) in sensitivity/subgroup analyses.

Methods
This review was conducted according to PRISMA guidelines20,21 and a full description of methods used can be found in the protocol. Our registered protocol22 had two additional aims which are not detailed in this paper: (1) to conduct pairwise meta-analyses for pooled treatment effects across intervention categories and estimate statistical heterogeneity and (2) to explore fatigability as a secondary outcome. Fatigability has been defined as ‘the magnitude or rate of change in a performance criterion relative to a reference value or given time of task performance or measure of mechanical output’.23 We have conducted the pairwise comparisons but as they overlap with previous publications, we have focused this paper on the novel NMA. We did not find any papers with fatiguability as an outcome.22,24 The inclusion/exclusion criteria are detailed in Table 1.

Studies were identified though a systematic online search of nine databases shown in Figure 1 undertaken in December 2015, and repeated in July 2017 and August 2018, using search terms presented in Supplementary Material A. As part of the updated searches, the same search strategy was used in the nine databases as the original search, but restricted to the date of the last search to capture any new hits since. Studies were also identified through a search of trial and grey literature databases, forward citation, reference lists of included studies and previous reviews, and contacting authors of published studies. Initial screening of titles/abstracts and full-text articles was performed independently by co-authors A.M.H. and R.S. and disagreements resolved by team consensus. The 2017 and 2018 updated searches and screening were conducted by M.L.v.d.L. and Georgia Andreopoulou (see Acknowledgements).

Data extraction and risk of bias assessment
A data extraction tool, including Cochrane Risk of Bias (RoB) tool, was developed a priori based on the Cochrane Handbook for Systematic Reviews25 and elements of the Template for Intervention Description and Replication (TIDieR).26 Data extracted were sample sizes, pre-post change scores between groups, means and standard deviations per arm at each post-randomisation assessment and related information (e.g. standard errors, confidence intervals and test statistics). Details of study participant characteristics and key intervention components or techniques were also extracted.

Data extraction and RoB assessments were performed by R.S. and S.G. for exercise and by A.M.H. and L.S.
Table 1. PICOS inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study design</strong></td>
<td>Randomised controlled trials (RCTs) or quasi-RCTs (feasibility, efficacy and naturalistic/pragmatic trials)</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td>Adults with any type of MS</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Behavioural and/or exercise interventions either fatigue-targeted or non-targeted versus any comparator (no intervention, usual care, medication, placebo treatment or another active intervention)</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td>Fatigue measured using validated patient-reported outcome measures (PROMs) of fatigue severity and/or impact as either a primary or secondary outcome</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>There were no language restrictions</td>
</tr>
</tbody>
</table>

Figure 1. PRISMA flowchart diagram.
for behavioural or combined trials. Agreement for RoB, based on 15 exercise and 15 behavioural or mixed studies, ranged from Kappa = 0.54–0.86 across each rater pairing indicating 'good' consistency. Agreement for RoB, based on 15 exercise and 15 behavioural or mixed studies, ranged from Kappa = 0.54–0.86 across each rater pairing indicating 'good' consistency. All remaining studies, including the update searches, were single-extracted by S.G., L.S., M.L.v.d.L., G.A., R.S. and A.M.H., and meta-analytic data and RoB were cross-checked by M.L.v.d.L., R.S. and A.M.H.

**Data analysis**

**Descriptive data and categorisation of studies.** All studies were tabulated, summarising demographics of the samples, fatigue outcome measures, end-of-treatment timepoint, attrition at follow-up, long-term follow-up if available and detailed intervention characteristics (see tables in Supplementary Material B). For fatigue-targeted studies, this information is available elsewhere.

The primary outcome was fatigue severity and/or impact at \( \leq 3 \) months post-treatment (end-of-treatment), while the secondary outcome was fatigue at longer follow-up, \( > 3 \) to 6 and \( > 6 \) months post-treatment. A three-level categorisation of interventions was used for the analysis (Figure 2; Table 2) based on a component breakdown conducted as part of a previous review. At the top-level interventions are coded as exercise, behavioural, combined and control treatments such as treatment as usual (TAU) and medication. The second level includes subgroups, defined in Table 2, to account for heterogeneity in intervention types and is the main unit for analysis. There were six subgroups of exercise, five for behavioural and two for combined. A further level of categorisation reflects heterogeneity in the nature of the intervention (e.g. intensity or environment). FACETS, consisting of CBT with some EC techniques for improved patient acceptability, was classified as EC, similar to two other interventions.

![Figure 2. Hierarchical categorisation of interventions included in the pairwise meta-analysis and network meta-analysis.](image-url)
Table 2. Description of intervention subtypes based on total number of studies identified (N=135).

<table>
<thead>
<tr>
<th>Intervention subtype</th>
<th>Descriptor</th>
<th>N studies included</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exercise</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerobic</td>
<td>Main fitness component in exercise in which the body’s large muscles move in a rhythmic manner for a sustained period of time. Aerobic activity, also called endurance activity, improves cardiorespiratory fitness. Examples include walking, running, cycling and swimming.</td>
<td>25</td>
</tr>
<tr>
<td>Both targeted and non-targeted interventions included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resistive (also referred to as resistance)</td>
<td>Main physical fitness component in exercise aiming at increasing skeletal muscle strength, power, endurance and mass (e.g. progressive resistance training (PRT) – fixed weights, PRT-free weights, resistance training and body-weight-resistance exercise).</td>
<td>8</td>
</tr>
<tr>
<td>Both targeted and non-targeted interventions included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexibility</td>
<td>Main fitness component in exercises aimed at increasing the range of motion possible at a joint. Flexibility is specific to each joint and depends on specific variables, including but not limited to the tightness of specific ligaments and tendons.</td>
<td>1</td>
</tr>
<tr>
<td><strong>Balance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both targeted and non-targeted interventions included</td>
<td>Main fitness component in static and dynamic exercises that are designed to improve individuals’ ability to withstand challenges from postural sway or destabilising stimuli caused by self-motion, the environment or other objects.</td>
<td>9</td>
</tr>
<tr>
<td>General exercise (aquatic and land-based)</td>
<td>Exercise involving combination of two or more of the above components, no dominant fitness component focus provided.</td>
<td>28 of which 6 aquatic</td>
</tr>
<tr>
<td>Both targeted and non-targeted interventions included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined exercise</td>
<td>Exercise explicitly aimed at more than one of the exercise components listed above, often using a progressive overload principle.</td>
<td>24</td>
</tr>
<tr>
<td><strong>Behavioural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy conservation</td>
<td>Energy effectiveness strategies or energy conservation (EC) education is defined as the ‘the identification and development of activity modifications to reduce fatigue through a systematic analysis of daily work, home and leisure activities in all relevant environments’ (p. 592). EC strategies include analysing and modifying activities to reduce energy expenditures; taking frequent rests; prioritising activities; planning; delegating some activities, using the body efficiently, organising tools, materials and work area, using assistive technologies to conserve energy; adopting good posture; leading a healthy lifestyle (regular exercise, healthy diet and stress management, examining and modifying standards and priorities).</td>
<td>12 of which 3 incorporated elements of CBT</td>
</tr>
<tr>
<td>Only fatigue-targeted interventions included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive behavioural therapy</td>
<td>CBT is founded on the premise that physiological, cognitive (thinking), emotional and behavioural responses influence one another in a reciprocal way within the context of the social environment, where change in any one of these responses may produce changes in others. CBT was originally developed to treat anxiety and depression. CBT has also been developed to treat fatigue specifically and MS-specific versions have been developed. CBT for fatigue is based on guided discovery where individuals identify which perpetuating factors may be relevant to them and are provided techniques to alter or manage these behavioural, cognitive, emotional and external factors. Protocols vary but tend to include creating consistent activity routines (including sleep wake cycles), graded increase of daily activity, identifying and managing unhelpful thoughts in relation to fatigue, and managing stress.</td>
<td>14, of which 4 low intensity</td>
</tr>
<tr>
<td>Both targeted and non-targeted interventions included</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
**Table 2. (Continued)**

<table>
<thead>
<tr>
<th>Intervention subtype</th>
<th>Descriptor</th>
<th>N studies included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurocognitive rehabilitation</td>
<td>Neurocognitive rehabilitation aims to reduce cognitive deficits, through learning and memory-based training to improve cognitive skills, like attention, or identification of alternative skills to compensate for cognitive deficits in daily living.</td>
<td>7</td>
</tr>
<tr>
<td>Relaxation and biofeedback</td>
<td>Relaxation techniques include ‘a number of practices such as progressive relaxation, guided imagery, biofeedback, self-hypnosis, and deep breathing exercises. The goal is similar in all: to produce the body’s natural relaxation response, characterized by slower breathing, lower blood pressure, and a feeling of increased well-being’.</td>
<td>5</td>
</tr>
<tr>
<td>Emotional expression therapy</td>
<td>The aim of emotional expression therapy is to facilitate and promote acceptance, expression, regulation and understanding of emotions.</td>
<td>3</td>
</tr>
<tr>
<td>Education or information</td>
<td>Education or information</td>
<td>8, predominantly as active control plus 3 active controls re-categorised as education for inclusion in the analysis</td>
</tr>
<tr>
<td>Combined (behavioural and exercise)</td>
<td>Behavioural techniques to promote physical activity/exercise or behavioural interventions with integrated exercise (e.g. EC plus aerobic exercise).</td>
<td>11</td>
</tr>
<tr>
<td>Physical rehabilitation</td>
<td>Occupational therapy and physiotherapy.</td>
<td>6</td>
</tr>
<tr>
<td>Control</td>
<td>Treatment as usual</td>
<td>82</td>
</tr>
<tr>
<td>Medication</td>
<td>Medication</td>
<td>2, of which Nedeljkovic et al., Methylprednisolone + TAU</td>
</tr>
</tbody>
</table>

Note. The number of studies for each intervention subtype captures studies where each intervention subtype was evaluated at least in one arm based on the total number of studies identified (N=135). In text, the number of studies and arms are reported for the top-level intervention groupings after exclusion of studies that did not provide sufficient data to be included in the analyses and combining/removing arms that were in the same intervention subcategory.
These interventions differ from what is considered a typical exemplar of this type of category and sensitivity analyses have been conducted to assess whether this affects the outcome. At post-treatment relative to TAU, EC with elements of CBT, SMD = −0.06, 95% CI = (−0.38, 0.27), appeared marginally less effective than EC alone, SMD = −0.23, 95% CI = (−0.41, −0.05). Whether methodological heterogeneity appears to result in statistical heterogeneity was explored in sensitivity analysis. Discrepancies between category assignment were resolved in discussion with M.L.d.v.L. and R.M.M. until consensus was reached.

Statistical analysis
All analyses were conducted using Stata 15.1. First, separate pairwise random-effects meta-analyses were conducted (using `admetan` package) including only studies comparing interventions against a TAU control arm (Supplementary Material C). Second, using restricted maximum likelihood by the `mvmeta` command and `network` packages, an NMA including all studies in a single analysis combining direct and indirect effect estimates was conducted. Effect sizes were expressed as standardised mean differences (SMD) between groups, calculated as Hedge’s $g$ with a small sample correction applied.\(^{29}\) SMDs were calculated using the between-group post-treatment mean change from baseline scores where available and post-treatment means otherwise. The consistency assumption, which is an extension of the pairwise meta-analytic assumption of statistical homogeneity,\(^{30}\) was tested using loop-specific and node-splitting approaches. The final model was estimated excluding studies that contributed to violations of the consistency assumption. Treatments were ranked according to estimates of the surface under the cumulative ranking curve (SUCRA), where this ranges between 0 and 1 and higher values indicate a greater likelihood of treatment being more effective relative to other treatments.

Interventions in two-arm studies categorised within the same subgroup (e.g. two aerobic exercise arms) could not be included in the same analysis. If a study included an additional arm/s not in the intervention subgroup (e.g. a common TAU control arm), the treatment arms within the same subgroup were combined by calculating the weighted mean and pooled standard deviation.\(^{31}\) Combining arms in this way does not impact on the treatment effect for that subgroup but ensures that the standard error is calculated correctly.\(^{32}\) If all arms were in the same subgroup, the control arm was re-categorised to allow study inclusion where possible or excluded from the analyses (the control arms for Mackay et al.\(^{40}\) and van Kessel et al.\(^{41}\) were both categorised as education arms, whereas originally the control arms mimicked the intervention arms but without biofeedback or without email support, respectively. In Straudi et al.,\(^{42}\) both arms were categorised as general exercise, and consequently, the control arm consisting of walking therapy was categorised as aerobic exercise for the purpose of analysis).

Results
Description of the included studies
Overall, 135 studies (292 arms) were identified from the literature search (Figure 1). These studies included 119 two-arm, 12 three-arm and 4 four-arm trials. Twenty-two studies did not provide sufficient data for inclusion in the pairwise or NMAs. After combining/removing arms that were in the same intervention subcategory, data from 6909 participants across 235 arms from 113 studies were included in the analyses (studies with multiple arms in the same category combined were Briken et al.,\(^{43}\) Garrett et al.,\(^{44}\) Hogan et al.,\(^{45}\) Seebacher et al.,\(^{46}\) Seebacher et al.\(^{47}\) and Shanazari et al.\(^{48}\) Both arms from Samaei et al.\(^{49}\) were aerobic exercise and could not be included in the analysis). There were 78 arms (51 studies) involving entirely or mainly exercise interventions, 56 arms (43 studies) involving entirely or mainly behaviour interventions, and 19 arms (19 studies) involving combined exercise and behavioural interventions. Table 2 shows the number of studies in each type of the intervention subtypes based on the total number of studies identified ($N$=135). Few studies directly compared an intervention involving an entirely or mainly exercise intervention with an entirely or mainly exercise intervention with an entirely or mainly behaviour intervention. No trials compared either CBT or EC to an exercise intervention. Figure 3
summarises the subcategory comparisons included in the network. The results focus on the NMA, but the results of the pairwise meta-analysis are available in Supplementary Material C.

**NMA of post-treatment effects**

The final NMA based on the model meeting the consistency assumption (inconsistency $\chi^2(26) = 34.93; p = 0.113$) involved data from 6430 participants across 213 arms from 103 studies (Figure 4). The median sample size was 20 per arm with an interquartile range from 14 to 37 (range = 4–204).

Ten of the 113 studies identified were excluded from the final model due to issues with assumption of consistency between direct and indirect treatment effect estimates in preliminary analysis ($\chi^2(29) = 43.5; p = 0.041$). The exclusion of six studies with treatment effect estimates that were clear outliers (outside 99% limits on funnel plot) reduced the overall test for inconsistency to non-significant.33–35,50–52 However, issues with loop-specific inconsistency remained for physical rehabilitation, relaxation/biofeedback and education/information until a further four studies were removed.53–56 The excluded studies involved general exercise, combined exercise, physical rehabilitation and relaxation/biofeedback as active control arms, where most other studies involved these treatments as the experimental arm. Estimates from the preliminary NMA model involving all 113 studies are presented in Supplementary Material D.

Treatment effect estimates relative to TAU from the final NMA model are presented in Figure 4. The ordering of the interventions in the graph is based on the SUCRA. The largest effect is observed for balance exercise (SMD = −0.84, 5 studies, 124 participants).

Along with CBT (SMD = −0.60, 15 studies, 594 participants) (separating CBT into high vs low intensity indicated no difference in the treatment effect estimates: high intensity SMD = −0.62, 95% CI = (−0.82, −0.41), low intensity SMD = −0.51, 95% CI = (−0.89, −0.13)), general exercise (SMD = −0.52, 16 studies, 373 participants) (separating general exercise into exercise undertaken in a non-aquatic vs aquatic environment indicated no difference in the treatment effect estimates: non-aquatic SMD = −0.50, 95% CI = (−0.80, −0.21), aquatic SMD = −0.39, 95% CI = (−0.80, 0.03). The setting of delivery for general exercise may therefore not be an important factor); resistive exercise (SMD = −0.42, 5 studies, 90 participants); combined exercise (SMD = −0.39, 15 studies, 531 participants); aerobic exercise (SMD = −0.38, 22 studies, 352 participants); relaxation/biofeedback (SMD = −0.32, 8 studies, 182 participants); behavioural plus exercise (SMD = −0.29, 12 studies, 633 participants); and EC (SMD = −0.19, 10 studies, 436 participants) were all estimated to have statistically significant moderate effects on fatigue.

Table 3 provides an integrated overview of treatment effects based on the pairwise and NMAs and GRADE rating of the certainty of the evidence. In most categories, certainty of the evidence was moderate and the effects for the pairwise and network analyses were very similar. The certainty of the evidence presented for CBT, behavioural plus exercise and EC was high and for combined exercise and rehabilitation, it was low.

Table 4 details the significant comparative efficacy between interventions from the NMA. Balance exercise performed significantly better compared to aerobic exercise, combined exercise, EC, neurocognitive rehabilitation, relaxation/biofeedback, emotional expression therapy, education/information, behavioural interventions including exercise or physical activity, and physical rehabilitation. CBT was estimated to perform significantly better than EC, education/information, behavioural interventions including exercise or physical activity, and physical rehabilitation.

**NMA of follow-up data**

Data from 33 studies (35 arms) were available at midterm follow-up (3–6 months) allowing for assessment of treatment effects for 9 treatment types. The effects were generally comparable to end-of-treatment, with some attenuation evident for most interventions (Figure 5). However, these estimates need to be interpreted with caution as the number of studies per intervention type is typically low and the effect estimates generally provide considerable uncertainty as to the
Table 3. Treatment effects based on pairwise and network meta-analyses by intervention subcategories taking into consideration certainty of the evidence based on GRADE assessment.

<table>
<thead>
<tr>
<th>Classification of interventions</th>
<th>Intervention</th>
<th>Pairwise meta-analysis</th>
<th>NMA&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Certainty of the evidence&lt;sup&gt;d&lt;/sup&gt; (GRADE)&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N&lt;sub&gt;comparisons&lt;/sub&gt;: &lt;i&gt;N&lt;/i&gt;&lt;sub&gt;participants&lt;/sub&gt; (median)</td>
<td>SMD&lt;sup&gt;a&lt;/sup&gt; (95% CI)</td>
<td>I&lt;sup&gt;2&lt;/sup&gt; (%)</td>
</tr>
<tr>
<td>Moderate to large effect</td>
<td>Balance</td>
<td>3; 168 (67)</td>
<td>−0.87 (−1.18, −0.55)</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Cognitive behavioural therapy</td>
<td>9; 775 (50)</td>
<td>−0.54 (−0.69, −0.40)</td>
<td>3.40</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>11; 544 (34)</td>
<td>−0.47 (−0.65, −0.30)</td>
<td>32.60</td>
</tr>
<tr>
<td>Small to moderate effect</td>
<td>Resitive</td>
<td>4; 161 (32.5)</td>
<td>−0.46 (−0.77, −0.15)</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Combined exercise</td>
<td>12; 767 (41.5)</td>
<td>−0.35 (−0.50, −0.20)</td>
<td>68.20</td>
</tr>
<tr>
<td></td>
<td>Aerobic</td>
<td>15; 475 (30)</td>
<td>−0.41 (−0.62, −0.22)</td>
<td>18.50</td>
</tr>
<tr>
<td></td>
<td>Relaxation and biofeedback</td>
<td>5; 217 (48)</td>
<td>−0.53 (−0.80, −0.26)</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Behavioural plus exercise</td>
<td>8; 728 (94.5)</td>
<td>−0.28 (−0.43, −0.13)</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Emotional expression therapy</td>
<td>1; 60 (NA)</td>
<td>NA&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NA&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Neurocognitive rehabilitation</td>
<td>6; 269 (41.5)</td>
<td>−0.24 (−0.51, 0.02)</td>
<td>12.30</td>
</tr>
<tr>
<td></td>
<td>Physical rehabilitation</td>
<td>4; 211 (43)</td>
<td>−0.16 (−0.44, 0.11)</td>
<td>0.00</td>
</tr>
<tr>
<td>Negligible to small effect</td>
<td>Energy conservation</td>
<td>6; 632 (108.5)</td>
<td>−0.19 (−0.35, −0.03)</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Education or information</td>
<td>1; 49 (NA)</td>
<td>NA&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NA&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
<tr>
<td>Overall groupings</td>
<td>Overall exercise intervention</td>
<td>46; 2153 (32)</td>
<td>−0.44 (−0.53, −0.35)</td>
<td>42.5</td>
</tr>
<tr>
<td></td>
<td>Overall behavioural intervention</td>
<td>28; 2002 (50)</td>
<td>−0.37 (−0.47, −0.28)</td>
<td>12.40</td>
</tr>
<tr>
<td></td>
<td>Overall combined intervention</td>
<td>12; 939 (85)</td>
<td>−0.25 (−0.38, −0.12)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>Standardised mean difference.  
<sup>b</sup>Network meta-analysis.  
<sup>c</sup>Surface under the cumulative ranking curve.  
<sup>d</sup>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.  
<sup>e</sup>Grading of Recommendations, Assessment, Development and Evaluations.  
<sup>f</sup>Total sample size is small.  
<sup>g</sup>Most information is from studies with inadequate allocation concealment or incomplete accounting for outcome data.  
<sup>h</sup>Wide variation in effect size exists across studies or large I<sup>2</sup> indicates large proportion of the variation in effect size due to among-study differences.  
<sup>i</sup>Not applicable.  
<sup>j</sup>Small sample effect on the estimated effect due to asymmetrical funnel plot.
<table>
<thead>
<tr>
<th>Exercise</th>
<th>Aerobic exercise</th>
<th>Balance exercise</th>
<th>Resistive exercise</th>
<th>Flexibility exercise</th>
<th>Combined exercise</th>
<th>General exercise</th>
<th>Combined interventions</th>
<th>Physical rehabilitation</th>
<th>Treatment as usual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.03, -0.09)</td>
<td>(0.03, 0.18)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
</tr>
<tr>
<td></td>
<td>(0.03, 0.18)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
</tr>
<tr>
<td></td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
</tr>
<tr>
<td></td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
<td>(0.03, -0.09)</td>
</tr>
</tbody>
</table>

Comparisons are for columns versus row (i.e., negative values column and positive values row). Significant (p < 0.05) differences denoted in bold. Please note that order of interventions is not ranked in this table.
Figure 5. Treatment effects relative to TAU at mid-term follow-up (3–6 months).
Estimates of the effect at longer-term follow-up versus TAU where only one study was available are not presented (balance exercise, combined exercise and emotional expression therapy). true effect sizes. Data at longer-term follow-up (7–12 months) were available in 9 studies, which was insufficient for analysis.

Moderator and sensitivity analyses
Planned analyses were conducted to assess whether effect sizes vary according to the following pre-specified characteristics.

Type of MS and total contact hours with HCP. Most studies included mixed MS samples or did not specify MS type. Samples of 19 studies consisted only of participants with RRMS. Of these, data were too sparse for NMA, with a maximum of five studies evaluating combined exercise interventions and only one CBT intervention. Similarly, there were only six studies with samples consisting exclusively of participants with PPMS, and therefore insufficient for NMA. The number of studies evaluating subcategories of interventions by MS subtype is detailed in Supplementary Material E.

Predominantly, interventions included over 80 minutes of HCP contact (except for 11 arms) and consequently, this precluded us from examining treatment effects by intensity of HCP contact.

Fatigue-targeted versus non-fatigue-targeted interventions. A common effect of fatigue-targeted versus non-fatigue-targeted interventions was estimated across all interventions, rather than on an intervention-by-intervention basis, due to some intervention categories having small numbers of either targeted or non-targeted interventions. The magnitude of the difference in the SMD for fatigue-targeted versus non-target interventions was negligible and non-significant ($g = -0.01, p = 0.940, 95\% \text{ CI} = (-0.23, 0.21)$).

Since some CBT interventions specifically targeted fatigue, the NMA was re-estimated splitting CBT into a targeted and non-targeted groups. For the targeted CBT interventions, the effect was larger (targeted $-0.78, p < 0.001, 95\% \text{ CI} = (-1.04, -0.53)$; non-targeted $-0.47, p < 0.001, 95\% \text{ CI} = (-0.68, -0.26)$); however, the difference was not statistically reliable (difference $g = 0.31, p = 0.062, 95\% \text{ CI} = (-0.02, 0.64)$).

RoB. A summary of RoB assessment as percentage across all included studies and for each included study is presented in Supplementary Material F. Behavioural studies had the lowest RoB relative to exercise studies, followed by combined intervention studies. Exercise studies had the highest RoB relative to behavioural and combined intervention studies. In assessing RoB, performance bias was not considered for the overall quality judgement as lack of blinding of participants and healthcare professional is an inherent limitation in studies of behavioural and exercise interventions. For the other five RoB domains, behavioural studies and combined behavioural and exercise studies had low RoB compared to exercise studies. In exercise studies (both fatigue-targeted and non-targeted interventions), the method for random allocation to arms was either unclear or inappropriate in 9.2% of studies ($n = 7$). Furthermore, 53.9% of studies ($n = 41$) did not provide sufficient information on allocation concealment, and 40.8% of studies ($n = 31$) had incomplete outcome data, whereas in behavioural studies and combined behavioural and exercise intervention studies, these ratings were 6.8% ($n = 3$), 29.5% ($n = 13$), 25.0% ($n = 11$) and 5.9% ($n = 1$), 29.4% ($n = 5$), 17.6% ($n = 3$), respectively. Reporting bias was identified in 10.5% of exercise ($n = 8$) studies compared to 13.6% of behavioural studies ($n = 6$) and 23.5% of combined intervention studies ($n = 4$). Planned sensitivity analysis was not possible to examine RoB on an intervention-by-intervention basis, so a common effect was estimated pooling across all studies. Studies with low RoB were estimated to have SMDs relative to TAU that were less favourable of the intervention compared to those that were not considered low RoB, though the difference was non-significant ($0.15, p = 0.257, 95\% \text{ CI} = (-0.11, 0.41)$).

Discussion
Summary of evidence
The overarching aim of the current meta-analysis was to elucidate which subtypes of exercise, behavioural and combined interventions are likely to be the most effective for the management of fatigue in MS using NMA. NMA reflects a novel and powerful quantitative
A total of 135 studies were identified from the literature search; 113 studies (including 235 arms) had sufficient data to be included in the NMA, but only 103 studies were retained based on the model meeting inconsistency assumptions. The largest intervention category was TAU (82 studies), followed by entirely or mainly exercise interventions (51 studies), entirely or mainly behavioural interventions (43 studies) and finally, combined exercise and behavioural interventions (19 studies). Studies evaluating behavioural and combined intervention were of better quality according to Cochrane’s RoB tool, compared to exercise studies. A sensitivity analysis across all interventions showed that studies which included fatigue as a primary outcome had similar effects to those where fatigue was a secondary outcome. Overall, sensitivity analysis indicated that high RoB studies resulted in larger estimates of the common effect across all studies; however, this was not statistically significant and the magnitude of the difference was small (difference $g=0.15, p=0.257, 95\% CI = (-0.11, 0.41)$). Notably, many exercise studies had inadequate allocation concealment (50.3%) or incomplete outcome data (40.8%), which, with respect to quality of evidence, lowered our confidence in the overall estimate of treatment effect of exercise interventions, and in particular, of aerobic, general and combined exercise interventions.

The findings of the current review that synthesised evidence on both fatigue-targeted and non-fatigue-targeted interventions using the NMA approach are similar to the findings of our previous review that focused exclusively on interventions aimed at fatigue using pairwise meta-analysis. However, the NMA approach allowed us to gain precision and additional insights into less studied interventions, such as balance exercise ($n=2, current\ review\ n=5$) and aerobic exercise ($n=3, current\ review\ n=22$).

In terms of end-of-treatment effects, balance exercise had the largest effect on fatigue when compared to TAU in the NMA ($SMD = -0.84$). This effect was significantly larger than all other types of exercise and behavioural interventions except CBT. CBT had the next largest effect ($SMD = -0.60$) and was significantly superior to all other behavioural treatments. The GRADE rating for certainty of the CBT effect was high. The rating for balance exercise was moderate so needs to be treated with some caution as it is based on small studies ($N$ range $= 12–88$) and differences in the nature of the interventions. Balance exercise interventions included hippotherapy, vestibular rehabilitation, and balance and eye movement exercises. As there are only end-of-treatment effects in this category, we do not know if these effects will sustain. Follow-up data on CBT showed that this ranked the highest of intervention types where follow-up ($3–6$ months) was available and the effect was still significant although somewhat attenuated ($SMD=0.39$). It also needs to be noted that heterogeneity was also identified in the CBT category, in terms of type and target of CBT (fatigue versus depression or stress), the mode of delivery (web, telephone and face-to-face delivery), amount of therapist contact and type of therapists. Pre-planned subgroup analysis suggests that CBT which specifically targets fatigue may have larger effects that CBT where fatigue is a secondary target.

General exercise (defined as including two or more of the key exercise types – balance, aerobic, strength and/or flexibility) also had a moderate to large effect. Other interventions which had significant moderate effects on fatigue at the end of treatment were resistance exercise, combined exercise, aerobic exercise, relaxation/biofeedback interventions and behavioural interventions which included an exercise or physical activity component. The certainty of the evidence in these categories was low or moderate except for behavioural plus exercise where the GRADE rating was high.

EC which was specifically developed to treat fatigue in MS had a marginally significant effect on fatigue at end-of-treatment and was ranked one of the least effective treatments, followed by flexibility exercise.
and education. The certainty of the small to negligible effect for EC is high. It is worth noting that EC had a significant small effect at longer-term follow-up. The study that contributed most to the follow-up effect combined EC with elements of CBT which may have enhanced the effect. Alternately, small positive effects for EC maybe be experienced later on. Behavioural interventions which included an exercise or physical activity component appeared to retain significant effects at longer-term follow-up, but aerobic and resistive exercise did not. No data were available on longer-term follow-up of combined exercise.

Evidence in favour of end-of-treatment effects of different subtypes of exercise interventions (except flexibility on its own) suggests there is no single optimal exercise modality for MS fatigue, but rather a choice of exercise subtype may depend on MS symptoms, level of disability, and the needs and preferences of pwMS. The mechanisms through which exercise improves fatigue will therefore differ by subtypes of exercise interventions. For example, impaired balance is common in MS, affecting 85% of pwMS, and has been linked to fatigue. Impaired postural control and increased likelihood of falls can lead to reduced ability to perform physical activity and further deconditioning. In addition, extreme and continuous efforts to maintain postural control may not only lead to further muscle strain and subsequently pain, spasticity and inefficient movement patterns, but also demand constant attention—all resulting in increased fatigue.

Further work should focus on tailoring exercise interventions to the needs and preferences of pwMS. Needs-based assessment should include an understanding of possible mechanisms of fatigue which may include the balance ones discussed above or physiological deconditioning due to fatigue related to activity. For this reason, it is of paramount importance for studies to specify MS subtypes within their samples and for future research to explore the treatment needs and systematic patterns in treatment effects by MS subtype. A focus is needed on maintaining effects at long-term follow-up. As interventions which combined exercise and behavioural treatments show sustained significant effects, combining approaches may be beneficial. Combining the two types of interventions may improve adherence to exercise and maintenance of benefits over time. Behavioural treatments may also tackle fatigue mechanisms not addressed by exercise such as poor sleep or disrupted circadian rhythms. To date, no studies have combined CBT with exercise. As CBT ranked highest in the behavioural category, an empirically validated theoretical underpinning—a combined CBT and exercise intervention may be one way to enhance existing treatment effects.

Based on the common effect across all intervention types, there was no significant difference between fatigue-targeted versus non-fatigue-targeted interventions; however, the limited number of fatigue-targeted interventions in some intervention categories precluded us from exploring the effect of treatment target on an intervention-by-intervention basis so this finding should be treated with caution. In terms of CBT, pre-planned moderator analysis suggested that CBT designed to treat fatigue had greater effects than CBT for depression or stress. This may be in part because CBT for fatigue includes a focus on grading or increasing activity levels, whereas CBT for depression focuses on activities which improve mood, suggesting that treatment recommendations should specify type of CBT. This has also been previously highlighted in a meta-analytic systematic review of non-pharmacological interventions for cancer fatigue.

Limitations
To our knowledge, this is the first systematic review of non-pharmacological interventions for MS fatigue that adopted a detailed approach to intervention coding and utilised NMA. Nonetheless, limitations of this review need to be noted. First, for the analyses, categorisation of interventions was necessary given the limited number of some subtypes of exercise, behavioural and combined interventions, for example, mindfulness interventions were grouped together with CBT and FACETS which combined CBT and EC was grouped under EC. Important nuances may have therefore been obscured because of data reduction. This may explain some of the heterogeneity in these subcategories. Any inferences about the effect size relate to the broad category and not necessarily to any specific type of intervention within that category. It is important to also note the heterogeneity related to the measurement of fatigue across studies with 13 different fatigue measures used, an inherent weakness of the fatigue literature. However, evidence in support of the convergent validity between the different self-report measures of fatigue, particularly the Modified Fatigue Impact Scale (MFIS) and Fatigue Severity Scale (FSS) (the most commonly used measures in the studies here, \( n = 49 \) and \( n = 43 \), respectively), and the use of standardised mean differences circumvents the uncertainty related to measurement heterogeneity.
Second, this review included feasibility, efficacy and naturalistic/pragmatic trials. The aims and scope are likely to differ between these types of trials, which may have important implications on treatment effects. Finally, statistical heterogeneity was high for most of the pairwise meta-analyses (Supplementary Material C), which relates to the NMA assumption of consistency. The final NMA model met consistency assumptions after the exclusion of 10 studies and, therefore, most treatment effect estimates can be interpreted without too much concern regarding the level of statistical heterogeneity initially observed. However, there was some indication of inconsistency in the estimates for the relaxation/biofeedback and education/information intervention treatment effect estimates. These estimates must be interpreted more cautiously.

Conclusion

The findings of this review provide insights and specific recommendations for treatment guidelines for fatigue in MS. These need to be more prescriptive in terms of type of behavioural intervention. Balance exercise and CBT interventions appear to be the most promising interventions for fatigue based on direct and indirect comparisons. CBT that is fatigue-focused should maximise effects. Other exercise modalities also appear effective, including aerobic and strength/resistive exercise and combined exercise (balance, aerobic, strength and flexibility) in the short-term, but more work is needed to sustain these effects. This may include allowing patient preference in terms of choice of exercise and behavioural methods to enhance regular exercise habits. More support to maintain exercise is needed over the trajectory of illness, particularly around how to adapt exercise for relapse and progression. EC, which is still offered as a treatment to pwMS, has minimal effects as does flexibility exercise on their own. Just educating people on fatigue is also not an effective intervention. To date, a combined exercise and CBT approach for MS fatigue has not been evaluated, but this warrants further exploration. Although there is a clear need for adequately powered trials with evaluation of treatment mechanisms, and a focus on maintaining treatment effects, there is sufficient pooled evidence of more than 100 trials to suggest that the methods mentioned above should become part of standard treatment for fatigue in MS. Future work should focus on how to effectively implement these treatments into routine care including the most cost-effective ways of delivering treatment.

Note: Studies that were included in the review but not mentioned in the main text are the following: 76–183

Acknowledgements

The authors express thanks to the MS Society for funding this research. They would also like to thank Dr Jane Petty, King’s College London, Stephanie Hanna, Carole Bennett and Kay-Anne Sheen, MS Society, patient and public involvement members, for their help and support with the broader project. With thanks also to Sam Goodliffe and Louise Sweeney for their support with data extraction, and Georgia Andreopoulou for her involvement in the updated search and screening process.

Author Contribution

All authors contributed to the design of the review, where previously R.M.M., T.M., C.W. and M.L.v.d.L. developed the funding proposal, and R.M.M., T.M. and M.L.v.d.L. were grant holders for the awarded MS Society UK grant; A.M.H. and R.S. conducted all searches, formal screening, data extraction and quality assessment of studies with support from research assistants (see Acknowledgements); S.N. undertook all statistical analyses and interpretation, with assistance from F.P.; A.M.H., R.S., F.P., S.N. and R.M.M. drafted and revised the manuscript with support from the other authors.

Declaration of Conflicting Interest

The author(s) declared the following potential conflict of interest with respect to the research, authorship, and/or publication of this article: R.M.M. has published RCTs of behavioural interventions for fatigue in MS, which were included in the current review. However, preliminary searches, formal screening of search results against eligibility criteria, data extraction, RoB assessment and data analysis were conducted independently of R.M.M. Remaining authors declare no financial or other conflicts of interest. R.M.M. acknowledges the financial support of the Department of Health via the National Institute for Health Research (NIHR) Specialist Biomedical Research Centre for Mental Health award to the South London and Maudsley NHS Foundation Trust (SLaM) and the Institute of Psychiatry at King’s College London. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the Multiple Sclerosis Society UK (Grant No. 26). The funder of the review had no role in study design, data collection, data analysis, data interpretation or writing of the report. The authors had access to all study data.
and final responsibility for the decision to submit for publication.

**Supplemental material**
Supplemental material for this article is available online.

**References**


48. Shanazari Z, Marandi SM and Minasian V. Effect of 12-week Pilates and aquatic training on fatigue in...


96. Geddes EL, Costello E, Raivel K, et al. The effects of a twelve-week home walking program on


